CHARACTERISTICS OF COW'S VOICES IN TIME AND FREQUENCY DOMAINS FOR RECOGNITION

Y. Ikeda[1] and Y. Ishii[1]
[1]Department of Environmental Science and Technology,
Graduate School of Agriculture, Kyoto University, Kyoto, 6068502 Japan
E Mail ; ikeda@e7sun-1.kais.kyoto-u.ac.jp

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 is determined by examining the voices 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 combination of the two parameters of the fundamental frequency, the slope of the straight line regressed from the log-log spectra of the amplitude envelope and the only one coefficient involved in the linear prediction filter can differentiate the two cows.

[Keywords] Voice Analysis, Cow, Linear Prediction Filter, MEM, Formant, Recognition

1. 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, 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.

The objectives of the present research are to characterize the cow 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.

2. Experimental Conditions

2.1 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 are bred in the barn, and breed is *Japanese Black*.

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

2.2 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 former was set at a distance of 60 cm from the fence of pen (whose height was 150 cm) and at height of 100 cm on 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 we could easily remove the metallic sound caused by bump of the fence and the cow.

2.3 Digitization ad 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 voices 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

3. Characteristics in Time Domain

3.1 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

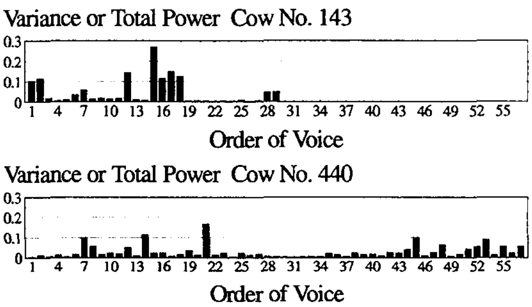

Fig. 1 Total Power of Voice Signal of Cattle

for 5 s, and the results are shown in **Fig. 1**. 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 were used for analysis in this research, that is seven voices for Cow 143 and five voices for Cow 440.

4. Voice Production Process

4.1 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. The present value $s(n)$ 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),

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

where, the coefficients $\hat{a}_i\,(i=1,\ldots,M)$ 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 1' is the system gain. The power spectra of the signal generated by this process are given by the mathematical expression using these coefficients $\hat{a}_i$. And we can suppose that the coefficients represent the degree of influence of the past values $s(n-i)$, for $i=1,\ldots,M$, 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 involve 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 by the 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.

4.2 Determination of Order of Linear Prediction Model

4.2.1 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, 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) 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, as far as the minimum values of the FPE are concerned, we could not determine the optimal number of the prediction coefficients. 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.

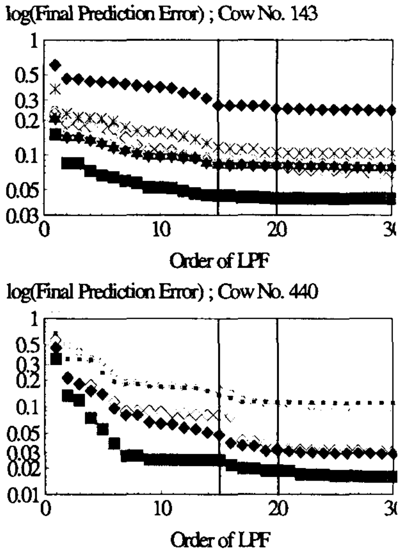

Fig. 2 Final Prediction Error (FPE) of 60th Order Linear Prediction Filter

4.2.2 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 at the certain order becomes almost zero, the order of the model 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) and (b) show the values of the prediction coefficients of the filter for the voices 02 and 16 of the cow 143, and the voices 45 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 30 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

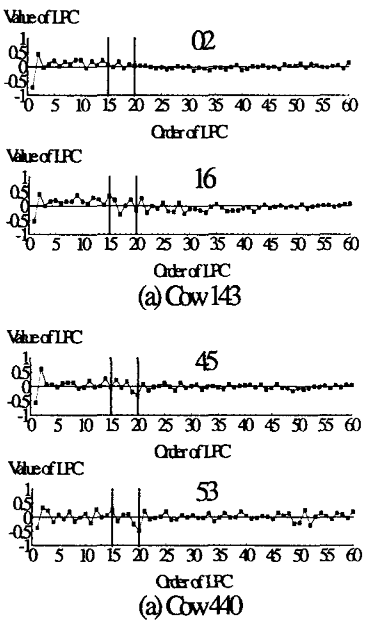

Fig. 3 Values of Prediction Coefficients of Linear Filter for the (a) Cow 143 and (b) Cow 440

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 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.

5. Characteristics in Frequency Domain

5.1 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 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 less 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.

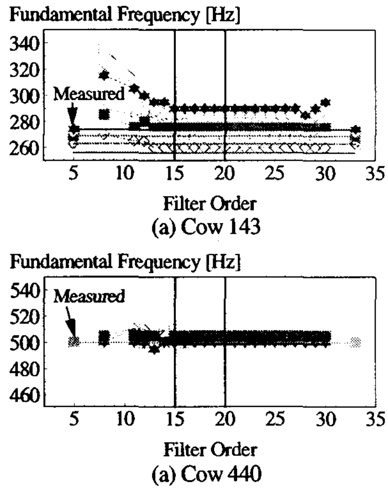

Fig. 4 Change in Estimated Fundamental Frequencies with Order of Linear Prediction Filter

5.2 Estimation of Power Spectra by FFT and LPF

The power spectra were estimated by FFT to examine the influence of the number of coefficients of the filter to similarity of two kinds of spectra. In FFT, the number of data points was 1024 and Akaike's spectral window was used 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

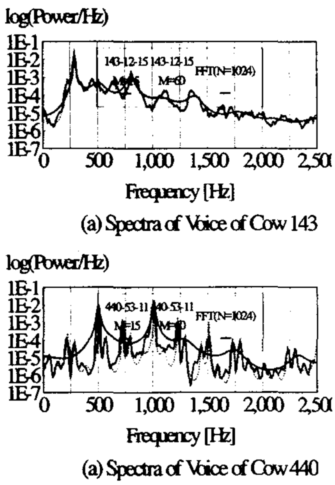

Fig. 5 Power Spectra by FFT and LPF

spectra did not coincide with the FFT spectra at the antiresonance frequencies. Considering the number of formant of the animal voice, 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.

5.3 Change in Formants among Voices In **Table 2,** the formant frequencies of each voice of individual cow are summerized. 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 say that the voice of Cow 440 was stabler than that of Cow 143

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

6. Fluctuation Characteristics of Amplitude Envelope

The amplitude envelope of the voice signal was approximated by the short-term variance of the voice signal. The length of the short-term was 5 ms, then the sampling interval of the sequence of variances was 5 ms. In **Fig. 6(a)**, the spectra show the aliasing effect in the high frequency band, then the frequency range higher than 50 Hz was truncated to increase the correlation coefficient of the straight line regressing the log-log spectra as shown in (b). **Fig. 7** shows the slopes of the regression lines, and it may be said that the slope can be used as the feature parameter for differentiation of the voices of two cows.

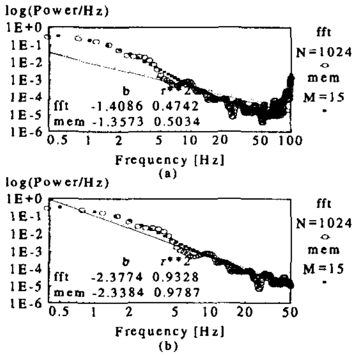

Fig.6 Spectra of Amplitude Envelope of Voice Signal

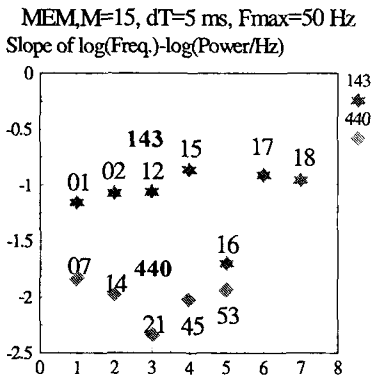

Fig. 7 Slope of log-log Spectra of Amplitude Envelope

7. Recognition of Individual Cow

7.1 Recognition by Fundamental Frequency 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. 8** shows the two-dimensional feature space, whose elements are the ninth prediction coefficient and the fundamental frequency. It is noticeable that two cows can be recognized perfectly by the hyper plane (that is straight line) on the 2-D plane.

7.2 Recognition by Prediction Coefficients and Slope

In **Fig. 9**, feature space consisting of the ninth prediction coefficients and the slope of log-log spectra of the amplitude envelope is shown, and the two cows can 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.

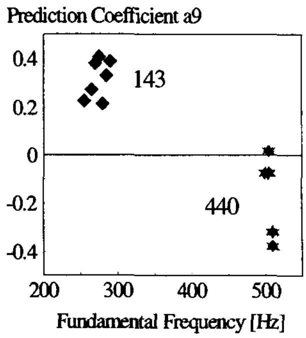

Fig. 8 Recognition of Two Cows by Fundamental Frequency and Prediction Coefficient

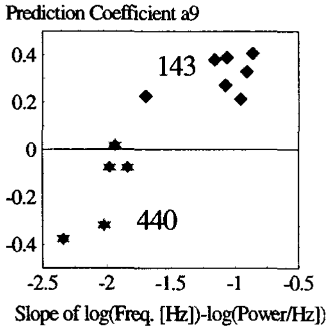

Fig. 9 Recognition of Two Cows by Slope of log-log Spectra and Prediction Coefficient

7. Conclusions

The voice characteristics of two adult cows *Japanese Black* were examined in the domains of 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 the total voice powers were different within cow.

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

2) Based on the FPE values themselves and the rate of decrease of them, the reasonable order of the linear prediction filter may be 15 for Cow 143 and 20 for Cow 440.
3) 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 for Cow 143 and 30 for Cow 440.
4) 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 two parameters of one prediction coefficient, the slope of the log-log spectra of the amplitude envelope, and the fundamental fre-

quency of the voice.

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 research, that is, the prediction coefficients, slope of the spectra of the amplitude envelope and fundamental frequency 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.

REFERENCES

[1] **Akaike,K.**, **Nakagawa,T.**(1975).Statistical Analysis and Control of Dynamic System(in Japanese) Associates, Inc.

[2] **Bradbury,L.** and Vehrencamp,S.L.(1998).Principles of Animal Communication, Sinauer

[3] **Deller,J.R.**, Proakis,J.G. and Hansen,J.H.L.(1993).Discrete-Time Processing of Speech Signals, Macmillan J. of JSAM, Vol.56, No.2, 167-172

[4] **Fukukawa,K.**(1994).New Technology for Livestock Breeding and Management(in Japanese),

[5] **Hino,M.**(1986).Spectral Analysis (in Japanese), Asakura Book Co., Springer

[6] **Hopp,S.L.**, Owren,M.J. and Evans,C.S.,Eds.(1998).Animal Acoustic Communication,

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

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

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

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

[11] **Onyango,C.M.**, Marchant,J.A., Ruff,B.P.(1995).Model Based Location of Pigs in Scenes, Comp. and Electronics in Agric., Vol.12, 261-273 Cambridge Univ. Press

[12] **Owings,D.H.** and Morton,E.S.(1998).Animal Vocal Communication-A New Approach-,

[13] **Rabinar,L.** and Juang,B.H.(1993).Fundamentals of Speech Recognition, Prentice Hall,

[14] **Tamaki,K.**, Susawa,K., Otani,R. and Amano,K.(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,

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

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