International Journal of Multimedia and Ubiquitous Engineering
Vol.11, No.2 (2016), pp. 197-206
http://dx.doi.org/10.14257/ijmue.2016.11.2.20

A Study on the Ways of Performing Voice Match with Voice Analysis Methods Using Spectral Contents

Han-Kil Kim[1], Dae-Jin Jung[2] and Hoe-Kyung Jung[2*]

[1]Korea University of Media Arts, 312, Daehak-gil, Sejong, Korea
[2]PaiChai University, 155-40, Baejae-ro, SeoGu, DaeJeon, Korea
khg0482@pro.ac.kr, daejin4u@nate.com, hkjung@pcu.ac.kr

Abstract

In order to verify the integrity of the original voice and determine the voice match, researches on analyzing speakers' own voice have been undergoing. However, spectrogram and waveform analysis technologies that are currently used have issues of lack of accuracy and reliability when determining the voice match. In this paper, ways of determining and performing voice match are suggested and proven by confirming the original voice frequency form of each individual with voice analysis methods using spectral contents. Also, the circuit for correcting or improving the quality of voice is designed and proposed. When the form of voice recording signals was analog, improved voice analysis data was secured by eliminating unnecessary noise and increasing property during PCM process. Thus, portability of voice analysis more enhanced than existing research results was verified and studied.

Keywords*: Correction Circuit, Spectral Content, Speech Analysis, Speech Match, Filter Circuit*

1. Introduction

Studies on analysis of the voice determination and device development have been conducted to solve the problems of theft of the voice operation and information with the flood of smart devices. However, there are still a lot of technical problems left [1-4]. The purpose of this paper is to improve the accuracy and reliability of the voice analysis by complementing existing problems, removing the noise and correcting it, throughout voice analysis using spectral contents [5-10]. Also, it is to confirm the special shape of individual's different voice by spectral contents and to suggest ways of same voice matching determination.

2. System Design of Spectral Contents Voice Analysis

In this chapter, it is proposed that voice analysis methods and voice match using spectral contents, and system design of voice analysis via the voice correction circuit.

2.1 Spectral Contents Audio Analysis Methods

It is to analyze original sound source and recorded sound source performed steps of analysis obtained source and sound correction, A / D converter, digital conversion system. Through the image screen of the spectral contents, it is determined by the particular curved shape of sound, words, sentences. All configurations of the spectral content speech analysis system are as follows: Figure 1.

[*] Corresponding Author

ISSN: 1975-0080 IJMUE

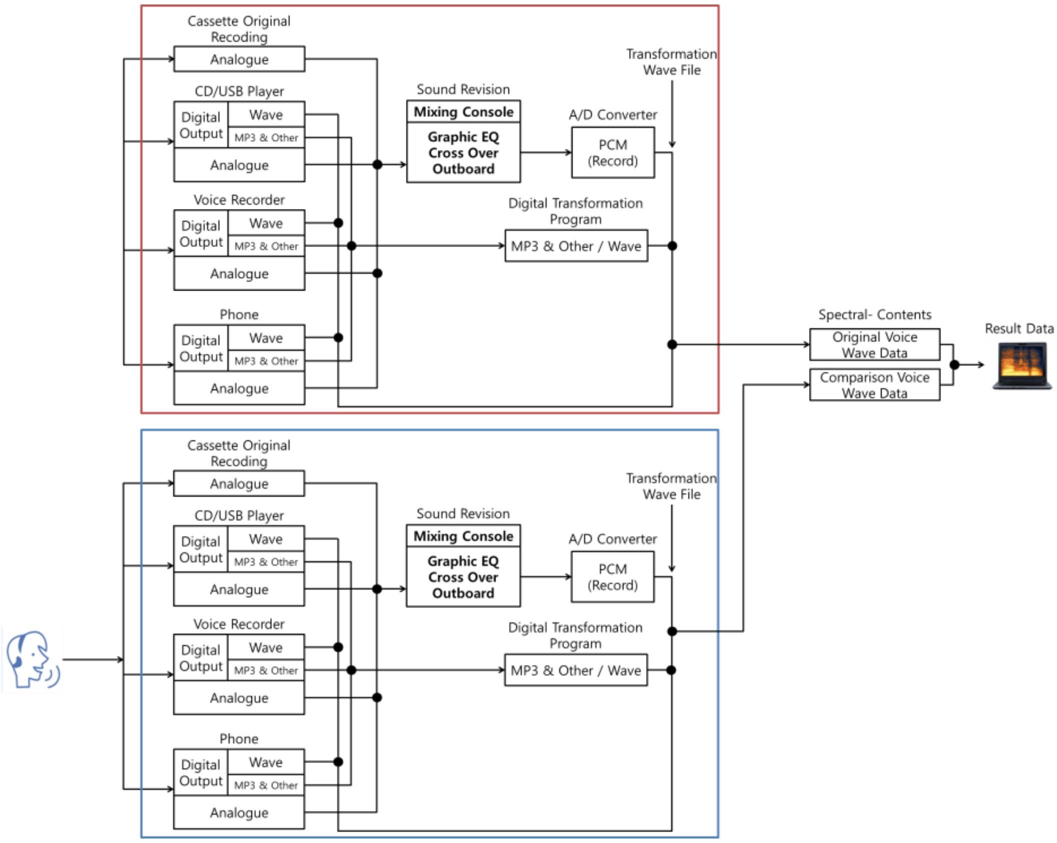


Figure 1. Spectral Audio Content Analysis System

2.2 System Design of the Spectral Content Speech Analysis

Sound sources are classified as an original audio source with another audio source to compare to. Each of the audio sources is composed of analog Wave Data, MP3 and other digital audio Data. Comparative analysis of the final sound analysis program is shown in the waveform image and analysis of the spectral content. To raise the same person voice match analysis, the sound quality of the recorded original file or the quality of comparing file should be high. It is important that the comparative audio files should be recorded in the same recording conditions.

When the voice analyzed wave file is acquired, it is to plug-in on the analysis system of spectral contents in order to analyze frequency wave form. An analysis method for converting the spectral image can be focused on, and then a sound, word, sentence sinks speech waveforms. The shape form is identified via three-dimensional image of the specific sound.

In this way, by comparing the data of the original image and the comparing image, voice match determination is performed.

2.3 Detailed Configuration of the Speech Spectral Content Analysis System

2.3.1 Wave Audio Source File Extraction and Configuration

It is ideal for sampling via the LPF for band limitation of the input signal to ensure a common signal restoration. Through EQ and gain of the analog sound mixing console of a cassette tape provided with a voice analysis of the sound, and performs the correction step, sampling the design method for converting an analog signal into a digital signal to be converted to an audio file Wave. If the gain and tone of the original audio signal

(Tone) are correct, the size of the signal change does not affect the frequency waveform of the original audio data. Since the original audio source is transmitted to the stage for comparison.

2.3.2 Wave File Recording Sampling of the Original Digital Files

If the original recording source are digital signals stored by a cassette or CD player, voice recorder, USB, mobile phones, it is extracted as a Wave file through the digital conversion system in Figure. 2.

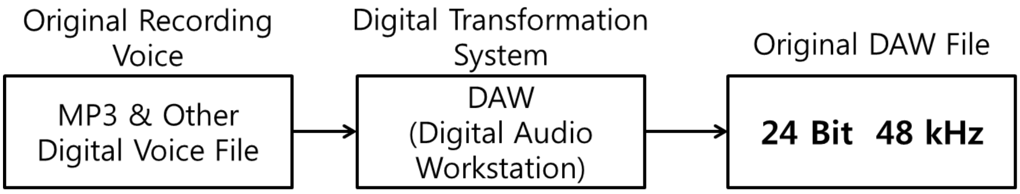


Figure 2. Wave File Conversion of Digital Audio Signals

2.3.3 Wave File Sample of the Original Analogue Recordings

If the original recording source are analogue signals stored by a cassette or CD player, voice recorder, USB, and mobile phones, the gain level of the amplifier and the noise removal process is performed on the mixing console and outboards of sound correction circuit(graphic EQ, crossover, *etc.*).

1) Sound Correction System

When the output of the original recorded sound source is coming to the mixing console, it is to adjust the input gain level by pressing the PFL switch and make noise or tone using the peakinge EQ and shelving EQ on the mixer. The analog voice correction circuit is in Figure 3.

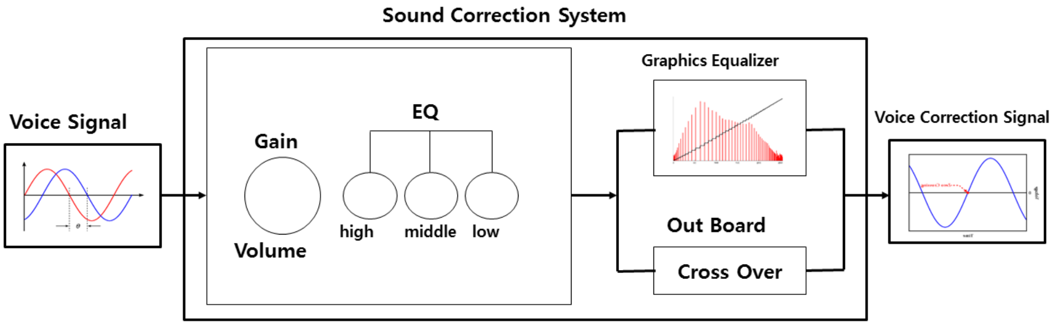


Figure 3. Analog Audio Correction Circuit

2) Circuit Design of the Sound Correction System

In order to make correction system, hardware of sound device is needed. Simple sound correction systems are ok to be used. The sound correction circuit is shown in Figure 4.

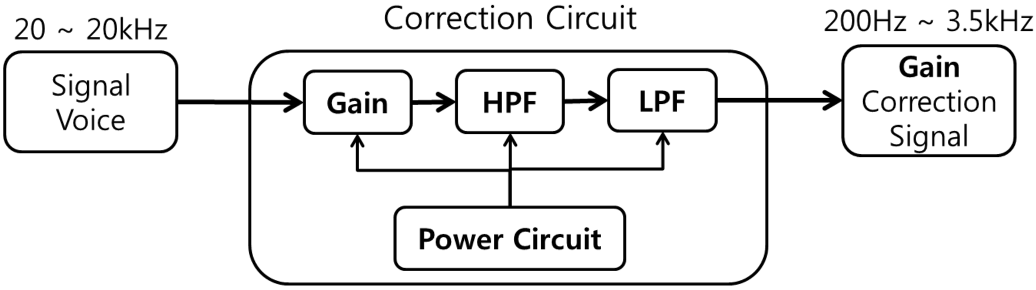


Figure 4. Sound Correction Circuit

Sound correction circuit configuration sends out the final audio signal via input amplifier circuit and HPF and LPF as shown in Figure 4.

The most important thing to design the circuit is selecting OP-AMP. By considering the frequency characteristics and economic aspects, the TL072 type is used. Spectral Contents Voice Analysis converts voice data compared with original audio data in Figure 5 to wafe/DAW files and analyze the waveforms and determine the match through the 3-dimensional image screen.

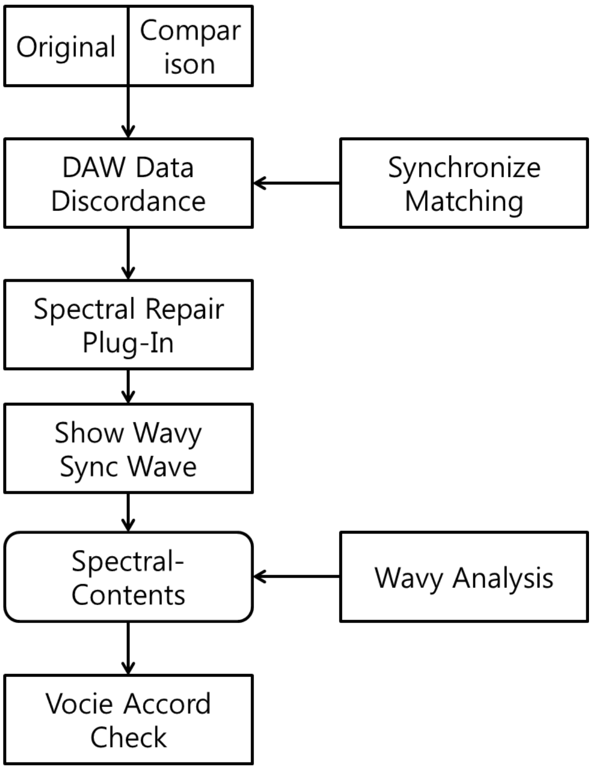


Figure 5. Voice Analysis Diagram

3. The Experiment of Voice Matches Analysis and Signal Enhancement of Spectral Contents

3.1 Settings of Spectral Contents Analyzing System

System Preferences are used for voice analysis of the spectral content is a digital audio workstation, Avid's Pro Tools HD10 and the RX plug-in of iZotope, the spectral contents analysis manufacturer. We used Apple’s iMac hardware. The specifications are

3.5GHz Core i7, 32GB Memory and it is a workstation that operates as a 64-bit operating systems.

3.2 Experiments on Spectral Contents Voice Matching

The data recorded in a mobile phone analysis were prepared in the same manner as in Table 1 below for the experiments on matching the data recorded by the DAW. As samples "a, e, i, o, u, ka, na, da, ra, daehanminkuk" 10 types of sound and words were chosen at random.

Table 1. The Environment for DAW Recording Samples

Summary	Contents
Recording Tool	Avid Technology – DAW Pro Tools HD10
Converting Frequency	24Bit, 48kHz
Converting Rate	Sample rate & Bitrate
Converting Method	Tweak Head Method
Experimental Preparation	Five Man, Five Woman
Experimental Word	a, e, i, o, u, ka, na, da, ra, ma, daehanminkuk
The distance between cellphone and microphone	10㎝ ~ 30㎝
Microphone Preamp	–30~40dBu
DAW Meter	About 3/4 ~ 1/2 point

3.3 Using the Spectral Content ‘a, e, i, o, u’ Pronunciation Analysis

When you pronounce ‘a, e, i, o, u’, five pillars is formed. By analyzing this shape, voice match can be determined. The samples were selected from 'A' to 'F'.

From A to F, according to spectral contents ‘a, e, i, o, u’ are same pronunciation while in a DAW spectral contents, they are not the same voice. The mobile phone can also be seen that there is no analytical data matches in the voice as well.

However, the analysis data of the mobile phone and DAW has the same results. If there’s no reliability of the analysis results in a continuous tone, you analyze one sound out of 5, and then you will get the result. Figure 6 is a comparison of ‘a, e, i, o, u’ with audio files on DAW and on the mobile phone.

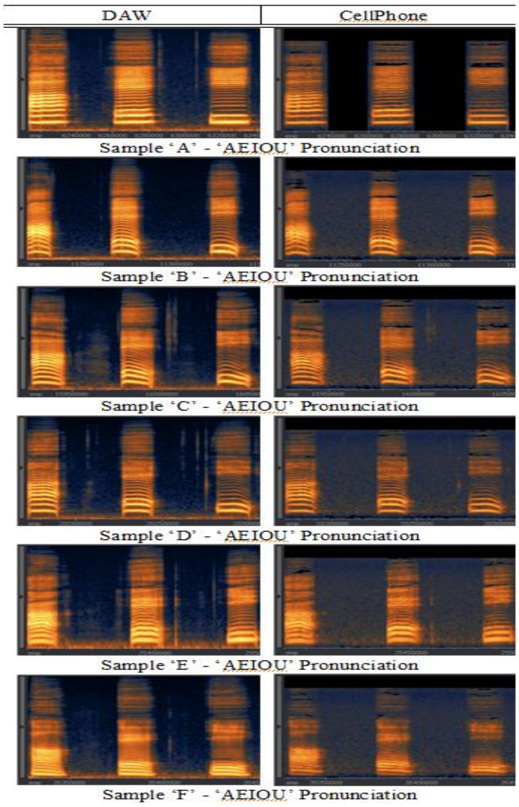


Figure 6. Comparison with DAW and the Mobile Phone

There are 2 types of voice recordings- mobile and DAW of 5men/5women. 16Bit 44.1kHz m4a recordings of phone recordings were compared with experimental data to a file recorded files to a 24Bit 48kHz via the DAW. A data match experiment was done and the result is shown in Figure 7.8.

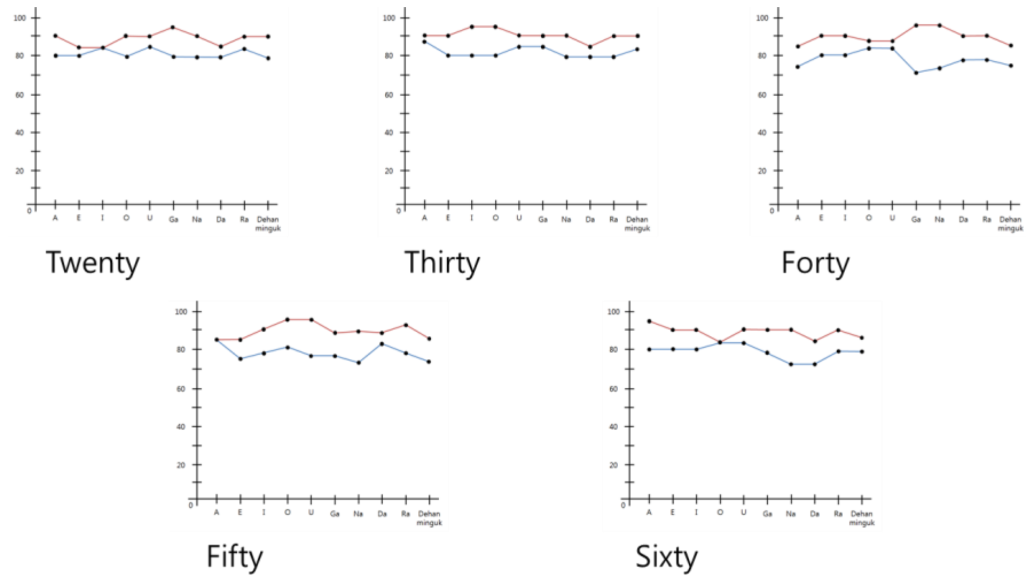


Figure 7. Men 20s-60s Spectral Contents Match Comparison of Single Words

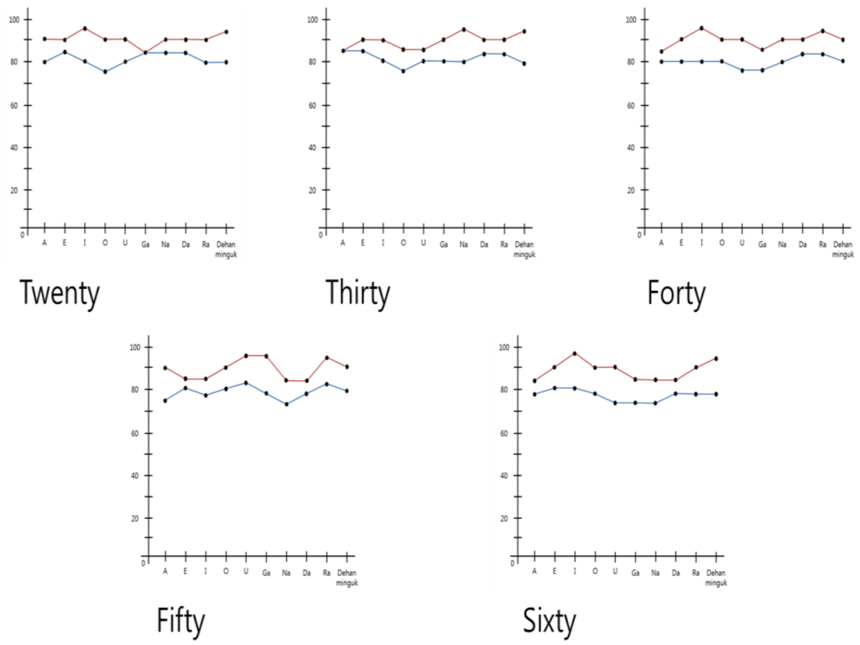


Figure 8. Women 20s-60s Spectral Contents Match Comparison of Single Words

Table 2. Words Match Comparison Table Before Correction

Words	Man1	Man2	Man3	Man4	Man5	Woman1	Woman 2	Woman 3	Woman 4	Woman 5
A	80%	85%	75%	85%	80%	80%	85%	80%	75%	80%
E	80%	80%	80%	75%	80%	85%	85%	80%	80%	85%
I	85%	80%	80%	80%	80%	80%	80%	80%	75%	85%
O	80%	80%	85%	85%	85%	75%	75%	80%	80%	80%
U	85%	85%	85%	80%	85%	80%	80%	75%	85%	75%
ka	80%	85%	75%	80%	80%	85%	80%	75%	80%	75%
na	80%	80%	80%	75%	75%	85%	80%	80%	75%	75%
da	80%	80%	85%	85%	75%	85%	85%	85%	80%	80%
ra	85%	80%	85%	80%	80%	80%	85%	85%	85%	80%
daehanminkuk	80%	85%	80%	75%	80%	80%	80%	80%	80%	80%

Table 3. Words Match Comparison Table Before Correction

Words	Man1	Man2	Man3	Man4	Man5	Woman1	Woman 2	Woman 3	Woman 4	Woman 5
A	90%	90%	85%	85%	95%	90%	85%	85%	90%	85%
E	85%	90%	90%	85%	90%	90%	90%	90%	85%	90%
I	85%	95%	90%	90%	90%	95%	90%	95%	85%	95%
O	90%	95%	85%	95%	85%	90%	85%	90%	90%	90%
U	90%	90%	85%	95%	90%	90%	85%	90%	95%	90%
ka	95%	90%	95%	90%	90%	85%	90%	85%	95%	85%
na	90%	90%	95%	90%	90%	90%	95%	90%	85%	85%
da	85%	85%	90%	90%	85%	90%	90%	90%	85%	85%
ra	90%	90%	90%	95%	90%	90%	90%	95%	95%	90%
daehanminkuk	90%	90%	85%	85%	85%	95%	95%	90%	90%	95%

Comparing Figures 7 and 8 above, the difference can be seen before and after the sound correction. After the correction, the result is improved to about 10%.

The experiment was a representative sample of 10 persons of men and women of 20s to 60s. Among them, 20 men and women, 30 men, 50 men were conducted by randomly selected.

4. Discussion and Conclusions

In this paper, we propose a way to analyze a word or a sentence through voice analysis using the spectral contents.

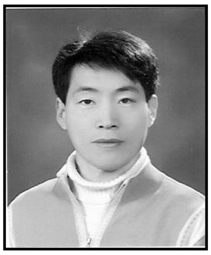
To compensate for the problem of the existing sound analysis, the noise of the audio data was removed and the accuracy and reliability of the data were improved by amplifying the data. Also, the particular shape of the spectral contents of each person was identified and the ways of determining same voice have been studied and demonstrated. As a result, we were able to validate 10% increased voice analysis and portability than traditional research. The results can be presented to disputes concerning voice related and also, can be used in various ways. The future research includes system development that produces an objective reading system of data match from voice analysis methods.

References

[1] Y. P. Kim and H. Y. Lee, "Studies on the method of improving speech recognition", Journal of the Korea Institute of Electronic and Communication Sciences, vol. 8, no. 1, **(2013)**, pp. 77-83.
[2] H. Yoo, "Estimation Using the audio coding method of a bit stream structure analysis". Journal of the Korea Institute of Information and Communication Engineers, vol.16, no.1, **(2012)**, pp.59-64.
[3] S. H. Choe and C. H. Choe, "Fluctuation analysis of voice disorders according to the audio signal classification, nonlinear dynamic analysis, the usefulness of the cepstrum analysis", Speech sciences, vol.6, no.3, **(2014)**, pp.63-72.
[4] Y. H. Kim, "Characterization of Korean vowel perception by the band restriction", Speech sciences, vol.6, no.1, **(2014)**, pp. 58-93.
[5] M. J. Lee, "Educational use of voice analysis content", Linguistics, vol. 27, **(2013)**, pp. 177-196.
[6] B. H. Kim, "Chest circumference prediction method using speech signal analysis techniques", Journal of the Korea Institute of Information and Communication Engineers, vol. 16, no. 9, **(2012)**, pp. 2031-2036.
[7] S. S. Park, "A study on the human emotion infer in meaningful place using HRV signal and text mapping", Proceedings of HCI Korea. Hanbit Media, Inc., **(2014)**.
[8] Y. S. Kim, "Study on the Korean vowel formants compared according to age and gender", Speech sciences, vol. 5, no.1, **(2013)**, pp. 73-80.
[9] C. S. Ahn and G. H. Choe, "Improve speech recognition performance using the gamma tone feature extraction acoustic models", Digital Policy, vol. 11, no.7, **(2013)**, pp. 209-214.
[10] W. I. Kim, "Features independently applicable compensation based speech enhancement techniques in a heterogeneous speech recognition system", Journal of the Korea Institute of Information and Communication Engineers, vol. 18, no. 10, **(2014)**, pp. 2367-2374.

Authors

Han Kil Kim, He received the B.S., M.S. degrees from the Department of Electronic Engineering of Hanbat National University, Korea, in 2006, and 2011, and the Ph. D. degree in 2015 from the Department of Computer Engineering of Paichai University, Korea. Since 2006, he worked for the Department of Music & Sound Technology at Korea University of Media Arts as a professor. His current research interests include information retrieval system, web 2.0, and the semantic web.



Dae Jin Jung, He received the M.S. in 2014 from the Department of Computer Engineering of PaiChai University, Korea. He is currently taking Doctorate degree in Department of Computer Engineering of Paichai University. From 2000 to 2007, he worked as a software engineer in Data Management Center in Korea Intellectual Property Office. Since 2014, he has worked as a software engineer in Realtimetech (In-Memory Database). His current research interests include Computer Graphics, Virtual Reality, Information & Communication Technology.

Hoe Kyung Jung, He received the B.S degree in 1987 and Ph. D. degree in 1993 from the Department of Computer Engineering of Kwangwoon University, Korea. From 1994 to 2005, he worked for ETRI as a researcher. Since 1994, he has worked in the Department of Computer Engineering at Paichai University, where he now works as a professor. His current research interests include multimedia document architecture modeling, information processing, information retrieval, and databases.

