Characteristics of Cow's Voices in Time and Frequency Domains for Recognition

Y. Ikeda, Y. Ishii

Abstract: On the assumption that the voices of the cows are produced by the linear prediction filter, we characterized the cows' voices. The order of this filter was determined by examining the voice characteristics both in time and frequency domains. The proposed order of the linear prediction filter is 15 for modeling voice production of the cow. The characteristics of the amplitude envelope of the voice signal was investigated by analyzing the sequence of the short time variance both in time and frequency domains, and the new parameters were defined. One of the coefficients of the linear prediction filter generating the voice signal, the fundamental frequency, the slope of the straight line regressed from the log-log spectra of the short time variance and the coefficients of the linear prediction filter generating the sequence of the short time variance of the voice signal can differentiate the two cows.

Keywords: Voice analysis, Cow, Linear prediction filter, Maximum entropy method, Formant, Recognition

Introduction

It is conjectured that the voices of the animals are produced for communication with the environmental others, and contains a good deal of valuable information about their mental conditions as well as physical states(Bradbury et al., 1998, Owings et al., 1998, Hopp et al., 1998), and experienced herdsman may identify problems through animal vocalization(Xin et al., 1989). It is pointed out that the mechanization of livestock farming should be realized under considering animal welfare, for example without stress and invasion to animals (Fukukawa, 1994). One powerful technique for monitoring animals without invasion may be to use machine vision (Stuyft et al., 1991, Onyango et al., 1995). Another method is to analyze the voice of animal since voice may contain information on animal states (Xin et al., 1989, Kashiwamura et al., 1985, Ikeda et al., 1991, Tamaki et al., 1993, Jahns et al., 1997, 1998).

Variation of voice characteristics of animal has two aspects, that is, one is among animals and another is within animal. The former could be utilized for recognition of individual animal, and the latter for monitoring of animal conditions such as hunger, oestrus and sickness. So, the voice may be utilized as the signal of the passive sensor for monitoring the animal conditions.

Therefore, understanding the animal voice exactly is very important for automatization and preciseness of animal husbandry to realize the animal welfare through the noninvasive and non-contact sensing and to release the farmers from the twenty-four hours labor. Moreover, the precision breeding of the animals may realize supply of the animal production of higher quality and safty.

The objectives of the present research are to characterize the cows' voices in the frequency domain as well as in the time domain, and to investigate applicability of the voice characteristics to recognition of the individual cow, using the cows of minimum number, that is *two*, for recognition. And the final target of this research is to understand the autonomous behavior of the animals having the central nervous system.

Experimental Conditions

1. Cattle Conditions used for Experiments

In this research, we measured the voices of six cows and used two cows of the high level voices for analysis as shown in Table 1. These cows were bred in the barn, and breed was *Japanese Black*.

The authors are **Yoshio Ikeda**, Professor, Dept. of Bio-production Technology, Div. of Environmental Science and Technology, Graduate School of Agricultural Science, Kyoto University, Japan. **Y. Ishii**, Research Engineer, Institute of Hyper-media, Sanyo Electric Ltd. Co., Japan. **Corresponding author**: Yoshio Ikeda, Professor, Dept. of Bio-production Technology, Div. of Environmental Science and Technology, Graduate School of Agricultural Science, Kyoto University, Kyoto, 6068502 Japan; e-mail:ikeda@e7sun-1.kais.kyoto-u.ac.jp

Table 1 Condition of Cows for Analysis

	ID No. of Cow	
	143	440
Age (Months)	133	78
Sexuality	Female	Female
Body Mass (kg)	717	540

2. Recording Voice Signal

The voices of cow were recorded from September 11 to 14 in 1995 at 7 to 8 o'clock in the morning before feeding, and then the length of total recording time was around 4 hours. The precision microphone (RION, NA-60) and digital audio tape recorder (Pioneer, DAT-05, 16 bits sampled at 96 kHz) was used, the microphone was set at a distance of 60 cm from the fence of pen (whose height was 150 cm) and at height of 100 cm above the floor. Cows might not be disturbed with these measuring devices and environment.

The weather were fine or cloudy, and so background noise was only the low level sound of circulating fans, and by inspection on the CRT monitor for reproducing the recorded sound, we could easily remove the metallic sound caused by bump of the fence and the cow from the voice signal.

3. Digitization and Rearrangement of Voice Signal

The digital data were converted to analog signal, and sampled again at the interval of 10 kHz as the digital signal of 8 bits. It is important to clip the complete voice signal from the stream of the recorded tape through detecting outset of the voice in order to realize the accurate measurement (Rabinar et al., 1993). By inspecting the reproduced analog signal on the CRT, the part of the voice signal was extracted manually, whose length was 5 s. This part contained whole voice and was converted to the digital data sampled at the interval of 0.0001 s. The computer program rearranged these raw data into the new data set sampled at the interval of 0.2 ms and clipped for 2 s. In this new data set, almost whole voice were encompassed.

Characteristics in Time Domain

1. Total Power of Voice

The loudness of the voice reflected the power of the voice. The value of variance of voice signal represents the total power of that signal (Deller et al., 1993).

The variances of the voice signal were calculated for the original digital data sequence sampled at 0.0001 s and clipped for 5 s, and the results are shown in Fig. 1 for two cows of the high level voice. In this figure, height of each column represents the total power of the one voce signal. This figure shows that the numbers of voices were different from animal to animal. The mean time between vocalization was about 8 min. for Cow 143 and 4 min. for Cow 440. The voices of the power higher than 0.1/s were used for analysis in this research, that is seven voices for Cow 143 and five voices for Cow 440.

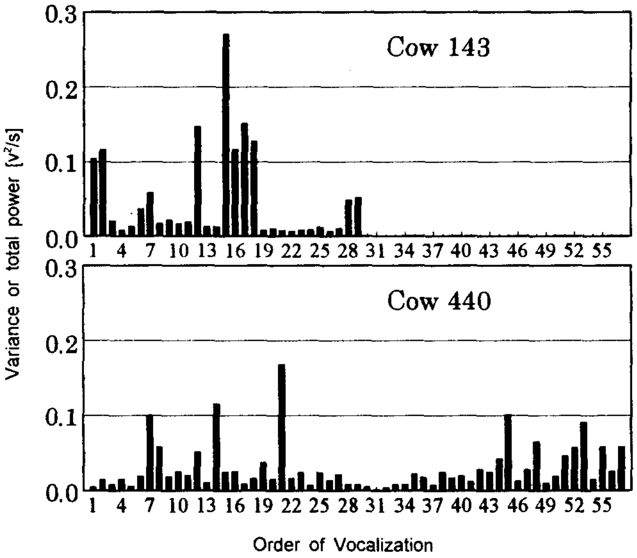

Fig. 1 Variance or total power of voice signal.

Voice Production Process

1. Linear Prediction Model for Voice Production

In acoustics, the all-pole model or autoregressive process describes the speech production process of mammals (Rabinar et al., 1993), and can be explained as the linear prediction filter. In this model, the present value of the signal $s(n)$ sampled at time n from the voice signal can be approximated with the linear combination of the past M sampled values as follows(Deller et al., 1993),

$$s(n) = \hat{a}_1 s(n-1) + \hat{a}_2 s(n-2) + \cdots\cdot + \hat{a}_M s(n-M)$$

where, the coefficients $\hat{a}_i$ (i=1,...M) are considered to be constant in the interval of analysis. Including the driving signal, the signal production process in the linear prediction method can be represented by

$$s(n) = \sum_{i=1}^{M} \hat{a}_i s(n-i) + \Theta' e(n) \ ,$$

where Θ' is the system gain. The power spectrum of the signal generated by this process is given by the mathematical expression using these coefficients $\hat{a}_i$ as follows[5].

$$p(f) = \frac{\Delta t P_M}{\left| 1 + \sum_{k=1}^{M} \hat{a}_k \, e^{j2\pi f k \Delta t} \right|^2},$$

where Δt is the sampling interval of the signal and P_M is the variance of the prediction error or the expected value of the prediction error given by

$$P_M = E[(x_t + \gamma_{M1} x_{t-1} + \gamma_{M2} x_{t-2} + \cdots + \gamma_{MM} x_{t-M})^2],$$

where γ_{Mi} is the estimated value of the i-th linear prediction coefficient when the order of prediction filter is M.

And we can suppose that the coefficients represent the degree of influence of the past values $s(n-i)$, for i=1,...,M, of the voice signal to the present value $s(n)$. The system gain can be interpreted as follows. The relative distance and attitude of the microphone to the sound source could not be controlled and not consistent, and the change in the signal level transferred to the microphone could be reflected on the system gain. Therefore, the system gain could compensate the effect of the measuring conditions such as distance and attitude. It might also be assumed that the values of the prediction coefficients were kept constant during given segment of one vocalization. Then, the dynamic characteristics of the voice production system may be described parametrically and the voice production system can be compared directly with the prediction filter coefficients $\hat{a}_i$. Moreover, the frequency characteristics of the voice signal can be represented by these parameters. In this paper, the maximum entropy method (MEM) and the Burg's algorithm estimated the prediction coefficients (Hino, 1986). Since it is important and difficult to determine the reasonable number M of terms of the autoregressive process (Deller et al., 1993) or filter order M, then in the next sections, we will discuss about this problem and attempt to determine the reasonable order of the linear prediction filter.

2. Determination of Order of Linear Prediction Model

(1) Final Prediction Error of Linear Prediction Filter

Usually, the order M of the linear prediction filter should be determined so as to minimize the final prediction error (FPE) of the measured signal and predicted one (Akaike et al., 1975). Fig. 2 shows the final prediction errors of the linear prediction filter. In this research, the voice signal was divided into some segments each of which had the length of 0.2s (equivalent to 1,000 data points), and for each segment, the variance was calculated. In this computing process, the adjacent segment was obtained by shifting 0.1s (equivalent to 500 data points), and the computation was repeated. So, the number of segments for one voice was 18. The computed results for the segment that had the maximum variance will be examined hereafter. For both the Cows 143 and 440, the values of FPE stopped decreasing at 15-th or 20-th order. After 21-st order, the FPE's scarcely decreased for some voices. Then, from the point of view that the proper order of the filter should be determined as the order at which the FPE reaches the minimum value, we could not determine the optimal order of the prediction filter. However, it may be reasonable to adopt the order between 15 and 20, considering the rate of decrease of the FPE's or increase of accuracy of prediction of the present output.

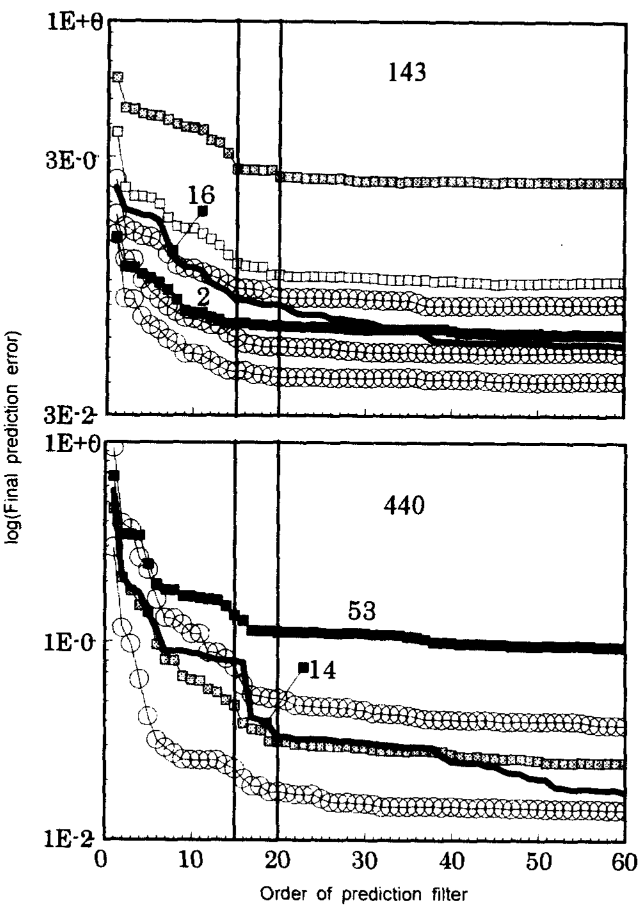

Fig. 2 Final prediction error (FPE) of 60th order linear prediction filter.

(2) Linear Prediction Coefficients

The values of coefficients of the filter can be interpreted as the degree of influence of the past output signals to the present one. Therefore, when the value of the coefficient on and after the certain order becomes almost zero, the order of the filter should be specified at that order. Based on this hypothesis, the behavior of the values of the coefficients may give some information on decision of the appropriate order of the model. Fig. 3(a) to (d) show the values of the prediction coefficients of the filter for the voices 02 and 16 of the cow 143, and the voices 14 and 53 of Cow 440, respectively. For the voice 02 of Cow 143, the values of the coefficients are small after the order of 20. However, in voice of 16, the around 40 past output signals have the effects to the present signal. For the voice 53 of Cow 440, the values of the prediction coefficient are not so small even beyond 20-th, and do not converge to zero at the 60-th order. That is, in case of voice 53 of the cow 440, the present output is affected by considerably past output values. From above discussion, the range of the filter order may exceed 30 significantly. From the viewpoint of the human speech analysis, in Nyquist band after sampling 3 to 5 formants are found in practice (Rabinar et al., 1993). Based on the vocal-tract transfer function, each pair of poles in the z-plane at complex conjugate locations roughly corresponds to a formant in the spectrum. If the poles in the z-plane are well separated, poles in the upper half of the z-plane give good estimates for formant frequencies (Deller et al., 1993). That is, the number of terms of the filter is twice of the number of formant. Then, in this context the reasonable number of order of the filter may be 6 to 10. On the other hand, in human speech-recognition application, it is generally acknowledged that the values of *M* are on the order of 8 to 10(Deller et al., 1993). It is said that as *M* increases, more of detailed properties of the signal spectrum are preserved in the LP spectrum, and that beyond some value of *M*, the details of the signal spectrum that are preserved are generally irrelevant ones. Based on these discussions, the order of the filter may be 20 at most in this research.

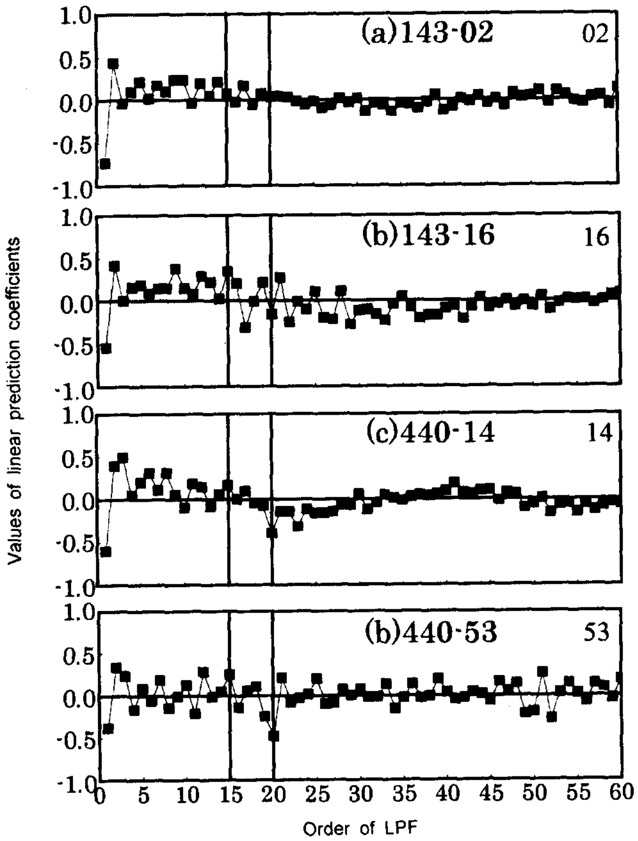

Fig. 3 Values of prediction coefficients of linear filter for (a) Cow 143 and (b) Cow 440.

Characteristics in Frequency Domain

1. Measured and Estimated Fundamental Frequencies

The longest noticeable wave length was measured on the auto-correlation function of the voice signal, and reciprocal of the length was considered as the measured fundamental frequency. The estimated fundamental frequency was defined as the lowest frequency where the power has the maximum or locally maximum value on the spectral distribution. The fundamental frequency on the estimated spectra was changed with the order of the linear prediction filter as shown in Fig. 4. For the order lower than 14, the estimated values were far from the measured values, and the range of spread of the values became wider for lower order. For the order higher than 25, the estimated values depart from the measured ones in case of Cow 143. The estimated values for Cow 440 were stable for the wide range of the filter order. From these figures, the order of the filter should be greater than 15, and in this research we chose the minimum number of 15 as the order of the linear prediction filter.

2. Estimation of Power Spectra by FFT and LPF

The power spectra were estimated by FFT to

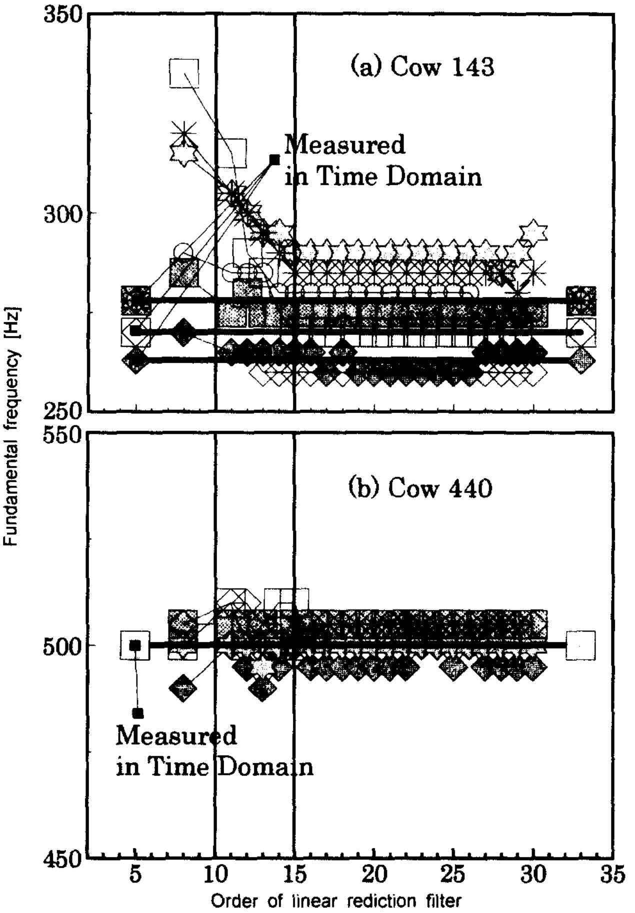

Fig. 4 Change in estimated fundamental frequencies with order of linear prediction filter.

examine the influence of the number of coefficients of the filter to similarity of spectra estimated by two different kinds method. In FFT, the number of data points was 1024 and Akaike's spectral window was used in the frequency domain to remove the ripples from the raw spectrum (Akaike et al., 1975). As shown in Fig. 5, the main peaks of the 15-th order LPF spectra coincided with those of the FFT spectra. The spectra estimated by the 60-th order filter traced the FFT spectra as a whole, however for Cow 440 the LPF spectra did not coincide with the FFT spectra at the antiresonance frequencies. Considering the number of formant of the animal voice as mentioned above, the 60-th order LPF may be excessively high order. In the followings, we will use the 15-th order LPF to discuss the voice characteristics of the cows.

3. Change in Formants among Voices In Table 2

The formant frequencies of each voice of individual cow are summarized. The range of variation of the 1-st and 2-nd formants of Cow 440 were narrow, that is, these formant frequencies were stable. The 1-st and 3-rd formants of Cow 143 are somewhat stable and 2-nd formant of Cow 143 not so stable. The harmonic structure could be acknowledged for both cows. The fundamental frequencies of Cow 440 had the varying range of 10 Hz and those of Cow 143 did 25 Hz. That is, it may be said that the voice of Cow 440 was more stable than that of Cow 143.

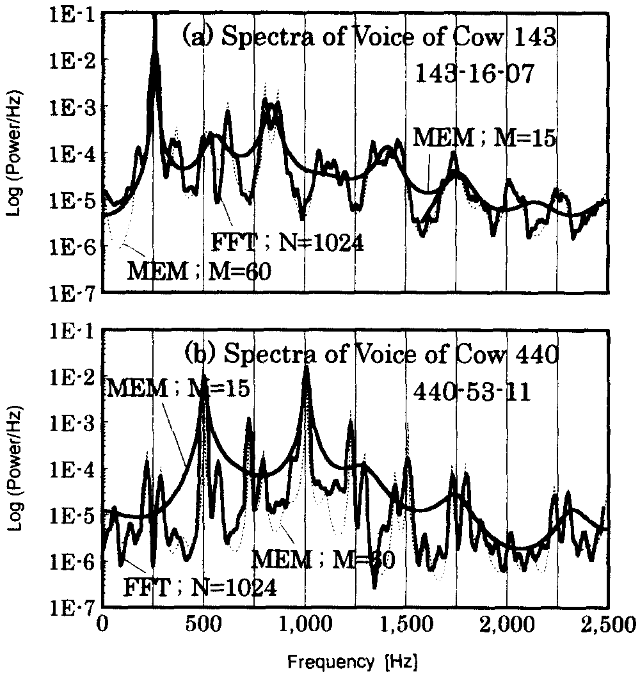

Fig. 5 Power spectra by FFT and LPF.

Table 2 Formant frequencies of vocalizations of cows

	Cow 143		
	1-st Formant	2-nd Formant	3-rd Formant
Mean	275Hz	505Hz	820Hz
f_i/f_1	1.0	1.8	3.0
Range	25Hz	100Hz	55Hz
	Cow 440		
Mean	505Hz	1010Hz	1480Hz
f_i/f_1	1.0	2.0	2.9
Range	10Hz	5Hz	60Hz

Fluctuation Characteristics of Short Time Variance

1. Characteristics in Time Domain

The temporal change of the short time variance (or total power) of the voice signal could approximate the upper half of the amplitude envelope of the voice signal. Consequently, the waveform of the temporal change may contain any information on the voice characteristics. Fig. 6 shows the examples of temporal change of the short time variances, and it can be observed that the waveforms are considerably different besides the difference of duration time of vocalization. The interval of integration for the variance was 2.5 ms and the sampling interval of the original voice data was 0.1 ms, then the data points of the short time was 25. This remarkable dissimilarity of the waveform may be used for recognition of two cows.

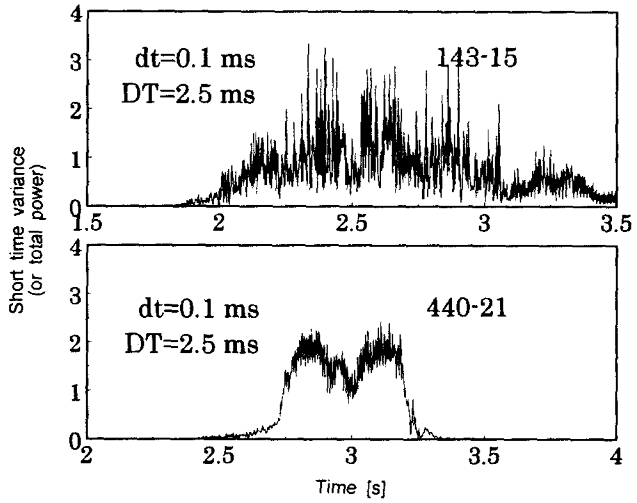

Fig. 6 Temporal change of short time variance.

To examine the characteristics in the frequency domain, the power spectra are estimated by FFT and MEM. For the computation, the data points sampled at the interval of 0.2 ms were used, and the interval for integration of the variance was 1 ms. The number of data points were 2048 for FFT and 2000 for MEM. Therefore, the data lengths were 2.048 s and 2 s, respectively. These data lengths contain the whole vocalization. To estimate the power spectrum by MEM, it is necessary to determine the order of the linear prediction filter. In this section, the order iswas determined by examining the similarity of the spectra estimated by FFT and MEM. As shown in Fig. 7, the spectra estimated with the 15-th order linear prediction filter can trace the FFT spectra more precisely than the 7-th order filter. Hereafter, to estimate the MEM spectrum, the 15-th order filter will be used for the analysis of the temporal change of the short time variance.

The 15 coefficients of the 15-th order filter are shown in Fig. 8(a), and the coefficients of the even-numbered order in (b). The 2nd, 4th, 6th, 8th and 12th prediction coefficients can be separated without overlap, that is these coefficients can be used to recognize two cows.

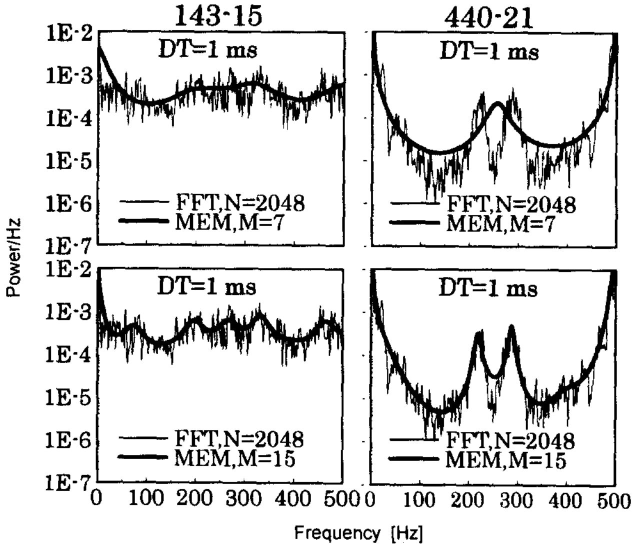

Fig. 7 Power spectra of temporal change of short time variace.

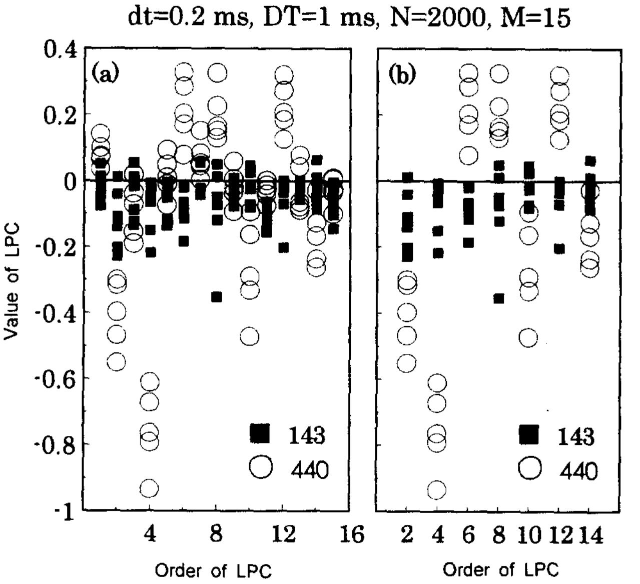

Fig. 8 Values of linear prediction coefficient.

2. Characteristics in Frequency Domain

The information based on the loudness or the amplitude of the voice signal should not be used, to avoid the effect due to the relative attitude and distance between the sound source and the microphone. The slope of the spectra drawn on the logarithmic coordinates of both frequency and power, however, can cancel the effect of the amplitude difference due to the relativity of the sound source and receiver.

In Fig. 9(a), the spectra show the aliasing effect in the high frequency band, therefore the frequency range higher than 50 Hz was truncated, and the data points between 0.4 Hz and 50 Hz were used to fit the straight line. This processing had an effect to increase the correlation coefficient of the straight line regression of the log-log spectra as shown in Fig. 9(b). Moreover, the slope of the regressed straight line reveals the complexity of the amplitude envelope. The negative infinite slope means that the spectrum contains the unique line spectrum at or near zero Herz, and the sequence of the short time variance does not fluctuate temporally, that is the amplitude envelope of the voice signal does not fluctuate. When the slope is zero, the temporal change of the short time variance is random. These two cases are the extremum two modes, and usually the amplitude of the voice signal fluctuate within these two modes. Fig. 10 shows the slopes of the regression lines, and it may be said that the slope can be used as the feature parameter for recognition of the voices of two cows.

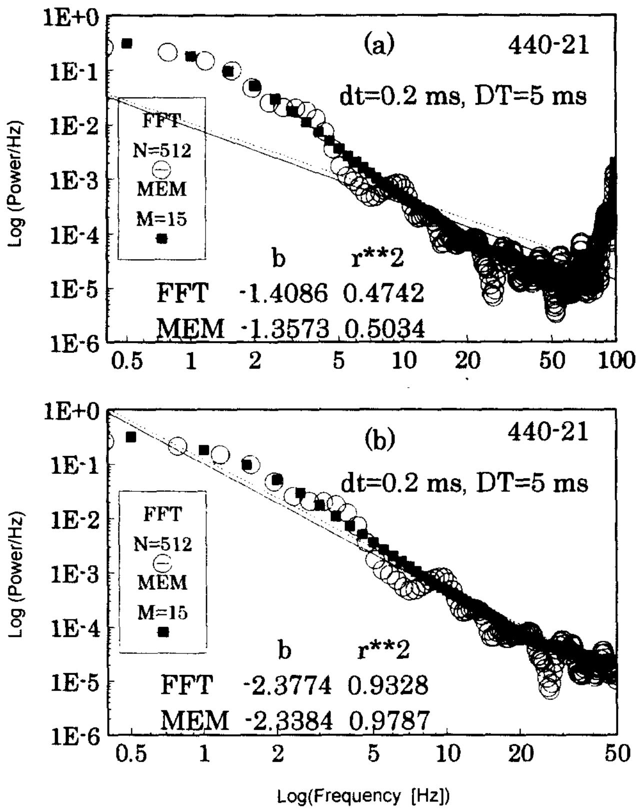

Fig. 9 Spectra of short time power of voice signal.

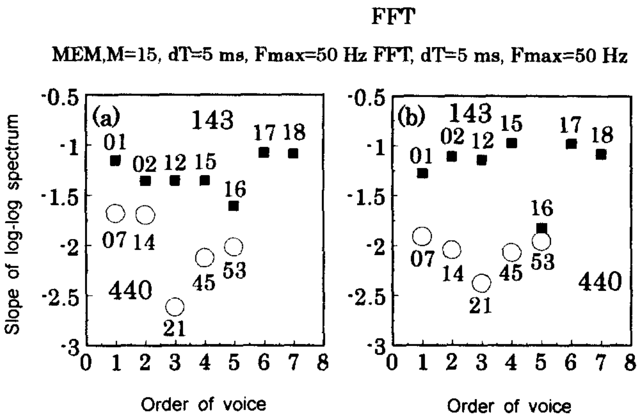

Fig. 10 Slope of log-log spectra of short time power.

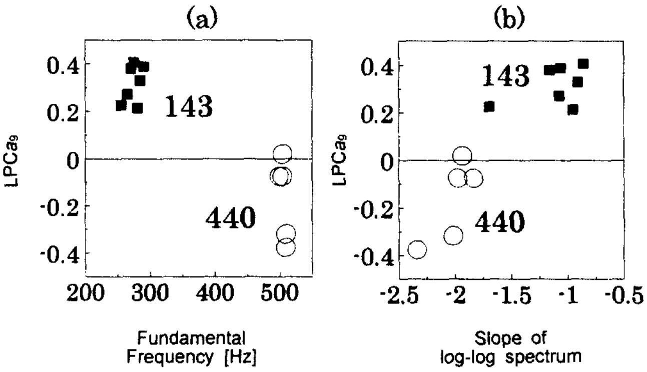

Fig. 11 Recognition of two cows by fundamental frequency, slope of log-log spectrum and prediction coefficient.

Recognition of Individual Cow

1. Recognition by Fundamental Frequency and Prediction Coefficients

As mentioned previously, the fundamental frequencies of two cows had the fairly different value, therefore these might be used as the possible element of the feature vector. Fig. 11(a) shows the two-dimensional feature space, whose elements are the 9th prediction coefficient and the fundamental frequency. It is noticeable that two cows can be recognized completely by the hyper plane (that is straight line) on the 2-D plane.

2. Recognition by Prediction Coefficients and Slope of log-log Spectra of Short Time Variance

In the frequency spectrum, the power corresponding the fundamental frequency was not necessarily highest as shown in Fig. 5(b), where the power was highest at the second formant. Therefore, it is necessary to examine another feature parameters. In Fig. 11(b), feature space consisting of the 9th prediction coefficient and the slope of log-log spectra of the temporal change of the short time variance is shown, and the two cows could be differentiated completely. These two feature parameters were not affected by the effect of the relative distance and attitude between the sound source and receiver.

3. Recognition by Prediction Coefficients of Short Time Variance

Since the 4th coefficients of the prediction filter describing fluctuation of the short time variance can be separated most clearly between two cows as shown in Fig. 8(b), the combination of these 4th coefficients and the other coefficients separated into two groups based on the individual cow can be the promising feature space for recognition. On both two feature spaces composed of the a_4-a_2 and a_4-a_6 coefficients as shown in Fig. 12, two cows can be separated completely.

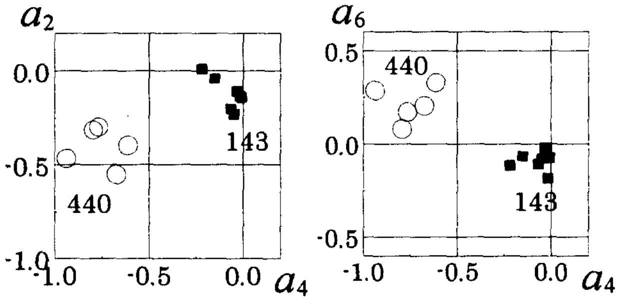

Fig. 12 Recognition by LPC of temporal change of short time variance.

Conclusions

The voice characteristics of two adult cows of Japanese Black were examined in the domains of both time and frequency. As a principle, the length of data sequence was 2 s and data sampling interval 0.2ms. The reasonable order of the linear prediction filter was examined based on the some aspects. The following results were obtained in the time domain.

1) The numbers of voices were different between cows, and

2) the total voice powers were different within a cow.

In the following discussion, the attention was focused on the segment of the maximum power or variance of length 0.1 s.

1) Based on the FPE values themselves and the rate of decrease of them, the reasonable order of the linear prediction filter may be between 15 and 20.

2) The values of the prediction coefficients stopped changing at the 20-th to 30-th depending on the voices and cows. Then, the reasonable order of the filter may be 20 at most.

3) Based on comparing the fundamental frequencies estimated on the autocorrelation function and the power spectra, the minimum order of the filter might be 15.

Considering the number of formants in the human speech recognition problem, the order of the filter in this research was determined to be 15. The results in the frequency domain are as follows.

1) As the filter order increased, the number of the peaks of spectra or formants also increased. The spectra estimated by LPF of order 60 traced those estimated by FFT except around the anti-formant frequencies.

2) The voices of two cows had three formants and they composed the harmonic overtones. The average fundamental frequencies were 275 Hz for Cow 143 and 505 Hz for Cow 440.

3) Two cows could be recognized with the prediction coefficient of the voice signals, the slope of the log-log spectra of the temporal change of the short time variance, the fundamental frequency of the voice, and the linear prediction coefficients for the temporal fluctuation of the short time variance.

The future problems to be solved in recognition are

1) to examine the voices of low level, 2) to investigate the more cows than three, and 3) to investigate applicability of the features examined in this paper, that is, the prediction coefficients, slope of the spectra of for the cases of the low level voices and a lot of cows.

Recognition of the physical conditions should be investigated by using the voices. In such a case, we should ask for the help of animal psychologists and animal ecologists, zoologist or theriatrics etc.

Remarks

This paper was revised and remade from the paper published on the Journal of the Japanese Society of Agricultural Machinery(Y. Ikeda et al., 2001)

References

Akaike, K. and T. Nakagawa. 1975. Statistical Analysis and Control of Dynamic System(in Japanese).

Bradbury, L. and S. L. Vehrencamp. 1998. Principles of Animal Communication, Sinauer Associates, Inc.

Deller, J. R., J. G. Proakis and J. H. L. Hansen. 1993. Discrete-Time Processing of Speech Signals, Macmillan.

Fukukawa, K. 1994. New Technology for Livestock Breeding and Management (in Japanese), J. of JSAM, Vol. 56, No. 2, 167-172.

Hino, M. 1986. Spectral Analysis (in Japanese), Asakura Book Co.

Hopp, S. L., M. J. Owren and C. S. Eds. Evans. 1998. Animal Acoustic Communication, Springer.

Ikeda, Y. and J. Shimada. 1991. Application of Acoustic Emission in Agricultural Mechanization -Frequency Characteristics of Livestock- (in Japanese), J. of JSAM Kansai Branch, No. 70, 123-124.

Ikeda, Y. and Y. Ishii. 2001. Appreciation of Cow's Voice by Computer, -Application of Sound in Agriculture (Special Edition), J. of JSAM, Vol. 63, No. 1, 4-9.

Jahns, G., W. Kowalczyk and K. Walter. 1997. An Application of Sound Processing Techniques for Determining Condition of Cow, Proc. of 4-th Intl. Workshop on Systems, Signals and Image Processing.

Jahns, G., W. Kowalczyk and K. Walter. 1998. Sound Analysis to Recognize Individuals and Animal Conditions, Proc. of VIII CIGR Congress on Angric. Engng.

Kashiwamura, F. and M. Yamamoto. 1985. The Classification and Analytical Method of the voice of Cattle (in Japanese), Livestock Management, Vol.21, 73-83.

Onyango, C. M., J. A. Marchant and B. P. Ruff. 1995. Model Based Location of Pigs in Scenes, Comp. and Electronics in Agric., Vol. 12, 261-273.

Owings, D. H. and E. S. Morton. 1998. Animal Vocal Communication -A New Approach-, Cambridge Univ. Press.

Rabinar, L. and B. H. Juang. 1993. Fundamentals of Speech Recognition, Prentice Hall.

Tamaki, K., K. Susawa, R. Otani and K. Amano. 1993. Characteristics of Cattle Voices and the Possibility of Their Discrimination (in Japanese), Research Bull. No. 158 of the Hokkaido Nat. Agric. Exp. Station, 1-11.

Van der Stuyft, E., C. O. Schofield, J. M. Randall, P. Wambacq and V. Goedseels. 1991. Development and Application of Computer Vision System for Use in Livestock Production, Comp. & Elec. in Agric., Vol. 6, 243-265.

Xin, H., J. A. DeShazer and D. W. Leger. 1989. Pig Vocalization Under Selected Husbandry Practices, Trans. of ASAE, Vol. 32, No. 6, 2181-2184.