The Audio Signal Classification System Using Contents Based Analysis

Kwang-Seok Lee, Young-Sub Kim, Hag-Yong Han and Kang-In Hur, *Member, KIMICS*

Abstract—In this paper, we research the content-based analysis and classification according to the composition of the feature parameter data base for the audio data to implement the audio data index and searching system. Audio data is classified to the primitive various auditory types. We described the analysis and feature extraction method for the feature parameters available to the audio data classification. And we compose the feature parameters data base in the index group unit, then compare and analyze the audio data centering the including level around and index criterion into the audio categories. Based on this result, we compose feature vectors of audio data according to the classification categories, and simulate to classify using discrimination function.

Index Terms—Signal Classification, Audio Characteristic Parameter, Audio Data Processing

I. INTRODUCTION

The large audio data base is various and it is getting more important gradually to manage the audio data base effectively, according to contents based analysis. Nevertheless, due to containing dynamic characteristics and variety of audio data, the research of the audio index and search of multimedia stream is poorer than that of contents based image and video data base. Because it is simple to segment data in image based processing for multimedia data owing to data meaning. However, the research of unified index and search of various multimedia factors is essential to solve in parsing video data. Therefore it is important to study contents based audio data. In this research, we expand basic classification category of audio form as speech and music, and we established a various sound data as speech with back music and song with back music in addition.[1]-[4],[7]

At first, after comparing and analyzing each category through the audio signal analysis for optimal index and search of the established data, and we propose a several feature parameters with upper scheme, compose feature parameters data base of proposed audio signal characteristics. And we simulated classification experiments for audio data using composed feature parameter data base. We introduce a various parameters of audio feature parameter based on audio signal content in section II, and also introduce extraction scheme due to computation method of feature parameter. In section III, we compose a new feature parameter data base, and explain data classification method using the static classification function classifier, and show simulation results. Finally, we conclude this research.

Manuscript received June 29, 2007.

Asterisk indicates corresponding author.

Kwang-seok Lee (phone: +82-55-751-3333, email: kslee@ jinju.ac.kr) is with the Department of Electronic Engineering, Jinju National University, Jinju, Korea.

Young-Sub Kim (phone: +82-51-200-6961, email: yskim770202@gmail.com)

Hag-yong Han (phone: +82-51-200-6961, email: hyhan88@ nate.com)

Kang-in Hur (phone: +82-51-200-7708, email: kihur@dau.ac.kr) is with the Department of Electronic Engineering, Dong-A University, Busan, Korea.

II. FEATURE PARAMETER OF AUDIO SIGNAL

We use the composed various feature parameters with audio signal characteristics in feature parameter data base in basic. In this section, also, we use general feature parameter in digital signal processing, and we specially propose a new parameter, i.e., harmonic degree (HD), frequency duration degree (FDD) and frequency convergence degree (FCD). And we simulated that how is effective to audio classification parameter.

A. Average energy in short duration

The audio signal energy is defined mean square energy in equation (1), and defined in equation (2) for signal containing threshold with margin.

$$E_n = \frac{1}{N}\sum_{m=0}^{N-1} X^2(m) \tag{1}$$

$$E_n(dB) = 10\log(E_n) \tag{2}$$

The audio signal energy has high value in a voiced and has low value in a unvoiced. i.e., energy is the effective factor containing threshold to classify signal data. Also, energy in audio signal containing music has higher value than speech signal. Therefore, we use energy as feature parameter to classify audio signal.[1]-[7]

B. Average zero crossing rates in short duration

Zero crossing rate (ZCR) is has a different sign in adjacent discrete signal, is defined as equation (3).

$$L_n = \frac{1}{2}\sum_{m=0}^{N-1}\left|\mathrm{sgn}[x(n-m)] - \mathrm{sgn}[x(n-m-1)]\right|$$

$$\text{where, } \mathrm{sgn}[s(n)] = 1, \quad s(n) \geq 0$$
$$= -1, \quad s(n) \prec 0 \qquad (3)$$

The audio signals compose a various source. It has very different characteristics in regularity of ZCR cover features, periodic, stability, amplitude limit, and dispersion. Specially, speech signal is able to classify with four conditions. The first condition is exclusive relation between ZCR and instant curve of energy as shown in Fig. 1. ZCR curve in speech segment has peak shape and concave shape in voiced-unvoiced, respectively. And energy curve is opposite to ZCR curve.

Therefore we use exclusive relation between two characteristics to classify speech signal. That is, we clip out of ZCR and energy curve in two-third point of the maximum amplitude, and have remained one peak section omitting low amplitude from ZCR and energy curve., and calculate inner product of two remained signal curve. In speech, this inner product has near the "0", because peak of ZCR and energy exist a different time. However, it has a large value in a different audio data.

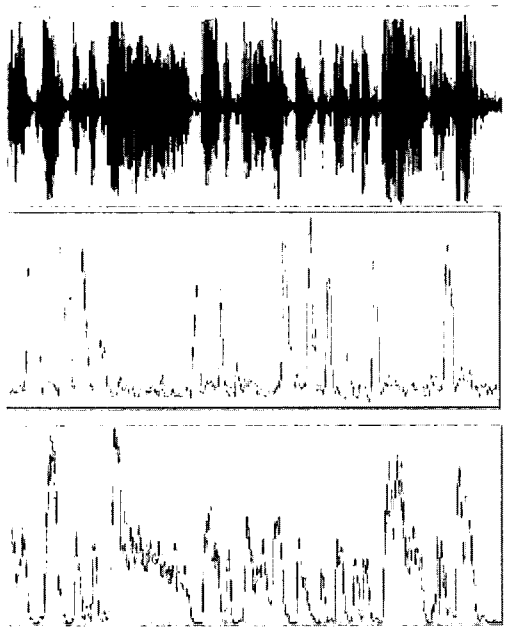
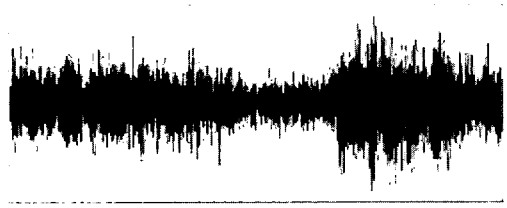


Fig. 1 Exclusive relation between ZCR and energy to speech signal



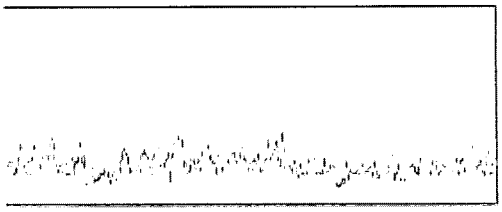


Fig. 2 The stable ZCR in music signal

C. Harmonic degree

The sound consists of basic frequency, multiple frequencies with harmonic and signal without harmonics, it shows Fig. (3) and Fig. (4), respectively. And speech consists of mixed harmonic voiced and non-harmonic unvoiced, while sound with musical instruments is generally harmonics. The existence/non-existence of harmonics for audio data distinguished from estimating signal spectrum using Auto Regressive (AR) model. The estimated spectrum by generated coefficients through AR model is envelope of the frequency spectrum, and has remarkable peaks. We estimate spectrum using algorithm proposed by Durvin, Berg and Yule-Walker, show Fig. (3) and Fig. (4). i.e., is frequency spectrum estimated by Durvin algorithm method. The shape of peak is more remarkable in harmonic than in non-harmonic.

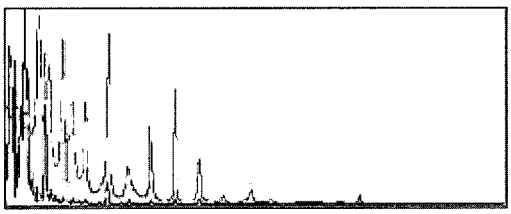


Fig. 3 Music signal with harmonics (in case of AR model order: 40)

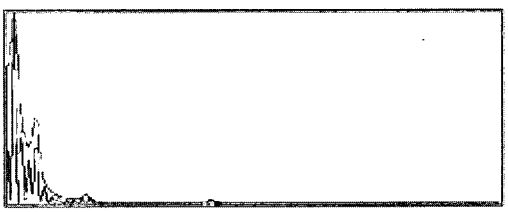


Fig. 4 Sound without harmonics (in case of AR model order: 40)

A periodic sharp peak within segments of audio signal estimated by AR model can classify sound with harmonics. If corresponding frame by analyzing short duration has harmonics, it can classify music elements, and can index as "1" otherwise is "0" Harmonic degree is defined by ratio of numbers of "0" and total indexed numbers in index stream. It has a little ratio for sound less containing music elements. And it is important parameter to distinguish music elements. The computation of harmonic degree shows in Fig. 5, its value has between "1" and "0".[5]

D. Maintenance degree of the basic frequency in short duration

The Maintenance degree of the basic frequency in short duration is defined by total numbers of the continuous duration, where, continuous duration has higher value than threshold frame value. The Maintenance degree of the basic frequency in short duration can use to classify portion containing music elements effectively.

E. Convergence degree of the frequency in short duration

The convergence degree of the frequency in short duration is defined by total frequency numbers of a higher accumulated mean than threshold within index group. To track the spectral peak, we used a direct FFT scheme.

The convergence degree of the frequency has a higher value in song signal without back music than others in contents based analysis.

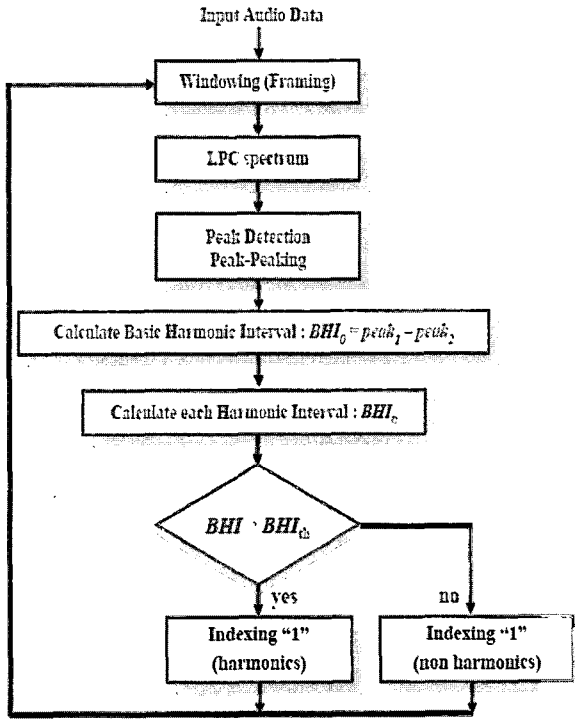


Fig. 5 Harmonic degree extraction algorithm

III. CLASSIFICATION USING FEATURE PARAMETER DATA BASE

A. Category for classification

The category for classification is consists of silence, music, speech, song, speech with back music, song with back music and effective sound.

B. Parameter data base

The feature parameter for simulation shows table 1.

C. Classification according to the classification function

The classification to audio classification category compose of seven order feature parameters using seven feature parameters, and simulate after logarithm it. We use a second order Bayesian discriminant for estimating with parameter, i.e. mean of probability density function and covariance.

$$g_i = -\frac{1}{2}(x-\mu_i)^T \Sigma_i^{-1}(x-\mu_i) - \frac{1}{2}\log(|\Sigma_i|) + \log(P(\omega_i)) \quad (4)$$

where, x is a input data, i.e., feature parameter of target to classify, μ_i is the average of training data to each class, Σ_i^{-1} is the inverse covariance matrix to each class of training data, $|\Sigma_i|$ is the covariance matrix of training data, $\log(P(\omega_i))$ is the prior probability of ω_i.

Table 1 Average analysis result of parameter D/B

Classifi -cation	Speech	Music	Song	Back +speech	Back +song
Energy	161.8	704.8	435.3	295.9	1,320.7
ZCR	20.4	25.4	27.5	21.9	27.3
ZCR range	93	40.2	87.7	60.625	82.4
ZCR Var./10,000	178	2	330	44	152
Inner product	248.7	12,866.6	2,111.3	1,424.45	8,653.1
FCD	90.9	122	12,301	90.95	60.1
HD	0.389	0.874	0.678	0.591	0.721

IV. SIMULATION AND RESULT

A. Analysis conditions

The data base for analyzing audio signal take from original sound track of movie "chin-gu" and "sound of music", and make 5seconds-204clips. The Analysis conditions shows table 2.

Table 2 Analysis condition of audio data

Speech format	PCM raw data
A/D conversion	16kHz, 16bit
Window	hamming window
Window length	256 points
Shifting period	60 points

We simulate to classify using each 40 index unit, where, index unit is a sampled data taking 2 seconds. And use API function to analyze sampled data.

B. Classification results

The silence classify using energy and ZCR threshold value in classification category, determine as silence in case energy is lower than threshold value and ZCR is

higher than threshold value. The simulation results show table 3~4.

Table 3 Classification result of audio data

Classfi -cation	Speech	Music	Song	Back +mus ic	Back +song	%
Speech	37	0	1	1	1	92. 5
Music	1	32	2	1	4	80
Song	6	0	31	1	2	77. 5
Back +music	0	4	3	29	4	72. 5
Back +song	3	0	4	2	31	77. 5

Table 4 Classification result of audio data

Classification	Song	Back +music	Back +song	%
Speech back+speech	78	6	0	97.5
Song back+song	9	71	0	88.75
Music	3	10	67	83.75

V. CONCLUSIONS

We proposed a new several parameters after analyzing effective feature parameter for implementation of real time audio index and search system as a basic research item. And we composed feature parameter data base and simulated to classify audio data using Bayesian formula.

In this paper, we established and expanded category containing music, speech, song, speech with back music, song with back music, and silent factor, and simulated to classify audio data using classification function classifier.

Consequently, we have a good performance in classification rates. Also, It shows that dimensioned feature parameter with a various parameters in feature parameter data base has a high classification ability.

REFERENCES

[1] H.Y Han, S.H Kim and K.I Hur, "Content based analysis using feature parameter of audio data," The Journal of the Acoustic Society of Korea. vol. 21, 2002, pp. 182-189.

[2] M. J. Carey, "A Comparison of feature for speech, music discrimination," Proc. ICASSP, vol. 1, 1999, pp. 145-152.

[3] J. Saunders, "Real time discrimination of broadcast speech/music," proc. ICASSP, vol. 2, 1996, pp. 141-144.

[4] K.Y Lee, B.S. Seo and J.Y Kim, "A Comparison of speech/music classification feature for audio indexing," The Journal of the Acoustic Society of Korea. vol. 20, 2001, pp. 10-15.

[5] Y. Medan, E. Vair and D. Chazan, "Super resolution pitch determination of speech signals," IEEE Trans. On Signal Processing, vol. 39, 1991, pp. 40-48.

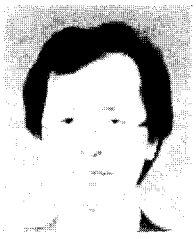
[6] H.Y Han, S.Y Koh and K.I Hur, "A syllable segmentation of korean speech," The Journal of the Acoustic Society of Korea. vol. 20, 2001, pp. 70-75.

[7] T Zgang and C.-C. J. Kuo, " Heurislic Approach for Generic Audio Data Segmentation and Annotation," proc. ACM Multimedia 99, Nov. 5, 1999, pp. 67-76.

Kwang-Seok Lee
Received B.S., M.S. and Ph.D. degrees of Electronic Engineering from Dong-A University in 1983, 1985 and 1992 respectively. In 1995, he joined the Jinju National University in Jinju, Korea, where is currently a professor and his research interests are in Intelligent System, Neural Network, Speech Processing, Speech Recognition and Synthesis and Biometrics.

Young-Sub Kim
Received B.S degree of computer Engineering from Dong-Myung University in 2005. And M.S. degree from Dong-A University in 2007. And his research interests are in Speech Signal Processing, Neural Network, Speech Recognition and Synthesis

Hag-Yong Han
Received B.S., M.S. and Ph.D. degrees of Electronic Engineering from Dong-A University in 1994, 1998 and 2004 respectively. He is currently Post-Doc researcher in the Imaging and Infomation Tech-nology Center in Pusan University. His research interests are in Biometrics, Pattern Recognition and DSP Applications.



Kang-In Hur
Received the B.S and M.S degrees of electronic engineering from Dong-A Uni-versity in 1980 and 1982 respectively. And Ph.D. degree from Kyung Hee University in 1990. In 1984, he joined the Dong-A University in Busan, Korea, where is currently an Professor. His research interests are in DSP, Speech Recognition, Synthesis and Neural Networks.