Analysis and perception of spectral 1/*f* characteristics of amplitude and period fluctuations in normal sustained vowels

Naofumi Aoki and Tohru Ifukube
Research Institute for Electronic Science, Hokkaido University, N12 W6 Sapporo, 060-0812 Japan

(Received 6 June 1998; revised 3 March 1999; accepted 29 march 1999)

Two kinds of fluctuations are always observed in the steady parts of normal sustained vowels. One is amplitude fluctuation, defined as the cyclic changes of maximum peak amplitudes. The other is period fluctuation, defined as the cyclic changes of pitch periods. The primary purpose of this paper is to present quantitative descriptions of amplitude and period sequences obtained from normal sustained vowels. These fluctuation sequences consisted of maximum peak amplitudes or pitch periods extracted successively from 512 consecutive pitch periods in the steady part. Results of the frequency analysis indicated that their frequency characteristics seemed to be subject to the spectral $1/f$ power law. In order to investigate the possibility that the frequency characteristics of the fluctuation sequences influence the voice quality of sustained vowels, psychoacoustic experiments were conducted. Amplitude and period sequences evaluated in the experiments were spectral $1/f^0$ (white noise), $1/f$, $1/f^2$, and $1/f^3$ sequences, respectively. The experimental results indicated that the subjective voice quality of synthesized sustained vowels could reflect the differences in the frequency characteristics of the fluctuation sequences. [S0001-4966(99)01507-6]

PACS numbers: 43.71.Bp, 43.70.Gr [WS]

INTRODUCTION

Normal sustained vowels always contain cyclic change of maximum peak amplitudes and pitch periods, even in their most steady parts where these values are considered to be quite stable (Lieberman, 1961; Hiki *et al.*, 1966; Hollien *et al.*, 1975; Horii, 1975). In this paper, they are labeled as amplitude fluctuation and period fluctuation, respectively. In addition, the sequences which consisted of maximum peak amplitudes or pitch periods extracted successively from the steady part of normal sustained vowels are labeled as amplitude sequence (AS) or period sequence (PS), respectively.

One of the major objectives of the research on these voice fluctuations is to examine their applicability for quantitative discrimination of pathological voices from normal cases. Since the characteristics of AS and PS obtained from sustained vowels produced by speakers with laryngeal pathology tend to deviate from those of normal sustained vowels, both the fluctuation sequences are considered to contain potentially significant information for diagnostic screening of pathological voices.

As is summarized in the literature (Pinto and Titze, 1990), a variety of measures has been devised for diagnostic screening. A number of the measures focus on the deviations of the successive values in AS and PS defined, respectively, as shimmer and jitter (Koike, 1973; Hollien *et al.*, 1975; Kitajima and Gould, 1976; Horii, 1979, 1980; Milenkovic, 1987; Titze *et al.*, 1987; Childers and Wu, 1990; Pinto and Titze, 1990; Scherer *et al.*, 1995; Bielamowicz *et al.*, 1996). Since the size of jitter and shimmer tends to be larger for pathological sustained vowels than for normal cases, these differences would give potentially useful information for the discrimination between pathological and normal sustained vowels (Lieberman, 1963; Hecker and Kreul, 1971; Koike, 1973; Kitajima and Gould, 1976; Kasuya *et al.*, 1983, 1986; Askenfelt and Hammarberg, 1986; Hillenbrand, 1987; Eskenazi *et al.*, 1990; Martin *et al.*, 1995).

On the other hand, several studies have pointed out that the differences in the dynamic characteristics of AS and PS could be a factor which causes the differences in voice quality between pathological and normal sustained vowels. The viewpoints of these studies are (1) temporal characteristics (Lieberman, 1961), (a) autocorrelation (Koike, 1969; von Ledan and Koike, 1970), (3) random fractal nature (Baken, 1990), (4) nonlinear dynamics (Herzel *et al.*, 1994), and (5) frequency characteristics of the fluctuation sequences (Endo and Kasuya, 1994).

The temporal characteristics of the fluctuation sequences obtained from normal sustained vowels are suggested to be not completely random processes. The current values in the fluctuation sequences are influenced by the previous values in the sequences (Lieberman, 1961). The randomness of the fluctuation sequences parametrized by the fractal dimension also suggested that they are not completely random (Baken, 1990).

In addition, the randomness of the fluctuation sequences was evaluated by visualizing their autocorrelation function (Koike, 1969; von Leden and Koike, 1970), attractor patterns in the phase space (Herzel *et al.*, 1994), and their frequency characteristics (Endo and Kasuya, 1994). These studies also indicated that the dynamic characteristics of the fluctuation sequences could be applicable for diagnostic screening, since the dynamic characteristics derived from pathological sustained vowels tended to be different from normal cases.

In order to establish the methodology of diagnostic

screening from the viewpoint of the dynamic characteristics of fluctuation sequences, further investigation is required. The relationship between the dynamic characteristics of the fluctuation sequences and the corresponding speech perception could be useful information to design measures for diagnostic screening.

Another major objective of the research on the voice fluctuations is the enhancement of synthesized sustained vowels (Childers and Hu, 1994; Klatt and Klatt, 1990; McCree and Barnwell, 1995). Synthesized sustained vowels would be perceived as buzzer-like without incorporating amplitude and period fluctuations. A number of studies have indicated that the size of the standard deviation of AS and PS significantly influences the voice quality of synthesized sustained vowels. It has been suggested that the buzzer-like voice quality can be improved, if the standard deviation of AS and PS incorporated into synthesized sustained vowels are optimized (Wendahl, 1963, 1966; Hiki *et al.*, 1966; Coleman, 1969, 1971; Rozsypal and Millar, 1979). On the other hand, it is also indicated that the large standard deviations of AS and PS are associated with rough voice quality as perceived in pathological cases. This finding could be consistent with studies which have indicated that the size of shimmer and jitter can be useful measures for the diagnostic screening of pathological voices from normal cases. The size of shimmer and jitter tends to be large as the size of the standard deviation of AS and PS becomes larger. The size of the standard deviation of AS and PS appears to be one of the important factors in enhancing the voice quality of synthesized sustained vowels.

Furthermore, it is also indicated that the dynamic characteristics of the sequences can be another significant factor to enhance the voice quality of synthesized sustained vowels. Several studies have suggested that the frequency characteristics of AS and PS would influence the voice quality (Kobayashi and Sekine, 1991; Komuro and Kasuya, 1991; Aoki and Ifukube, 1996). For the enhancement of the voice quality of synthesized sustained vowels, it seems potentially useful to investigate what features of the frequency characteristics of the fluctuation sequences contribute to the enhancement.

The primary purpose of the present study was to describe statistically the characteristics of AS and PS obtained from normal sustained vowels. Speech analysis was conducted to gain information which would be potentially useful for modeling the fluctuation sequences. The size of the standard deviation and the frequency characteristics of the fluctuation sequences were included in the speech analysis. In addition, psychoacoustic experiments investigated how the synthesized sustained vowels characterized by chosen frequency characteristics of AS and PS were perceived. The experiments aimed to examine whether or not the differences in their frequency characteristics were associated with subjective differences in the voice quality of synthesized sustained vowels.

I. SPEECH ANALYSIS

This section describes several statistical characteristics of AS and PS obtained from normal sustained vowels. The speech analysis included the investigation of the size of the standard deviation and the frequency characteristics of both fluctuation sequences. In addition, their stationarity, distribution, and the correlation between AS and PS were investigated to obtain the information for modeling the fluctuation sequences.

A. Speech samples

Ten male subjects between 22 and 26 years of age who did not suffer from any laryngeal disorders were selected in order to obtain normal sustained vowels. Each subject was requested to phonate the sustained vowel /a/ as steadily as possible in a soundproof anechoic room (Rion, audiometry room) toward an electret condenser microphone (Audiotechnica, AT822) at a distance of about 15 cm from the mouth. The sustained vowels were directly recorded onto a hard disk by way of a microphone mixer (Mackie, microseries 1202-VLZ), a low-pass filter (8th-order Bessel characteristic), and an analog-to-digital converter (Digidesign, audiomedia II). The sampling rate and quantization level were 44.1 kHz and 16 bits, respectively. The cutoff frequency of the low-pass filter was set to 5 kHz.

Speakers phonated the vowels at a pitch and loudness that was comfortable. The duration of the phonation was requested to be approximately 10 s. All records contained a steady portion of at least 512 pitch periods lasting over approximately 4 s, in which the mean pitch period was found to range from 7.6 to 9.1 ms. The calculated mean pitch period of all speech samples was 8.3 ms. The sound-pressure level (SPL) was also measured by a precision noise meter using the C weighting condition (Brüel & Kjær, type 2209), which was placed about 15 cm from the mouth. Measured SPL ranged from 80 to 86 dB for all subjects. The gain of the microphone mixer was adjusted for each subject for an optimal recording level. Twenty speech samples were taken per subject, since at least 15 speech samples were required to guarantee statistical significance in the fluctuation sequences of normal sustained vowels (Scherer *et al.*, 1995). Two hundred speech samples in total (20 utterances $\times$ 10 subjects) were obtained.

B. Extraction of AS and PS

Since updated values for each cycle of both fluctuations were required to form fluctuation sequences, each value of AS and PS was extracted from the digitized speech samples using a peak-picking method and a zero-crossing method, respectively (Hollien *et al.*, 1973; Hori, 1975, 1979, 1980; Titze *et al.*, 1987; Doherty and Shipp, 1988; Titze and Liang, 1993). The resolution of the extraction for AS was set to be better than a 0.1% accuracy level. It is reported that this accuracy level is guaranteed if the maximum peak amplitudes are represented by more than 9 bits (Titze *et al.*, 1987). This requirement was satisfied by adjusting proper recording levels in the sampling session. Maximum peak amplitudes were represented by more than 13 bits, so that this condition would be considered sufficient for satisfying the required accuracy level. In this study, a parabolic interpolation technique was also employed in order to improve the accuracy of the extraction (Titze *et al.*, 1987).

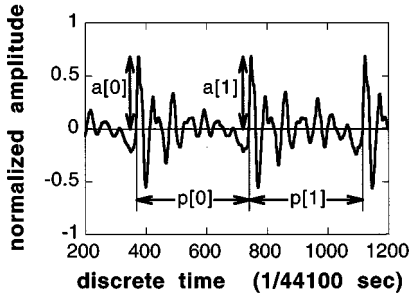

FIG. 1. Definition of amplitude sequence (AS) and period sequence (PS). Amplitudes of the speech sample are normalized to vary from −1 to 1, which corresponds to −32 768 to 32 767 (16 bits quantization level). From consecutive 512-pitch periods, the AS denoted by $a[n]$, $n=0,1,...,511$ and the PS denoted by $p[n]$, $n[0],1,...,511$ were successively extracted.

The resolution which is better than a 0.1% accuracy level requires an extremely high sampling rate for the extraction of PS (Horii, 1979; Milenkovic, 1987; Titze *et al.*, 1987; Titze and Liang, 1993). It is reported that more than 500 samples are required to represent each cycle in order to satisfy this accuracy level. According to this criterion, a sampling rate greater than 65.8 kHz is necessary if the pitch period is 7.6 ms, as it is our highest case. Although such a high sampling rate is desirable for accurate extraction of PS, it is also reported that the zero-crossing method with a linear interpolation can be a remedy to decrease the required sampling rate without sacrificing the accuracy (Titze *et al.*, 1987). In this study, the linear interpolation scheme was employed to prevent the degradation of the accuracy of the extraction caused by utilizing a 44.1-kHz sampling rate.

The extraction of fluctuation sequences was performed in the following order. First, the middle portion of each speech sample was extracted by using an editing program for acoustic signals (Digidesign, SOUND DESIGNER II) after visually and acoustically inspecting that particular unsteadiness was not detected in this portion. The extracted portion was referred to as a steady part. From the steady part, PS was then extracted successively from consecutive 512 pitch periods. AS was also extracted from the same 512 pitch periods. Figure 1 illustrates the definition of AS and PS, where the amplitudes of the speech sample were normalized to range from −1 to 1, which corresponds to −32 768 to 32 767, the quantization level of 16 bits. Figure 2 shows an example of a pair of AS and PS obtained from one of the speech samples.

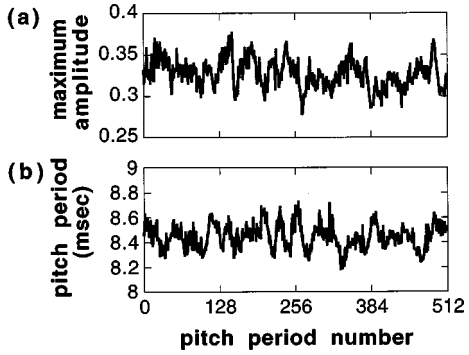

FIG. 2. Examles of (a) AS and (b) PS obtained from one of the speech samples. The AS shown in (a) is represented by using the same normalization as in Fig. 1.

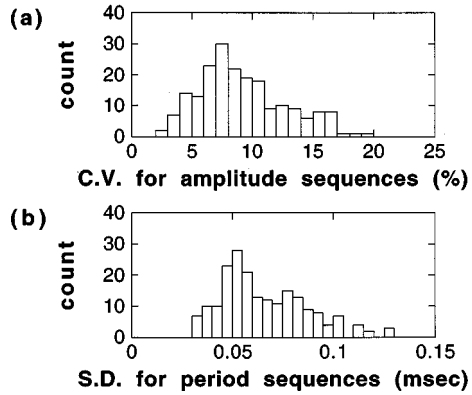

FIG. 3. Distributions of (a) the coefficient of variation (C.V.) of AS and (b) standard deviation (s.d.) of PS obtained from all speech samples.

C. Distributions of the coefficient of variation of AS and the standard deviation of PS

As mentioned earlier, the size of the standard deviation of AS and PS is considered to be a significantly important factor that influences the voice quality of sustained vowels. The standard deviations of both fluctuation sequences were statistically analyzed in order to investigate their valid size for normal sustained vowels.

Since the gain of the microphone mixer was adjusted depending on the loudness of the subject, the magnitudes of AS were represented by arbitrary units. Therefore, coefficient of variation (C.V.) was chosen as a measure for the size of AS (Pinto and Titze, 1990). C.V. is a measure which represents the standard deviation of a sequence normalized by the mean. The distribution of the C.V. of AS is shown in Fig. 3(a). It ranged from 2% to 20% and its mode was found at around 7.5%. The mode as within the normative range for the normal AS, which was reported to be 6.68%±3.03% (mean ± standard deviation %) (Scherer *et al.*, 1995).

Standard deviation (s.d.) itself was employed as a measure to compare the size of PS. As shown in Fig. 3(b), the s.d. of PS ranged from 0.03 to 0.13 ms and its mode was around 0.05 ms. In addition, the C.V. of PS was also calculated for the comparison with the normative range reported in the literature (Scherer *et al.*, 1995). The C.V. of PS ranged from 0.5% to 1.6% and its mode was found at around 0.8%. The mode was within the normative range 1.05±0.40% for the normal PS (Scherer *et al.*, 1995).

The correlation of the size of AS and PS, both of which were obtained from an identical speech sample, was also investigated for the development of the model of the fluctuation sequences. Figure 4 shows the scattergram plotted for C.V. of AS versus s.d. of PS. Although a moderate positive correlation coefficient ($r=0.64$) was obtained from the scattergram as average tendency (Bendat and Piersol, 1971), there was a variety of the combinations which obscures the meaning of the average correlation coefficient. For example, one of the cases showed that the C.V. of AS was small, while the s.d. of PS was large, and vice versa. Such deviations

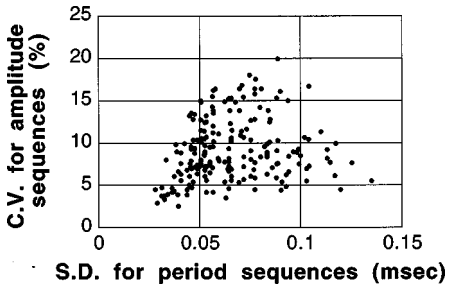

FIG. 4. The scattergram obtained from all speech samples shows the correlation of the size between AS and PS obtained from the identical speech sample. The average correlation coefficient is 0.64.

were found among individual speech samples, even from the same subject.

The size of fluctuation sequences investigated in terms of the C.V. for AS and the s.d. for PS would be useful information when we develop the model of the fluctuation sequences. Since it was difficult to make any meaningful conclusion that AS and PS were correlated or independent in their size, the size of PS for AS or AS for PS could be chosen rather arbitrarily for the preliminary model.

D. Stationarity of AS and PS

For developing the model of the fluctuation sequences, it is useful to examine whether the size of fluctuation sequences changes as the duration of the sequences changes. In order to clarify this issue, the stationarity of AS and PS was investigated. In this study, stationarity was defined as the time invariance of the mean and the variance of the sequence (Bendat and Piersol, 1971).

A runs test was employed in order to judge whether a fluctuation sequence was a stationary or nonstationary process. The test examined whether the changes of short-time mean and variance of a sequence were acceptable as those of a stationary process (Bendt and Piersol, 1971). Table I shows the results of the tests. The severity of the tests defined by the level of significance was chosen to be $\alpha=0.01$. The numbers of (A) AS and (B) PS for each subject in Table I represent the fluctuation sequences which were acceptable as stationary out of 20 sequences. Almost all of AS (92%) as well as PS (96%) were acceptable as stationary processes.

TABLE I. The results of the runs test which determined whether (A) AS and (B) PS were stationary or nonstationary processes. The level of significance was set to $\alpha=0.01$ for the test. The table represents the numbers of the fluctuation sequences acceptable as stationary processes out of the 20 sequences from each subject.

Subject	(A)	(B)
A.H.	16	19
M.Y.	18	18
I.M.	18	20
S.Y.	17	20
Y.T.	20	19
H.S.	17	20
M.I.	18	18
M.S.	20	19
S.S.	19	18
T.S.	20	20
total	183/200 (92%)	191/200 (96%)

TABLE II. The results of the *chi*-squared test which determined whether the distributions of (A) AS and (B) PS were Gaussian. The level of significance was set to $\alpha=0.01$ for the tests. The table represents the numbers of the sequences acceptable as Gaussian out of the 20 sequences from each subject.

Subject	(A)	(B)
A.H.	12	15
M.Y.	11	9
I.M.	16	12
S.Y.	15	10
Y.T.	12	16
H.S.	10	17
M.I.	11	8
M.S.	14	10
S.S.	9	7
T.S.	11	15
total	121/200 (61%)	119/200 (60%)

In conclusion, AS and PS extracted from the steady part of normal sustained vowels would be regarded as stationary processes. The size of fluctuation sequences would not change, even though their duration changed. This result shows one of the features of the fluctuation sequences to be taken into account for their model.

E. Distributions of AS and PS

The distribution of fluctuation sequences is one of the important features for developing their model. It has been reported that the distributions of AS and PS are seemingly regarded as Gaussian (Komuro and Kasuya, 1991; Aoki and Ifukube, 1996). This tendency was reexamined in this study.

Table II shows the results of the *chi*-squared test which examined whether the distribution of a sequence was classified as Gaussian (Bendat and Piersol, 1971). The numbers of (A) AS and (B) PS for each subject in Table II represent the fluctuation sequences which were acceptable as Gaussian out of 20 sequences. The level of significance was chosen to be $\alpha=0.01$ for the test.

The results indicated that the Gaussian distribution was considered to be one of the possible choices in modeling, since more than half of the distributions of AS (61%) and PS (60%) were acceptable as Gaussian. However, the results did not clearly confirm that the Gaussian distribution was always necessary for their model, since it appeared that a number of the distributions were not acceptable as Gaussian. Further investigation of their distribution in order to develop more precise models of the fluctuation sequences were left for future study.

F. Correlation between AS and PS

The correlation between AS and PS also influences the model of the fluctuation sequences. Correlation coefficients were calculated from all the pairs of AS and PS (Bendat and Piersol, 1971). Since individual tendency was not particularly different from subject to subject, pooled distribution of the correlation coefficients was obtained from all subjects.

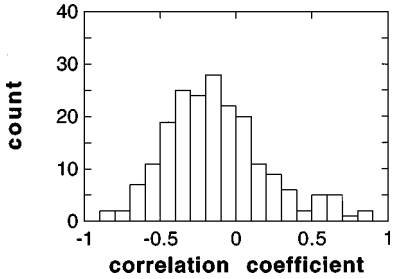

FIG. 5. Distribution of correlation coefficients between AS and PS obtained from the identical speech samples. The mean and the standard deviation of the distribution are −0.12 and 0.32, respectively.

The result is shown in Fig. 5. The mean and standard deviation of the distribution were −0.12 and 0.32, respectively. Since the correlation coefficients tended to center approximately at zero, both fluctuation sequences are likely to be modeled as processes as independent of each other.

G. Frequency characteristics of AS and PS

This study investigated the dynamic characteristics of the fluctuation sequences from the viewpoint of their frequency characteristics (Endo and Kasuya, 1994). The frequency characteristics of AS and PS were estimated using the 512-point fast Fourier transform (FFT) with Hamming window (Bendat and Piersol, 1971). As a result, it was found that the gross approximation of the frequency characteristics was subject to the spectral $1/f^{\beta}$ power law (Mandelbrot, 1977; Voss and Clarke, 1978; Keshner, 1982; Musha and Yamamoto, 1995), although the details might deviate from this approximation. This tendency was consistent among all fluctuation sequences.

Figure 6 shows examples of the frequency characteristics of AS and PS. The value of the exponent β was estimated by the least-squares line fitting (Bendat and Piersol, 1971). The value of β is equivalent to the gradient of the fitted line in the frequency characteristics represented in the log–log scale. The value of β of this example was 0.99 for AS and 0.96 for PS.

The average frequency characteristics of AS and PS are shown in Fig. 7. It was found that the mean value of β was

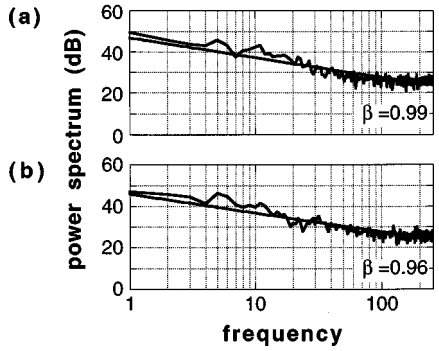

FIG. 6. Examples of the frequency characteristics of (a) AS and (b) PS obtained from a speech sample. The highest frequency is represented as 256 due to the 512-point FFT. The least-squares line fitting indicates that $\beta = 0.99$ for the AS and $\beta = 0.96$ for the PS in modeling these frequency characteristics by spectral $1/f^{\beta}$ power law.

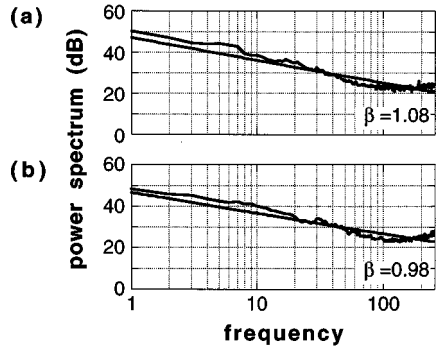

FIG. 7. Mean frequency characteristics of (a) AS and (b) PS obtained from all speech samples. As in the case of Fig. 6, the value of β is close to 1; $\beta = 1.08$ for the AS and $\beta = 0.98$ for the PS.

1.08 for AS and 0.98 for PS. The calculated standard deviation of β was 0.13 for AS and 0.17 for PS, respectively. Since the value of β tended to center approximately at 1, the results of frequency analysis would indicate that both fluctuation sequences can be modeled as spectral $1/f$ processes as a preliminary choice.

H. SUMMARY

The results of the speech analyses are summarized as follows. The size of the fluctuation sequences varied in the range of 2% to 20% in terms of C.V. for AS, for 0.03 to 0.13 ms in terms of s.d. for PS. The mode of C.V. for AS was 7.5%, and s.d. for PS was 0.05 ms. The correlation of the size between AS and PS was not particularly characterized by the average correlation coefficient. The results of the runs test indicated that both fluctuation sequences could be modeled as stationary processes. As a preliminary choice, the distribution of both fluctuation sequences can be modeled as Gaussian, since more than half of the distributions of AS and PS were considered to be Gaussian. Furthermore, AS and PS could be modeled as processes independent of each other, since AS and PS did not consistently show a strong correlation. Although the results of the frequency analysis suggested much finer models, both fluctuation sequences would be roughly modeled as spectral $1/f^{\beta}$ processes, where the β is equivalent to 1 as a gross approximation.

II. PSYCHOACOUSTIC EXPERIMENTS

As mentioned earlier, several studies have suggested that the frequency characteristics of AS and PS influence the voice quality of synthesized sustained vowels (Kobayashi and Sekine, 1991; Komuro and Kasuya, 1991; Aoki and Ifukube, 1996). In order to explore the influence of the frequency characteristics of fluctuation sequences on speech perception, a series of psychoacoustic experiments was conducted. The purpose of the experiments was to investigate how the differences in their frequency characteristics caused the subjective differences in the voice quality of synthesized sustained vowels.

A. Stimuli

Stimuli for the psychoacoustic experiments were sustained vowels /a/ produced by a partial autocorrelation

(PARCOR) synthesizer (Rabiner and Schafer, 1978; Kondoz, 1994). They were characterized by the different combinations of AS and PS.

The filter coefficients of the PARCOR synthesizer were derived from one of the speech samples whose AS and PS showed the representative characteristics of all fluctuation sequences investigated in this study. The C.V. of the AS was 7.5% and the s.d. of the PS was 0.05 ms. Both fluctuation sequences were acceptable as stationary processes. Their distributions were also acceptable as Gaussian. Since the correlation coefficient between the AS and PS was -0.06, they were considered not strongly correlated with each other. Their frequency characteristics were approximated by the spectral $1/f^{\beta}$ power law, in which β was 0.99 for the AS and 0.96 for the PS. Both frequency characteristics are, respectively, shown in Fig. 6. Furthermore, the mean pitch period calculated from the PS was 8.4 ms, which was close to the average of all speech samples. Both fluctuation sequences were employed in synthesizing stimuli.

The filter order of the PARCOR synthesizer was set to 40 for the condition of a 44.1-kHz sampling rate (Rabiner and Schafter, 1978; Kondoz, 1994). It was visually inspected that the frequency characteristics of the PARCOR filter in this condition appropriately represented four dominant formants below the cutoff frequency of the low-pass filter of 5 kHz. The coefficients of the PARCOR filter were not altered during the synthesis. This condition was based on the assumption that the characteristics of the vocal tract filter for normal sustained vowels do not substantially change during the phonation. In order to synthesize sustained vowels, impulse trains, which are conventionally used for synthesizing voiced speech, were employed as source signals of the PARCOR synthesizer (Rabiner and Schafer, 1978; Kondoz, 1994). Period fluctuations in the stimuli were implemented by the pulse-position modulation (Lathi, 1968). It employed PS as the modulating signals. In order to guarantee the accuracy of the modulation at a 44.1-kHz sampling rate, impulse trains represented in the analog form were passed through a -6-dB/oct low-pass filter and sampled at a 44.1-kHz sampling rate. This low-pass filter theoretically consisted of both -12-dB/oct glottal characteristics and $+6$-dB/oct radiation characteristics from the mouth (Rabiner and Schafer, 1978; Kondoz, 1994). After synthesizing sustained vowels by the PARCOR filter, cyclic gain adjustments defined by AS were employed to incorporate amplitude fluctuations.

Each stimulus consisted of 128 pitch periods. Since the mean pitch period was set to 8.4 ms, the duration of each stimulus was approximately 1 s. A linearly increasing or decreasing gate function whose duration was 10 ms was employed at the beginning and the end of each stimulus in order to prevent click-like sounds.

The psychoacoustic experiments investigated four different conditions in regard to AS and PS. In conditions 1 and 2, the frequency characteristics of AS were manipulated, while all of the stimuli employed the PS obtained from the speech sample. On the other hand, PS was changed from stimulus to stimulus in conditions 3 and 4, while all of the stimuli employed the AS obtained from the speech sample.

TABLE III. Fluctuation sequences characterizing the stimuli.

	Stimulus	Conditions 1 and 2	Conditions 3 and 4
	a	AS (speech sample)	PS (speech sample)
	b	no AS	no PS
	c		
$\beta 0$	*d*	AS ($1/f^0$ sequences)	PS ($1/f^0$ sequences)
	e		
	f		
$\beta 1$	*g*	AS ($1/f$ sequences)	PS ($1/f$ sequences)
	h		
	i		
$\beta 2$	*j*	AS ($1/f^2$ sequences)	PS ($1/f^2$ sequences)
	k		
	l		
$\beta 3$	*m*	AS ($1/f^3$ sequences)	PS ($1/f^3$ sequences)
	n		

Thus, conditions 1 and 2 focused on how fluctuations in AS influenced perception, while conditions 3 and 4 focused on how perception was influenced by the different frequency characteristics of PS.

Fourteen stimuli labeled from ''*a*'' to ''*n*'' were produced for each condition. Fluctuation sequences employed to characterize the stimuli are summarized in Table III. Stimulus ''*a*'' employed the AS and the PS obtained from the speech sample. Although stimulus ''*a*'' was not the speech sample itself, its voice quality was considered to reflect the characteristics of the AS and the PS of the speech sample. Since stimulus ''*a*'' was used as a reference stimulus in evaluating all stimuli including stimulus ''*a*'' itself, it was also labeled the reference stimulus. Comparisons between the reference stimulus and stimulus ''*a*'' were the control for the experiment. Stimulus ''*b*'' was produced without amplitude or period fluctuation. This stimulus was aimed to examine whether amplitude or period fluctuation was an important factor for speech perception. In addition, four stimulus groups, each of which consisted of three stimuli, were produced. The three stimuli of each stimulus group employed AS or PS whose frequency characteristics were classified in the same category, while the fluctuatiaon sequences themselves were different from each other, since randomization used for producing the fluctuation sequences was different. The four stimulus groups were labeled as $\beta 0$, $\beta 1$, $\beta 2$, and $\beta 3$ according to the frequency characteristics of the fluctuation sequences. Stimulus group $\beta 0$, which consisted of stimulus ''*c*,'' ''*d*,'' and ''*e*,'' employed spectral $1/f^0$ sequences (white noise). Stimulus group $\beta 1$, which consisted of stimulus ''*f*,'' ''*g*,'' and ''*h*,'' employed spectral $1/f$ sequences. Stimulus group $\beta 2$, which consisted of stimulus ''*i*,'' ''*j*,'' and ''*k*,'' employed spectral $1/f^2$ sequences. Stimulus group $\beta 3$, which consisted of stimulus ''*l*,'' ''*m*,'' and ''*n*,'' employed spectral $1/f^3$ sequences.

The value of the exponent β in the spectral $1/f^{\beta}$ sequences for each stimulus group was considered to be a rather preliminary choice. The integer values from 0 to 3 were employed, since there was no *a priori* knowledge about the relationship between speech perception and the values of β. This condition also aimed to examine whether or not the perceptual effects were categorized by the values of β.

The C.V. of all AS was set to 7.5% for condition 1 and 15% for condition 2, in which the C.V. of the AS obtained from the speech sample was also readjusted. Condition 2 aimed to examine whether or not the larger C.V. of AS influenced the experimental result compared with condition 1. The s.d. of PS was set to 0.05 ms for all stimuli throughout conditions 1 and 2.

On the other hand, the s.d. of PS was set to 0.05 ms for condition 3 and 0.10 ms for condition 4. Condition 4 aimed to examine whether or not the larger s.d. of PS influenced the experimental result compared with condition 3. The C.V. of AS was set to 7.5% for all stimuli throughout conditions 3 and 4.

All of AS and PS employed in the stimulus groups were fractional Brownian motions produced by the FFT method (Saupe, 1988; Voss, 1989). Gaussian white noise was first transformed to the frequency domain, then passed through the low-pass filter characterized by the spectral $1/f^{\beta}$ power law. The result was transformed back into the time domain. Although the speech analysis indicated that the distributions of AS and PS were not necessarily Gaussian, the fluctuation sequences employed in this study were simply assumed as Gaussian.

The power spectrum of a spectral $1/f^{\beta}$ sequence is represented as

$$S_v(f)=|T(f)|^2 S_w(f)\propto|T(f)|^2, \tag{1}$$

where $S_v(f)$ is the power spectrum of a spectral $1/f^{\beta}$ sequence $v(t)$, $T(f)$ is the frequency characteristics of a spectral $1/f^{\beta}$ filter, and $S_w(f)$ is the power spectrum of Gaussian white noise.

Thus, the spectral $1/f^{\beta}$ filter is required to be

$$|T(f)|=1/f^{\beta/2}. \tag{2}$$

The typical sequences produced by this method are shown in Fig. 8. These are the examples of spectral $1/f^0$ (white noise), $1/f$, $1/f^2$, and $1/f^3$ sequences, respectively. The smoothness of the sequences increase as the value of β increases. Simultaneously, the sequences prove to be nonstationary process. These are attributable to the dominance of the low-frequency components in the sequences for a larger β.

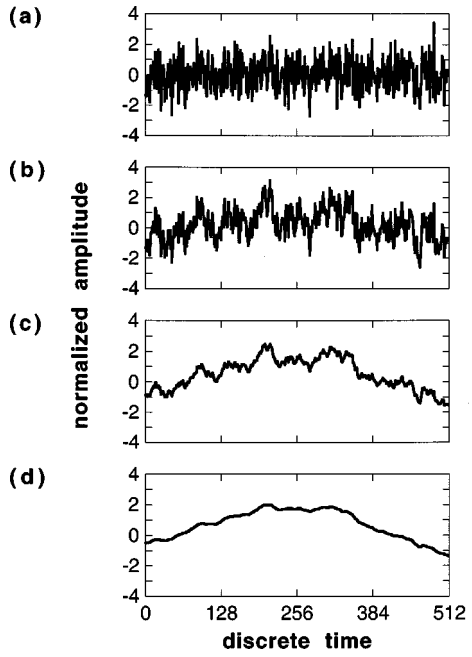

FIG. 8. Variations of $1/f^{\beta}$ sequence. The values of the exponent β are (a) zero, (b) 1, (c) 2, and (d) 3. The mean and the standard deviation are normalized to be zero and 1.

B. Subjects

Twenty subjects consisting of 12 males and eight females participated in the experiment. Their age ranged from 20 to 26 years. None of them was experienced in psychoacoustic experiments. All reported having no hearing problems.

C. Procedures

All stimuli were synthesized using a personal computer (Apple, Macintosh Quadra 800). The stimuli were passed through a digital-to-analog converter (Digidesign, audiomedia II) and then low-pass filtered (8th-order Bessel characteristic). The sampling rate and quantization level were 44.1 kHz and 16 bits, respectively. The cutoff frequency of the low-pass filter was set to be 10 kHz. The stimuli were presented through a monitor speaker (Denon, USC-101) which was attached to a premain amplifier (Sansui, AU-α 507XR). The speaker was placed 1.5 m in front of the subject in a soundproof anechoic room (Rion, audiometry room). The SPL of the stimuli was set to 65 dB upon presentation.

Each subject took part individually in the experiment of all four conditions. Each condition consisted of 14 paired-comparison trials to evaluate the similarity between the preceding stimulus *A* and the succeeding stimulus *B*. The stimulus *A* was the reference stimulus. The stimulus *B* was one of the 14 stimuli produced for each condition. The order of the presentation in regard to stimulus *B* was randomized.

The stimulus *A* and *B* were presented to the subject twice in the form of an *AB* pair. There was a 1-s silent interval between stimulus *A* and *B*, and a 2-s silent interval between the first and second *AB* pair. For the judgment, a 6-s interval was given to the subject after listening to the two *AB* pairs.

The subject was asked to judge whether or not the voice quality of stimulus *B* was perceived as the same as that of stimulus *A*. They were forced to select one of the following three choices, (1) ''same,'' (2) ''undecided,'' or (3) ''different.'' The three-point scale aimed to examine whether or not the subject could correctly distinguish the voice quality between a stimulus of artificially produced fluctuation sequences and a stimulus of the fluctuation sequences obtained from the speech sample. The five- or seven-point scale, which is conventionally used to grade the differences in the voice quality, was not employed (Kreiman *et al.*, 1993). In order to compare the experimental results, the above three choices were translated into the numerical measured called similarity, which was defined as (1) 100%, (2) 50%, and (3) 0% corresponding to the three choices (Ifukube *et al.*, 1991).

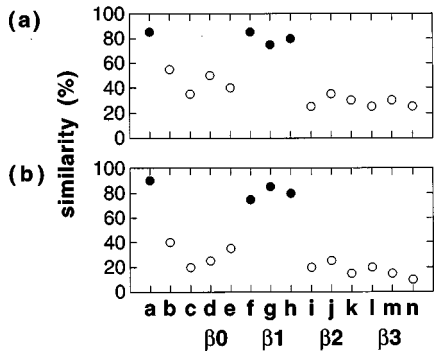

FIG. 9. The similarity of each stimulus is summarized in (a) for condition 1 and (b) for condition 2. Filled circles represent the results of χ^2 test that examined whether or not the similarity of each stimulus was acceptable as being in the same category of the control stimulus *a*. All stimuli of the stimulus group $\beta 1$ were acceptable as in the same category of the stimulus *a*. The level of significance for the test was set to $\alpha = 0.01$.

D. Results of practice trials

Prior to the experiment of four conditions, each subject took part in the practice stage consisting of 12 paired-comparison trials. Half of the 12 trials were practice runs for conditions 1 and 2. The other half of the trials were for conditions 3 and 4. Each half included a comparison between the reference stimulus and the control stimulus ''*a*'' that was the same as the reference stimulus itself. Since they were the same stimuli, responses for this comparisons were expected to be same. The errors with regard to this comparison occurred with five subjects as to AS and three subjects as to PS out of 20 subjects. This result suggested that most of the subjects could correctly detect the sameness between the same stimuli. In addition, most of the subjects judged these as the same even though they were not familiar with such kind of psychoacoustic comparison tests. The subjects were not informed of the correctness of their responses. Any criteria which might influence their judgments were not given to the subjects before or after the practice trials.

E. Results of conditions 1 and 2

The results of conditions 1 and 2 are summarized in Fig. 9(a) and (b), respectively. The similarity of each stimulus is represented by either an open or filled circle as the average of the results over all subjects.

It appeared that the control stimulus *a* and the stimulus group $\beta 1$ tended to be evaluated as more similar to the reference stimulus rather than the other stimuli throughout conditions 1 and 2. These results indicated that most of the subjects could not distinguish the voice quality of the stimulus group $\beta 1$ from that of the reference stimulus. The larger C.V. of AS examined in condition 2 did not substantially influence this tendency.

As for the other stimuli, most of the subjects reported that the voice quality of the stimulus group $\beta 0$ was rougher than that of the reference stimulus. Particular changes in the loudness were perceived in the stimulus group $\beta 2$ and $\beta 3$, while such features were not perceived in the reference stimulus. Some of the subjects also reported that the loudness changes of stimulus *b* were perceived as flat compared with the reference stimulus. These differences could be a subjective clue for the discrimination between the reference stimulus and these stimuli.

In order to evaluate the experimental results objectively, the independence between the control stimulus *a* and the other stimuli with regard to the response distributions in the three-point scale was examined by the *chi*-squared test (Bendat and Piersol, 1971). The level of significance for the test was chosen to be $\alpha = 0.01$.

The filled circles shown in Fig. 9 represent the stimuli whose response distributions were acceptable as in the same category as that of the stimulus *a*. Since all the stimuli of the stimulus group $\beta 1$ were acceptable as the same throughout conditions 1 and 2, it can be indicated that the voice quality of the stimulus group $\beta 1$ could be perceived as the same as the stimulus *a*, namely the reference stimulus.

F. Results of conditions 3 and 4

The results of conditions 3 and 4 are summarized in Fig. 10(a) and (b), respectively. It appeared that the control stimulus *a* and the stimulus group $\beta 1$ tended to be evaluated as more similar to the reference stimulus rather than the other stimuli throughout conditions 3 and 4. These results indicated that most of the subjects could not distinguish the voice quality of the stimulus group $\beta 1$ from that of the reference stimulus. The larger s.d. of PS examined in condition 4 did not substantially influence this tendency.

As for the other stimuli, most of the subjects reported that the voice quality of the stimulus group $\beta 0$ was rougher than that of the reference stimulus. Unstable changes in the pitch were perceived in the stimulus group $\beta 2$ and $\beta 3$, while such features were not perceived in the reference stimulus. Furthermore, most of the subjects reported that the stimulus *b* was perceived as buzzer-like compared with the reference stimulus. These differences could be a subjective clue for the discrimination between the reference stimulus and these stimuli.

The filled circles shown in Fig. 10 represent the stimuli whose response distributions in the three-point scale were acceptable as in the same category as that of the control stimulus *a* in terms of the *chi*-squared test (Bendat and Piersol, 1971). The level of significance for the test was chosen

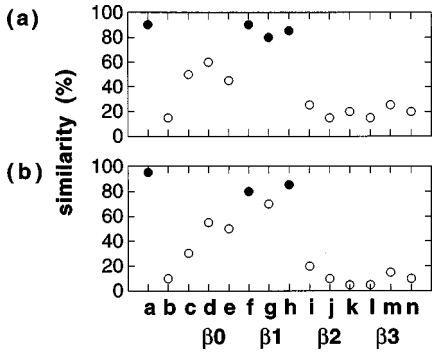

FIG. 10. The similarity of each stimulus is summarized in (a) for condition 3 and (b) for condition 4. Filled circles represent the results of χ^2 test. Most of the stimuli of the stimulus group $\beta 1$ (five cases out of six) are acceptable as in the same category of the control stimulus *a*. The level of significance for the test was set to $\alpha = 0.01$.

to be $\alpha=0.01$. Since five cases out of six were acceptable as the same throughout conditions 3 and 4, it can be indicated that the voice quality of the stimulus group $\beta 1$ could be classified as in almost the same category of the stimulus *a*, namely the reference stimulus.

G. SUMMARY

In spite of the different conditions, the stimulus group $\beta 1$ and the control stimulus *a* tended to be evaluated as similar to the reference stimulus throughout the four conditions. On the other hand, the other stimulus groups and stimulus *b* were not evaluated as being similar.

Since the similarity of stimuli *a* and *b* tended to be judged as low, it can be concluded that AS and PS play significant roles in the speech perception of sustained vowels, as mentioned in the previous literature (Wendahl, 1963, 1966; Hiki *et al.*, 1966; Coleman, 1969, 1971; Rozsypal and Millar, 1979). The results of the other stimuli suggest that the differences in the frequency characteristics of AS and PS can significantly influence speech perception. High similarity between the stimulus group $\beta 1$ and the reference stimulus was obtained. These results could be attributable to the similarity of the frequency characteristics of the fluctuation sequences between stimulus group $\beta 1$ and the reference stimulus.

Since there were no large differences in the similarity among all the stimuli of the stimulus group $\beta 1$, the randomization for producing the fluctuation sequences could have little effect on the speech perception. In addition, it seemed that the similarity of the stimulus group $\beta 1$ could not be significantly influenced by the differences in the size of the fluctuation sequences. The similarity of the stimulus group $\beta 1$ was consistently evaluated high throughout conditions 1 and 2 or conditions 3 and 4.

III. DISCUSSION

The results of the psychoacoustic experiments suggest that the frequency characteristics of fluctuation sequences could be a significant factor which influences the voice quality of sustained vowels. Compared with other stimulus groups, the stimuli which were characterized by spectral $1/f$ sequences tended to be evaluated as more similar to the reference stimuli. This result could be attributable to the fact that the frequency characteristics of the spectral $1/f$ sequences were close to those of the AS or the PS employed in the reference stimuli.

The spectral $1/f$ power law is often observed in a variety of natural phenomena, including fluctuation sequences obtained from biomedical signals, such as heart-rate fluctuation (Mandelbrot, 1977; Voss and Clarke, 1978; Keshner, 1982; Musha and Yamamoto, 1995). Although the spectral $1/f^{\beta}$ sequences are generally classified as nonstationary processes with the condition of $1<\beta$, the quasistationarity is observed in spectral $1/f$ sequences (Mandelbrot, 1977; Voss and Clarke, 1978; Keshner, 1982; Saupe, 1988; Voss, 1989; Wornell, 1996). The mean and the mean-square value of spectral $1/f^{\beta}$ sequences are subject to the following relationship with the condition of $1<\beta<3$ and $\beta=2H+1$ (Saupe, 1988; Voss, 1989; Wornell, 1996):

$$\langle v(rt)\rangle = r^{H}\langle v(t)\rangle, \qquad \langle v^{2}(rt)\rangle = r^{2H}\langle v^{2}(t)\rangle, \tag{3}$$

where $v(t)$ is a spectral $1/f^{\beta}$ sequence, r is the resolution factor of the sequence, H is the Hurst exponent, and $\langle\cdot\rangle$ is the expectation operator.

Since the mean and the variance derived from Eq. (3) are considered to change as the resolution factor r changes, $1/f^{\beta}$ fluctuations are classified as nonstationary processes with the condition of $1<\beta$. However, the mean and the variance of $1/f$ sequences, which are derived from Eq. (3) with the condition of $\beta\rightarrow 1$, namely $H\rightarrow 0$, are statistically invariance even though the resolution factor r changes. This nature of spectral $1/f$ sequences is known as self-similarity, which guarantees the quasistationarity of the sequences (Mandelbrot, 1977; Voss and Clarke, 1978; Keshner, 1982; Saupe, 1988; Voss, 1989; Wornell, 1996).

Taking these characterisics into consideration, AS and PS modeled as either spectral $1/f^{2}$ or $1/f^{3}$ sequences are classified as nonstationary processes. The typical examples of spectral $1/f^{2}$ and $1/f^{3}$ sequences shown in Fig. 8(c) and (d) exemplify their nonstationarity. Compared with spectral $1/f^{0}$ and $1/f$ sequences shown in Fig. 8(a) and (b), nonstationary changes in the short-time mean of spectral $1/f^{2}$ and $1/f^{3}$ sequences are easily detected.

The nonstationary changes in the loudness and the pitch by spectral $1/f^{2}$ or $1/f^{3}$ sequences are considered to have caused subjective differences in the voice quality from the reference stimulus which was characterized by stationary AS and PS. Taking account of the result of the speech analysis in which almost all fluctuation sequences would be acceptable as stationary processes, spectral $1/f^{2}$ and $1/f^{3}$ sequences could not be appropriate models for AS and PS of normal sustained vowels.

Compared with these nonstationary sequences, spectral $1/f$ sequences are considered to be more appropriate models for AS and PS due to their quasistationarity. The psychoacoustic experiment also subjectively supported the validity of the models. Considering the result of the speech analysis in which the fluctuation sequences were suggested to be approximated as spectral $1/f$ processes, it could be concluded that the models of spectral $1/f$ sequences are potentially useful choices for AS and PS of normal sustained vowels.

In this study, gross approximation of the frequency characteristics of fluctuation sequences was its only focus. However, the result of the frequency analysis suggested the possibility of much finer models. For example, it was found that the frequency characteristics in the high-frequency region tended to elevate greater than the spectral $1/f$ power law, as shown in Figs. 6 and 7. It will be of interest to investigate how finer models influence speech perception. In addition, it will also be of interest to examine how the perceptual differences are caused by the value of β as it gradually changes

from 1. These issues are currently being investigated by the authors for developing more detailed models of the fluctuation sequences.

Some previous studies also developed the model of PS for normal sustained vowels from the viewpoint of its frequency characteristics (Kobayashi and Sekine, 1991; Komuro and Kasuya, 1991). These models have represnted the frequency characteristics of PS by autoregression (AR) (Kobayashi and Sekine, 1991) or autoregressive moving-average (ARMA) (Komuro and Kasusya, 1991) forms of digital filter. Similarly to the results presented in this paper, these studies also indicated that the frequency characteristics of PS would be a key factor for the voice quality of sustained vowels. The gradual decreasing characteristics found in the high-frequency region were suggested to be a feature of PS which could be related to the voice quality of normal sustained vowels.

The decreasing characteristics of AS as well as PS were also pointed out by other previous studies (Kakita *et al.*, 1986; Hirama and Kakita, 1989; Titze and Liang, 1993; Endo and Kasuya, 1994). A speech analysis of normal sustained vowels showed decreases of high-frequency components of both AS and PS (Kakita *et al.*, 1986; Hirama and Kakita, 1989). These fluctuation sequences were extracted from three Japanese vowels /a/, /i/, and /u/ phonated by one adult male speaker. Such decreasing characteristics were also suggested by the speech analysis of PS obtained from normal sustained vowels /a/ (Titze and Liang, 1993). It is also reported that the decreasing characteristics of AS and PS might be one of the features of normal sustained vowels compared with pathological cases (Hirama and Kakita, 1989; Endo and Kasuya, 1994).

From these two viewpoints: (1) speech perception caused by fluctuation sequences, and (2) speech analysis of fluctuation sequences, it might be suggested that the decreasing characteristics are considered to be one of the features of the fluctuation sequences obtained from normal sustained vowels. Such decreasing characteristics could be a significant factor for the voice quality of normal sustained vowels.

IV. CONCLUSIONS

The present study statistically showed several aspects of AS and PS of normal sustained vowels. The speech analysis indicated that both fluctuation sequences would be approximated as spectral $1/f$ sequences in terms of their frequency characteristics.

In addition, the psychoacoustic experiments indicated that voice quality of sustained vowels appeared to be influenced by the frequency characteristics of the fluctuation sequences. The results of the present study are considered to provide potentially useful information for exploring the speech perception caused by fluctuation sequences and for developing their appropriate models for enhancing the voice quality of synthesized sustained vowels.

ACKNOWLEDGMENTS

We thank Dr. Strange at the University of South Florida and three anonymous reviewers for their useful comments. We also thank Dr. Katagiri and Dr. Pruitt at ATR, Japan and Dr. Takaya at the University of Saskatchewan, Canada for their useful advice in revising this paper.

Aoki, N., and Ifukube, T. (**1996**). ''Two $1/f$ fluctuations in sustained phonation and their roles on naturalness of synthetic voice,'' in *Proceedings ICECS 96* (IEEE, Rodos), pp. 311–314.

Askenfelt, A. G., and Hammarberg, B. (**1986**). ''Speech waveform perturbation analysis: A perceptual-acoustic comparison of seven measures.'' J. Speech Hear. Res. **29**, 50–64.

Baken, R. J. (**1990**). ''Irregularity of vocal period and amplitude: A first approach to the fractal analysis of voice,'' J. Voice **4**, 185–197.

Bendat, J. S., and Piersol, A. G. (**1971**). *Random data: Analysis and Measurement Procedures* (Wiley, New York).

Bielamowicz, S., Kreiman, J., Gerratt, B. R., Dauer, M. S., and Berke, G. S. (**1996**). ''Comparison of voice analysis systems for perturbation measurement,'' J. Speech Hear. Res. **39**, 126–134.

Childers, D. G., and Wu, K. (**1990**). ''Quality of speech produced by analysis-synthesis,'' Speech Commun. **9**, 97–117.

Childers, D. G., and Hu, H. T. (**1994**). ''Speech synthesis by glottal excited linear prediction,'' J. Acoust. Soc. Am. **96**, 2026–2036.

Coleman, R. F. (**1969**). ''Effect of median frequency levels upon the roughness of jittered stimuli,'' J. Speech Hear. Res. **12**, 330–336.

Coleman, R. F. (**1971**). ''Effect of waveform changes upon roughness perception,'' Folia Phoniatr. **23**, 314–322.

Doherty, E. T., and Shipp, T. (**1988**). ''Tape recorder effects on jitter and shimmer extraction,'' J. Speech Hear. Res. **31**, 485–490.

Endo, Y., and Kasuya, H. (**1994**). ''Spectral analysis of fundamental period perturbation and its modeling,'' Jpn. J. Logop. Phoniatr. **35**, 193–198 (in Japanese).

Eskenazi, L., Childers, D. G., and Hicks, D. M. (**1990**). ''Acoustic correlates of vocal quality,'' J. Speech Hear. Res. **33**, 298–306.

Hecker, M. H. L., and Kruel, E. J. (**1971**). ''Descriptions of the speech of patients with cancer of the vocal folds. Part I: Measures of fundamental frequency,'' J. Acoust. Soc. Am. **49**, 1275–1282.

Herzel, H., Berry, D., Titze, I. R., and Saleh, M. (**1994**). ''Analysis of vocal disorders with methods from nonlinear dynamics,'' J. Speech Hear. Res. **37**, 1008–1019.

Hiki, S., Sugawara, K., and Oizumi, J. (**1966**). ''On the rapid fluctuation of voice pitch,'' J. Acoust. Soc. Jpn. **22**, 290–291.

Hillenbrand, J. (**1987**). ''A methodological study of perturbation and additive noise in synthetically generated voice signals,'' J. Speech Hear. Res. **30**, 448–461.

Hirama, J., and Kakita, Y. (**1989**). ''Characteristics of a pathological rough voice based on the power spectrum of fluctuations,'' Jpn. J. Logop. Phoniatr. **30**, 225–230 (in Japanese).

Hollien, H., Michel, J., and Doherty, E. T. (**1973**). ''A method for analyzing vocal jitter in sustained phonation,'' J. Phonetics **1**, 85–91.

Horii, Y. (**1975**). ''Some statistical characteristics of voice fundamental frequency,'' J. Speech Hear. Res. **18**, 192–201.

Horii, Y. (**1979**). ''Fundamental frequency perturbation observed in sustained phonation,'' J. Speech Hear. Res. **22**, 5–19.

Horii, Y. (**1980**). ''Vocal shimmer in sustained phonation,'' J. Speech Hear. Res. **23**, 202–209.

Ifukube, T., Hashiba, M., and Matsushima, J. (**1991**). ''A role of 'waveform fluctuation' on the naturality of vowels,'' J. Acoust. Soc. Jpn. **47**, 903–910 (in Japanese).

Kakita, Y., Hirama, J., Hamatani, K., Ohtani, M., and Suzuki, M. (**1986**). ''Fluctuation of the fundamental frequency and the maximum amplitude,'' Trans. Comm. Speech Res., Acoustic. Soc. Jpn. **S85–103**, 805–812 (in Japanese).

Kasuya, H., Kobayashi, Y., and Kobayashi, T. (**1983**). ''Characteristics of pitch period and amplitude perturbations in pathologic voice,'' in *Proceedings ICASSP 83* (IEEE, Boston), pp. 1372–1375.

Kasuya, H., Ogawa, S., and Kikuchi, Y. (**1986**). ''An acoustic analysis of pathological voice and its application to the evaluation of laryngeal pathology,'' Speech Commun. **5**, 171–181.

Keshner, M. S. (**1982**). ''$1/f$ noise,'' Proc. IEEE **70**, 212–218.

Kitajima, K., and Gould, W. J. (**1976**). ''Vocal shimmer in sustained phonation of normal and pathologic voice,'' Ann. Otol. Rhinol. Laryngol. **85**, 377–381.

Klatt, D. H., and Klatt, L. C. (**1990**). ''Analysis, synthesis, and perception of

voice quality variations among female and male talkers,'' J. Acoust. Soc. Am. **87**, 820–857.
Kobayashi, T., and Sekine, H. (**1991**). ''The role of fluctuations in fundamental period for natural speech synthesis,'' J. Acoust. Soc. Jpn. **47**, 539–544 (in Japanese).
Koike, Y. (**1969**). ''Vowel amplitude modulations in patients with laryngeal diseases,'' J. Acoust. Soc. Am. **45**, 839–844.
Koike, Y. (**1973**). ''Application of some acoustic measures for the evaluation of laryngeal dysfunction,'' Stud. Phonol. **7**, 17–23.
Komuro, O., and Kasuya, H. (**1991**). ''Characteristics of fundamental period variation and its modeling,'' J. Acoust. Soc. Jpn. **47**, 928–934 (in Japanese).
Kondoz, A. M. (**1994**). *Digital Speech* (Wiley, New York).
Kreiman, J., Gerratt, B. R., Kempster, G. B., Erman, A., and Berke, G. S. (**1993**). ''Perceptual evaluation of voice quality: Review, tutorial, and a framework for future research,'' J. Speech Hear. Res. **36**, 21–40.
Lathi, B. P. (**1968**). *Communication Systems* (Wiley, New York).
Lieberman, P. (**1961**). ''Perturbations in vocal pitch,'' J. Acoust. Soc. Am. **33**, 597–603.
Lieberman, P. (**1963**). ''Some acoustic measures of the fundamental periodicity of normal and pathologic larynges,'' J. Acoust. Soc. Am. **35**, 344–353.
Mandelbrot, B. B. (**1977**). *The Fractal Geometry of Nature* (Freeman, New York).
Martin, D., Fitch, J., and Wolfe, V. (**1995**). ''Pathologic voice type and the acoustic prediction of severity,'' J. Speech Hear. Res. **38**, 765–771.
McCree, A. V., and Barnwell III, T. P. (**1995**). ''A mixed excitation LPC vocoder model for low bit rate speech coding,'' IEEE Trans. Speech Audio Process. **3**, 242–250.
Milenkovic, P. (**1987**). ''Least mean square measures of voice perturbation,'' J. Speech Hear. Res. **30**, 529–538.
Musha, T., and Yamamoto, M. (**1995**). ''1/f-like fluctuations of biological rhythm'' Noise Physical Syst., 22–31.
Pinto, N. B., and Titze, I. R. (**1990**). ''Unification of perturbation measures in speech signals,'' J. Acoust. Soc. Am. **87**, 1278–1289..
Rabiner, L. R., and Schafer, R. W. (**1978**). *Digital Processing of Speech Signals* (Prentice-Hall, Englewood Cliffs, NJ).
Rozsypal, A. J., and Miller, B. F. (**1979**). ''Perception of jitter and shimmer in synthetic vowels,'' J. Phonetics **7**, 343–355.
Saupe, D. (**1988**). ''Algorithms for random fractals,'' in *The Science of Fractal Images*, edited by H.-O. Peitgen and D. Saupe (Springer, New York), pp. 71–136.
Scherer, R. C., Vail, V. J., and Guo, C. G. (**1995**). ''Required number of tokens to determine representative voice perturbation values,'' J. Speech Hear. Res. **38**, 1260–1269.
Titze, I. R., Horii, Y., and Scherer, R. C. (**1987**). ''Some technical considerations in voice perturbation measurements,'' J. Speech Hear. Res. **30**, 252–260.
Titze, I. R., and Liang, H. (**1993**). ''Comparison of $F0$ extraction methods for high-precision voice perturbation measurements,'' J. Speech Hear. Res. **36**, 1120–1133.
von Leden, H., and Koike, Y. (**1970**). ''Detection of laryngeal disease by computer technique,'' Arch. Otolaryngol. **91**, 3–10.
Voss, R. F., and Clarke, J. (**1978**). '' '1/f noise' in music: Music from 1/f noise,'' J. Acoust. Soc. Am. **63**, 258–263.
Voss, R. F. (**1989**). ''Random fractals: Self-affinity in noise, music, mountains, and clouds,'' Physica D **38**, 362–371.
Wendahl, R. W. (**1963**). ''Laryngeal analog synthesis of harsh voice quality,'' Folia Phoniatr. **15**, 241–250.
Wendahl, R. W. (**1966**). ''Laryngeal analog synthesis of jitter and shimmer auditory parameters of harshness,'' Folia Phoniatr. **18**, 98–108.
Wornell, G. W. (**1996**). *Signal Processing with Fractals* (Prentice-Hall, Englewood Cliffs, NJ).