2013년 12월 전자공학회 논문지 제 50 권 제 12 호
Journal of The Institute of Electronics Engineers of Korea Vol. 50, NO. 12, December 2013 http://dx.doi.org/10.5573/ieek.2013.50.12.238

논문 2013-50-12-29

다중 옥타브 밴드 기반 음악 장르 분류 시스템

(Musical Genre Classification System based on Multiple-Octave Bands)

변 가 람*, 김 무 영**

(Karam Byun and Moo Young Kim©)

요 약

음악 장르 분류를 위해서 다양한 종류의 특징 벡터들이 이용되고 있다. 대표적인 short-term 특징 벡터들로는 mel-frequency cepstral coefficient (MFCC), decorrelated filter bank (DFB), octave-based spectral contrast (OSC) 등이 있으며, 이들의 long-term variation이 함께 이용된다. 본 논문에서는 OSC 특징을 추출하는데 있어서 하나의 옥타브 밴드 뿐만 아니라 다중 옥타브 밴드를 동시에 이용하여 옥타브 밴드 간 상관관계를 함께 반영할 수 있도록 하였다. 2012년도 music information retrieval evaluation exchange (MIREX) 평가회의 mixed 장르 분류 분야에서 4위를 한 알고리즘에 다중 옥타브 밴드를 이용한 결과, GTZAN과 Ballroom 데이터베이스에 대해서 각각 0.40% 포인트와 3.15% 포인트의 성능 향상을 얻을 수 있었다.

Abstract

For musical genre classification, various types of feature vectors are utilized. Mel-frequency cepstral coefficient (MFCC), decorrelated filter bank (DFB), and octave-based spectral contrast (OSC) are widely used as short-term features, and their long-term variations are also utilized. In this paper, OSC features are extracted not only in the single-octave band domain, but also in the multiple-octave band one to capture the correlation between octave bands. As a baseline system, we select the genre classification system that won the fourth place in the 2012 music information retrieval evaluation exchange (MIREX) contest. By applying the OSC features based on multiple-octave bands, we obtain the better classification accuracy by 0.40% and 3.15% for the GTZAN and Ballroom databases, respectively.

Keywords : Music information retrieval, music genre classification, MFCC, DFB, OSC, SVM

Ⅰ. 서 론

인터넷의 발전으로 음원을 다운로드할 수 있는 웹서비스가 활발히 이루어지게 되었다. 음원들의 보급이 보편화 되면서, 방대한 양의 음악을 효과적으로 관리 및 검색하는 방법들이 필요하게 되었다. 따라서 최근 들어 음악 정보 검색 (music information retrieval)과 관련된 연구 분야가 주목을 받고 있다[1]. 음악 정보 검색과 관련된 커뮤니티로는 International Society for Music Information Retrieval (ISMIR) 학회와 Music Information Retrieval Evaluation eXchange (MIREX) 등이 있다[2~3]. ISMIR 학회에서는 매년 다양한 주제의 음악 정보 검색 알고리즘이 발표되고 있다. 더불어 2005년도부터 시작된 MIREX contest는 다양한 음악 정보 검색 연구 주제에 대한 알고리즘들을 비교, 평가하여 ISMIR 학회에서 발표하고 있다. 음악 정보 검색 연구로는 음악의 beat를 찾는 알고리즘부터 장르/무드/

* 학생회원, ** 정회원, 세종대학교 정보통신공학과
(Department of Information and Communication Engineering, Sejong University)
© Corresponding Author(E-mail: mooyoung@sejong.ac.kr)
※ “본 연구는 미래창조과학부 및 정보통신산업진흥원의 대학 IT연구센터 육성지원사업의 연구결과로 수행되었음” (NIPA-2013-H0301-13-4007)
접수일자: 2013년11월4일, 수정완료일: 2013년11월27일

작곡가 분류[4], singing/ humming을 이용한 타이틀 검색[5~6] 등을 포함한다. 장르는 blues, rock, pop 등과 같은 label로 표현되며, 무드는 사람의 감정과 유사하게 음악의 즐거움, 슬픔 등으로 표현된다. 하지만, 장르와 무드의 정의를 명확히 내리기 어려워 분류하기에 어려움이 있다. 그럼에도 불구하고 많은 연구들이 진행되고 있다.

음악 장르 분류의 성능은 음악의 특징을 얼마나 잘 표현하는가와 어떠한 패턴 인식 방법을 사용하는가에 따라 크게 영향을 받는다. 본 논문에서는 새로운 특징 벡터를 추출함으로써 장르 분류 시스템의 성능 개선을 이루고자 하였다.

본 논문의 구성은 다음과 같다. Ⅱ장에서 다중 옥타브 밴드를 이용한 장르 분류 시스템에 대하여 소개하고, Ⅲ장에서는 실험을 통해 기존 방식과 제안한 방식의 성능을 평가하였다. 마지막으로 Ⅳ장에서는 최종 결론을 서술하겠다.

Ⅱ. 본 론

음악의 장르 분류 시스템은 그림 1과 같이 특징 추출 (feature extraction), 특징 선별 (feature selection), 모델링 (modeling), 분류 (classification) 과정으로 구성되어 있다. 장르 분류를 위한 baseline 알고리즘으로는 2012년도 MIREX 평가회의 mixed 장르 분류 분야에서 4위를 한 알고리즘을 사용하였다[12].

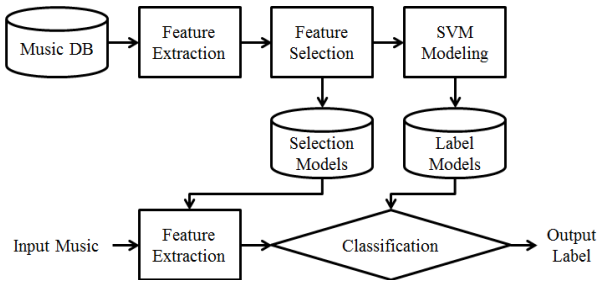

그림 1. 음악 장르 분류 시스템의 블록도
Fig. 1. Block-diagram of music genre classification system.

1. 특징 추출

음악의 특징 추출 방법에는 다양한 방법들이 있다. 본 논문에서는 기존에 제안된 시스템에 실제 악기들의 연주 주파수 범위에 의거한 특징 벡터를 결합하여 장르 간의 차이를 두르러지게 구분할 수 있는 방법을 제안하였다. 추출된 특징은 표 1과 같다.

표 1. 특징 벡터와 dimension:
(a)기존 시스템, (b)제안 시스템
Table 1. Feature vectors and their dimension:
(a) baseline system, (b) proposed system.

Feature vectors				Dimension (a)	Dimension (b)
Texture window	Mean		MFCC	13	13
			DFB	13	13
			OSC	16	22
	Variance		MFCC	13	13
			DFB	13	13
			OSC	16	22
	Max		MFCC	13	13
			DFB	13	13
			OSC	16	22
	Min		MFCC	13	13
			DFB	13	13
			OSC	16	22
Octave-based modulation spectrum	MSFM		OBS	8	11
	MSCM		OBS	8	11
	Mean	MSC/MSV	OBS	16	22
	Var	MSC/MSV	OBS	16	22
Feature-based modulation spectrum	FMSFM		MFCC	13	13
			DFB	13	13
			OSC	16	22
	FMSCM		MFCC	13	13
			DFB	13	13
			OSC	16	22
	Mean	FMSC/ FMSV	MFCC	26	26
			DFB	26	26
			OSC	32	44
	Var	FMSC/ FMSV	MFCC	26	26
			DFB	26	26
			OSC	32	44
Total vector dimension				468	546

가. Short-term feature

(1) Mel-frequency cepstral coefficient (MFCC)

MFCC는 사람의 청각 모델을 기반으로 한 특징으로 화자 인식이나 음성 인식에 많이 사용된다[7]. MFCC는 mel-scale band-pass filter와 discrete cosine transform

(DCT)에 의해 추출된다. B개의 mel-scale band에 대해서, b번째 band의 파워 스펙트럼 값이 $S[b]$라 하면, k번째 MFCC는 다음과 같이 구한다.

$$MFCC[k] = \sum_{b=0}^{B-1} \log S[b] \cos\left(\frac{(b+0.5)k\pi}{B}\right) \quad (1)$$

(2) Decorrelated filter bank (DFB)

DFB는 MFCC와 마찬가지로 화자 인식이나 음성 인식에 주로 사용되지만[8], DCT 대신 high-pass filter를 이용해 추출된다. MFCC와 마찬가지로 band별 가중치 합 $S[k]$를 구한 후, 다음과 같이 high-pass filter를 적용하여 추출한다.

$$DFB[k] = \log S[k+1] - \log S[k] \quad (2)$$

MFCC가 DCT에 의해 band별 특징을 나타낸다면, DFB는 각 band간 차이를 나타내는 특징이다.

(3) Octave-based spectral contrast(OSC)

OSC 추출 시에는 MFCC와 DFB에서 사용하는 청각 모델인 mel-scale band 대신에 octave-scale band를 사용한다[10]. 표 2에 보듯이, 22050Hz 샘플링 주파수를 기준으로 octave-scale band-pass filter는 8개의 single-octave band로 구성된다(0~100Hz, 100~200Hz, 200~400Hz, 400~800Hz, 800~1600Hz, 1600~3200Hz, 3200~6400Hz, 6400~11025Hz). 본 논문에서는 single-octave band를 사용하는 baseline 시스템[12]에 multiple-octave band 별로도 특징을 추출하는 방법[11]을 추가적으로 적용하였다. 표 2에서 보듯이, 본 논문에서는 multiple-octave band 추출을 위해서, 4개의 octave씩 묶은 band 2개 (0~800Hz, 800~11025Hz)와 전체 주파수 범위인 global octave를 사용한 band 1개 (0~11025Hz)를 추가적으로 이용하였다.

표 2. Single-octave band와 Multiple-octave band의 주파수 범위

Table 2. Frequency range of single-octave band and multiple-octave band.

Band		Frequency range (Hz)
Single-octave band	1	[0~100]
	2	[100~200]
	3	[200~400]
	4	[400~800]
	5	[800~1600]
	6	[1600~3200]
	7	[3200~6400]
	8	[6400~11025]
Multiple-octave band	1	[0~800]
	2	[800~11025]
	3	[0~11025]

스펙트럼은 위와 같은 band를 이용하여 각 band의 구간으로 나누어진 후, 내림차순으로 정리한다. k번째 band의 스펙트럼을 $x[k]$이라 하면, 내림차순으로 정리된 스펙트럼은 $\{x'[1], x'[2], \cdots, x'[N_k]\}$이다. 이때, k번째 octave-scale band의 peak와 valley를 각각 $P[k]$, $V[k]$라 할 때, spectral contract는 다음과 같이 구한다.

$$SC[k] = P[k] - V[k] \quad (3)$$

일반적으로 OSC는 $\{V[k], SC[k]\}$로 정의된다. 대부분의 음악에서 강한 peak는 harmonic 부분과 연관되며, 강한 valley는 non-harmonic 부분과 연관된다. 따라서 OSC는 octave-scale sub-bands에서 spectral peak와 spectral valley의 spectral contrast를 고려함으로써, 음악의 harmonic과 non-harmonic 성분을 반영하는 특징이다.

나. Long-term feature

Long-term feature는 주어진 short-term feature의 시간축 변화율을 고려한 특징이다.

(1) Texture window

Texture window를 이용한 특징 추출 방법은 특징들의 통계적 값을 사용한다. 한 프레임을 analysis window라고 정의한다면, texture window는 N개의 analysis window로 정의된다. 즉, texture window에서 추출한 short-term feature의 mean, variance, 그리고 covariance 등과 같은 통계적 값을 특징벡터로 추출하게 된다[12]. Texture window를 사용하는 이유는 음악이 short-time 스펙트럼의 time series로 구성되어 있으며, 통계적으로 연관성을 가지기 때문이다.

(2) Modulation feature

Modulation feature는 modulation 스펙트럼을 이용하

여 추출한 특징이다. 각 texture window에 대한 analysis frame의 총 개수를 T라하고, t번째 analysis frame의 b번째 sub-band 내 Fourier 스펙트럼의 합을 $S_t[b]$라 하면, n번째 modulation frequency에 대한 modulation 스펙트럼은 다음과 같이 구한다[13].

$$M_n[b] = \sum_{t=1}^{T} S_t[b]\exp(-j2\pi nt/T) \quad (4)$$

- Octave-based modulation spectrum (OMS)

OMS란 식 (4)에서 $S_t[b]$ 대신 octave-scale band-pass filter에 의한 b번째 sub-band 내 Fourier 스펙트럼의 합을 사용한다.

OSC는 octave-scale band-pass filter를 이용하여 계산된 band별 스펙트럼에 대한 spectral contrast 값이다. 반면, octave-based modulation spectral contrast (OMSC)는 각 octave band energy의 modulation 스펙트럼을 구하고, modulation 스펙트럼에 대한 spectral contrast를 구한 특징이다.

OMS를 바탕으로 계산한 modulation spectral flatness/crest measure (MSFM/MSCM)는 octave band energy의 modulation 스펙트럼에 대해 spectral flatness/crest measure (SFM/SCM)를 구한 특징이다[13~14]. MFSM의 작은 값과 큰 값은 평균 modulation 스펙트럼의 peakiness와 flatness를 나타낸다. MSCM은 MSFM과 반대 특성을 나타낸다. 만약 b번째 MSFM이 0에 가까운 값을 가진다면, octave-band b번째 modulation frequency가 반복되는 패턴을 가진다. 이는 beat가 강하다는 것을 나타내는 특징이다. 또한 본 논문에서는 modulation spectral contrast/valley (MSC/MSV)도 사용하였다[15~16].

- Feature-based modulation spectrum (FMS)

Feature-based modulation flatness/crest measure (FMSFM/FMSCM)는 feature vector에 기반한 modulation 스펙트럼을 사용하여, 음악의 time-varying 특징을 나타낸다[4]. K번째 dimension의 t번째 analysis frame의 short-term feature vector가 $F_t[k]$라 하면, n번째 modulation frequency에 대한 FMS는 다음과 같이 구한다.

$$FMS_n[k] = \sum_{t=1}^{T} F_t[k]\exp(-j2\pi nt/T) \quad (5)$$

식 (5)는 식 (4)의 변형된 형태로써, $S_t[b]$ 대신 $F_t[k]$를 사용했다. 본 논문에서는 MFCC, DFB, 그리고 OSC를 short-term feature로 사용하였다. 또한 feature-based MSC/MSV (FMSC/FMSV)를 계산하여 특징으로 사용하였다[12].

2. 특징 선별

추출된 feature vector들은 특징 선별 알고리즘에 의해서 각 장르에 적합한 특징들만을 골라 분류에 사용한다. 특징 선별 알고리즘을 사용하면 메모리와 계산량을 줄일 수 있고, 선별된 특징들을 이용하여 성능 향상을 이룰 수 있다. 특징 선별 과정은 학습 과정에서 이루어진다. 따라서 테스트 과정에서는 선별된 특징들만이 계산되기 때문에 계산 시간 또한 줄일 수 있다.

특징 선별 알고리즘에는 principal component analysis, linear discriminant analysis, non-negative matrix factorization, and support vector machine (SVM) ranker 등이 있다[18~19]. 본 논문에서는 SVM ranker를 사용하여 특징 선별을 하였다.

3. 모델링 및 분류

Feature vector를 추출하는 것도 중요하지만, 이를 이용하여 패턴 분류를 어떻게 할지도 중요한 부분이다. 분류를 위해서는 먼저 분류할 class별 모델을 만들어야 하는데, 패턴 인식 방법마다 다른 모델링 방법으로 모델을 만든다. 패턴 인식 방법으로는 통계적인 접근법[7], 신경망을 이용한 접근법[18] 등 다양한 방법이 있다. 각 분류기마다 특징이 있어 같은 feature vector를 사용하더라도 인식률은 달라진다.

본 논문에서는 음악 장르 분류를 위해서 SVM을 이용하여 모델링하고 분류하였다[20]. SVM은 Vapnik와 AT&T Bell 연구소에서 제안한 방법으로 구조적인 위험 최소화를 사용하여 분류 에러를 줄이는 방법이다.

Ⅲ. 실 험

장르 분류 시스템의 성능을 평가하기 위해 GTZAN과 Ballroom 데이터베이스를 사용하였다. GTZAN은 blues, classical, country, disco, metal, hiphop, jazz, pop, reggae, rock의 총 10개 장르를 포함하고 있으며 한 장르 당 100곡, 한 곡 당 30초로 구성되어 있다.

표 3. 기존 방식과 제안 방식의 장르 분류 정확도
Table 3. Genre-classification accuracy of the baseline and proposed systems.

		GTZAN	Ballroom
Accuracy (%)	기존 방식	85.00	70.20
	제안 방식	85.40	73.35
Feature-vector dimension		160	160

Ballroom은 dance 음악 스타일의 cha-cha, jive, quickstep, rumba, samba, tango, viennese waltz, waltz 의 총 8개 장르를 포함하고 있다. 총 698곡이지만, 각 장르 당 곡의 수는 유니폼하게 분포되어 있지 않으며, 한 곡 당 30초로 구성되어 있다.

Ⅱ-1장에서 서술한 방식으로 feature를 추출하였고, Ⅱ-2장에서 설명한 SVM ranker를 특징 선별 알고리즘으로 적용하였다. 분류기로는 선형 커널 함수를 적용한 SVM에 대하여 1-against-1 방식을 사용하였다. 또한 실험의 신뢰성을 위해 10-fold cross-validation (CV) 실험을 하였다. 10-fold CV는 전체 데이터베이스의 90%를 이용하여 모델을 학습하고, 나머지 10%를 이용하여 장르 분류를 테스트하는 과정이다.

Baseline system으로는 single-octave band를 이용하여 다양한 특징을 추출한 [12]를 사용하였고, 제안 방식은 multiple-octave band를 이용하여 다양한 특징을 추출한 방식을 사용하였다. Feature vector별 dimension은 표 1에서 확인할 수 있다. Single-octave band 이외에 multiple-octave band를 추가적으로 이용하였기 때문에 OSC를 이용한 특징들의 dimension이 증가하였다. 특징 선별 알고리즘을 적용하기 전의 특징이 갖는 dimension은 기존과 제안 방식에 대해서 각각 468과 546이었다.

표 3에 보듯이, 제안 방식에 특징 선별 알고리즘을 적용한 후의 dimension은 160으로 기존 방식과 동일하였으며, 따라서 SVM 분류 시의 계산량은 동일함을 알 수 있다.

장르 분류 정확도 측면에서는 제안한 방식이 기존 방식에 비해서 향상된 성능을 보였다. GTZAN 데이터베이스에 대해서는 0.40% 포인트의 성능 향상을, Ballroom 데이터베이스에 대해서는 3.15% 포인트의 성능 향상을 보였다. Ballroom 데이터베이스에 대해서 SVM ranker를 이용한 특징 선별 알고리즘을 수행한 결과, multiple-octave band에 기반한 feature vector들이 상당수 선택되었음을 알 수 있었다. 따라서 제안한 feature vector는 악기별 octave 범위에 근거하여 설계된 multiple-octave band를 이용하였기 때문에 Ballroom 데이터베이스에 포함된 cha-cha, jive, quickstep, rumba 등과 같이 여러 악기를 사용한 dance 음악 장르들의 구분에 유용하게 사용될 수 있음을 알 수 있었다.

Ⅳ. 결 론

본 논문에서는 특징 벡터 추출 시 single-octave band 이외에 multiple-octave band를 추가적으로 이용함으로써 음악 장르 분류 인식률을 향상 시키는 알고리즘을 제안하였다. 그 결과 GTZAN 데이터베이스에 대해서는 0.4% 포인트의 성능 향상을, Ballroom 데이터베이스에 대해서는 3.15% 포인트의 성능 향상을 보였다. Octave band-pass filter를 좀 더 정교하게 설계할 필요가 있으며, 악기 별 특징을 고려하여 band-pass filter를 설계하는 방법에 대한 연구가 필요하다.

REFERENCES

[1] Z. Fu, G. Lu, K. M. Ting, and D. Zhang, “A survey of audio-based music classification and annotation,” IEEE Trans. Multimedia, vol. 13, no. 2, pp.303-319, 2011.

[2] J. S. Downie, “The music information retrieval evaluation exchange (2005-2007): a window into music information retrieval research,” Acoustical Science and Technology, vol. 29, no. 4, pp. 247-255, 2008.

[3] J. S. Downie, A. F. Ehmann, M. Bay, and M. C. Jones, “The music information retrieval evaluation exchange: some observations and insights,” Advances in Music Information Retrieval, vol. 274, pp. 93-115, 2010.

[4] S.-C. Lim, S.-J. Jang, S.-P. Lee, and M. Y. Kim, “Music genre/mood classification using a feature-based modulation spectrum,” in Proc. IEEE Int. Conf. Mobile IT Convergence, pp. 133-136, 2011.

[5] G. P. Nam, K. R. Park, S.-P. Lee, E. C. Lee, M.-Y. Kim, K. Kim, “Intelligent query by humming system,” in Proc. IEEE Int. Conf. Ubiquitous Information Technologies Applications, pp. 22-23, 2009.

[6] K. Kim, K. R. Park, S.-J. Park, S.-P. Lee, and M. Y. Kim, “Robust query-by-singing/humming system against background noise environments,” IEEE Trans. Consumer Electronics, vol. 57, no. 2, pp. 720-725, 2011.
[7] D. A. Reynolds and R. C. Rose, “Robust text independent speaker identification using gaussian mixture speaker model,” IEEE Trans. Speech, Audio Process., vol. 3, no. 1, pp. 72-83, 1995.
[8] J. Ming, “Robust speaker recognition in noisy conditions,” IEEE Trans. Audio, Speech, Language Process., vol. 15, no. 5, pp. 1711-1726, 2007.
[10] D. N. Jiang, L. Lu, H. J. Zhang, J. H. Tao, and L. H. Cai, “Music type classification by spectral contrast feature,” in Proc. IEEE Int. Conf. Multimedia and Expo., pp. 113-116, 2002.
[11] S.-C. Lim, S.-J. Jang, S.-P. Lee, and M. Y. Kim, “Multiple octave-band based genre classification algorithm for music recommendation,” KIICE, vol. 15, no. 7, pp. 1487-1494, 2011.
[12] S.-C. Lim, J.-S. Lee, S.-J. Jang, S.-P. Lee, and M. Y. Kim, “Music-genre classification system based on spectro-temporal features and feature selection,” IEEE Trans. Consum. Electron., vol. 58, no. 4, pp. 1262-1268, 2012.
[13] D. Jang and C. D. Y, “Music genre classification using novel features and a weighted voting method,” in Proc. IEEE Int. Conf. Multimedia and Expo., pp. 1377-1380, 2008.
[14] D. Jang and C. D. Y, “Music information retrieval using novel features and a weighted voting method,” in Proc. IEEE Int. Symposium on Industrial Electronics, pp. 1341-1346, 2009.
[15] C-H. Lee, J-L. Shih, K-M. Yu, and J-M Su, “Automatic music genre classification using modulation spectral contrast feature,” in Proc. IEEE Int. Conf. Multimedia and Expo., pp. 204-207, 2007.
[16] C.-H. Lee, J.-L. Shih, K.-M. Yu, and H.-S. Lin, “Automatic music genre classification based on modulation spectral analysis of spectral and cepstral features,” IEEE Trans. Multimedia, vol. 11, no. 4, pp. 670-682, 2009.
[17] C. A. de los Santos, “Nonlinear audio recurrence analysis with application to music genre classification,” M.S. thesis, Univ. Pompeu Fabra, 2010.
[18] E. Benetos and C. Kotropoulos, “Non-negative tensor factorization applied to music genre classification,” IEEE Audio, Speech, Language Process., vol. 18, no. 8, pp. 1955-1967, 2010.
[19] I. guyon, J. Weston, S. Barnhill, and V. Vapnik, “Gene selection for cancer classification using support vector machines,” Machine Learning, no. 1-3, vol. 46, pp. 389-422, 2002.
[20] Y. Wang, “A Tree-Based Multi-class SVM Classifier for Digital Library Document”, in Proc. IEEE Int. Conf. Multimedia and Information Technology, pp. 15-18, 2008.

저 자 소 개

변 가 람(학생회원)

2012년 세종대학교 정보통신공학과 학사 졸업

2012년～현재 세종대학교 정보통신공학과 석사과정

<주관심분야 : 음악정보검색, 잡음 제거, 음성 인식, 화자 인식>

김 무 영(정회원)-교신저자

1993년 연세대학교 전자공학과 학사 졸업

1995년 연세대학교 전자공학과 석사 졸업

1995년～2000년 삼성종합기술원 전문연구원

2001년～2004년 Royal Institute of Technology (KTH, 스웨덴) Dept. Signals, Sensors, Systems, 박사 졸업

2004년～2005년 Royal Institute of Technology (KTH, 스웨덴) Dept. Signals, Sensors, Systems, PostDoc

2005년～2006년 Ericsson Research (스웨덴), Senior Research Engineer

2006년～현재 세종대학교 정보통신공학과, 부교수

<주관심분야 : 음악정보검색, 음성/오디오 신호처리 및 코딩, 패턴인식, 정보이론.>