Analysis of speech signal based on frequency distribution and intonation

Naoya OHTA and Akio NOZAWA

Meisei University, 2-1-1 Hodokubo, Hino, Tokyo, JAPAN (Tel: +81-02-591-5895; Fax: +81-42-591-5895) (akio@ee.meisei-u.ac.jp)

Abstract: In this study, the atmosphere as a communication mode of human was defined by time information of speech. The objective of this study was the development of quantitative measurement technique of the atmosphere as KANSE I in a group. The communication mode of the group was indexed by using a time frequency distribution analysis on con versation voice. In particular, a speech duration time was focused. And the correlation between the time series of the sp eech duration time and a communication modality was evaluated. Consequently, the difference of the communication m ode between man and woman was seen in the speech duration time. And a difference of the characteristics of speech wa s found in the duration time and the frequency.

Keywords: atmosphere, communication, speech duration time, frequency

I. INTRODUCTION

The atmosphere of field was unexpressed explicitly in language of emotional relationship in situation of interpersonal relationship and social group, power relationship, and interest. Atmosphere was idiomatic phrase of Japanese which indicating many elements of relationship without expression. Awaking atmosphere is implicit. This ability is called social intelligence by psychology [1]. Reading atmosphere is cognition of communication mode. In particular, subtlety in communication is technique which learned in interpersonal psychology. That is, the skill of human relations is not innate target in interpersonal psychology.

In recent years, student evaluation system has been introduced into many schools [2]. The objective of the system is to find the problem of the class in order to refine the class. However, the questionnaire was written by the student whose grade was in the hands of teachers [3]. The evaluation by those students may tend to be rather high. Thus, the evaluation doesn't have much value for the objective. The introduction of the system in school is seen as just the evidence of an enthusiasm for refining the class. That is, an alibi. A basis of the evaluation has variety. The relationship between the effect of learning and the satisfaction for the class doesn't always has a correlation. Sometimes, they could have an inverse relationship. Because teachers tend to have the priority to the satisfaction of students and ignore the effect of learning. This is like putting the cart before the horse. The student evaluation system has the

aspects of understandability, an effect of learning, an ambition and an attitude of teacher, which are evaluated by students. Because of these factors, objective evaluation of the class becomes increasingly important. The activity of the class was evaluated as atmosphere. Understanding of refining of student evaluation system is considered become one of indicators.

In recent years, the various applications that was introduced the sound recognition technology was developed[4]. Speech dialogue system as target to information guidance was practical use. Many of those speech dialogue is remain handle only language information contained in speech. Therefore, uniform response is shown against any opponent. Dialogue between humans is and non speech information obtained from ears and eyes. There are also non-verbal information contained in speech. They are born deep interaction by use together. That is similar about recognition of communication mode including atmosphere. Factor to form an atmosphere was interpersonal relationship and emotional relationship and power relationship and interests relationship. Atmosphere is difficult captured properly by machine. In addition, atmosphere is difficult captured properly by human with no prior knowledge. But human is can catch sensuously an atmosphere. Because it is interpret atmosphere as communication mode of group based at nonverbal information of facial expression, tone, gesture not only explicit language information in content. Interpret of nonverbal information by machine is studied extensively by area of human-machine system.

Table1. Sta	ats of	speech	duration	time
-------------	--------	--------	----------	------

object	Freq/times	Total/s	Avg/s	Max/s	volume
news(M)	130.82	190.96	1.47	5.03	131.08
news(F)	152.14	195.41	1.34	4.70	98.45
Intr(F-M)	246.00	177.79	0.76	5.79	64.33
Intr(F-F)	235.13	174.66	0.80	6.01	66.79

Study example as object for communication mode of group is little. In addition interpret of verbal information concerning communication mode is become premise to interpretation of context whole of communication. Therefore complexity isn't avoided. This study, atmosphere as communication mode of between humans was defined by time information of speech. Purpose of this study is develop quantitative measurement technique of atmosphere as sensitivity of group. As result of this study is improvement of student evaluation system. Quantitative evaluation of the atmosphere is considered become one of evaluation indicator concerning communication mode as group. In addition, by communication mode visualization and quantify is function as one of indicator grasp familiarity density in talks. This is brings the efficiency of information transmission. This study was focused in relationship of speech duration time and speech frequency. This study is analysis of speech signal based on frequency distribution and intonation.

II. EXPERIMENT METHODOLOGY

In this study, conversation sound was used as analysis object. Recorded audio signal at PC was analyzed as digital data. Audio signal was converted by PCM method to A/D. Audio signal was recorded in file at WAVE form. sampling frequency was 22.050kHz. Recorded file was quantized at 16bit. In addition, the number of channels was one (monaural). Speech sound of record object was not included acoustic of sound effect except speech. The conversation to become object is idealized the speech of professionals trained concerning speech. But securement of subjects was difficult. So, audio source sample was acquired equivalent audio source from internet. Audio source sample is read sound of male announcer in radio news (News(M)) of internet delivery. And, read sound of female announcer in radio news (News(F)) of internet delivery. And, interview from female to male (Intr(F-M)) in internet delivery program. And, interview from female to female (Intr(F-F))in internet delivery program.



Fig.2. time series of speech duration time (News(F))

III. ANALYSIS PROCEDURE

Audio Source sample used at analysis was News(M) of 10kinds. And News(F) of 7 kinds. And Intr(F-M) of 6 kinds. And Intr(F-F) of 8 kinds. Audio source sample of news recitation is 17kinds in total. Audio source sample of interview is 14 kinds in total. Absolute value processing was administered for recorded voice signal. And get at envelope of peak value. Speech duration time and the corresponding speech time was recorded at time series. Voice signal per one sound source was 300 seconds. 10 seconds of first is silence part for environmental noise level evaluation. Extract of speech is performed as standard of environmental noise level. Threshold was about 3 times of the average environmental noise level. And tried visualization of communication mode by conversation sound analysis. This study was focused in speech duration time. The conversation breaks were detected as standards silence part of constant time. This study conversation break was regarded silence part more than 0.2 second.

IV. RESULT AND DISCUSSION

Table.1 is result of the stats concerning speech duration time. Fig.1 is result of time series of speech duration time (News(M)). Fig.2 is result of time series of speech duration time (News(F)). Fig.3 is result of time series of speech duration time (Intr(F-M)). Fig.4 is



Fig.4. time series of speech duration time(Intr(F-M))

result of time series of speech duration time (Intr(F-F)). Fig.5, 6 is result of Time frequency distribution of speech of news recitation and interview. Fig.7, 8, 9, 10 is result of each frequency analysis. From result of table 1, number of speech of men of the news recitation is 130.82 times. Number of speech of female of the news recitation is 152.14 times. Number of speech of female is lot of than men. Summation of time series of speech duration time is 190.96 second by news recitation of male. Summation of time series of speech duration time is 195.41 second by news recitation of female. Average of speech duration time is 1.47 second by news recitation of male. Average of speech duration time is 1.34 second by news recitation of female. Maximum of speech duration time is 5.03 second by news recitation of male. Maximum of speech duration time is 4.70 second by news recitation of female. Average of volume of speech duration time is 131.08 db by news recitation of male. Average of volume of speech duration time is 98.45 db by news recitation of female. Female is seen tend extension of speech duration time than men. For result of male of news recitation of time series of fig.1, amplitude is seen fit into within constant. In addition, male of news recitation of fig.1 is bigger than amplitude female. For result of female of news recitation of time series of fig.2, amplitude is seen fit into within constant. Amplitude of female of fig.2 is seen to smaller than amplitude of male of fig.1 Summation of speech duration time concerning news recitation wasn't seen both male and female significant difference. Both



difference of news recitation is showing at distribute of speech duration time. This report, news contents is not control. Therefore, difference between male and female in distribute of speech duration time is cannot disaffirmance possible of factor of contents. Result of news recitation of both male and female is seen constant amplitude within the range. Speech of news recitation is spoken by constant rhythm easy to hear. So, amplitude is considered become constant. Number of speech of interview of female for male is 5.79 times. Number of speech of interview of female for female is 6.01 times. Summation of time series of speech duration time of interview of female for male is 177.79 second. Summation of time series of speech duration time of interview of female for female is 174.66 second. Average of speech duration time of interview of female for male is 0.76 second. Average of speech duration time of interview of female for female is 0.80 second. Max of speech duration time of interview of female for male is 5.79 second. Max of speech duration time of interview of female for female is 6.01 second. Average of volume of speech duration time of interview of female for male is 64.33 db. Average of volume of speech duration time of interview of female for female is 66.79 db. Number of speech and max time of



Fig.8. frequency analysis (News(F))

interview is greater than the news recitation. But summation and average of speech duration time is smaller than the news recitation. Principal speech of conversation of interview is 2 man. And conversation of interview is accrual communication. And conversation is hard to break. And speech duration time is inferred extending. Accretion of principal speech is not bring accretion of speech time. Information concerning speech duration time as per the communication mode is indicating difference arise. From result of time frequency distribution of fig.5, 6, speech of short time is increased frequency of occurrence. Conversely long speech of duration time is low frequency of occurrence. Time frequency distribution of speech is inferred take on fractal structure. From result of frequency analysis of fig.7, 8, 9, 10, power spectrum density is inverse proportion f of frequency. And this is become relationship in 1/f fluctuation.

V. CONCLUSION

In this study, focusing at non-verbal part of speech. Object is to visualize communication mode from implicit called atmosphere. Conversation sound as analysis target was applied time frequency distribution. In particular, focused at speech duration time. And evaluate is concerning of time series of the speech duration time and the communication mode. As result, difference of speech mode concerning to sex was



acknowledged. In addition, speech frequency and speech duration time was ascertained possible indicate a fractal structure. Fractal dimension of news recitation of speech as analysis target is constant. News recitation is thinking became constant fractal dimension by spoke trained announcer. Fractal dimension in interview is multiplicity. Therefore Fractal dimension in interview is considered possible to be one of the indicators to capture the communication mode.

REFERENCES

[1]John D. Mayey, Peter Salovey, David R.Caruso : "Emotional Inteligence : Thaeory, Finding and Implications," Psychological Inquiry, Vol15, No3, pp.197-205(2004)

[2]Yumiko M, Kensaku S, Toshio O, Akihumi J, Akiko I, The Enforcement and the Tabulation of the Class Evaluation Questionnaire(in Japanese). Tokyo Management College, Vol.17, pp.47-58(2009)

[3]Yoshiaki H, Anonymous course evaluation and final evaluation of each class registered (in Japanese). International financial and administrative Vol.23, pp.47-58(2009)

[4]Yosuke A, Kiyoshi S, Atsushi K, A Study for a Scalability Evaluation Model of Spoken Dialogue System(in Japanese), Information Processing Society of Japan, Vol. 46, No.9, pp.2269-2278(2005)