



특집

MIREX 기술 동향

김무영 (세종대학교), 이석필 (KETI)

I. 서론

디지털 음악 콘텐츠의 보급이 보편화 되면서 대용량 음악 데이터베이스로부터 음악 콘텐츠를 가공하여 사용자의 요구에 부응하는 서비스를 제공할 수 있는 음악 정보검색 (Music Information Retrieval) 기술의 필요성이 절실해지고 있다^[1~2]. 음악 정보검색 기술은 Singing/Humming 을 이용한 음원 검색 기술, 유사곡 추천 및 표절 곡 탐색 기술, 나이/감정/취미 등의 사용자 프로파일에 따른 음악 추천 기술 등에 사용된다. 또한, 대용량의 동영상상을 사용자가 효과적으로 탐색하기 위해서는 연산량이 적은 오디오 정보만을 이용하여 중요구간을 우선적으로 탐색한 후 비디오 정보 등을 이용하여 상세 탐색을 실행하는 기술이 개발되고 있다.

음악 정보검색 기술은 다학제간 연구 분야로써 신호처리, 전산학, 음향학, 음악학, 도서관학, 심리음향학 등 다양한 학문 분야의 공동 연구가 필요하다. 음악 정보검색은 텍스트가 아닌 음악 자체를 query로 이용하여 검색을 실시함으로써 기존의 텍스트 기반 검색 기술이 가지는 한계를 극복하기 위해서 대두되었다. 음악 query를 이

용한 대표적인 방식에는 사용자의 노래 소리나 허밍을 이용하여 곡을 찾는 Query-by-Singing/Humming 기술이나 MP3 음원을 query로 이용하여 비슷한 장르나 무드의 곡을 찾는 Query-by-Example 기술 등이 있다.

음악 정보검색 기술에 대한 최근 동향은 International Society for Music Information Retrieval (ISMIR) Conference에서 매년 발표되고 있다^[3]. 또한, 기술별 최고 성능을 가지는 알고리즘은 Music Information Retrieval Evaluation eXchange (Mnge) 평가회를 통해 확인할 수 있다. Mnge는 음악 정보검색과 관련된 알고리즘들을 공식적으로 평가하는 커뮤니티 기반의 프레임워크이다. 음악 정보검색 기술에 대한 국제적인 컨테스트인 ange (평가회는 J. Stephen Downie 교수 (International Music Information Retrieval Systems Evaluation Laboratory (IMIRSEL), University of Illinois at Urbana-Champaign, 미국) 주관 하에 2005년부터 매년 1회씩 실시되고 있다.

MIREX 평가회는 2004년 Universitat Pompeu Fabra의 Music Technology Group (MTG)이 실시한 Audio Description Contest

(ADC)에 기원을 두고 있다^[4]. 또한, ADC와 MIREX의 평가 방법은 Text Retrieval Conference (TREC)를 기반으로 하고 있다^[5~6].

MIREX 평가회의 해당 연도별 task와 평가 방법은 커뮤니티 메일링 리스트와 위키 홈페이지인 <http://music-ir.org/mirexwiki>를 통한 토의를 거쳐서 정해진다^[1]. 매년 초에 해당 연도 task를 정하기 위한 토의가 시작되며, 커뮤니티 내 연구 그룹들은 각자의 관심 분야를 공개적으로 표명하게 된다. 관심 분야별 task leader는 위키 홈페이지를 개설하고, 관심 있는 연구 그룹들은 이곳에 테스트 데이터 수집 방법부터, 평가방법, 제출 알고리즘의 입출력 포맷, 평가가 이루어질 task의 구체화 등을 논의하게 된다. 이 때, 평가회 제출을 선언한 알고리즘의 종류가 3가지 이상이 되는 task는 MIREX의 당해년도 공식 task로 인정받게 된다. 테스트는 매년 7-8월 사이에 IMIRSEL에서 진행되며, 최종 결과는 ISMIR 학회 직전인 8-9월에 MIREX 위키에 공개된다. ISMIR 학회 기간 동안에는 MIREX 포스터 세션이 열리고, MIREX 결과에 대한 평가와 차년도 테스트 계획 등을 논의하기 위해서 plenary meeting 또한 개최된다.

본 기고서의 II장에서는 2005년부터 시작된 MIREX 평가회의 기술 동향 및 기술 분야별 대표적인 알고리즘 들을 요약해 보고, III장에서 음악 정보검색 기술의 주요 응용처를 살펴본 후, IV장에서 결론을 맺고자 한다.

II. 기술 분야별 MIREX 평가 결과

[1,7]을 기반으로 2005년부터 2009년 까지 MIREX 평가회에 참여한 알고리즘 수를 기술 분야별로 <표 1>에 정리하였다. 각 task에 대한 상

세 설명 및 대표 알고리즘들은 [2,8]를 포함한 다양한 참고문헌들을 참고하기 바란다. 총 23개의 서로 다른 task가 수행 되었으며, task에 따라서는 여러 개의 subtask를 수행하기도 하였다.

<표 1> MIREX 평가회에 참여한 알고리즘 수

TASK	2005	2006	2007	2008	2009
Audio Artist Identification	7		7	1	30
Audio Classical Composer Identification			7	11	
Audio Beat Tracking		5			22*
Audio Cover Song Identification		8	8	8	6*
Audio Chord Detection				15*	18*
Audio Drum Detection	8				
Audio Genre Classification	15		7	26*	64*
Audio Key Finding	7				
Audio Melody Extraction	10	10*		7	12
Audio Music Mood Classification			9	13	33
Audio Music Similarity and Retrieval		6	12		15
Audio Onset Detection	9	13	17		12
Audio Tempo Extraction	13	7			
Audio Tag Classification				11	30#
Multiple Fundamental Frequency Estimation & Tracking			27*	28*	26#
Music Structure Segmentation					5
Real-time Audio to Score Alignment (a.k.a Score Following)				4	
Score Following		2			
Symbolic Genre Classification	5				
Symbolic Key Finding	5				
Symbolic Melodic Similarity	7	18#	8		
Query-by-Singing/Humming		23*	20*	17*	6*
Query-by-Tapping				5	6*
연도별 task 총합	86	92	122	156	285

* : task comprised two subtasks
: task comprised three subtasks

이 중에서 Symbolic Genre Classification, Symbolic Key Finding, Symbolic Melodic Similarity만이 심볼릭 도메인에서 이루어졌으며, 나머지 task들은 모두 신호처리 도메인에서 이루어졌다. 그 이유는 대부분의 음악신호처리 연구자들이 신호처리 background를 가지고 있기 때문이다. 또한, 심볼릭 데이터에 비해서 CD나 MP3 데이터가 쉽게 구할 수 있기 때문이다.

Audio Music Similarity and Retrieval, Query-by-Singing/Humming, Audio Artist Identification, Audio Mood Classification, Audio Genre Classification처럼 음악 정보검색 시스템의 성능을 평가하는 task들도 있었지만, 많은 task들이 Audio Melody Extraction, Audio Onset Detection, Multiple Fundamental Frequency Estimation/Note Tracking, Audio Tempo Extraction 등과 같이 기초기술과 관련되어 있었다.

II장에서는 대표적인 task들에 대한 2009년도 MIREX 성능 평가 방법 및 평가 결과를 요약해 보고자 한다.

1. Multiple Fundamental Frequency Estimation & Note Tracking

음악 정보검색 기술의 근간이 되는 피치 또는 기본 주파수를 검출하는 task로 크게 2가지 subtask를 수행하고 있다. 복잡한 음악 신호는 기본 주파수의 contour로 분할하여 표현할 수 있으며, 복수개의 기본 주파수를 추정하여 연속된 노트와 팀버를 찾아낼 수 있다. 따라서, 10ms 프레임별로 복수개의 기본 주파수 후보들을 찾아내는 subtask와 이를 기반으로 노트를 트래킹

하는 subtask를 포함하게 된다. 프레임별로 추정된 피치와 노트는 각각 손으로 마킹한 ground-truth 데이터와 비교해서 half semi-tone 이내에 들어오면 올바르게 추정된 것으로 간주하였다. 노트의 경우는 onset과 offset의 정확도 또한 평가하였다.

사용한 데이터베이스는 우선 Beethoven's Variations for String Quartet, Op.18, No. 5에 대해서 플룻, 오보에, 클라리넷, 호른, 바순 부분을 따로 녹음하여 사용하였다. 또한, RWC 미디어/샘플 중 클래식과 재즈 컬렉션, 폴리포닉 피아노 샘플 등을 사용하였다. 2009년 평가 결과는 <표 2>와 <표 3>과 같으며, 피치추정은 Yeh와 Roebel이 제안한 알고리즘이 0.688로 accuracy 관점에서 최고점을 받았으며, 노트 트래킹은 Nakano, Egashira, Ono, Sagayama가 제안한 알고리즘이 최고점을 받았다.

<표 2> Multiple Fundamental Frequency Estimation 평가결과

참가자	Accuracy
Yeh & Roebel (2)	0.688
Yeh & Roebel (1)	0.646
Duan, Han & Pardo (2)	0.569
Duan, Han & Pardo (1)	0.548
Nakano, Egashira, Ono & Sagayama (2)	0.492
Zhang & Liu	0.488
Nesi, Pantaleo & Argenti (2)	0.483
Nesi, Pantaleo & Argenti (1)	0.437
Nakano, Egashira, Ono & Sagayama (1)	0.354
Raczynski & Sagayama (2)	0.272
Raczynski & Sagayama (1)	0.261

〈표 3〉 Multiple Fundamental Frequency Note Tracking 평가결과

참가자	F-Measure
Nakano, Egashira, Ono & Sagayama (2)	0.319
Yeh & Roebel	0.308
Nesi, Pantaleo & Argenti (2)	0.227
Duan, Han & Pardo (1)	0.222
Raczynski & Sagayama (3)	0.159
Raczynski & Sagayama (1)	0.158
Raczynski & Sagayama (4)	0.150
Raczynski & Sagayama (6)	0.149
Raczynski & Sagayama (2)	0.148
Raczynski & Sagayama (5)	0.144
Nesi, Pantaleo & Argenti (1)	0.136
Nakano, Egashira, Ono & Sagayama (1)	0.037
Bertin, Vincent & Badeau	0.003

2. Audio Onset Detection

Audio Onset은 오디오 신호의 진폭이 정해진 한계치보다 크게 증가하는 순간을 나타내며 다양한 음악 이벤트의 시작 시점을 의미한다. 사용한 데이터베이스는 총 85곡으로 이 중에서 30곡은 솔로 드럼, 30곡은 모노포닉 피치를 가지는 악기, 10곡은 폴리포닉 피치를 가지는 악기, 15곡은 다양한 악기들이 동시에 연주된 형태이다. 각 곡에 대해서 3-5명의 청취자가 Audio Onset 위치를 판정하여 평가에 사용하였다. 2009년 평가 결과는 <표 4>와 같으며, Robel이 제안한 알고리즘이 0.796으로 F-measure 관점에서 최고점을 받았다.

〈표 4〉 Audio Onset Detection 평가결과

참가자	F-Measure
Robel(2)	0.796
Robel(1)	0.790
Robel(5)	0.783
Pertusa & Inesta	0.768
Robel(4)	0.765
Robel(3)	0.761
Tan, Zhu & Chaisorn(1)	0.744
Tan, Zhu & Chaisorn(2)	0.734
Tan, Zhu & Chaisorn(3)	0.686
Tan, Zhu & Chaisorn(5)	0.682
Tan, Zhu & Chaisorn(4)	0.679
Tzanetakis	0.595

3. Audio Melody Extraction

본 기술은 반주 등을 포함하는 폴리포닉 음악에서 멜로디를 추출하는 기술로써 Query by Singing/Humming, Audio Similarity and Retrieval, Music Recommendation 등에 사용된다. 이 task는 voicing 검출과 피치 검출로 나누어서 생각할 수 있다. Voicing 검출은 특정 시간영역 프레임에 멜로디 피치 정보가 있는지 여부를 판단하는 것이고, 피치 검출은 voicing 프레임에서 멜로디 피치를 추출하는 것이다. 사용된 데이터베이스로는 MIR-1K Dataset (가라오케에서 녹음한 중국 노래 1000곡으로 구성되어 있으며, singing voice와 반주를 각각 왼쪽과 오른쪽 채널에 별도로 녹음함), MIREX08 database (북부 인디아 지방 노래로써 singing

<표 5> Audio Melody Extraction 평가결과

참가자	Accuracy
Dressler	73,35%
Durrieu & Richard(1)	66,86%
Durrieu & Richard(2)	66,17%
Rao & Rao	65,22%
Cancela	62,88%
Joo, Jo & Yoo	56,64%
Cao & Li(2)	55,19%
Tachibana, Ono & Sagayama	55,08%
Wendelboe	55,07%
Cao & Li(1)	52,19%
Hsu, Jang & Chen(1)	50,49%
Hsu, Jang & Chen(2)	49,01%

voice와 다양한 인디아 악기들을 함께 녹음했음), MIREX05 database (Rock, R&B, 재즈, 클래식 피아노 등의 장르를 포함하는 노래) 등이며 각각 10ms 마다 손으로 표시한 Voicing 정보와 피치 정보를 포함하고 있다. 성능 평가 결과는 <표 5>에 있다.

4. Query by Singing/Humming/Tapping

Query by Singing/Humming/Tapping은 사람의 노래소리, 허밍, 태핑을 이용하여 오디오 파일을 검색하는 기술이다. Query by Singing/Humming을 위한 음악 데이터베이스로는 Jang's corpus와 IOACAS corpus를 사용하였고, 노이즈 데이터로는 Essen collection을 사용하였다.

Jang's corpus는 음악검색의 대상이 되는 48개의 모노포닉 미디 그룹과 118명으로부터 녹음한 4431개의 Singing/Humming 그룹으로 구성되어 있다. 미디 그룹은 미디파일과 미디파일로부터 추출된 wav파일, 그리고 각 미디음의

semitone이 기록되어 있는 pv파일로 구성된다. Singing/Humming 그룹은 실제 녹음된 허밍 wav 파일, 수작업으로 찾은 semitone을 기록한 파일, 그리고 미디파일로 구성된다. Jang's collection의 허밍음은 전화, 마이크로폰 등 다양한 환경에서 녹음하였고, 모두 노래의 시작부분만을 이용하여 만들었다.

IOACAS corpus에는 총 106개의 MIDI 파일이 존재한다. Query로 사용되는 것은 106명의 사람이 녹음한 총 355개의 Singing /Humming wav 파일이다. 이 Query는 Jang's corpus와는 다르게 노래의 시작 부분만을 사용하지 않고 임의 부분을 사용하였다. 2009년에는 192개 MIDI 파일과 404개 Query로 구성된 IOACAS corpus2를 추가적으로 사용하였다.

세번째 DB인 Essen collection은 Jang's DB와 IOACAS corpus에 포함되지 않은 20000곡의 noise MIDI 파일이다. Essen collection은 독일, 폴란드, 중국 등 다수 국가의 노래들을 담고 있다.

테스트는 Jang's corpus, IOACAS corpus, Essen collection을 이용하는 subtask와, Jang's corpus와 IOACAS corpus를 입력 query로 이용하고 데이터베이스로는 ground-truth 미디의 다양한 변형을 사용하는 두 번째 subtask로 구성하였다. 각 subtask의 성능평가 기준은 다음 표에서 보듯이 top-10 후보곡을 출력하여 mean reciprocal rank (MRR)를 계산하는 방법과 top-10 내에 올바른 매칭이 이루어지면 1점씩 count하는 방식을 사용하였다. <표 6>과 <표 7>에서 보듯이 두가지 subtask 모두에 대해서 Jang이 제안한 알고리즘들이 가장 우수한 성능을 나타내었다. Query by Singing /Humming의 경우 2006년부터 매년 평가가 이루어지고 있



〈표 6〉 Query-by-Singing/Humming 평가결과 (Known-Item)

참가자	MRR
Jang(2)	0.81
Hanna, Allali, Ferraro & Robine	0.67
Jang(1)	0.66

〈표 7〉 Query-by-Singing/Humming 평가결과 (Variations)

참가자	Count
Jang(1)	8.48
Hanna, Allali, Ferraro & Robine	6.98
Jang(2)	6.97

으며 평균 20개의 알고리즘이 경합을 벌이고 있는 분야이나, 2009년에는 참가 알고리즘 수가 적어서 좀 더 다양한 평가가 이루어지지 못했다.

5. Audio Artist/Genre/Mood Classification

본 기술 분야에서는 입력 오디오 신호로부터 가수이름, 장르, 무드 등을 분류하고자 하였다. 2009년도 task 중에서 가장 많은 알고리즘들이 참여했으며, 가수이름, 장르, 무드별로 30, 64, 33개의 알고리즘들이 경합을 벌였다.

2009년도 Artist 데이터베이스로는 베토벤, 바하, 브람스, 쇼팽, 하이든 등 11명 클래식 작곡가의 2772곡을 사용하였다. 평가결과는 〈표 8〉에서 보듯이 Wack, Gaus, Laurier, Meyers, Marxer, Bogdanov, Serra, Herrera가 제안한 알고리즘이 62.05%의 정확도로 최고 점수를 얻었다.

장르 검색을 위해서는 MIREX2007 collection과 Latin music database를 이용하였다. MIREX2007 collection은 블루스, 재즈, 컨트

〈표 8〉 Audio Classical Composer Identification 평가결과 (상위 15위)

참가자	Accuracy
Wack, Gaus, Laurier, Meyers, Marxer, Bogdanov, Serra & Herrera(2)	62.05%
Cao & Li(1)	60.97%
Cao & Li(2)	60.03%
Xu, Zhen & Zheng(RUC2)	57.18%
Wang(1)	56.35%
Burred & Peeters(1)	55.66%
Greco, Lidy & Rauber(1)(full)	55.34%
Burred & Peeters(2)	54.76%
Wack, Gaus, Laurier, Meyers, Marxer, Bogdanov, Serra & Herrera(1)	54.73%
Liu, Zheng & Gao(RUC1)	54.40%
Lidy, Greco, Rauber, Pertusa, Leon & Inesta (1) (WMM)	53.57%
Lidy, Greco, Rauber, Pertusa, Leon & Inesta (2) (BWWW)	53.57%
Xu, Liu, Zheng, Zhen & Gao(RUC1)	53.54%
Wang(2)	53.10%
Seyerlehner & Schedl	52.56%

리, 바로크, 클래식, 로맨틱, 일렉트로니카, 힙합, 록, 하드록/메탈 등 10개 장르의 7000곡을 포함한다. Latin music database는 MIREX2007 collection과는 전혀 다른 탱고, 볼레로 등의 장르 음악을 포함하고 있다. 두가지 데이터베이스를 이용한 테스트 모두에 대해서 Cao와 Li가 제안한 알고리즘이 가장 좋은 성능을 보였으며, MIREX 2007collection에 대해서는 73.33%, Latin Music Database에 대해서는 74.66%의 정확도를 보였다.

음악 심리학 분야에서 무드는 음악을 가장 잘 표현하는 방법으로 인식되고 있다. 또한, 사용자의 음악 검색에서도 무드를 이용하는 방식이 가장 많이 사용되고 있다^[9,10]. 본 평가에서 무드 검색은 고전적인 트레이닝-테스트 알고리즘을 이용하였다. 무드에 대한 정의와 분류는 연구자

〈표 9〉 Audio Genre Classification 평가결과
(MIREX 2007 collection: 상위 15위)

참가자	Accuracy
Cao & Li(2)	73.33%
Cao & Li(1)	73.23%
Greco, Lidy & Rauber(1)(full)	71.23%
Burred & Peeters(1)	70.63%
Wack, Gaus, Laurier, Meyers, Marxer, Bogdanov, Serra & Herrera(5)	70.44%
Xu, Zhen & Zheng(RUC2)	69.36%
Xu, Liu, Zheng, Zhen & Gao(RUC1)	68.93%
Lidy, Greco, Rauber, Pertusa, Leon & Inesta (1) (WMV)	68.84%
Burred & Peeters(2)	68.51%
Liu, Zheng & Gao(RUC1)	68.29%
Tsunoo, Tzanetakis, Ono & Sagayama	67.89%
Tzanetakis(2)(stereo)	67.87%
Lidy, Greco, Rauber, Pertusa, Leon & Inesta (2) (BWWV)	67.39%
Seyerlehner & Schedl	66.60%
Wang(1)	65.99%

들마다 다르며 평가하는 데이터베이스도 일치하지 않으므로 알고리즘간의 비교가 불가능하다. 따라서, IMIRSEL에서는 AMG 무드 레이블을 이용해서 아래와 같이 5가지 무드 클러스터를 정의하였다^[11].

- Cluster 1: passionate, rousing, confident, boisterous, rowdy
- Cluster 2: rollicking, cheerful, fun, sweet, amiable/good natured
- Cluster 3: literate, poignant, wistful, bittersweet, autumnal, brooding
- Cluster 4: humorous, silly, campy, quirky, whimsical, witty, wry
- Cluster 5: aggressive, fiery, tense/nxious, intense, volatile, visceral

〈표 10〉 Audio Mood Classification 평가결과

참가자	Accuracy
Cao & Li(1)	65.67%
Cao & Li(2)	65.50%
Peeters	63.67%
Wack, Gaus, Laurier, Meyers, Marxer, Bogdanov, Serra & Herrera(5)	62.83%
Wang(2)	61.67%
Liu, Zheng & Gao(RUC1)	61.67%
Wang(1)	61.33%
Greco, Lidy & Rauber (1) (full)	60.83%
Feng, Chen & Yang(1)	60.33%
Lidy, Greco, Rauber, Pertusa, Leon & Inesta (2) (BWWV)	60.17%
Xu, Zhen & Zheng(RUC2)	60.00%
Wack, Gaus, Laurier, Meyers, Marxer, Bogdanov, Serra & Herrera(3)	59.83%
Burred & Peeters(2)	59.67%
Wack, Gaus, Laurier, Meyers, Marxer, Bogdanov, Serra & Herrera(6)	59.50%
Tzanetakis(1)(mono)	59.33%

또한, 총 1250곡의 APM collection (www.apmmusic.com)을 APM 메타데이터를 이용하여 5가지 무드 클러스터 중 한가지로 분류하여 테스트에 사용할 수 있도록 하였다. APM 메타데이터가 옳게 분류되었는지 확인하기 위해서 3명이 들어보고 2명 이상이 옳다고 생각하는 무드 클러스터로 재분류를 실행하였다. <표 10>은 이렇게 만들어진 데이터를 이용하여 알고리즘을 트레이닝/테스트하여 성능 평가를 실시한 결과이다. <표 10>에 보듯이 무드 분류에 있어서도 Cao와 Li가 제안한 알고리즘이 65.67%로 가장 좋은 성능을 보였다.

6. Audio Tag Classification

본 기술은 음악을 듣고 연상되는 단어(장르,

스타일, 악기, 템포, 무드, 언제듣기 좋은 곡인지, 가수의 목소리 스타일)를 검출하여 검색에 이용하는 기술이다. 예를 들어, '기타가 들어간 행복한 노래'란 사용자의 요청에 응답하는 시스템을 디자인 할 때 사용될 수 있다. 하나의 오디오 클립은 여러 개의 태그를 가질 수 있으며, 데이터베이스 구축을 위해서 2,300곡에 대해서 MajorMinor 게임을 이용하여 총 73,000개의 태깅을 얻었다. 2009년 테스트를 위해서 USPOP을 포함하여 3,469곡에 대해서 <표 11>의 태그 그룹 분류에 기반 해서 18개의 무드 태그를 중복을 허용하며 할당하였다^[12].

실험 결과, <표 12>와 <표 13>에 보듯이 두가지 subtask 모두에 대해서 Lo와 Wang이 제안

<표 12> Audio Tag Classification 평가결과 (Major Minor)

Participant	F-Measure
Lo, Wang & Wang(2)	0.311
Tzanetakis(2)	0.293
Burred & Peeters(2)	0.29
Tzanetakis(1)	0.29
Lo, Wang & Wang(1)	0.289
Burred & Peeters(1)	0.277
Cao & Li(4)	0.263
Cao & Li(2)	0.241
Cao & Li(1)	0.209
Cao & Li(3)	0.17
Hoffman, Blei & Cook	0.044

한 알고리즘이 최고의 성능을 보였다.

<표 11> 오디오 태그 그룹 분류표

ID	Tags
G12	calm, comfort, quiet, serene, mellow, chill out, calm down, calming, chillout, comforting, content, cool down, mellow music, mellow rock, peace of mind, quietness, relaxation, serenity, solace, soothe, soothing, still, tranquil, tranquility, tranquillity
G15	sad, sadness, unhappy, melancholic, melancholy, feeling sad, mood: sad - slightly, sad song
G5	happy, happiness, happy songs, happy music, glad, mood: happy
G32	romantic, romantic music
G2	upbeat, gleeful, high spirits, zest, enthusiastic, buoyancy, elation, mood: upbeat
G16	depressed, blue, dark, depressive, dreary, gloom, darkness, depress, depression, depressing, gloomy
G28	anger, angry, choleric, fury, outraged, rage, angry music
G17	grief, heartbreak, mournful, sorrow, sorry, doleful, heartache, heartbreaking, heartsick, lachrymose, mourning, plaintive, regret, sorrowful
G14	dreamy
G6	cheerful, cheer up, festive, jolly, jovial, merry, cheer, cheering, cheery, get happy, rejoice, songs that are cheerful, sunny
G8	brooding, contemplative, meditative, reflective, broody, pensive, pondering, wistful
G29	aggression, aggressive
G25	angst, anxiety, anxious, jumpy, nervous, angsty
G9	confident, encouraging, encouragement, optimism, optimistic
G7	desire, hope, hopeful, mood: hopeful
G11	earnest, heartfelt
G31	pessimism, cynical, pessimistic, weltenschmerz, cynical/sarcastic
G1	excitement, exciting, exhilarating, thrill, ardor, stimulating, thrilling, titillating

〈표 13〉 Audio Tag Classification 평가결과
(Mood Set)

Participant	F-Measure
Lo, Wang & Wang(2)	0.219
Tzanetakis(1)	0.211
Tzanetakis(2)	0.209
Lo, Wang & Wang(1)	0.204
Burred & Peeters(1)	0.195
Burred & Peeters(2)	0.193
Cao & Li(4)	0.183
Cao & Li(2)	0.18
Cao & Li(1)	0.172
Cao & Li(3)	0.147
Peeters	0.084
Hoffman, Blei & Cook	0.063

7. Audio Cover Song Identification

커버송의 정의는 “a recording of a song or tune which has previously been recorded by someone else”이다. 즉, 원곡의 팀버, 템포, 리듬, 가사, 언어 등을 바꿔서 부른 곡을 의미한다. 따라서, 커버송 인식은 Audio Similarity and Retrieval 기술과 유사한 분야이다.

테스트에 사용된 데이터베이스는 49곡의 마주르카 무곡 커버송 중에서 11곡을 임의로 선정하여 구성하였다. 두번째 데이터베이스는 각각 11곡의 커버송을 가지는 30곡에 대해서 테스트를 실시하였다. 이 경우는 클래식, 재즈, 가스펠 등 다양한 장르를 포함하고 있다.

테스트 결과 <표 14>와 <표 15>에 보듯이 Serra, Zanin, Andrzejak이 제안한 알고리즘이 각 데이터베이스에 대해서 0.96과 0.75의 정확도로 가장 우수한 성능을 보였다.

〈표 14〉 Audio Cover Song Identification 평가결과
(Mazurkas)

참가자	Precision
Serra, Zanin & Andrzejak	0.96
Ravuri & Ellis	0.91
Ahonen	0.56

〈표 15〉 Audio Cover Song Identification 평가결과
(Mixed)

참가자	Precision
Serra, Zanin & Andrzejak	0.75
Ravuri & Ellis	0.66
Ahonen	0.2

8. Audio Similarity and Retrieval

본 task는 다양한 음악 신호 간의 유사도를 측정하는 기술로써, Audio Music Mood Classification, Audio Artist Identification, Audio Genre Classification, Audio Cover song identification과 비슷한 기술 분야이다. 데이터 베이스로는 10개 장르 7000곡의 오디오 클립을 이용하였으며, 평가방법에는 사람이 직접 들어보고 유사도를 결정하는 방법 (not similar, somewhat similar, very similar 중 한 가지를 선택하거나 0-10 스케일로 유사도를 평가)과 객관적으로 측정하는 방법이 있다. 객관적 측정법의 경우 장르/작곡가/앨범 매치 등을 이용하여 number of times similar curve를 그려서 측정한다. 평가 결과, <표 16>에 보듯이 Pohle와 Schnitzer이 제안한 알고리즘이 2007년 평가에 있어서 가장 우수한 알고리즘으로 선정되었다. 또한, Pohle와 Schnitzer이 2007년에 제안한 알고리즘은 본인들이 2009년에 제안한 알고리즘에 있어서 2위를 차지했다.

<표 16> Audio Music Similarity 평가결과

참가자	Fine Score
Pohle & Schnitzer(2009)	6.46
Pohle & Schnitzer(2007)	5.75
Bogdanov, Serra, Wack & Herrera (hybrid)	5.73
Lidy & Rauber	5.47
Anonymous	5.39
Cao & Li (2)	5.39
Tzanetakis	5.34
Bogdanov, Serra, Wack & Herrera (class)	5.14
Hubler(1)	5.04
Hubler(2)	4.93
Fields(mfcc10)	2.59
Maillet & Eck(sda)	2.59
Cao & Li(1)	2.53

9. Audio Chord Detection

본 기술은 하모닉 특성과 리듬 특성 등을 이용하여 chord를 검출하는 기술로, 오디오 장르 분류나 오디오 유사도 검사 등에 사용된다^[4]. 다양한 chord 검출 알고리즘이 존재하지만 사용하고 있는 데이터베이스와 측정 방법의 차이로 상호 비교가 어려우므로, 본 task에서는 이를 해결하고자 하였다.

데이터베이스로는 Christopher Harte의 비틀즈 데이터셋을 이용하였다. 또한, 2009년에는 Mauch가 제공한 Queen과 Zweieck 데이터셋도 추가적으로 사용하였다. 실측 정보인 Ground-truth 데이터는 다음과 같이 시작시간, 끝시간, 그리고 chord 레이블의 형태로 표현되었다^[6].

41.2631021 44.2456460 B

44.2456460 45.7201130 E
 45.7201130 47.2061900 E:7/3
 47.2061900 48.6922670 A
 48.6922670 50.1551240 A:min/b3

<표 17>과 <표 18>에 보듯이, 트레이닝 후 테스트를 진행한 경우에 대해서는 Weller, Ellis, Jebara 알고리즘이, 트레이닝 없이 테스트를 한 경우에는 Mauch과 Dixon의 알고리즘이 가장 우수한 성능을 보였다.

<표 17> Audio Chord Detection 평가결과 (Train-Test)

참가자	Overlap
Weller, Ellis & Jebara(4)	74.20%
Weller, Ellis & Jebara (2)	72.30%
Weller, Ellis & Jebara(3)	72.30%
Weller, Ellis & Jebara(1)	70.40%
Reed, Ueda, Siniscalchi, Uchiyama, Sagayama & Lee	70.10%

<표 18> Audio Chord Detection 평가결과 (No Training)

참가자	Overlap
Mauch & Dixon	71.20%
Oudre, Grenier & Fevotte(2)	71.10%
Khadkevich & Omologo(2)	70.80%
Oudre, Grenier & Fevotte(1)	70.60%
Ellis	69.70%
Khadkevich & Omologo(1)	69.70%
Pauwels, Varewyck & Martens(1)	68.20%
Papadopoulos & Peeters	67.30%
Pauwels, Varewyck & Martens(2)	65.40%
Rocher, Robine, Hanna & Strandh(1)	62.90%
Harte	57.60%
Rocher, Robine, Hanna & Strandh(2)	54.50%

10. Audio Beat Tracking

본 task에서는 오디오 신호에서 비트 위치를 검출하고자 하였다. 데이터베이스로는 Chopin의 Mazurka 367 pieces와 MIREX2006에서 사용했던 McKinney 데이터를 이용하였다. McKinney 데이터는 30초짜리 160곡으로 구성되어 있으며 대부분의 경우에 40명의 청취자가 들어보고 비트 위치를 손으로 표시하였다. 이 중에서 20곡은 트레이닝을 위해서 참가자에게 미리 제공되며,

<표 19> Audio Beat Tracking 평가결과 (Mazurka)

참가자	F-Measure
Davies, Robertson & Plumbly (3)	0.678
Peeters (2) (VE)	0.547
Peeters (1) (VE)	0.542
Davies, Robertson & Plumbly (2)	0.484
Peeters (4) (CE)	0.454
Davies, Robertson & Plumbly (4)	0.453
Lee	0.449
Peeter (3)(CE)	0.447
Oliveira, Gouyon & Martins (R)	0.321
Oliveira, Gouyon & Martins (SC)	0.307
Davies, Robertson & Plumbly (1)	0.279

<표 20> Audio Beat Tracking 평가결과 (McKinney)

참가자	F-Measure
Peeters(1)(VE)	0.548
Peeters(3)(CE)	0.546
Peeters(4)(CE)	0.545
Peeters(2)(VE)	0.537
Lee	0.5
Davies, Robertson & Plumbly(4)	0.48
Davies, Robertson & Plumbly(3)	0.47
Oliveira, Gouyon & Martins(R)	0.415
Oliveira, Gouyon & Martins(SC)	0.397
Davies, Robertson & Plumbly(2)	0.354

나머지 140곡은 테스트에 사용하였다. <표 19>와 <표 20>에 보듯이, Mazurka와 McKinney 데이터베이스에 대해서 각각 Davies, Robertson, Plumbly 알고리즘과 Peeters 알고리즘이 가장 좋은 성능을 보였다. Peeters 알고리즘은 Mazurka 데이터베이스에 대해서도 2등 그룹을 형성하였다.

11. Structural Segmentation

본 task는 Single channel/CD-quality 오디오 신호의 구조를 segment 별로 검출하는 기술로써, “<onset_time> <offset_time> <label>”과 같이 Chris Harte의 코드 레이블링 파일 형태의 출력을 요구하였다. 220곡에 대해서 Ground truth 데이터가 존재하며, 그 중 177곡은 비틀즈 노래이다. 실제 평가에서는 label이 옳은지 보다는 onset-offset에 대한 평가를 주로 하게 되며, <표 21>에 보듯이 평가결과는 Mauch, Noland, Dixon 알고리즘이 가장 좋은 성능을 보였다.

<표 21> Structural Segmentation 평가결과

참가자	F-Measure
Mauch, Noland & Dixon	0.6
Anonymous(1)	0.582
Anonymous(2)	0.577
Paulus & Klapuri	0.54
Peeters	0.533

12. Music Recommendation

본 기술은 2009년에 처음 제안된 task로 참가자가 부족해서 테스트가 진행되지는 못했다. 음

악추천 기술은 애플 아이튠즈의 Genius가 제공하는 것과 같이 seed 곡으로부터 추천곡을 찾거나, 개인용 라디오 (personalized radio)에서 쓰이는 플레이 리스트를 제공하는 것을 목표로 한다. 본 기술은 신호처리적인 접근 방법 뿐만 아니라 특정 음악 구매자의 다른 음악 구매 빈도수나 사용자의 특정 음악 청취 빈도수 등을 측정한 사용자 프로파일을 이용한 collaborative filtering 방법 등이 많이 사용되고 있다.

III. 음악 정보검색 기술의 응용처

음악 정보검색 (Music Information Retrieval) 기술은 세 가지 주요 응용처가 있으며, 기반 기술들과의 상호 연관성이 크므로 기반 기술 단계부터 조직적으로 개발할 필요가 있다.

Singing/Humming/Tapping을 이용한 음원 검색 기술은 사람의 노래, 허밍, 또는 태핑 신호에서 멜로디, 박자, 가사 등의 정보를 추출하여 이를 기반으로 사용자 요구에 맞는 최적의 음원을 검색하는 차세대 오디오 검색 기술이다. 해당 기술을 대용량의 동영상 브라우징 서비스 등에 적용할 경우, 연산량이 적은 오디오 정보만을 이용하여 중요 구간을 우선적으로 탐색하고, 비디오 정보 등을 추가적으로 이용하여 상세히 탐색하는 MPEG-7의 고속 멀티미디어 데이터베이스 검색 기술 분야에 적용이 가능하리라 여겨진다. 또한, 모바일 단말기, 셋톱박스, 노래방 기기 및 미디어 서비스 등 다양한 응용처에 이용될 수 있을 것으로 기대된다. 본 기술과 관련된 MIREX task로는 Audio Beat Tracking, Audio Melody Extraction, Audio Onset Detection,

Audio Tempo Detection, Multiple Fundamental Frequency Estimation /Note Tracking 등이 있다.

또한, 디지털 음악 컬렉션이 증가함에 따라서, 유사곡 또는 표절곡을 탐색하는 기술 개발이 중요해지고 있다. 본 기술은 음반 업체들에서 표절곡 탐색 등에 쓰이거나, 작곡가 입장에서 본인이 작곡한 곡의 표절 여부를 조기 진단하도록 함으로써 digital rights management (DRM) 분야에 기여할 수 있다. 특히, 한류와 더불어 증가하고 있는 국내 음원의 해외 표절을 찾아 내는데 적용될 수 있다. 본 기술과 관련된 MIREX task로는 Audio Beat Tracking, Audio Chord Detection, Audio Melody Extraction, Audio Onset Detection, Multiple Fundamental Frequency Estimation /Note Tracking, Audio Cover Song Identification, Audio Similarity and Retrieval 등이 있다.

세 번째, Music Recommendation 기술은 음악 신호로부터 가수가 누구인지 (Audio Artist Identification), Jazz, Classical, Hip-Hop, Rock 등 어떤 장르의 곡인지 (Audio Genre Classification), 그리고 지루한, 흥분되는, 행복한, 신경질적인, 평화로운, 기쁜, 슬픈, 졸린 등 어떤 무드의 곡인지 (Audio Mood Classification)를 자동 분류한 후 사용자 프로파일에 따라서 적절한 곡을 추천하는 기술이다. 본 기술을 적용하면 감성기반 오디오 인터페이스 기술 개발을 통해 고정형 단말기, 모바일 단말기, 온라인 뮤직 스토어, 가정용 오디오 기기 및 포털 서비스 등 다양한 정보전자기기 사용자에게 개인화 추천 서비스 관점에서 한 차원 진보된 서비스를 제공할 수 있으리라 기대된다.

IV. 결론

최근 들어 사용자가 접하는 음악 콘텐츠의 양이 기하급수적으로 증가함에 따라, 음악 내의 다양한 정보를 가공하여 사용자의 요구에 부응하는 서비스를 제공하는 음악 정보검색 (Music Information Retrieval) 기술의 필요성이 대두되고 있다. 음악 정보검색 기술에 대한 최근 동향은 ISMIR Conference에서 매년 발표되고 있으며, 기술별 최고 성능을 가지는 알고리즘은 MIREX 평가회를 통해 확인할 수 있다. 본 기고에서는 MIREX 평가회의 기술 동향을 살펴 보았고, 음악 정보검색 기술의 응용처를 기술하였다. MIREX 평가회에 참가한 알고리즘 각각에 대한 설명은 [8]을 참고하기 바란다.

MIREX 평가회에서 다뤄지고 있는 주요 task는 크게 응용기술과 기반기술로 나뉘어 진다. 음악 정보검색을 이용한 응용기술은

- Audio Artist/Classical Composer Identification
- Audio Cover Song Identification
- Audio Genre Classification
- Audio Mood Classification
- Audio Similarity and Retrieval
- Audio Tag Classification
- Music Recommendation
- Query by Singing/Humming/Tapping
- Symbolic Genre Classification
- Symbolic Melodic Similarity

와 같으며, 기반기술은

- Audio Beat Tracking

- Audio Chord Detection
- Audio Drum Detection
- Audio Key Finding
- Audio Melody Extraction
- Audio Onset Detection
- Audio Tempo Detection
- Multiple Fundamental Frequency Estimation/Note Tracking
- Symbolic Key Finding
- Structural Segmentation

와 같다. 미래에 개발될 음악 정보검색 응용 시스템을 위해서는 기반기술의 개선이 필수적이다. 예를 들어서, Audio Onset Detection과 Multiple Fundamental Frequency Estimation 기술을 결합한 Audio Melody Extraction 기술의 근본적인 개선은 Query by Singing/Humming을 비롯한 다양한 응용 및 서비스 분야 기술의 성능을 한 단계 향상시킬 수 있을 것이다.

참고문헌

- [1] 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.
- [2] N. Orio. *Music retrieval: A tutorial and review*. Now Publishers Inc., 2006.
- [3] <http://www.ismir.net>
- [4] P. Cano, E. Gomez, F. Gouyon, P.

- Herrera, M. Koppenberger, B. Ong, X. Serra, S. Streich, and N. Wack. ISMIR 2004 audio description contest. Technical Report MTG-TR-2006-02, MTG, Pompeu Fabra University, 2006.
- [5] J. S. Downie, "The scientific evaluation of music information retrieval systems: foundations and future," *Computer Music Journal*, Vol.28, No.2, pp.12-23, 2004.
- [6] E. M. Voorhees and D. K. Harmon. *TREC: Experiment and Evaluation in Information Retrieval*. MIT Press, 2005.
- [7] <http://www.music-ir.org/mirex/2009>
- [8] http://www.music-ir.org/mirex/2009/index.php/MIREX2009_Results
- [9] S. J. Cunningham, M. Jones, and S. Jones, "Organizing digital music for use: An examination of personal music collections," in *Proc. Int. Conf. Music Information Retrieval*, pp.447-454, 2004.
- [10] S. J. Cunningham, D. Bainbridge, and A. Falconer, "More of an art than a science: Supporting the creation of playlists and mixes," in *Proc. Int. Conf. Music Information Retrieval*, Victoria, Canada, pp.240-245, 2006.
- [11] X. Hu and J. S. Downie, "Exploring mood metadata: Relationships with genre, artist and usage metadata," in *Int. Conf. Music Information Retrieval*, pp.23-27, 2007.
- [12] X. Hu, J. S. Downie, A. Ehmman, "Lyric Text Mining in Music Mood Classification," in *Int. Conf. Music Information Retrieval*, pp.411-416, 2009.
- [14] C. A. Harte and M. B. Sandler, "Automatic chord identification using a quantised chromagram," in *Proc. Audio Engineering Society's Convention*, pp.28-31, 2005.

저자소개



김 무 영

1993년 2월 연세대학교 전자공학과 학사
 1995년 2월 연세대학교 전자공학과 석사
 2004년 11월 스웨덴 왕립공과대학 박사
 1995년 2월~2000년 12월 삼성종합기술원 전문연구원
 2004년 12월~2005년 2월 스웨덴 왕립공과대학 PostDoc
 2005년 2월~2006년 8월 스웨덴 에릭슨 중앙연구소 선임연구원
 2006년 8월~현재 세종대학교 조교수

주관심 분야 : 음성/오디오 신호처리, 패턴인식, 정보이론

저자소개



이 석 필

1990년 2월 연세대학교 전기공학과 공학사
1992년 8월 연세대학교 전기공학과 공학석사
1997년 8월 연세대학교 전기전자공학과 공학박사
1997년~2002년 대우전자 영상 연구소, 선임연구원
2002년~현재 KETI 디지털미디어연구센터, 센터장
주관심 분야 : 디지털방송통신융합시스템, A/V 신호처리