

---

# 리듬정보를 이용한 악보생성 시스템 구현

소 두 석\* , 최 재 원\* , 이 종 혁\*

Implementation of Musical Note Generation System using Rhythm Information

Du-Sek So\* , Jae-Weon Choe\* , Jong-Hyeok Lee\*

---

이 논문은 2002년도 Brain Busan 21 사업에 의하여 지원되었음

---

## 요 약

본 논문에서는 기존의 오디오검색 관련 시스템의 텍스트 정보기반의 음악 검색의 한계를 극복하기 위해 사용자가 입력한 음악데이터에서 리듬정보를 추출해 내어 데이터 베이스 내의 음악 정보 검색이 가능하게 하는 내용기반 검색 시스템의 일종인 리듬 추출 시스템을 제안하였다. 본 논문에서는 음악정보에서 리듬의 특징 정보를 추출하여 특징 파라미터를 생성하고 신경회로망을 사용하여 악기 인식을 통한 악보생성 시뮬레이션 하였다.

## ABSTRACT

Traditional indexing mechanism are based on the song's metadata such as the title and the composer and so on. However, these system have a major limitation that users have to know the metadata of the songs they want to retrieve.

In order to solve these limitation, we proposed a rhythm extraction system that allows users to retrieve music information efficiently from a large music database using the rhythm that is defined as the parts of the music.

## 키워드

리듬정보, 자동악보 생성, 정보 검색, 신경회로망

## 1. 서 론

전통적으로 도서관과 음악정보센터는 음악과 관련된 책, 인쇄 또는 악보, 녹음자료, 시청각 자료 등을 소장하고 이에 관한 목록을 준비하여 이용자들이 작곡가나 곡명, 연주가 등의 서지 기술 사항을 통해 음악자료를 검색할 수 있도록 서비스를 제공하고 있다. 최근에는 디지털 기술의 발전으로 인해 많은 음악정보원들을 디지털 형태로 축

적할 수 있게 되었다. 즉, 악보를 디지털 이미지 형태로 축적하거나 아날로그 형태의 음악을 CD같은 매체를 통해 디지털 형태로 저장하는 기술이 보편화되면서 전자형태의 음악 정보원들이 생겨나게 되었다.[1]

멀티미디어 데이터베이스나 파일시스템은 다수의 오디오 녹음을 가질 수 있는데, 이들은 효과음 자료에서부터 보관된 뉴스 필름의 음성 녹음 부분에 이르기까지 무엇이든 가질 수 있다. 이런 자

---

\* 경성대학교 전기전자컴퓨터공학부

접수일자 : 2003. 8. 21

료들의 색인이나 데이터의 생성 시 메타데이터의 내용이 작업자의 주관에 의해서 수행되므로 모호하게 기록되는 경우가 허다하다.

유용한 오디오 데이터를 검색하기 위한 방법에는 일반적으로 텍스트기반 검색기법과 내용기반 검색기법이 있다. 텍스트 기반 검색기법은 텍스트가 가지는 데이터 표현의 한계로 인해 검색 시 원하는 검색 결과를 얻기 어렵다. 내용기반 검색기법은 텍스트기반 검색기법의 이러한 단점을 극복하기 위하여 오디오 데이터가 가지는 여러 특징들을 이용하고 있다. 이러한 방법은 검색효율을 향상시키며, 사용자에게 다양한 검색 방법을 제공할 수 있다.

현재 상용화되어 있는 검색 엔진들은 멀티미디어 데이터가 가지는 방대함과 모호성으로 인하여 검색에 여러 가지 어려움을 가지고 있다. 음악 정보의 검색도 멀티미디어 데이터검색과 관련되어 대단히 흥미 있는 분야중의 하나로 등장하고 있다.

기존의 오디오 관련 검색 시스템은 주로 음악정보가 가지는 제목이나 작곡가와 같은 메타 데이터를 이용한 색인을 구성하였으나 이와 같은 색인 기법을 이용한 검색에서는 사용자가 자신이 찾고자 하는 오디오정보의 메타 데이터를 반드시 인지하고 있어야 검색이 가능하다는 한계가 있으므로 최근에는 이러한 문제를 해결하기 위해 오디오의 내용기반 검색 알고리즘이 제안되고 있다.[2][3][4]

본 논문에서는 오디오 신호에서 특징 파라미터를 추출하고, 신경망을 이용하여 리듬을 담당하는 악기를 인식하여 오디오 데이터를 악보화 시키므로 추후 내용기반 검색 시스템에 이용될 수 있도록 하고자한다.

## II. 이론적 배경

본 장에서는 음악의 구성요소, 음악정보의 디지털화, 드럼 및 리듬의 특징 정보에 대해 알아본다.

### 2.1 음악의 구성요소

우리는 흔히 리듬, 멜로디, 하모니를 음악의 중요한 세 가지 요소라고 하는데, 잘 만들어진 음악

에서는 이 세 가지가 적절히 조화를 이루고 있다. 음악에서 각각의 마디에는 다른 음보다 특별히 더 강조되는 음이 있다. 이처럼 강조되는 음들을 시간의 순서로 배열하는 것이 바로 리듬이다.[5]

멜로디는 음악에 명확한 성격을 부여하는 것으로, 계속되는 반복과 형태의 변화 때문에 음악을 이루는 요소들 중 가장 두드러지게 드러난다. 대부분의 작곡가들은 하나의 멜로디를 보다 더 정교하고 복잡하게 만들어 한 작품 내에서 계속해서 반복하고 다양하게 변화시킨다. 대부분의 경우 멜로디에는 화음이 따르는데, 이런 결합을 통해 만들어지는 소리를 하모니라고 한다.

### 2.2 음악의 디지털화

최근 디지털 도서관에 대한 관심이 증가하면서 도서관의 소장자료들을 디지털화하기 위한 많은 연구가 진행되고 있다. 이러한 연구는 텍스트 또는 이미지 기반의 문헌을 대상으로 한 것으로서 전문(Full-text)검색을 위해 문헌을 ASCII기반의 텍스트로 전환하는 방법, 구조화 마크 업 언어를 이용하여 텍스트를 구조화하는 방법, 문헌 자체를 스캐너로 읽어서 이미지 형태로 디지털 화하는 방법 등을 이용하여 문헌을 디지털화 한다. 그러나, 소리에 관한 정보들을 담고 있는 음악정보는 음악을 일상적인 인간의 언어로 표현하기 어려우므로 일반적인 텍스트 문헌의 디지털 처리 방법과는 다른 새로운 방법이 필요하다.

디지털 형태로 음악을 표현하는 방법으로는 악보를 디지털 화하는 방법과 소리를 축적한 사운드 레코드를 디지털 화하는 방법이 있다. 악보를 디지털 형태로 저장하는 방법에는 첫째, 악보를 스캐너로 읽어서 이미지 형태로 디지털 화하는 방법과 둘째, 악보에 해당하는 각각의 기호를 코드 화하여 컴퓨터가 해석할 수 있도록 하는 인코딩 방법이 있다.[6]

### 2.3 드럼 및 리듬의 특징 정보

리듬을 형성하는 중요한 악기는 드럼이다. 드럼의 특성을 살펴보면 대부분 곡에 나타나고 표현되는 주파수가 거의 일정하다. 구성하는 악기는 베이스 드럼, 스네어, 탐탐, 심벌이 대표적이다. 드럼의

고유 주파수 대역은 베이스드럼의 경우는 60Hz~80Hz, 스네어는 220Hz ~ 240Hz, 심벌의 경우는 7k Hz이상에서 나타나고 탐탐은 베이스와 스네어의 중간쯤인 100 ~ 200Hz의 주파수대역을 가지고 있다.

### III. 악보 생성 시스템 구현

본 장에서는 악보 생성 시스템을 구현하기 위한 하드웨어/소프트웨어적인 환경과 악보생성 시스템의 전체 구성과 주파수 및 시간영역 파라미터를 생성하여 신경망을 학습시키는 과정 및 악보생성 시스템의 시뮬레이션 결과에 대해 알아본다.

#### 3.1 구현환경 및 악보생성 시스템

##### 3.1.1 구현환경

시스템 구현을 위해서 개인용 컴퓨터 (Pentium-III 700MHz, 512RAM)에 오디오 매직II 사운드/오디오 카드를 설치하였다. 운영체제는 Windows2000으로 하였으며, 사용언어는 전반적인 사용자 인터페이스 부분은 JAVA SDK\_1.4를 사용하였고, 신경회로망은 속도향상을 위해 Microsoft Visual C++ 6.0으로 구현하였다. WAV파일을 위해서는 미국 Syntrillium Software사의 윈도우즈용 디지털 오디오 편집기인 Cool Edit 2000을 이용하였으며, 악보의 기보를 위해서는 국내의 오딧세이 소프트웨어사의 매직 스코어 2000을 이용하였다.

표 1은 본 논문에서 사용한 음악들이며, 이는 세트 드럼을 배우기 위한 교본 중에서 각 장르별로 가장 일반적인 리듬으로 구성된 음악들이다.

표 3. 악보생성 시스템에 사용한 음악  
Table 1. Music list in note generation system

번호	곡명	번호	곡명
01	BLUES1.WAV	10	FUSION1.WAV
02	BLUES2.WAV	11	HVYMETAL.WAV
03	BOSANOVA.WAV	12	MAMBO.WAV
04	DANCE1.WAV	13	REGGAE1.WAV
05	DANCE2.WAV	14	REGGAE2.WAV
06	DANCE3.WAV	15	SHUFFLE.WAV
07	DANCE4.WAV	16	SLOWRCK1.WAV
08	FUNKY.WAV	17	SWING.WAV
09	FUNKYF~2.WAV	18	WALTZ.WAV

표 1에 제시한 음악을 청취해 본 결과, 사용된 대표적인 악기를 알 수 있었으며, 표 2에 사용된 각 악기들과 이의 코드를 나타내었다.

표 4. 악기정보의 구성  
Table 2. List of instrument information

BB	SS	B1	B2
베이스드럼	스네어	베이스+심벌(1)	베이스+심벌(2)
BT	HT	MT	LT
베이스+탐탐	하이탐탐	중간탐탐	로우탐탐

##### 3.1.2 악보생성 시스템

악보생성 시스템은 Wav파일의 전처리 과정, 특징 추출 과정, 악기 인식과정, 악보생성 과정으로 나누어지며 이를 그림 1에 나타내었다.

#### 3.2 전처리과정

전처리 과정(Pre-processing)은 환경 적응, 시작점 검출, 반향 제거, 잡음 제거와 같은 오디오 신호를 본격적인 분석과정에 들어가기 전에 처리하는 과정이다.

본 시스템에서는 시작점 검출을 위해 절대 에너지 값을 사용하였다. 512개의 표본 수를 한 프레임으로 하였으며, 일레로 BLUES1의 에너지 패턴을 그림 2에 나타내었다.

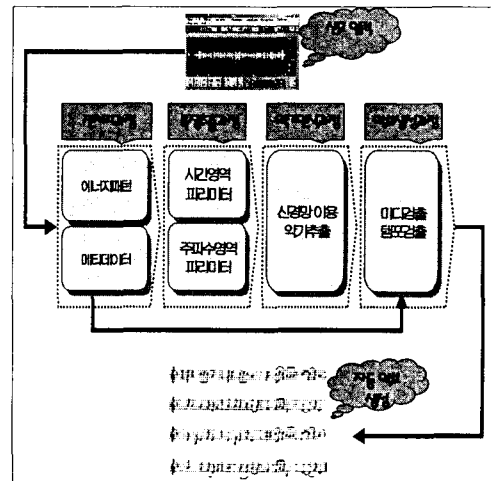


그림 5. 악보생성 시스템의 블록도  
Fig. 1. Block diagram of note generation system

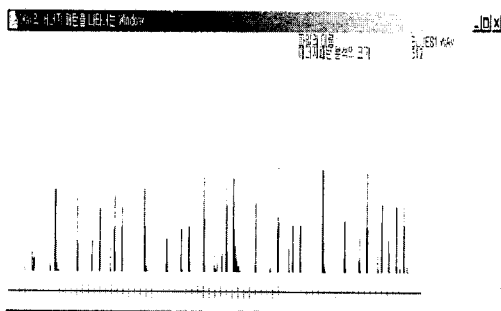


그림 2. BLUES1의 에너지 패턴  
Fig. 2. Energy pattern of BLUES1

오디오정보의 메타데이터에는 이름, 파일 형식, 표본화율 등과 같은 가장 원시적인 영역만 가진다. 본 연구에서는 RIFF 포맷에 의거하여 WAVE 파일의 내부에 포함되어져 있는 메타데이터를 추출하여 전체 곡의 특성을 나타내는 데이터로 활용한다. WAVE의 포맷에 따라 해당 파일에서 처음 44Byte의 헤더정보를 추출하였다.

### 3.3 특징추출

특징추출이란 일반적으로 디지털 오디오 신호의 용량이 크므로 인식과정에서 필요한 정보만을 추출하는 것을 말한다. 시간영역 특징 파라미터와 주파수 영역 특징 파라미터를 추출하여 악기인식에 사용하였다.

#### 3.3.1 시간영역 특징 파라미터

시간영역 특징 파라미터는 에너지 패턴, Local Max Peak 및 신호패턴의 조합으로 하였다. 시작점 검출을 통해 찾아낸 음의 시작점에서 128point를 한 프레임으로 하고 연속된 3프레임에서 에너지와 Local Max Peak를 구하여 시간영역 특징 파라미터의 처음 6개의 정보로 하였다.

신호패턴을 위해서 음의 시작점에서 540 point를 대상으로 하였다. 540 point를 30 point씩 18개의 구간으로 나누고 N-point Filter를 거친 후 각 구간에서 Filter의 출력 값 18개를 신호패턴으로 하였다. 그림 3은 시간영역 특징 파라미터의 예를 나타내었다.

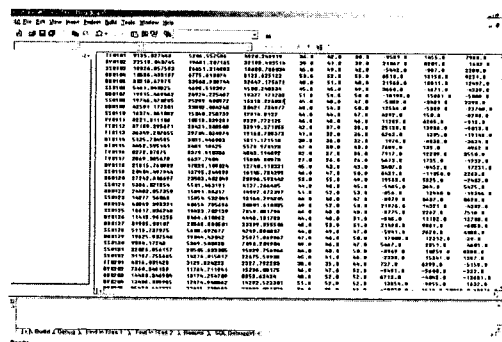


그림 3. 시간 영역 특징 파라미터  
Fig. 3. Characteristic parameter in time domain

#### 3.3.2 주파수 영역 특징 파라미터

음의 시작점에서 1024 point를 취하여 FFT[7] 처리를 하였다. 리듬 악기의 주파수 정보는 저주파 쪽에 편중되어 있으므로 다음과 같은 방법을 거쳐 최종 20차의 특징 파라미터로 하였다.

- ① 0에서 ~ 11 point는 1:1 매칭을 통해 12차의 파라미터를 생성한다.
- ② 12에서 ~ 511 point에서는 전체를 62 point씩 8개의 구간으로 나누고 각 구간의 평균값을 구하여 8차의 파라미터를 생성하였다. 그림 4에 주파수 영역 파라미터의 예를 나타내었다.

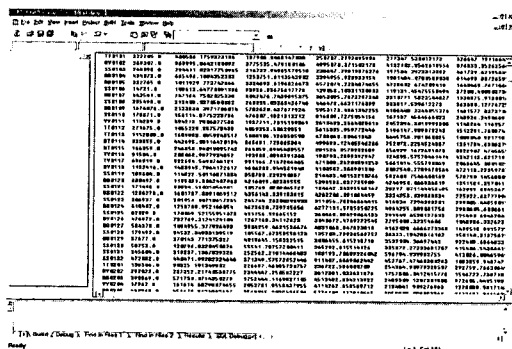


그림 4. 주파수 영역 특징 파라미터  
Fig. 4. Characteristic parameter in freq. domain

### 3.4 악기인식 신경회로망

악기 인식 신경회로망에서 입력 층의 노드 수는 주파수 영역 파라미터를 위해 20개, 시간영역 파라미터를 위해 24개 총 44개를 두었으며, 출력 층의 노드 수는 표 2에서 나타난 악기정보에 따라

8개로 하였다. 그리고 은닉 층은 20개로 하였다.[8]

기초 실험을 통하여 오류 역전파 알고리즘에서 학습률 gain은 1로 설정하고, 모멘텀 변수  $\alpha$ 는 0.6으로 두고, 이득률  $\eta$ 는 0.1로 설정하였다. 반복횟수는 1000번과 3000번 5000번으로 지정하여 학습한 것 중 학습시와 테스트시에 각각 가장 높은 인식률을 보인 3000번으로 지정하여 학습하였다. 본 논문에서 제안한 악기 인식 신경회로망의 구성은 그림 5와 같다.

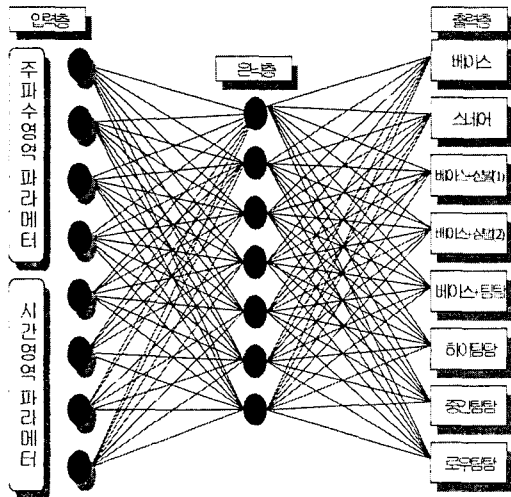


그림 5. 악기 인식 신경 회로망의 구성  
Fig. 5. Neural network for instrument recognition

신경회로망에서 악기 인식을 위하여 각 음악 장르의 대표격인 11가지의 장르 총 18개의 음악을 대상으로 하였다. 실험 시 사용한 데이터를 살펴보면, 학습 시 사용한 데이터는 각 장르별로 대표적인 리듬을 나타내는 음악 11가지이고, 테스트 시에 사용한 데이터는 그 외의 유사한 음악 7가지에 대해서 실험하였다. 실험결과를 표 3과 표 4에 나타내었으며 학습에 사용한 데이터의 경우 91%의 인식률을 보였고, 학습에 사용하지 않은 데이터에 대해서는 84%의 인식률을 보였다.

표3. 학습 데이터 인식결과  
Table 3. Recognition result at learned data

음 악	총계	正	誤	(%)
BLUES1.WAV	32	27	5	84
BOSANOVA.WAV	20	16	4	80
DANCE1.WAV	48	46	2	95
FUNKY.WAV	33	31	2	93
FUSION1.WAV	52	47	5	90
HVYMETAL.WAV	49	40	9	81
REGGAE2.WAV	39	38	1	97
SHUFFLE.WAV	56	45	9	80
SLOWRCK1.WAV	12	12	0	100
SWING.WAV	32	32	0	100
WALTZ.WAV	47	46	1	97
< 총 계 >	420	380	38	91

표 4. 미 학습 데이터 인식결과  
Table 4. recognition result at nonlearned data

음 악	총계	正	誤	(%)
BLUES2.WAV	40	38	2	95
DANCE2.WAV	41	34	7	83
DANCE3.WAV	22	20	2	91
DANCE4.WAV	10	4	6	40
FUNKYF2.WAV	14	12	2	86
REGGAE1.WAV	49	40	9	82
SOFTROCK.WAV	39	30	9	77
< 총 계 >	176	148	28	84

### 3.5 결과분석

표 3, 4에서 보는 것과 같이, 전체적으로 학습에 사용된 데이터의 인식률은 91%로 높은데 반하여 학습에 사용되지 않은 테스트용 데이터의 인식률은 84%로 상대적으로 저조했다. 전체적인 실험에서 잘못 인식된 데이터들을 표 5에 나타내었으며 특히, 베이스, 베이스+심벌에서와 스네어, 하이 탐탐에서의 인식결과가 저조한 것으로 나타났다.

표 5. 誤인식된 악기정보  
Table 5. Error instrument information list

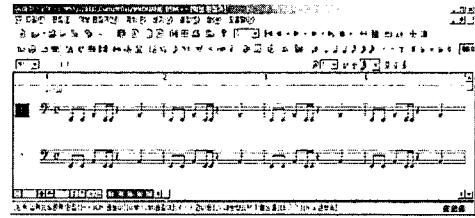
원래의 악기	악기인식결과	갯수
HT	SS	12
B2	BB	19
B2	SS	2
HT	B2	1
HT	BB	1
BB	B2	17
SS	HT	11
BT	BB	1
BB	SS	2
		66

이는 각 악기별 고유의 주파수가 베이스(BB) 및 베이스+심벌(B2)과 하이 탐탐(HT) 및 스네어(SS)의 주파수 및 시간영역 특징 정보가 거의 유사하기 때문인 것으로 판단되며, 이는 전체 誤인식의 89%를 차지한다. 또한 시작점 추출과정에서 이전 음의 마지막음과 시작 음이 겹치는 부분에서 해당 악기의 잔음의 영향으로 誤인식된 경우로 보여진다.

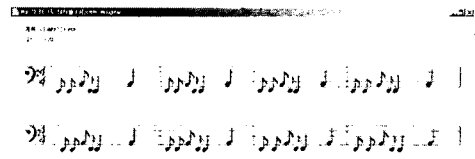
대표적인 리듬을 가진 2가지의 음악에 대해 원래의 악보와 시뮬레이션 한 결과를 그림 6,7에 나타내었다.

위 시뮬레이션에 쓰여진 음악들은 우리가 흔히 접할 수 있는 리듬으로 구성되어져 있는 대표적인 곡으로써 악기구성은 드럼과 베이스, 탐탐으로 이루어져있다. 전체 곡들은 심벌이 기본박자를 연주하고, 베이스와 스네어가 첫 박자와 4번째 박자에서 나오며, 한 동기가 끝날 때 탐탐이 나오면서 동기의 끝을 알리는 형식으로 이루어졌다.

시뮬레이션 결과 베이스와 스네어는 거의 대부분을 찾아내었으며, 탐탐 및 빠른 리듬의 연주에서는 세밀한 부분까지는 찾아내지 못한 것을 알 수 있다. 이러한 결과를 보이는 이유는 음의 시작점 검출 시 스네어 및 베이스는 여음이 거의 남아있지 않는 것에 반해 탐탐 및 심벌은 여음이 많이 남아있음으로써 생기는 문제인 것으로 판단된다.

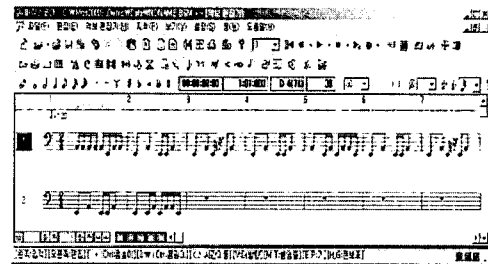


(a) 원래의 악보



(b) 시뮬레이션 결과

그림 6. DANCE1에 대한 시뮬레이션 결과  
Fig. 6 Simulation result of DANCE1



(a) 원래의 악보



(b) 시뮬레이션 결과

그림 7. Heavy Metal에 대한 시뮬레이션 결과  
Fig. 7. Simulation result of Heavy Metal

#### IV. 결론

기존의 오디오 관련 검색 시스템은 주로 음악 정보가 가지는 제목이나 작곡가와 같은 메타 데이

터를 이용한 색인을 구성하였으나 이와 같은 색인 기법을 이용한 검색에서는 사용자가 자신이 찾고자 하는 오디오 정보의 메타 데이터를 반드시 인지하고 있어야 검색이 가능하다는 한계를 가지고 있다.

본 논문에서는 이러한 한계를 극복하기 위해서 사용자가 입력한 음악데이터에서 리듬정보를 추출해 내어 데이터 베이스 내의 음악 정보 검색이 가능하게 하는 내용기반 검색 시스템의 일종인 리듬 추출 시스템을 제안하였다.

제안된 리듬 추출 시스템이 음악으로부터 리듬을 추출할 수 있음을 보여주기 위해 악보생성 과정을 시뮬레이션 하였다. 두 가지 특징 파라미터 즉, 시간 영역 특징 파라미터와 주파수 영역 특징 파라미터를 이용하여 다층 신경망을 학습시키고 그 결과로 악기정보를 추출하도록 하였다. 시간 영역 특징 파라미터는 Local Max Peak 와 에너지 패턴 그리고 신호 패턴으로 구성하여 24차로 하였으며, 주파수 영역 특징 파라미터는 FFT의 결과를 사용하여 20차로 하였다. 시뮬레이션 결과 평균 인식율은 91%를 나타내었다.

#### 참고 문헌

- [1] 박종도, "NIFF를 이용한 음악정보 검색 시스템연구", 중앙대학교 석사 학위논문집, 1998.
- [2] 지정규, 오해석, "선율을 이용한 음악 정보 검색 시스템의 설계 및 구현", 한국정보처리학회 정보 처리논문지, Vol.5, No.1, 1995.
- [3] A. Ghias, J. Login, D.Chamberlin and B.C.Smith, "Query By Humming Musical Information Retrieval in an Audio Database", ACM Multimedia 95, November 1995.
- [4] Niblack, W., et al., "QBIC Project: Querying Images by Content, Using Color, Texture, and Shape," Proceedings of the Conference Storage and Retrieval for Image and video Databases, Feb.2-3, 1993.
- [5] O'Keefe, Georgia, "음악의 심리학적 이해", Univ. Minnesota Art Museum Minneapolis, MN USA
- [6] 김명호, 이운준, 「개정 멀티미디어 개념 및

응용」, 한국과학기술원 전산학과, 1997.

- [7] 이강승, 「기초이론 디지털 신호처리」, 21세기, 1999.
- [8] 조현욱.이종혁, "음성과 영상정보를 이용한 우리말 숫자음 인식", 경성대학교 논문집, 2002년 2월

#### 저자 소개

##### 소두석(Du Sek So)

- 2001년 2월 경성대학교 음악학과(음악사)  
 2003년 8월 경성대학교 컴퓨터공학과(공학석사)  
 ※ 관심분야 : 정보통신, 멀티미디어, 컴퓨터응용



##### 최재원(Jae Weon Choe)

1988년 2월 고려대학교 컴퓨터 공학과(공학사)

- 1990년 8월 미시간주립대학교 컴퓨터공학과(공학석사)  
 1995년 8월 건국대학교 전자공학과(공학박사)  
 1990년 10월 ~ 1997년 8월 삼성전자 정보통신 연구소 선임연구원  
 1997년 9월 ~ 현재 경성대 컴퓨터공학과 부교수  
 ※ 관심분야 : 정보통신망, 이동통신, 운영체제



##### 이종혁(Jong Hyeok Lee)

1975년 2월 부산대 전자공학과(학사)  
 1980년 2월 부산대 전자공학과(석사)  
 1991년 2월 부산대학교 전자공학과(공학박사)

- 1990년 3월 ~ 현재 경성대 컴퓨터공학과 교수  
 1998년 7월 ~ 1999년 6월 미국 Beckman Institute, University of Illinois, 객원 연구원  
 ※ 관심분야 : 인공지능, 음성인식, 신호처리