International Journal of Advanced Intelligence
Volume 4, Number 1, pp.133-154, December, 2012.

AIA
International
Advanced
Information
Institute

A New Optimization Method of the Geometric Distance using Weighted Random Numbers

Michihiro Jinnai

Faculty of Human Life and Environmental Sciences,
Nagoya Women's University
3-40 Shioji-cho, Mizuho-ku, Nagoya, 467-8610, Japan
mjinnai@nagoya-wu.ac.jp

Satoru Tsuge

School of Informatics, Daido University
10-3 Takiharu-cho Minami-ku, Nagoya, 457-8530, Japan
tsuge@daido-it.ac.jp

Shingo Kuroiwa

Department of Information and Image Science, Chiba University
1-33 Yayoi-cho Inage-ku, Chiba, 263-8522, Japan
kuroiwa@faculty.chiba-u.jp

Fuji Ren

Faculty of Engineering, University of Tokushima
2-1 Minami-josanjima, Tokushima, 770-8506, Japan
ren@is.tokushima-u.ac.jp

Minoru Fukumi

Faculty of Engineering, University of Tokushima
2-1 Minami-josanjima, Tokushima, 770-8506, Japan
fukumi@is.tokushima-u.ac.jp

Received (4, October, 2012)
Revised (9, December, 2012)

We have proposed a new similarity measure called the Geometric Distance. In the conventional geometric distance algorithm, we have determined the optimum variance value of a normal distribution using the "clean vowels in the continuous speech" for vowel recognition. However, there is a shortcoming with the above optimization method because only the clean vowels are used. In this paper, to improve the shortcoming, we propose a new optimization method using the weighted random numbers generated by the computer and five patterns of long vowels, instead of the "clean vowels in the continuous speech". By using our proposed method, we have checked the relationship between the variance of the normal distribution and the vowel recognition accuracy, and estimated the optimum variance value. Also, by using the estimated value, we have performed evaluation experiments for the "long vowels with actual noise of 5 dB SNR" and achieved the vowel recognition accuracy of 80.3%. We have verified the effectiveness of the proposed method.

Keywords: Similarity scales; Distance functions; Pattern matching; Noise robust.

1. Introduction

Human beings, dogs, cats, and other such animals have "the sense of similarity" in hearing and sight. To realize "the sense of similarity" using an algorithm called "similarity measure" is an important subject for developing computer intelligence. In recent years, various similarity measures have been researched in speech recognition,[1,2,3,4,5,6,7,8,9] pattern classification,[10,11,12] image retrieval,[13,14,15,16] and detection of abnormal vibration.[17] In our previous papers,[18,19] we proposed a new similarity measure called the Geometric Distance. A similarity measure is a concept that should intuitively concur with the human concept of similarity in hearing and sight. Therefore, we developed a mathematical model incorporating the following two characteristics for the similarity measure.
<1> The distance metric must show good immunity to noise.
<2> The distance metric must increase monotonically when a difference increases between peaks of the standard and input patterns.
Then, we proposed an algorithm based on one-to-many point mapping to realize the mathematical model. Within the algorithm, the difference in shapes between the standard and input patterns is replaced by the shape change of a reference pattern having the initial shape of a normal distribution, and the magnitude of this shape change is numerically evaluated as a variable of the moment ratio. In such a case, from its principle, it is important to optimize the shape (variance σ^2) of the normal distribution that covers the standard and input patterns. Until now, we have determined the optimum variance value of the normal distribution using the "clean vowels in the continuous speech" for vowel recognition.[18,19]

However, there is a shortcoming with the above optimization method. That is, the characteristic <1> of the above mathematical model is ignored because only the clean vowels are used. The optimization needs to be made to maximize the effect of the characteristics $<1>$ and $<2>$ of the mathematical model simultaneously. Besides, since the optimum variance value of the normal distribution needs to be recalculated each time the speaker changes, a low processing overhead is also required to calculate the optimum value. To improve the shortcoming and to satisfy the requirement, we have studied the optimization method of the geometric distance for various sounds.[20]

In this paper, we propose a new method to determine the optimum variance value of the normal distribution for vowel recognition, where we consider both characteristics <1> and <2> of the mathematical model and reduce the processing overhead. We perform an experiment to estimate the optimum value by using our proposed method. Also, we perform evaluation experiments of vowel recognition by using the estimated value that we have calculated. These experiments use the same voice data and feature parameters as those used in our previous papers.[18,19] The paper consists of the following sections. Section 2 describes the shortcoming that is found in the conventional optimization method of the geometric distance. Section 3 describes the new optimization method of the geometric distance, and describes

the optimization experiment using the weighted random numbers generated by the computer and five patterns of long vowels. Section 4 describes the evaluation experiments of vowel recognition that have been carried out by using the calculated optimum value (estimated value), and describes the effectiveness of the proposed method. Section 5 describes the conclusions and touches on future work.

2. Conventional Optimization Method

Up to this stage, we have checked the relationship between the variance of the normal distribution and the vowel recognition accuracy, using the "clean long vowels having the variability with time of 12 weeks" and the "clean vowels in the continuous speech".[18,19] From the results of vowel recognition experiments, we have found that the recognition accuracy reaches 100% in a wide variance value range of the normal distribution in the variability with time below 4 weeks if the "clean long vowels having the variability with time" are used. In such a case, we have a problem determining the location of the maximum recognition accuracy. This means that we will find it difficult to determine the optimum variance value of the normal distribution by using the "clean long vowels produced in a short period". Meanwhile, if the "clean vowels in the continuous speech" are used, the power spectrum of the vowel changes minimally even if the voices are produced in a short period. Therefore, the location of the maximum recognition accuracy is most obvious. Owing to the above reason, the conventional optimization method estimates the optimum variance value of the normal distribution using the "clean vowels in the continuous speech". And the evaluation experiments of vowel recognition are performed for the "clean long vowels" and the "long vowels with actual noise" using the estimated value.

However, there is the shortcoming in the above optimization method where the characteristic <1> of the above mathematical model is ignored because only the clean vowels are used. The optimization needs to be made to maximize the effect of the characteristics <1> and <2> of the mathematical model simultaneously. In this case, the shortcoming seemed to be able to be solved by optimization using the "long vowels with actual noise". In other words, optimization is achieved under conditions where the "wobble" caused by the actual noise corresponds to the characteristic <1> of the mathematical model, and the "difference" between the formants of the standard and input patterns corresponds to the characteristic <2>. In this method, however, it is necessary to record all of actual noise in the daily life, create the voice data of long vowels including the actual noise each time the speaker changes, and calculate the optimum value using such voice data. This requires a huge processing overhead, and practical problems remain. As an improvement, we propose a new method that can determine the optimum value with a low processing overhead in the next section. This method simulates the actual noise in the daily life with a small amount of synthetic noise generated by the computer. Note that the "long vowel" is abbreviated as the "vowel" hereafter.

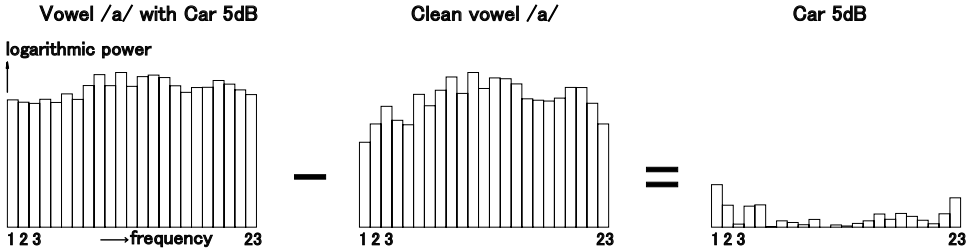

Fig. 1. Subtraction of clean vowel from vowel with car 5dB.

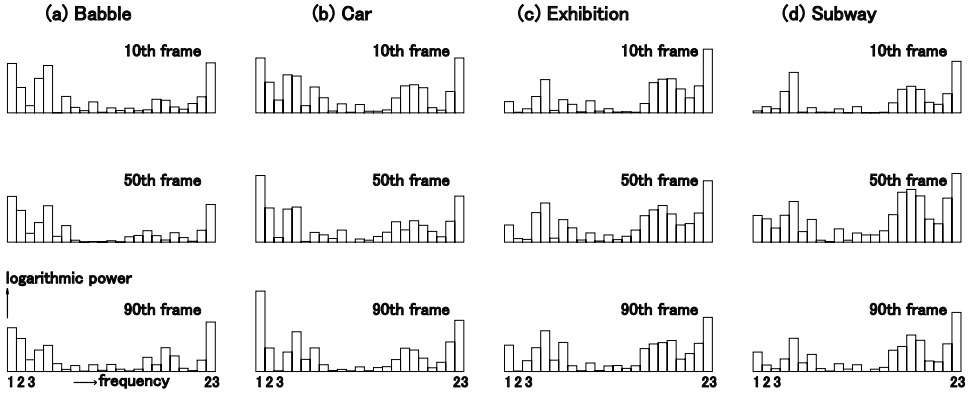

Fig. 2. Difference patterns of actual noises.

3. New Optimization Method

In this paper, we have adopted a method to add "wobble" directly to the pattern (the logarithmic power spectrum) whose shape is compared in order to apply the geometric distance to the general pattern recognition. Generally, in the study of speech recognition, the microphone output signal of the actual noise equivalent to the SNR is added to the microphone output signal of the clean vowel, and the voice data is created. Then, this voice data is multiplied by the window function (the "Hamming window" in this research) to calculate the logarithmic power spectrum. If the effect of the window function is considered, this is approximately equivalent to the calculation of the logarithmic power spectrum after adding the power spectrum of the actual noise equivalent to the SNR to the power spectrum of the clean vowel. It is replaced by the direct addition of "wobble" caused by the actual noise to the logarithmic power spectrum of the clean vowel. The proposed method uses weighted random numbers generated by the computer instead of the "wobble" caused by the actual noise. This means that the weighted random numbers generated by the computer are added to the logarithmic power spectrum of the clean vowel and it is used as the input pattern. Also, the logarithmic power spectrum of the clean vowel is used as the standard pattern. In this case, both the characteristics <1>

and $<2>$ of the mathematical model are well considered. In this section, we check the relationship between the variance of the normal distribution and the vowel recognition accuracy, using both the standard and input patterns as created above and the algorithm[19] of the geometric distance d_A. Then, we determine the optimum variance value of the normal distribution. In this section, we carry out the optimization experiment using the same voice data as described in our previous papers.[18,19]

3.1. *Difference pattern of actual noise*

In order to determine the best weighted random numbers to be added instead of the "wobble" caused by the actual noise, we check the "wobble" of the logarithmic power spectrum caused by the actual noise. An example is shown at the left and center of Fig. 1. They are the logarithmic power spectrum arrays of the 23rd dimensional Mel filter bank output (abbreviated as "logarithmic power spectrum" hereafter).[21] Note that the bar graph at the left of Fig. 1 shows the logarithmic power spectrum that is extracted from the voice data created by adding the microphone output signal of Car noise equivalent to the SNR of 5 dB to the microphone output signal of the clean vowel /a/. Also, the bar graph at the center of Fig. 1 shows the logarithmic power spectrum that is extracted from the clean vowel /a/. Then, the bar graph at the right of Fig. 1 shows a difference pattern that is created by subtracting the latter logarithmic power spectrum from the former logarithmic power spectrum. This difference pattern shows the "wobble" of the logarithmic power spectrum caused by the actual noise. Furthermore, Figs. 2(a)–(d) show the difference patterns which have been calculated by the above method, using the 10th, 50th and 90th frames of the central 100 frames of the clean vowel /a/ produced for a period of 2 seconds, and using the actual noises of Babble, Car, Exhibition and Subway. From Fig. 2, we can understand that the difference pattern of the actual noise changes randomly with time while maintaining a constant shape.

3.2. *Addition of weighted random numbers*

The m-th dimensional logarithmic power spectrum of the clean vowel /a/ is shown at the center of Fig. 1, where $m=23$. If the i-th logarithmic power spectrum values (where, $i=1,2,\cdots,m$) of a clean standard vowel and a clean input vowel are s_i and x_i, respectively, we create a standard pattern vector $\boldsymbol{s}$ having s_i components, and an input pattern vector $\boldsymbol{x}$ having x_i components, and represent them as follows. In Eq.(1), the function of "T" means a transposed matrix.

$$\begin{aligned} \boldsymbol{s} &= (\, s_1, s_2\, , \cdots , s_i, \cdots , s_m\,)^T \\ \boldsymbol{x} &= (x_1, x_2, \cdots , x_i, \cdots , x_m)^T \end{aligned} \tag{1}$$

Fig. 3 shows six types of m-th dimensional noise patterns as Noise 1 to Noise 6. They have been generated as a typical example of difference patterns of the actual

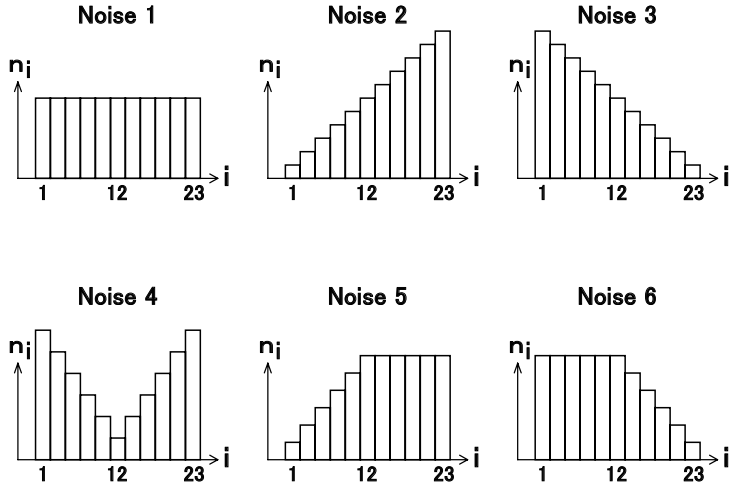

Fig. 3. 23rd dimensional noise patterns.

Table 1. Function n_i of Noise 1 to Noise 6.

Noise 1	$n_i = \alpha_1$	$(\ 1 \le i \le 23)$
Noise 2	$n_i = \alpha_2\, i$	$(\ 1 \le i \le 23)$
Noise 3	$n_i = \alpha_3\,(24 - i)$	$(\ 1 \le i \le 23)$
Noise 4	$n_i = \alpha_4\,(13 - i)$ $n_i = \alpha_4$ $n_i = \alpha_4\,(i - 11)$	$(\ 1 \le i \le 11)$ $(i = 12)$ $(13 \le i \le 23)$
Noise 5	$n_i = \alpha_5\, i$ $n_i = \alpha_5 \times 12$	$(\ 1 \le i \le 11)$ $(12 \le i \le 23)$
Noise 6	$n_i = \alpha_6 \times 12$ $n_i = \alpha_6\,(24 - i)$	$(\ 1 \le i \le 12)$ $(13 \le i \le 23)$

noise as explained in Figs. 2(a)–(d). Also, if the i-th value ($i = 1, 2, \cdots, m$) of the noise pattern shown in Fig. 3 is n_i, Table 1 shows n_i as the function of i. Note that values α_1 to α_6 are the constants which are calculated by the experiment described in the next section. Here, we create a noise pattern vector $\boldsymbol{n}$ having n_i components, and represent it as follows.

$$\boldsymbol{n} = (n_1, n_2, \cdots, n_i, \cdots, n_m)^T \tag{2}$$

Next, if variable *Rnd* is random numbers uniformly distributed within the range of 0.0 to 1.0, as shown in the following equations, we assign s_{oi} to the component

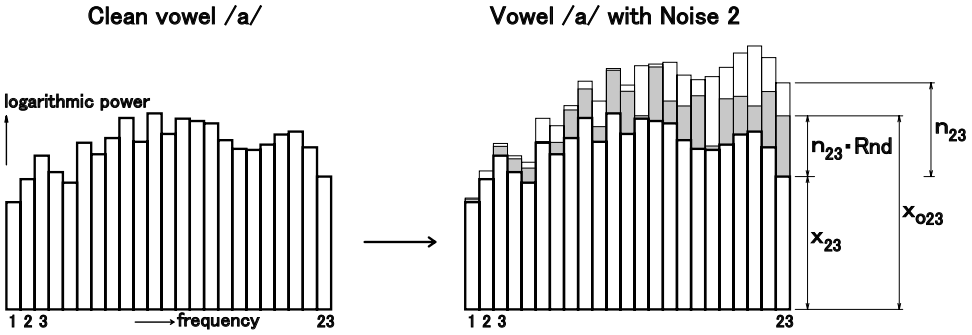

Fig. 4. Addition of weighted random numbers to clean vowel.

value s_i of standard pattern vector, and assign x_{oi} to the addition of the component value x_i of input pattern vector and the weighted random numbers $n_i{\cdot}Rnd$.

$$\begin{aligned} s_{oi} &= s_i \\ x_{oi} &= x_i + n_i{\cdot}Rnd \qquad\qquad (i = 1, 2, 3, \cdots, m) \end{aligned} \tag{3}$$

Then, we create an original standard pattern vector $\boldsymbol{s_o}$ having s_{oi} components, and an original input pattern vector $\boldsymbol{x_o}$ having x_{oi} components, and represent them as follows.[19]

$$\begin{aligned} \boldsymbol{s_o} &= (\, s_{o1},\, s_{o2}, \cdots, s_{oi}, \cdots, s_{om}\,)^T \\ \boldsymbol{x_o} &= (x_{o1}, x_{o2}, \cdots, x_{oi}, \cdots, x_{om})^T \end{aligned} \tag{4}$$

$\boldsymbol{s_o}$ is the original standard pattern vector which has been created from the logarithmic power spectrum of clean standard vowel, and $\boldsymbol{x_o}$ is the original input pattern vector which has been created from the logarithmic power spectrum of clean input vowel, added by the weighted random numbers generated by the computer. Fig. 4 shows the shape of the second formula of Eq. (3) using the noise pattern of Noise 2. The bar graph at the left of Fig. 4 shows the shape of input pattern vector $\boldsymbol{x}$ given by Eq. (1), and the bar graph at the right of Fig. 4 shows the shape of original input pattern vector $\boldsymbol{x_o}$ given by Eq. (4).

3.3. *Calculation of component value n_i of noise pattern vector*

In our previous papers,[18,19] the microphone output signals of Babble, Car, Exhibition and Subway noise were added to those of the clean vowel with the 20 dB, 10 dB and 5 dB SNR, and the voice data was created. From these voice data, the logarithmic power spectrum was calculated, and the input pattern was created. Then, the shapes were compared between the standard and input patterns. On the other hand, in this paper, as shown in Eq. (3), the input pattern is created by the direct addition of the weighted random numbers $n_i{\cdot}Rnd$ to the logarithmic power spectrum value x_i of the clean vowel, and their shapes are compared. Therefore,

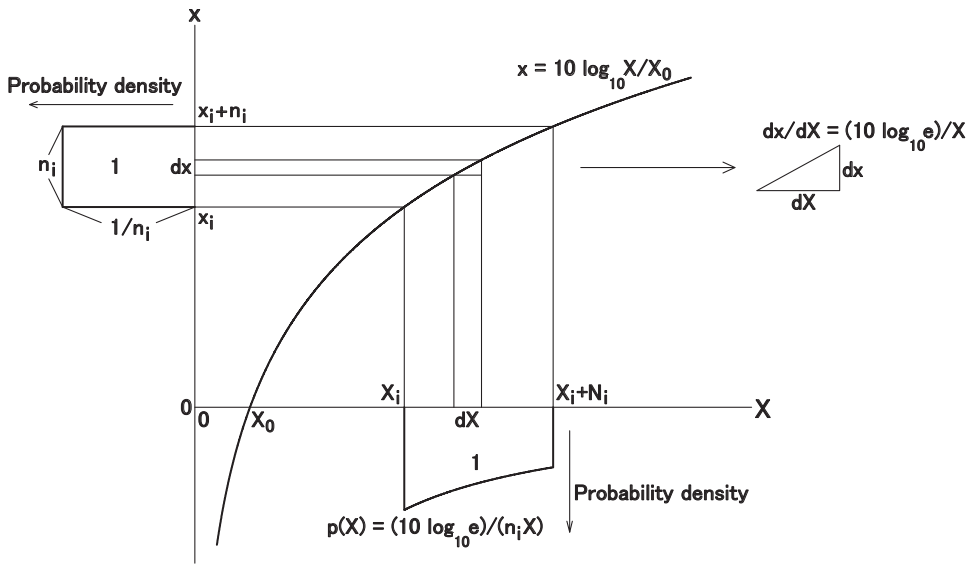

Fig. 5. Relationship between power spectrum and logarithmic power spectrum.

we need to calculate each component value n_i of the noise pattern vector that is equivalent to each SNR used in our previous papers.[18,19] In other words, in Fig. 3 and on Table 1, we need to calculate values α_1 to α_6 that are equivalent to the above SNR. The following explains their calculation.

When the microphone output signal of the clean vowel is passed through the Mel filter bank with the m frequency bands, we assume that the power spectrum array X_i $(i = 1, 2, \cdots, m)$ is obtained. If the reference value of power spectrum is X_0, the logarithmic power spectrum array x_i $(i = 1, 2, \cdots, m)$ that corresponds to X_i can be calculated from the first formula of the following equation. Also, if the component value n_i $(i = 1, 2, \cdots, m)$ of noise pattern vector is added to this logarithmic power spectrum array x_i $(i = 1, 2, \cdots, m)$, value $x_i + n_i$ $(i = 1, 2, \cdots, m)$ is obtained. The relationship between the value $x_i + n_i$ and its corresponding power spectrum array $X_i + N_i$ $(i = 1, 2, \cdots, m)$ can be represented as the second formula of the following equation.

$$\begin{aligned} x_i &= 10 \log_{10} \frac{X_i}{X_0} && (n_i > 0) \\ x_i + n_i &= 10 \log_{10} \frac{X_i + N_i}{X_0} && (i = 1, 2, 3, \cdots, m) \end{aligned} \tag{5}$$

Fig. 5 shows the relationship between X_i and x_i between $X_i + N_i$ and $x_i + n_i$ given by Eq. (5) for the i-th frequency band of the filter bank. This section aims to calculate the value n_i that is equivalent to the SNR of 5 dB. The following equation can be

obtained as an inverse function of Eq. (5).

$$\begin{aligned} X_i &= X_0 \cdot 10^{x_i/10} \\ X_i + N_i &= X_0 \cdot 10^{(x_i+n_i)/10} \qquad (i = 1, 2, 3, \cdots, m) \end{aligned} \tag{6}$$

In Eq. (6), we can obtain the following equation by substituting the first formula into the second formula.

$$N_i = X_0 \cdot 10^{x_i/10} \, (10^{n_i/10} - 1) \qquad (i = 1, 2, 3, \cdots, m) \tag{7}$$

In Eq. (3), if the variable *Rnd* is random numbers uniformly distributed within the range of 0.0 to 1.0, $x_{oi} = x_i + n_i \cdot Rnd$ and, therefore, x_{oi} uniformly distributes within the range of x_i to $x_i + n_i$. Fig. 5 shows the probability density function of the flat shape which has function value $1/n_i$ in range $[x_i, x_i + n_i]$ on axis x. As shown in Fig. 5, if we only focus on the i-th frequency band of the filter bank, it is appropriate to express the weighted random numbers $n_i \cdot Rnd$ as the uniformly distributed random numbers $n_i \cdot Rnd$. The weighted random numbers $n_i \cdot Rnd$ means the multiplication of different weight n_i to each of the i-th frequency band. In this section, we use them in differently ways as necessary. Because the gradient of logarithmic curve $x = 10 \log_{10} X/X_0$ is $dx/dX = (10 \log_{10} e)/X$, the probability density function $p(X)$ on axis X, which corresponds to the probability density function $1/n_i$ on axis x, is described by the following equation.

$$p(X) = \frac{10 \log_{10} e}{n_i X} \qquad (i = 1, 2, 3, \cdots, m) \tag{8}$$

Thus, Fig. 5 shows the probability density function which has function value $p(X) = (10 \log_{10} e)/(n_i X)$ in range $[X_i, X_i + N_i]$ on axis X. From the following equation, we can confirm that the total area of probability density function $p(X)$ is equal to 1. Here, we can obtain the fifth formula of Eq. (9) by substituting Eq. (5) into the fourth formula of Eq. (9).

$$\begin{aligned} \int_{X_i}^{X_i+N_i} p(X)\, dX &= \int_{X_i}^{X_i+N_i} \frac{10 \log_{10} e}{n_i X}\, dX \\ &= \frac{10 \log_{10} e}{n_i} \int_{X_i}^{X_i+N_i} \frac{1}{X}\, dX \\ &= \frac{10 \log_{10} e}{n_i} \Big\{ \log_e (X_i + N_i) - \log_e X_i \Big\} \\ &= \frac{1}{n_i} \Big\{ 10 \, \log_{10} \frac{X_i + N_i}{X_0} - 10 \, \log_{10} \frac{X_i}{X_0} \Big\} \\ &= \frac{1}{n_i} \Big\{ (x_i + n_i) - x_i \Big\} \\ &= 1 \qquad (i = 1, 2, 3, \cdots, m) \end{aligned} \tag{9}$$

Where, if the uniformly distributed random numbers $n_i \cdot Rnd$ are added to the logarithmic power spectrum x_i of the clean vowel on axis x, we assume that the power

spectrum on axis X, which corresponds to $x_i+n_i\cdot Rnd$, is X. Now, expected value $E_i[X]$ of the power spectrum X can be calculated by the following equation.

$$\begin{aligned} E_i[X] &= \int_{X_i}^{X_i+N_i} X\cdot p(X)\,dX \\ &= \int_{X_i}^{X_i+N_i} X\cdot \frac{10\log_{10} e}{n_i X} dX \\ &= (10\log_{10} e)\cdot \frac{1}{n_i}\cdot N_i \qquad\qquad (i=1,2,3,\cdots,m) \end{aligned} \tag{10}$$

We can obtain the following equation by substituting Eq. (7) into Eq. (10).

$$E_i[X] = (10\log_{10} e)\cdot X_0\cdot 10^{x_i/10}\cdot \frac{10^{n_i/10}-1}{n_i} \qquad (i=1,2,3,\cdots,m) \tag{11}$$

On axis X of Fig. 5, the average energy of power spectrum of the clean vowel is X_i, and the average energy of power spectrum, which corresponds to the uniformly distributed random numbers $n_i\cdot Rnd$, is $E_i[X]-X_i$. Therefore, the signal-to-noise ratio (SNR) of the entire frequency band can be calculated by the following equation.

$$\begin{aligned} SNR &= 10\ \log_{10} \frac{\sum_{i=1}^{m} X_i}{\sum_{i=1}^{m}(E_i[X]-X_i)} \\ &= 10\ \log_{10} \frac{\sum_{i=1}^{m} X_i}{\sum_{i=1}^{m} E_i[X] - \sum_{i=1}^{m} X_i} \end{aligned} \tag{12}$$

We can obtain the following equation by substituting Eqs. (6) and (11) into Eq. (12).

$$\begin{aligned} SNR &= 10\ \log_{10} \frac{X_0 \sum_{i=1}^{m} 10^{x_i/10}}{(10\ \log_{10} e)\cdot X_0 \sum_{i=1}^{m} 10^{x_i/10}\cdot \frac{10^{n_i/10}-1}{n_i} - X_0 \sum_{i=1}^{m} 10^{x_i/10}} \\ &= 10\ \log_{10} \sum_{i=1}^{m} 10^{x_i/10} \\ &\quad -10\ \log_{10}\left\{ (10\ \log_{10} e) \sum_{i=1}^{m} 10^{x_i/10}\cdot \frac{10^{n_i/10}-1}{n_i} - \sum_{i=1}^{m} 10^{x_i/10} \right\} \end{aligned} \tag{13}$$

Furthermore, we assign $\psi(n_1, n_2, \cdots, n_m)$ to the right side of Eq. (13) that is subtracted by the left side, and represent it as follows.

$$\psi(n_1, n_2, \cdots, n_m) = 10 \log_{10} \sum_{i=1}^{m} 10^{x_i/10} \tag{14}$$
$$-10 \log_{10} \left\{ (10 \log_{10} e) \sum_{i=1}^{m} 10^{x_i/10} \cdot \frac{10^{n_i/10} - 1}{n_i} - \sum_{i=1}^{m} 10^{x_i/10} \right\} - SNR$$

In Eq. (14), x_i is the logarithmic power spectrum value of the clean vowel, and we can set its value using the voice data. Therefore, Eq. (14) is the function of n_i $(i=1, 2, \cdots, m)$.

Next, we show that $\psi(n_1, n_2, \cdots, n_m)$ decreases monotonically when each n_i $(i=1, 2, \cdots, m)$ increases. For that purpose, we assign $\phi_1(n_i)$ to term $(10^{n_i/10} - 1)/n_i$ of Eq. (14) as follows, and we check its increase or decrease.

$$\phi_1(n_i) = \frac{10^{n_i/10} - 1}{n_i} \qquad (i = 1, 2, 3, \cdots, m) \tag{15}$$

Here, we can obtain the following equation by differentiating Eq. (15) by n_i.

$$\begin{aligned} \phi_1'(n_i) &= \left(\frac{10^{n_i/10} - 1}{n_i} \right)' \\ &= \frac{(\log_e 10^{1/10})\, n_i\, 10^{n_i/10} - 10^{n_i/10} + 1}{n_i^2} \\ &= \frac{\phi_2(n_i)}{n_i^2} \qquad (i = 1, 2, 3, \cdots, m) \end{aligned} \tag{16}$$

Furthermore, we assign $\phi_2(n_i)$ to the numerator of Eq. (16) as follows, and we check its positive or negative.

$$\phi_2(n_i) = (\log_e 10^{1/10})\, n_i\, 10^{n_i/10} - 10^{n_i/10} + 1 \qquad (i = 1, 2, 3, \cdots, m) \tag{17}$$

For that purpose, we calculate Eq. (17) if $n_i = 0$ and its derived function as follows.

$$\phi_2(\,0\,) = 0 \tag{18}$$
$$\phi_2'(n_i) = (\log_e 10^{1/10})^2\, n_i\, 10^{n_i/10} > 0 \qquad (\,n_i > 0\,) \qquad (i = 1, 2, 3, \cdots, m) \tag{19}$$

From Eqs. (18) and (19), it is clear that $\phi_2(n_i) > 0$. Then, from Eq. (16), it is clear that $\phi_1'(n_i) > 0$ and, therefore, Eq. (15) is a monotonically increasing function. From the above, it is clear that the value of Eq. (14) decreases monotonically when each n_i $(i=1, 2, \cdots, m)$ increases.

In this paper, each n_i $(i=1, 2, \cdots, m)$ is related to each other by the parameter α_k $(k=1, 2, \cdots, 6)$ as shown on Table 1. In the case of Noise 1 to Noise 6 shown on Table 1, each n_i increases monotonically when each α_k increases and, therefore, the

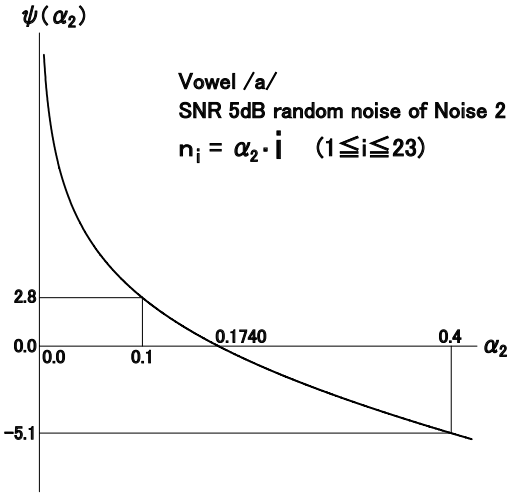

Fig. 6. Graph of function $\psi(\alpha_2)$.

Table 2. Solution α_2 of $\psi(\alpha_2)=0$ (Noise 2 : $n_i = \alpha_2 i$).

α_2	/a/	/i/	/u/	/e/	/o/
SNR 5 dB	0.1740	0.1642	0.1769	0.1701	0.1843
SNR 3 dB	0.2484	0.2340	0.2519	0.2426	0.2623
SNR 1 dB	0.3421	0.3216	0.3457	0.3337	0.3595

value of Eq. (14) decreases monotonically. In particular, $n_i=\alpha_2 i$ $(i=1,2,\cdots,m)$ for Noise 2, and Eq. (14) can be rewritten as follows.

$$\psi(\alpha_2) = 10\ \log_{10}\sum_{i=1}^{m} 10^{x_i/10} \tag{20}$$
$$-10\ \log_{10}\left\{(10\ \log_{10} e)\sum_{i=1}^{m} 10^{x_i/10}\cdot\frac{10^{\alpha_2 i/10}-1}{\alpha_2 i} - \sum_{i=1}^{m} 10^{x_i/10}\right\} - SNR$$

Fig. 6 shows a relational graph between α_2 and $\psi(\alpha_2)$ obtained through numerical analysis of Eq. (20). Note that we assumed that SNR=5 in Eq. (20). Also, we have substituted the mean value of each logarithmic power spectrum, calculated from the central 100 frames of the clean vowel /a/, into x_i $(i=1,2,\cdots,m)$. As shown in Fig. 6, Eq. (20) is a monotonically decreasing function, and it is clear that we can uniquely determine a solution α_2 of $\psi(\alpha_2)=0$ through numerical analysis. As described above, we could obtain solution $\alpha_2=0.1740$ of $\psi(\alpha_2)=0$ from Fig. 6. Table 2 shows the values of α_2 which are obtained for each vowel and for each SNR when SNR=5, SNR=3 and SNR=1 and if the noise pattern of Noise 2 and "01Clean"[18] of each vowel are used. "01Clean" is the first "clean vowel" that was produced among 72 sounds in 12 weeks.

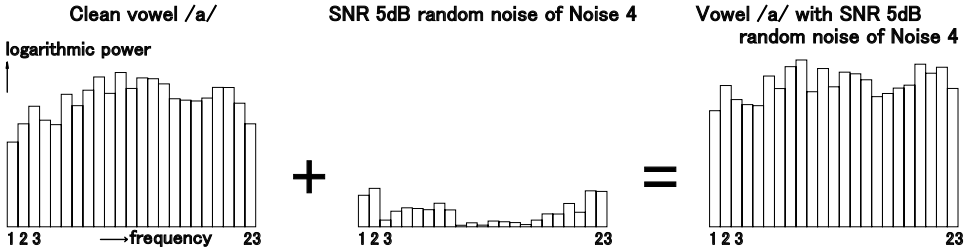

Fig. 7. Addition of random noise to clean vowel.

The above calculation procedure is summarized below. First, in Eq. (6), power spectra X_i and X_i+N_i on axis X shown in Fig. 5 are expressed by logarithmic power spectra x_i and x_i+n_i on axis x. Also, in Eq. (11), expected value $E_i[X]$ of power spectrum X on axis X, which corresponds to $x_i+n_i \cdot Rnd$ on axis x, is expressed by x_i and n_i. Then, we calculate the SNR on axis X using Eq. (12), substitute Eqs. (6) and (11) into Eq. (12). Therefore, the SNR is expressed by x_i and n_i in Eq. (13). We substitute the mean value of the logarithmic power spectra of the clean vowel into x_i. Now, Eq. (13) is an equation of m variables with unknowns n_i $(i=1,2,\cdots,m)$. In this paper, each n_i $(i=1,2,\cdots,m)$ is related by the parameter α_k $(k=1,2,\cdots,6)$ as shown on Table 1. Therefore, Eq. (13) is rewritten by Eq. (20). Eq. (20) is an equation of single variable with unknown α_2. And we calculate solution α_2 and obtain value n_i that is equivalent to the SNR of 5 dB.

By using the above calculation procedure, the value of each α_k $(k=1,2,\cdots,6)$ is calculated for the noise patterns of Noise 1 to Noise 6, and Table 2 of each noise pattern is obtained. Then, the weighted random numbers $n_i \cdot Rnd$, which is equivalent to the SNR, is generated by the computer. Fig. 7 shows the process where the weighted random numbers $n_i \cdot Rnd$ $(i=1,2,\cdots,m)$ equivalent to the SNR of 5 dB are added to the logarithmic power spectrum x_i $(i=1,2,\cdots,m)$ of the clean vowel /a/, using Noise 4 and Eq. (3), and then the component value x_{oi} $(i=1,2,\cdots,m)$ of the original input pattern vector is created. It is clear that the shape of the weighted random numbers, shown at the center of Fig. 7, is similar to the difference pattern of the actual noise shown in Fig. 2.

Finally in this section, we discuss the relationship between the area (or energy) of the weighted random numbers generated by the computer and that of the difference pattern of actual noise. After calculating the average area of the weighted random numbers of 5 dB SNR and that of the difference pattern of 5 dB SNR, using the central 100 frames of each vowel produced for a period of 2 seconds, we have found that the former value is 16.2% greater than the latter value. We suppose that there are two causes for that as follows. First, in the calculation of the weighted random numbers, we substituted the mean value of the logarithmic power spectra, calculated from the central 100 frames of each vowel produced for a period of 2 seconds, into Eq. (20), and obtained solution α_k $(k=1,2,\cdots,6)$. These frames are overlapped for the 25 msec frame width and 10 msec frame period. In the calculation of the

Table 3. Logarithmic power spectra for optimizing normal distribution.

		/a/	/i/	/u/	/e/	/o/
	01 Clean	100	100	100	100	100
	Standard pattern	1	1	1	1	1
{1}	01 Clean with SNR 5dB random noise of Noise 1					
	Input pattern	100×50	100×50	100×50	100×50	100×50
{2}	01 Clean with SNR 5dB random noise of Noise 2					
	Input pattern	100×50	100×50	100×50	100×50	100×50
:	:	...	...	...	...	...
{6}	01 Clean with SNR 5dB random noise of Noise 6					
	Input pattern	100×50	100×50	100×50	100×50	100×50

difference pattern, we calculated the SNR using the microphone output signal of the entire interval of 2-second vowel. We suppose that those average areas are different because the calculation intervals of SNR differ between them. Second, we obtained the logarithmic power spectrum value x_i ($i=1,2,\cdots,m$) of the clean vowel using the Hamming window, and substituted this value into Eq. (20) in order to obtain solution α_k. Therefore, we suppose that an effect of the Hamming window appears as described at the beginning of Section 3. In Section 4.2, based on our experiments, we will discuss the estimation error of optimum value caused by the above area difference.

3.4. *Creation of original pattern vectors*

Here, we use the α_k ($k=1,2,\cdots,6$) values obtained in the previous section, and create the original standard pattern vector and original input pattern vector given by Eq. (4), by applying the α_k values to the same voice data as those used in our previous papers.[18,19] Note that the original standard pattern vector is abbreviated as "the standard pattern", and the original input pattern vector is abbreviated as "the input pattern" hereafter. Table 3 shows the type and the number of the 23rd dimensional logarithmic power spectrum that has been used for the standard and input patterns in the optimization experiment. The logarithmic power spectra, each consisting of 100 frames shown on the first row of Table 3, have been extracted from "01Clean" of each vowel. Then, the median[18] is determined from the above 100 frames and it is used as the standard pattern of each vowel. The logarithmic power spectra, each consisting of one frame shown on the second row of Table 3, are the standard patterns that have been determined for each vowel.

Also, the logarithmic power spectra, each consisting of 100×50 frames shown in {1} to {6} of Table 3, have been created by adding the weighted random numbers to

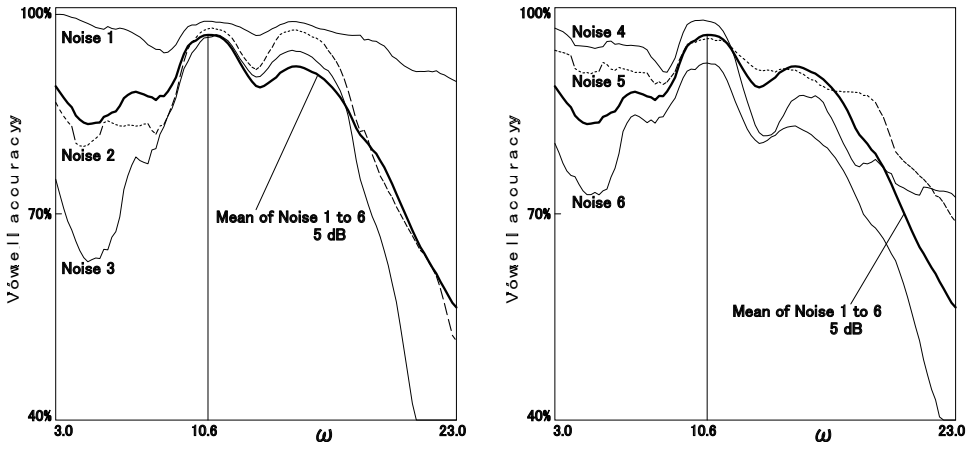

Fig. 8. Vowel recognition accuracy and optimum value ω.

the logarithmic power spectra, each consisting of 100 frames of the above "01Clean", using Eq. (3) and the noise patterns of Noise 1 to Noise 6 shown in Fig. 3 when SNR=5. During this time, the uniformly distributed random numbers *Rnd* are generated repeatedly and the logarithmic power spectra, each consisting of 100×50 frames, are created. Then, the logarithmic power spectra of these 6×100×50×5 frames are used as the input patterns.

As described above, in the optimization experiment, we create the standard pattern and the input pattern by using the weighted random numbers generated by the computer and five patterns of "clean vowel 01Clean".

3.5. *Variance optimization of normal distribution*

We determine the optimum value of the variance σ^2 of the normal distribution (the optimum value of ω)[18] using both the standard and input patterns created in the previous section and the algorithm[19] of the geometric distance d_A. Similar to the vowel recognition experiments of the previous papers,[18,19] the value ω is incremented by 0.2 from 3.0 to 23.0, and the recognition accuracy of the input pattern is calculated by using 100×50×5-frame input patterns shown in {1} to {6} of Table 3. Fig. 8 shows the calculated relationship between the value ω and the recognition accuracy by six thin lines, respectively. Also, these six curves are averaged and the average recognition accuracy is shown by thick lines in Fig. 8. From Fig. 8, it is discovered that the recognition accuracy curve of Noise 1 is higher than each curve of Noise 2 to Noise 6 in the all ω value range. We suppose the cause as follows. Within the geometric distance algorithm, the "wobble" caused by the random numbers is replaced by the shape change of the reference pattern having the initial shape of the normal distribution. During this time, the shape of the noise pattern of Noise 1 is flat (or uniform) as shown in Fig. 3 and, therefore, we suppose that the "wobble" is absorbed effectively. Furthermore, from Fig. 8, it is discovered

Table 4. Vowel recognition accuracy with geometric distance d_A. ($\omega = 10.6$)

	Babble	Car	Exhibition	Subway	Mean
Clean					99.98%
SNR 20 dB	99.92%	99.86%	99.22%	99.56%	99.64%
SNR 10 dB	98.52%	97.94%	88.16%	93.97%	94.65%
SNR 5 dB	91.68%	82.13%	66.85%	80.44%	80.28%

that the peak of recognition accuracy is at the same location for each of the Noise 1 to Noise 6 curves. We can see that the average recognition accuracy of Noise 1 to Noise 6 becomes maximum if ω=10.6. Thus, we determine ω=10.6 as the optimum value and use it in the following evaluation experiments. When we have performed the optimization experiment using the input pattern, each consisting of 100×10 frames, instead of the input pattern, each consisting of 100×50 frames shown in {1} to {6} of Table 3, we could obtain almost the same curves as the recognition accuracy curves shown in Fig. 8. The optimum value was ω=10.6. This shows that we can reduce the processing overhead to obtain the optimum value.

4. Evaluation Experiments of Vowel Recognition

To check the effectiveness of optimization method described in the previous section, we have performed the evaluation experiments for the "clean vowel" and the "vowel with actual noise" using the value ω=10.6 determined in the previous section and the algorithm[19] of the geometric distance d_A. The value ω=11.0 is used in our previous paper,[19] but the value ω=10.6 is used for the evaluation experiments in this section. Except for this value, we have performed the evaluation experiments of vowel recognition using the same voice data and the method as those used in our previous paper.[19]

4.1. *Evaluation experiments and their results*

In the optimization experiment of the previous section, we determined the optimum value (estimated value) of ω=10.6 by using only the "clean vowel 01Clean" that was produced first among 72 sounds in 12 weeks as shown on Table 3. Similar to the vowel recognition experiments of the previous paper,[19] in the evaluation experiments of this section, the median was determined from 100 frames of the above "clean vowel 01Clean" and it was used as the standard pattern of each vowel. On the other hand, the "clean vowel 02Clean to 72Clean" produced in the 2nd to 72nd sounds were used as the input patterns. In addition, the actual Babble, Car, Exhibition and Subway noises were added to these "clean vowel 02Clean to 72Clean" with the 20 dB, 10 dB and 5 dB SNR, and the input patterns were created.

Table 4 shows the result of evaluation experiments. As shown on Table 4, the average recognition accuracy of the "vowel with actual noise of 5 dB SNR" is 80.28% in the evaluation experiment where the optimum value (estimated value) of ω=10.6 is used.

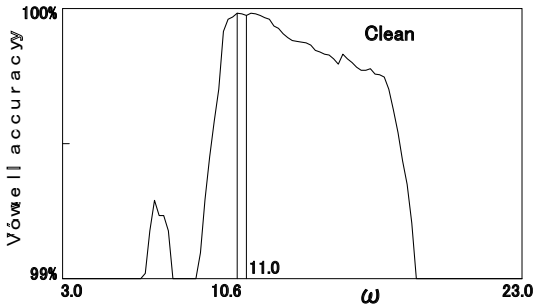

Fig. 9. Recognition accuracy of clean vowel.

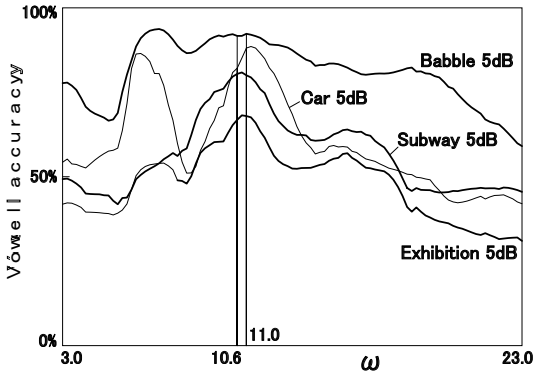

Fig. 10. Recognition accuracy of vowel with actual noise.

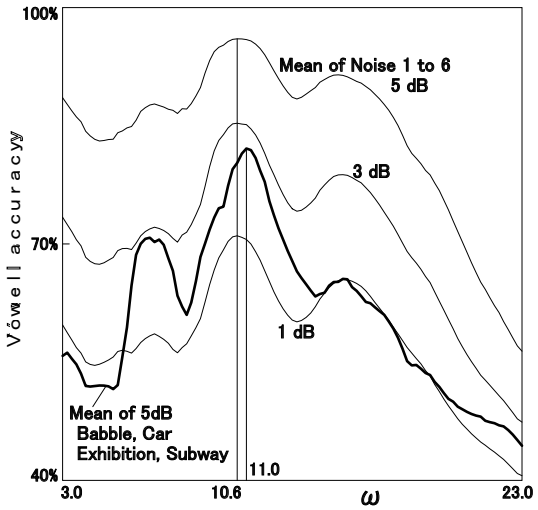

Fig. 11. Vowel recognition accuracy and optimum value ω.

4.2. *Verification of optimum value*

Table 4 shows the result of recognition accuracy using the optimum value (estimated value) of ω=10.6 that we have determined from Fig. 8. Here, in order to verify that the value ω=10.6 is truly the optimum value, the value ω is incremented by 0.2 from 3.0 to 23.0 and the recognition accuracy of the "clean vowel" and the "vowel with actual noise of 5 dB SNR" is calculated. Figs. 9 and 10 show the calculated relationship between the value ω and the recognition accuracy for the input patterns of the "clean vowel" and the "vowel with Babble 5dB, Car 5dB, Exhibition 5dB, and Subway 5dB", respectively. From Figs. 9 and 10, we can find that the recognition accuracy is almost maximum in the value ω=10.6.

Furthermore, the four curves of actual noise, shown in Fig. 10, are averaged and this average recognition accuracy is shown by a thick line in Fig. 11. In the calculation of the average recognition accuracy for Noise 1 to Noise 6 shown by thick lines in Fig. 8, the values of SNR=5, SNR=3 and SNR=1 are used respectively, and their results are shown by three thin lines in Fig. 11. Note that the average recognition accuracy curves of 5 dB SNR shown by the thick lines in Fig. 8, are the same as that shown by the thin line in Fig. 11. In Fig. 11, the recognition accuracy curves of the optimization experiments using the "vowel with weighted random numbers" are shown by three thin lines, but the recognition accuracy curve of the evaluation experiment using the "vowel with actual noise" is shown by one thick line. From Fig. 11, it is clear that the four curves of recognition accuracy have the same features and that the locations of the maximum recognition accuracy almost match each other. This means that we can estimate the optimum variance value of the normal distribution, using the "vowel with weighted random numbers" instead of the "vowel with actual noise". From Fig. 11, it is also clear that the weighted random numbers of 3 dB SNR is equivalent to the actual noise of 5 dB SNR for the average recognition accuracy. We suppose the cause as follows. Within the geometric distance algorithm, the "wobble" of input pattern is replaced by the shape change of the reference pattern having the initial shape of the normal distribution. During this time, the "wobble" caused by the random numbers is more random than the actual noise and, therefore, we suppose that the "wobble" is absorbed effectively.

At the end of Section 3.3, we described the difference between the area (or energy) of the weighted random numbers of 5 dB SNR and that of the difference pattern of actual noise of 5 dB SNR. Next, we discuss this. In Fig. 11, we can obtain the value ω=10.6 even when we use any of the recognition accuracy curves, shown by three thin lines, in the optimization experiment. Now, on Table 2, the value α_2 of 1 dB SNR is almost 2 times that of 5 dB SNR. In other words, the area of noise pattern of 1 dB SNR is almost 2 times larger than that of 5 dB SNR case. This is similar to other α_k values. When compared with this change, the 16.2% difference shown in Section 3.3 is small. They show that the difference between their areas does not affect the estimation of optimum value.

From the average recognition accuracy curve of the "vowel with actual noise of 5 dB SNR" shown by thick line in Fig. 11, it is discovered that the recognition accuracy becomes maximum if ω=11.0. It is discovered that the recognition accuracy is 80.28% if ω=10.6 and the recognition accuracy is 82.11% if ω=11.0. The difference between them is 1.83% and it is small. From the recognition accuracy curve of the "clean vowel" shown in Fig. 9, it is discovered that the recognition accuracy is 99.98% if ω=10.6 and the recognition accuracy is 99.97% if ω=11.0. The difference between them is small. This shows that we can determine the optimum value of ω using the "vowel with weighted random numbers".

In this paper, as shown in Figs. 8 and 11, we have used the vowel recognition accuracy as the objective function in order to estimate the optimum variance value. Meanwhile, we used a statistic T of "Welch's T-test" as the objective function and performed the optimization experiment for bird vocalisations.[20] If we compare the two results, we find that the former objective function curves and the latter objective function curve have the same features.

5. Conclusions and Future Work

We have proposed a new optimization method of the geometric distance to determine the optimum variance value of the normal distribution, using the weighted random numbers generated by the computer and five patterns of vowels. At this time, we have performed the vowel recognition experiments using the "vowel with weighted random numbers" and the "vowel with actual noise", respectively, and checked the relationship between the variance of the normal distribution and the vowel recognition accuracy. The results have shown that the curves of their vowel recognition accuracy have the same features and that the locations of the maximum recognition accuracy almost match each other. This means that we can estimate the optimum variance value of the normal distribution using the "vowel with weighted random numbers" instead of the "vowel with actual noise". Then, we have used the estimated value obtained from the "vowel with weighted random numbers" and performed the evaluation experiments for the "vowel with actual noise of 5 dB SNR", and verified the effectiveness of our proposal.

Finally, we describe future work. This paper shows that we have obtained the estimated value of ω=10.6 using each noise pattern of Noise 1 to Noise 6. On the other hand, we have found that the true optimum value is ω=11.0 in the evaluation experiments where we used four types of actual noises of Babble, Car, Exhibition, and Subway. In order to reduce the difference between them, we will perform the optimization experiments using more types of noise patterns and will perform the evaluation experiments using more types of actual noises. We will compare the results of those experiments, find out the type of noise pattern to be required at minimal for optimization, and improve our optimization method so that we can determine a more accurate estimation value and reduce the processing overhead by using less types of noise patterns. We will apply the results of the algorithm

proposed in this paper and the emotional expression analysis of text[22,23] to our project named Recognizing Human Emotion and Creating Machine Emotion.[24,25] Also, we will perform the optimization experiments using the normal random numbers, instead of the uniformly distributed random numbers, and will compare the results of these experiments.

Acknowledgments

This research has been partially supported by New Energy and Industrial Technology Development Organization (NEDO) of the Japanese Government under Grant No. 10HC7011, by Queensland Parks and Wildlife Service of the Australian Government under Coxen's Fig-parrot Recovery Plan, and funded by Mitsubishi Heavy Industries, Ltd. of Japan and Tokyo Gas Co., Ltd. of Japan, by West Nippon Expressway Engineering Shikoku Company Limited of Japan.

References

1. K.K. Paliwal. *Effect of Preemphasis on Vowel Recognition Performance*, Speech Communication, **3**, pp.101-106, 1984.
2. L.R. Rabiner and B.H. Juang. *Fundamentals of Speech Recognition*, Prentice Hall, Englewood Cliffs, New Jersey, 1993.
3. F. Itakura and S. Saito. *An Analysis-Synthesis Telephony Based on Maximum Likelihood Method*, Proc. 6th Int. Congr. Acoustics, C-5-5, 1968.
4. F. Itakura. *Minimum Prediction Residual Principle Applied to Speech Recognition*, IEEE Trans. Acoust., Speech and Signal Processing, **23**, pp.67-72, 1975.
5. S. Furui. *Digital Speech Processing, Synthesis, and Recognition (Electrical and Computer Engineering)*, Marcel Dekker, Inc., NewYork, 1989.
6. K. Shikano and M. Sugiyama. *Evaluation of LPC Spectral Matching Measures for Spoken Word Recognition*, Trans. IECE, 565-D, **5**, pp.535-541 1982.
7. D. Klatt. *Prediction of Perceived Phonetic Distance from Critical Band Spectra: A First Step*, Proc. ICASSP 82, **2**, pp.1278-1281, 1982.
8. D. Mansour and B.H. Juang. *A Family of Distortion Measures Based upon Projection Operation for Robust Speech Recognition*, IEEE Trans. Acoustics, Speech and Signal Processing, ASSP-37, **11**, pp.1659-1671, 1989.
9. N. Nocerino, F.K. Soong, L.R. Rabiner and D.H. Klatt. *Comparative Study of Several Distortion Measures for Speech Recognition*, Speech Communication, **4**, pp.317-331, 1985.
10. R.O. Duda, P.E. Hart and D.G. Stork. *Pattern Classification, second ed.*, Wiley, NewYork, 2000.
11. S.-H. Cha and S.N. Srihari. *On Measuring the Distance between Histograms*, Pattern Recognition, **35**, pp.1355-1370, 2002.
12. M. Jinnai, N. Boucher, J. Robertson and S. Kleindorfer. *Design Considerations in an Automatic Classification System for Bird Vocalisations using the Two-dimensional Geometric Distance and Cluster Analysis*, Proc. 20th Int. Congr. Acoustics, 130, 2010.
13. J.-K. Kamarainen, V. Kyrki, J. Ilonen and H. Kälviäinen. *Improving Similarity Measures of Histograms using Smoothing Projections*, Pattern Recognition Lett., **24**, pp.2009-2019, 2003.
14. F.-D. Jou, K.-C. Fan and Y.-L. Chang. *Efficient Matching of Large-size Histograms*, Pattern Recognition Lett., **25**, pp.277-286, 2004.
15. F. Serratosa and A. Sanfeliu. *Signatures versus Histograms: Definitions, Distances and Algorithms*, Pattern Recognition, **39**, pp.921-934, 2006.
16. V.V. Strelkov. *A New Similarity Measure for Histogram Comparison and its Application in Time Series Analysis*, Pattern Recognition Lett., **29**, pp.1768-1774, 2008.

17. M. Jinnai, Y. Akashi, S. Nakaya, F. Ren and M. Fukumi. *Recognition of Abnormal Vibrational Responses of Signposts using the Two-dimensional Geometric Distance and Wilcoxon Test*, Proc. 6th Int. Congr. IEEE NLP-KE, pp.166-173, 2010.
18. M. Jinnai, S. Tsuge, S. Kuroiwa, F. Ren and M. Fukumi. *New Similarity Scale to Measure the Difference in Like Patterns with Noise*, IJAI, Volume 1, Number 1, pp.59-88, November, 2009.
19. M. Jinnai, S. Tsuge, S. Kuroiwa and M. Fukumi. *A New Geometric Distance Method to Remove Pseudo Difference in Shapes*, IJAI, Volume 2, Number 1, pp.119-144, July, 2010.
20. M. Jinnai, N. Boucher, M. Fukumi and H. Taylor. *A New Optimization Method of the Geometric Distance in an Automatic Recognition System for Bird Vocalisations*, Proceedings of the Acoustics 2012 Nantes Conference, 105
21. HTK Team in Cambridge University Engineering Department. *HTK Speech Recognition Toolkit (The Hidden Markov Model Toolkit)*, http://htk.eng.cam.ac.uk/
22. F. Ren. *From Cloud Computing to Language Engineering, Affective Computing and Advanced Intelligence*, IJAI, Volume 2, Number 1, pp.1-14, July, 2010.
23. C. Quan and F. Ren. *Sentence Emotion Analysis and Recognition Based on Emotion Words Using Ren-CECps*, IJAI, Volume 2, Number 1, pp.105-117, July, 2010.
24. F. Ren. *Invited paper, Robotics Cloud and Robotics School*, Proc. 7th Int. Congr. IEEE NLP-KE, pp.1-8, 2011.
25. F. Ren. *Affective Information Processing and Recognizing Human Emotion*, Electronic Notes in Theoretical Computer Science, Vol.225, No.2009, pp.39-50, 2009.

Michihiro Jinnai (Member)

He received the B.S. degree in seismology from Kyoto University, Japan, the M.E. and Ph.D. degrees in speech recognition from Kobe University, Japan, in 1976, 1980, and 1983, respectively. From 2006 to 2012, he was a Professor of the Department of Electro-Mechanical Systems Engineering at Kagawa National College of Technology. Since 2012, he has been with Nagoya Women's University, Japan, where he is currently a Professor of Faculty of Human Life and Environmental Sciences. His research interests include similarity measure and pattern matching. He has been developing the application software with geometric distance. It is used for detecting bird call, bat call, and whale call in Australia.

Satoru Tsuge

Satoru Tsuge received his B.E., M.E., and Dr. Eng. degrees from the University of Tokushima, Tokushima in 1996, 1998, and 2001, respectively. From 1997 to 1999, he was an intern researcher at ATR Interpreting Telecommunications Research Laboratories, Kyoto. From 2000 to 2009, he was a lecturer at the University of Tokushima. Since 2010, he has been with Daido University, Japan, where he is currently an Associate Professor of Department of Information Systems. His current research interests include speech recognition, speaker recognition, and information retrieval. He is a member of IPSJ and ASJ.

Shingo Kuroiwa

He received the B.E., M.E. and D.E. degrees in electrocommunications from the University of Electro Communications, Tokyo, Japan, in 1986, 1988, and 2000, respectively. From 1988 to 2001 he was a researcher at the KDD R & D Laboratories. From 2001 to 2007, he was an Associate Professor of Institute of Technology and Science at the University of Tokushima, Japan. Since 2007, he has been with Chiba University, Japan, where he is currently a Professor of Graduate School of Advanced Integration Science. His current research interests include speech recognition, speaker recognition, natural language processing, and information retrieval. He is a member of the IEICE, IPSJ, and ASJ.

Fuji Ren (Member)

He received the Ph.D. degree in 1991 from Faculty of Engineering, Hokkaido University, Japan. He worked at CSK, Japan, where he was a chief researcher of NLP. From 1994 to 2000, he was an associate professor in the Faculty of Information Sciences, Hiroshima City University. From 2001 he joined the faculty of engineering, the University of Tokushima as a professor. His research interests include Natural Language Processing, Artificial Intelligence, Language Understanding and Communication. He is a member of IEICE, CAAI, IEEJ, IPSJ, JSAI, AAMT, and a senior member of IEEE. He is a fellow of the Japan Federation of Engineering Societies.

Minoru Fukumi

Minoru Fukumi received the B.E. and M.E. degrees from the University of Tokushima, in 1984 and 1987, respectively, and the doctor degree from Kyoto University in 1996. Since 1987, he has been with the Department of Information Science and Intelligent Systems, University of Tokushima. In 2005, he became a Professor in the same department. He received the best paper award from the SICE in 1995 and best paper awards from some international conferences. His research interests include neural networks, evolutionary algorithms, image processing and human sensing. He is a member of the IEEE, SICE, IEEJ, IPSJ and IEICE.