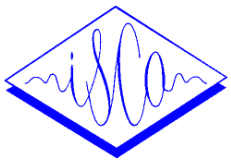


3rd International Conference on
Spoken Language Processing (ICSLP 94)
Yokohama, Japan
September 18-22, 1994

SYNTHESIS OF PATHOLOGICAL VOICE BASED ON A STOCHASTIC VOICE SOURCE MODEL

Yasuo. Endo and Hideki. Kasuya

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

ABSTRACT

This paper proposes a stochastic voice source model to synthesize pathological voice as well as normal voice. The voice signal is assumed to consist of a harmonic signal and an additive laryngeal noise signal. Perturbation of fundamental period, which is the key characteristic of pathological voice, is represented by an autoregressive moving average (ARMA) model. Suitability of the model is tested based on a spectral flatness measure by applying it to the normalized period sequence of 52 pathologic voice samples. Relationship between the model parameters and perceived roughness quality is also investigated. Based on this model, we construct an analysis–conversion–synthesis system of pathological voice. The system is applicable not only to perceptual experiments to explore acoustic correlates of pathological voice qualities, but also to the simulation of the voice quality variations of the voice treatment at voice clinics.

I. INTRODUCTION

Much research has been made with respect to the statistical relationships between perceived quality and acoustic characteristics of pathological voice [1–4]. In order to systematically understand contribution of individual acoustic parameters to the voice quality difference, however, some synthetic method is indispensable, whereby one can perform perceptual test under well controlled experimental conditions. In synthetic methods [5–7], conventional speech synthesizers based on normal voice production have been used. We propose an analysis–conversion–synthesis system of pathological voice.

Included in acoustic characteristics associated with pathological voice quality are perturbation in the fundamental period / amplitude sequence (jitter / shimmer) and laryngeal noise. In order to represent stochastic properties of the jitter and shimmer, we propose an ARMA model of the fundamental period and peak amplitude sequence.

II. HARMONIC ANALYSIS

A brief overview of a method to separate an additive laryngeal noise signal from a harmonic signal is presented [8]. Let $s_m(n)$, $h_m(n)$ and $w_m(n)$ be respectively a voice signal, a harmonic signal and a laryngeal noise signal of the m–th period, then the following relationship holds :

$$s_m(n) = h_m(n) + w_m(n). \tag{1}$$

A harmonic signal is expanded into Fourier series.

$$\begin{aligned} h_m(n) &= \sum_{k=1}^{K} \Big\{ c_k(m) \cos k\frac{2\pi}{T_0(m)} n \\ &+ d_k(m) \sin k\frac{2\pi}{T_0(m)} n \Big\}, \end{aligned} \tag{2}$$

where $T_0(m)$ is the fundamental period of the m–th period of the signal and K is the number of harmonics to be taken into consideration. Harmonic coefficients, $c_k(m), d_k(m)$ are estimated from the signals of L consecutive periods so as to minimize the following equation [8],

$$E_0 = \sum_{j=0}^{L-1} \sum_{n=0}^{T_0(m)-1} \{s_{m+j}(n) - h_m(n)\}^2, \tag{3}$$

where $L > 1$ and it is typically set to 2. Given an estimated harmonic signal $\hat{h}_m(n)$, the laryngeal noise signal $\hat{w}_m(n)$ is computed by subtracting $\hat{h}_m(n)$ from $s_m(n)$:

$$\hat{w}_m(n) = s_m(n) - \hat{h}_m(n). \tag{4}$$

III. PERTURBATION MODEL

A trend sequence included in the fundamental period / amplitude sequence $u(m)$ is approximated by the third order polynomial,

$$\hat{u}(m) = \sum_{i=0}^{3} \beta_i m^i, \tag{5}$$

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

$$x(m) = \frac{u(m) - \hat{u}(m)}{\bar{u}}, \tag{6}$$

where $\bar{u}$ is the average value of the original sequence.

Normalized sequence is modeled by an ARMA process as follows.

$$\begin{aligned} x(m) &= a_1 x(m-1) + a_2 x(m-2) \\ &+ G(e(m) - e(m-1)), \end{aligned} \tag{7}$$

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

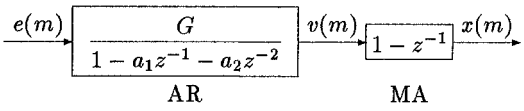


Fig. 1 An ARMA model to represent the normalized sequence $x(m)$

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

$$f_p = \frac{1}{2\pi} \arg s_p \qquad b_p = -\frac{1}{\pi} \log_e |s_p|. \tag{8}$$

Since the MA part has been introduced to simulate the removal of the trend sequence, the zero frequency is fixed at 0(normalized frequency). We use G, f_p and b_p to represent frequency characteristics of the normalized sequence.

IV. ANALYSIS–CONVERSION –SYNTHESIS SYSTEM

Based on the above model, we can construct an analysis–conversion–synthesis system as shown in Fig. 2. The part of analysis subsystem consists of fundamental period extraction, perturbation analysis and harmonic analysis. From the analysis results, we obtain perturbation parameters (gain G, pole frequency f_p and pole bandwidth b_p), average fundamental period $\bar{T}_0$, polynomial coefficients of a trend sequence $\beta_0, \cdots, \beta_3$, harmonic coefficients $c_k(m), d_k(m)$ and laryngeal noise signal $w(n)$. After some of these parameters and a signal are converted, pathological voice signal is synthesized by them. This system can synthesize various pathological voice by controlling the parameters and a signal which may contribute to perceptual quality.

1. Perturbation synthesis

Given the perturbation parameters, normalized fundamental period sequence $x(m)$ is represented as follows,

$$\begin{aligned} x'(m) &= 2e^{-\pi b'_p}(\cos 2\pi f'_p)x'(m-1) \\ &-e^{-2\pi b'_p}x'(m-2) + G'(e(m) - e(m-1)) \end{aligned} \tag{9}$$

With an average fundamental period $\bar{T}_0$ and trend sequence $\hat{T}_0$, the fundamental period sequence $T'_0(m)$ is finally computed as follows :

$$T'_0(m) = \hat{T}_0(m) + \bar{T}_0 x'(m). \tag{10}$$

2. Harmonic synthesis

Using a converted fundamental period sequence $T'_0(m)$ and harmonic coefficients $c_k(m), d_k(m)$, we give harmonic signal as follows :

$$\begin{aligned} h'_m(n) &= \sum_{k=1}^{K} \left\{ c_k(m) \cos k \frac{2\pi}{T'_0(m)} n \right. \\ &+ \left. d_k(m) \sin k \frac{2\pi}{T'_0(m)} n \right\}, \end{aligned} \tag{11}$$

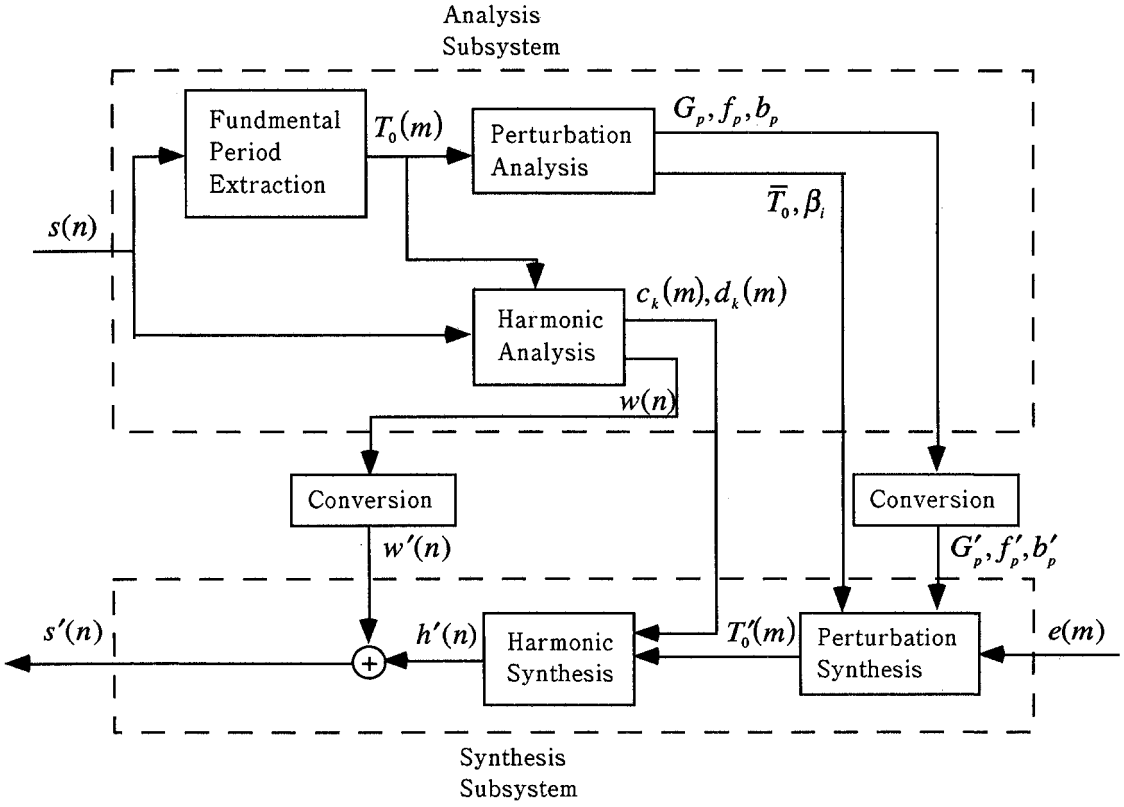


Fig.2 An analysis–conversion–synthesis system

V. EXPERIMENTS

Figure 3 indicates examples of the spectra of normalized period sequences. Solid and dotted lines denote the DFT and ARMA spectra, respectively. The ARMA spectra seems to approximate the DFT spectra well.

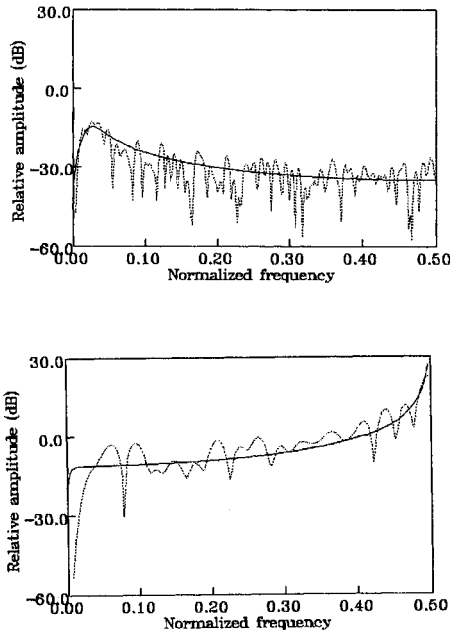
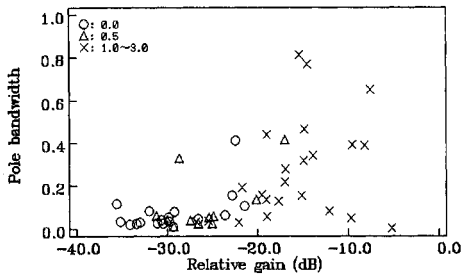


Fig.3 Examples of spectra of normalized period sequences.

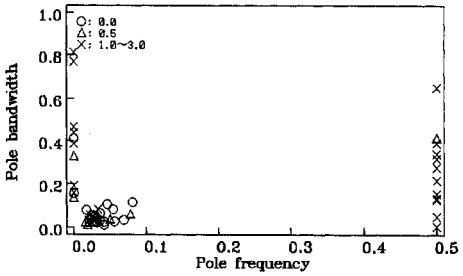
Suitability of the ARMA model to characterize jitter was tested with a spectral flatness measure[9].

The normalized period sequences $x(m)$ was measured from 52 Japanese vowel /e/ samples sustained by patients with laryngeal cancer, recurrent nerve paralysis and vocal fold polyp before and after the treatment. The flatness measure was computed from $e(m)$ of eq.(7), which was obtained by using the AR parameters estimated from $x(m)$. The average value of the flatness measure was 0.53 while the value of the flatness measure of a Gaussian white noise is Euler's constant of $0.5772\cdots$. We then can conclude that the ARMA model is satisfactory to represent the stochastic properties of the jitter sequence.

The relationship between the model parameters (gain, pole frequency and pole bandwidth) and perceptual scale (roughness) was investigated based on the voice samples. Figure 4 illustrates two scatter plots of the model parameters where circles, triangles and crosses are for no roughness, weak roughness and severe roughness, respectively.



(a) Gain and bandwidth.



(b) Pole frequency and bandwidth.

Fig. 4 Scatter plots of the model parameters for 52 voice samples.

Samples with relatively large gains ($G \geq$ -20dB) are almost rated as severely rough (Fig. 4 (a)). Therefore they are samples of high pole frequencies or large bandwidths (Fig. 4(b)). As a result, all the model parameters were strongly relative to the perceived quality.

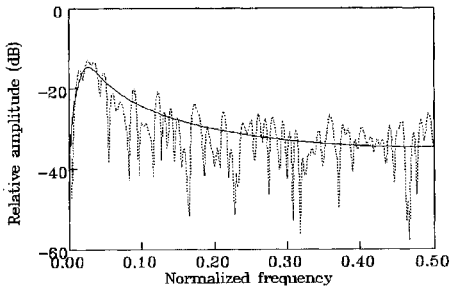
VI. DISTRIBUTIONAL CHARACTERISTICS OF THE PERTURBATION

We haven't so far paid much attention to the distributional characteristics of the normalized period sequence, by assuming only the distribution of the sequence to be normal. Pinto et. al. , however, have reported the examples in which the distribution seems abnormal[10]. In order to investigate the deviation of distributions from the normal distribution, we compute the skewness S and kurtosis K given by

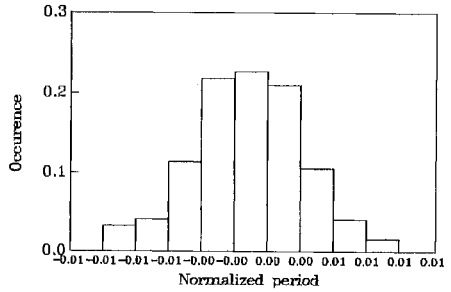
$$S = \frac{\frac{1}{M}\sum_{m=0}^{M-1} x^3(m)}{\sigma^3} \qquad K = \frac{\frac{1}{M}\sum_{m=0}^{M-1} x^4(m)}{\sigma^4} - 3, \quad (12)$$

where σ is the standard deviation of the sequence. S and K are measures on asymmetry and sharpness of the distribution, respectively, and are both zero if the distribution is ideally normal. In the 52 voice samples, 84% and 61% have the absolute values of the skewness and kurtosis less than 1, respectively. We may conclude that in most cases, the distributions are close to the normal. An example of the period sequence with the distribution close to the normal is

illustrated together with its two spectra in Fig. 5.



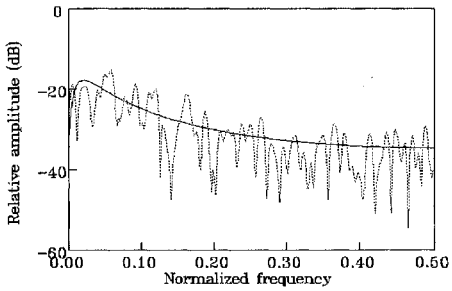
(a) Magnitude spectrum



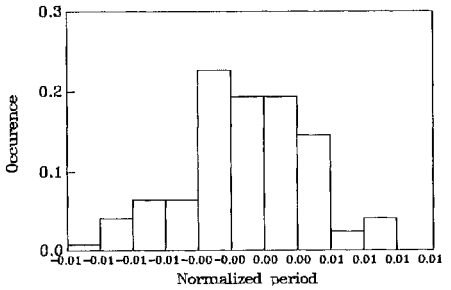
(b) Distribution

Fig. 5 An example of the normalized period sequence with the distribution close to the normal

This sequence has S=-0.36 and K=0.30. Figure 6 represents the distribution and two spectra of the sequence which has been synthesized with the ARMA model so as to simulate the spectral and statistical characteristics of the original sequence shown in Fig. 5.



(a) Magnitude spectrum



(b) Distribution

Fig. 6 The statistical distribution and spectra of the period sequence produced by the ARMA model.

The synthesized sequence seems to show the similar characteristics as those of original sequence.

VII. CONCLUSION

We proposed an analysis – conversion – synthesis system based on a stochastic perturbation model. The suitability of the model is validated by using a spectral flattness measure. We illustrated the statistical contribution of the perturbation model parameters to the perceived roughness quality for 52 pathologic voice samples. Although the distribution of our pathological voice data has been shown to be very close to the normal, the proposed ARMA model needs further improvement with respect to the statistical distributional characteristics. Experiments of the stochastic model parameters are now underway.

REFERENCES

[1] S. Imaizumi, "Acoustic measures of roughness in pathological voice," J. of Phonetics, Vol. 14, pp.457 – 462, 1986.

[2] O. Komuro and H. Kasuya, "Analysis of harmonic perturbation in pathologic vowel waveforms," JASJ, Vol. 47, No. , pp. 85–91, 1991 (in Japanese).

[3] L. Eskenazi and D. G. Childers, "Acoustic correlates of vocal quality," JSHR, pp. 298–306, 1990.

[4] Y. Endo and H. Kasuya, "Comparative study of several perturbation parameters in terms of their relationship with perceptual judgments," JJLP, Vol. 34, No. 4, pp. 338–341, 1993 (in Japanese) .

[5] J. Hillenbrand, "Perception of aperiodicities in synthetically generated voices," JASA, Vol. 83, No. 6, pp. 2361–2371.

[6] S. Imaizumi and J. Gauffin, "Acoustical and perceptual characteristics of pathological voices : rough, creak, fry and diplophonia," RILP, No. 25, pp. 109–119, 1991.

[7] J. Hirama and Y. Kakita, "Effects of F_0 fluctuation on the category of two pathological voice qualities : rough and voice tremor – when an FM model is used for fluctuation of F_0," JJLP, Vol. 33, No. 1, pp. 2–10, 1992 (in Japanese) .

[8] C. Yang and H. Kasuya, "A method to detect laryngeal noise from continuous speech," Proc. 14th ICA, Beijing, G2–10, 1992.

[9] J. D. Markel and A. H. Gray, "Linear prediction of speech," Springer – Verlag Berlin / Heidelberg / New York, 1980.

[10] N. B. Pinto and I. R. Titze, "Unification of perturbation measures in speech signals," JASA, Vol. 87, No. 3, pp. 1278–1289, 1990.