

工學博士學位 請求論文

주파수 분석을 이용한 한국전통음악의  
인식 및 분류

Recognition and Classification of Korean Traditional  
Music using Spectral Analysis

2006년 8월

仁 荷 大 學 校 大 學 院

電子工學科(電子通信工學專攻)

金 載 天

工學博士學位 請求論文

주파수 분석을 이용한 한국전통음악의  
인식 및 분류

Recognition and Classification of Korean Traditional  
Music using Spectral Analysis

2006년 8월

指導教授 郭慶燮

이 論文을 工學博士學位 論文으로 提出함

仁 荷 大 學 校 大 學 院

電子工學科(電子通信工學專攻)

金 載 天

## 요 약

급증하는 멀티미디어 데이터의 처리와 효율적인 검색서비스를 위하여 음성, 영상 인식의 필요성이 증가하고 있다. 본 연구는 한국전통악기와 장르의 공학적인 분석에 대한 관심에서 출발하였다. 기존의 멀티미디어 데이터는 파일 제목의 텍스트를 기준으로 분류하였으며 파일의 내용과 다른 제목이 붙여졌을 경우 이를 확인할 방법이 없었다. 음악파일의 내용을 인식하고 이를 자동으로 분류하게 된다면 급증하는 음악파일의 처리에 도움이 될 것이다. 인간의 음악인식과정을 이해하고 그 모델을 적용하였으며 한국전통음악의 장르에 대한 문화, 역사, 악기의 구성에 대하여 이해하고 4개의 장르 군으로 분류하였다.

본 연구는 청각모델을 한국전통음악에 적용한 최초의 시도로서 한국전통음악 장르구분이라는 목표를 설정하였다. 이를 위하여 최종실험을 위한 기본실험을 수행하였다. 청각시스템의 이해를 위하여 달팽이관에서 이루어지는 주파수 분해 및 청신경변환 과정을 단계별로 도표화 하였다. 이와 같이 음악인식시스템의 기본 구성은 음악파일을 주파수영역으로 푸리에 변환하게 된다. 주파수영역으로 변환된 데이터로부터 음악의 특성, 즉 리듬, 질감, 에너지등을 지표화 할 수 있는 특성함수들을 사용하여 특성값을 추출하여 특성값벡터를 얻는다. 분류알고리즘을 사용하여 특성값벡터를 분류하여 입력데이터를 분류한다.

본 연구에서는 특성값을 추출하기위한 특성값함수들 중에서 평균피크함수를 제안하여 효용성을 실험으로 입증하였고 룰오프 포인트함수의 최적 파라미터를 구하는 실험을 수행하였다. 현재 음악인식 및 분류연구에서 가장 많이 사용되고 있는 k-최근접이웃법(k-NN), 최근접이웃법(NN), 베이즈법(Bayes rule)의 세가지 알고리즘의 성능을 서양악기와 한국전통악기를

사용하여 비교평가를 수행했다. 확률과정법인 베이즈법보다 벡터거리법인 k-NN과 NN법이 우수함을 확인하였다. 최종실험으로는 악기의 분류실험과 한국전통음악의 장르 분류실험을 수행하였다. 악기분류실험에서는 서양악기와 한국전통악기를 분류하여 94.4%의 성공률을 얻었다. 장르구분요소로는 악기의 구성, 리듬, 음의진행, 문화, 사회적인 요소들이 있다. 본 연구에서는 한국전통음악을 악기의 구성 및 음향특성별로 4개의 장르로 분류한 다음, 기본실험에서 입증된 결과들을 적용하여 94.5%의 한국전통음악 장르 분류실험 결과를 얻었다.

한국전통음악의 형태와 음역, 사용악기 등에서 유사한 중국, 일본음악의 연구에도 공통적으로 적용할 수 있을 것이라 기대한다. 공학적인 기법과 인간의 의학적, 심리적 음악인식과정에 대한 연구가 병행된다면 더욱 발전된 시스템을 구현할 수 있을 것이다.

# **ABSTRACT**

To effectively process and search for multimedia data, the need for multimedia file recognition is increasing. This research was started because of an engineer's interest in Korean traditional instruments and the genre. The present searching systems discriminates against multimedia files by using titles of the files. If the file is incorrectly titled, there is no way to know the contents of the file. It would be helpful in processing music files and greatly increase the speed at which they are found, if there were a system that can recognized the contents of music files and classified music files automatically. The model of a human hearing system is applied in this thesis. Finally, Korean traditional music is classified into four genres in view of culture, history, and instruments of Korean traditional music.

The purpose of this research is the recognition and classification of Korean traditional music. This is the first attempt to recognize and classify Korean traditional music. To accomplish this goal, a few preliminary experiments were conducted. To understand the cochlea system, every step of the cochlea response was plotted using impulse signals and taegum playing. For this process, the basic structure of the music recognition system uses a fourier transformation. From the fourier transformed data, features were extracted using feature functions to index characteristics, rhythm, energy and so on. These features were classified using an algorithm then the results were produced.

In this research, among the feature functions, the average peak function was suggested and it's effectiveness was proved by experiment. Also the

optimal roll-off point feature function was determined. Three classification algorithms, k-NN, NN, and Bayes rule were compared to pick up the most effective one using western instruments and Korean traditional instruments. From the results of the experiments, the k-NN, and NN algorithms show excellent performance in music classification. In the musical instrument classification experiment, the kayagum, taegum, and violin are classified 94.4% correctly. Elements in music genre are composition of instruments, rhythm, sound, culture and other factors. The Korean music genre categorized into four genres according to composition of instruments and playing style. Then the results of experiment showed a 94.5% success rate.

Since Korea, Japan and China have the same musical roots, both in instruments and playing style, analyzing Korean traditional music can be helpful in the understanding of Far East Asian traditional music. Science, psychology and musicology should be incorporated to realize a more advanced system.

# 순 서

요 약.....	i
ABSTRACT.....	iii
순 서.....	v
그림순서.....	viii
표순서.....	ix
제1장 서론.....	1
제2장 음악인식 시스템.....	7
2.1 음악인식 시스템의 개요.....	7
2.2 음악신호처리.....	9
2.2.1 음악신호의 푸리에 변환.....	10
2.2.2 물리적 특성과 인지적 특성.....	15
2.3 음악장르 구분에 사용되는 특성.....	17
2.3.1 멜-주파수 캡스트럴 계수(MFCCs).....	19
2.3.2 영교차율(average zero-crossing rate).....	21
2.3.3 단시간 에너지(short-time energy).....	24
2.3.4 리듬과 비트(rhythm and beat).....	25
2.3.5 음색(timbre).....	25
2.4 청각시스템.....	27
2.5 청각응답 모델링.....	30
2.5.1 단일채널 응답 모델.....	31
2.5.2 다중채널 응답 모델.....	33
2.5.3 악기 대금의 다중채널 청각응답.....	38

<b>제3장 음악데이터 분류알고리즘과 특성함수.....</b>	<b>44</b>
3.1 분류시스템의 구성요소.....	44
3.1.1 베이저언 추정이론.....	48
3.1.2 은닉 마코브 모델.....	53
3.1.3 최근접이웃법.....	54
3.1.4 k-최근접이웃법.....	55
3.2 악기음원을 이용한 분류알고리즘 성능비교 실험.....	56
3.2.1 국악기 음원을 이용한 알고리즘 성능비교.....	57
3.2.2 바이올린과 첼로 음원을 이용한 알고리즘 성능비교.....	66
3.3 바이올린 음원을 이용한 최적의 롤-오프 포인트 검출 실험.....	71
3.4 실험결과 및 해석.....	76
 <b>제4장 한국전통음악의 분류.....</b>	 <b>78</b>
4.1 한국전통악기의 분류.....	79
4.1.1 악기신호의 분석.....	79
4.1.2 특성값과 분류알고리즘.....	83
4.2 악기분류 실험.....	86
4.3 한국전통음악의 장르 분류 실험.....	90
4.3.1 한국전통음악의 장르.....	91
4.3.2 장르구분 실험.....	94
4.4 실험결과 및 해석.....	97
 <b>제5장 결론.....</b>	 <b>98</b>
 <b>참고문헌.....</b>	 <b>102</b>



# 그 립 순 서

그림 2.1	음악인식 시스템의 과정.....	8
그림 2.2	입력된 데이터집합으로부터 특성값 벡터의 생성.....	8
그림 2.3	(a) 직교좌표 (b) 극좌표.....	10
그림 2.4	정현파의 합성.....	11
그림 2.5	그림 2.4 신호의 푸리에 분석.....	12
그림 2.6	세가지 장르의 스펙트로그램.....	18
그림 2.7	MFCCs 처리순서.....	20
그림 2.8	국악(그림 2.6(a))의 MFCCs 플롯.....	21
그림 2.9	영교차율을 이용한 기본주파수의 계산.....	22
그림 2.10	청각기관의 구조.....	28
그림 2.11	달팽이관의 구조.....	29
그림 2.12	내이의 신호처리 과정.....	30
그림 2.13	기저막에 입력된 임펄스 신호.....	31
그림 2.14	기저막의 단일 채널 임펄스 응답.....	32
그림 2.15	청신경에 전달되는 일련의 자극열.....	33
그림 2.16	기저막의 멀티채널 임펄스 응답.....	35
그림 2.17	반과정류된 내유모세포 임펄스응답.....	36
그림 2.18	청신경에 전달되는 일련의 자극열.....	37
그림 2.19	대금산조의 오디오신호.....	38
그림 2.20	대금산조의 기저막 멀티채널 응답.....	40
그림 2.21	반과정류된 대금의 내유모세포 응답.....	41
그림 2.22	대금산조 연주에 의해 청신경에 전달되는 일련의 자극열 .....	42

그림 3.1 (a) 학습 데이터에 최적화된 판단경계. (b) 미지의 데이터를 위해 타협점 설정.....	47
그림 3.2 $d$ 개의 입력 $x_i$ 와 $c$ 개의 판별함수 $g_i(x)$ 로 구성된 분류시스템의 구조.....	50
그림 3.3 실험순서.....	72
그림 3.4 롤-오프 포인트 변화에 따른 분류 성공률.....	74
그림 4.1 시간영역에서의 가야금 신호.....	80
그림 4.2 (a) 가야금의 PSD 분포 (b) 해금의 PSD.....	81
그림 4.3 장고와 피리의 PSD 분포 비교: 실선; 장고, 점선; 피리.....	82
그림 4.4 실험순서.....	88

## 표 순 서

표 3.1 추출된 특성값의 예.....	59
표 3.2 Bayes rule 실험결과, 분류성공률 96.66%.....	60
표 3.3 NN 실험결과, 분류성공률 100%.....	60
표 3.4 k-NN 시험결과, 분류성공률.....	60
표 3.5 Bayes rule, 군별 세부 정확도.....	62
표 3.6 NN, 군별 세부 정확도.....	62
표 3.7 k-NN, 군별 세부 정확도.....	62
표 3.8 Bayes rule, 전체 정확도.....	64
표 3.9 NN, 전체 정확도.....	64
표 3.10 k-NN, 전체 정확도.....	65
표 3.11 추출된 특성값의 예.....	66
표 3.12 Bayes rule 실험결과, 분류성공률 77.08%.....	67
표 3.13 NN 실험결과, 분류성공률 88.54%.....	67
표 3.14 k-NN 시험결과, 분류성공률 89.58%, 2개의 표본데이터는 미분류.....	67
표 3.15 Bayes rule, 군별 세부 정확도.....	68
표 3.16 NN, 군별 세부 정확도.....	68
표 3.17 k-NN, 군별 세부 정확도.....	68
표 3.18 Bayes rule, 전체 정확도.....	69
표 3.19 NN, 전체 정확도.....	69
표 3.20 k-NN, 전체 정확도.....	70
표 3.21 추출된 특성값의 예.....	73
표 3.22 스펙트랄 롤오프 포인트 0.85에서 분류결과.....	75

표 3.23 전체 정확도.....	75
표 4.1 추출된 특성값의 예.....	87
표 4.2 분류 실험결과, 분류성공률 94.44%.....	88
표 4.3 군별 세부 정확도. ....	89
표 4.4 전체 정확도.....	89
표 4.5 연주형태에 따른 한국전통음악의 장르구분.....	95
표 4.6 k-최근접이웃법 실험결과, 분류성공률 95.4%.....	95
표 4.7 k-NN, 군별 세부 정확도.....	95
표 4.8 전체 정확도.....	96

## 제1장 서론

현재 통신망의 광대역화로 멀티미디어 데이터는 급격한 증가 추세에 있다. 멀티미디어 데이터의 효율적인 관리 및 검색은 앞으로 더욱 중요해질 것이다. 인터넷, 이동통신 분야에서도 영상, 음악관련 서비스를 선점하기 위한 경쟁이 진행 중이다. 이러한 상황에서 음악인식분야는 큰 잠재력을 가지고 있으며 정보기술 분야의 경쟁력확보에 필수적이다. 음악인식분야는 최근에 시작되어 아직 연구가 부족한 부분이 많이 있다. 상당부분 음성인식 분야의 연구도구를 차용하여 연구가 진행되고 있으며 주파수성분 분석을 통한 연구가 주류를 이루고 있다.

음악인식과정은 매우 난해한 과정이다. 우선 음악은 일반적으로 복합적인 다성의 소리이다. 지난 십여년 동안 많은 연구성과가 축적되어 왔으나 대부분의 연구는 실생활에는 존재하지 않는 단음에 대한 연구이다. 최근에 와서 복합적인 다성의 음악을 분석하고 연구하기 시작했다. 음악은 본질적으로 감각적인 지각의 결과물이다. 주기적인 공기의 진동 인간의 고막을 진동시키는 물리적측면의 신호는 많이 연구되어 왔지만 어떻게 지각하는지에 대해서는 완전히 알려져 있지 않다. 따라서 인간의 감각을 완전히 구현하는 컴퓨터 시스템을 구현하기에는 아직 어려움이 많이 있다.

다른 모든 과학 분야와 마찬가지로 음악인식 분야 역시 용어의 정의가 필요하다. 그러나 음악인식분야에서는 다른 분야와는 다르게 적절한 정의 없이 용어가 사용되고 있다. 음악의 특성상 음의 높이와 빠르기와 같은 주관적 용어를 사용한다. 그리고 인간의 인식은 감정상태, 문화, 개인에 따라 다르기 때문에 정확한 정의가 더욱 어렵다. 음악 또는 악기 인식은 크게 보아 음성인식의 한 분야이다. 음악 및 악기 인식은 소리로부터 그 소리를 특징 짓는 특성값을 추출하여 이 특성값을 사용하여 다른 음악 및

악기를 분류하는 일련의 과정이다.

컴퓨터를 이용하여 디지털신호를 처리하고 평균, 특정 레벨, 데이터 크기의 평균 등을 구하는 것보다 이를 이해하고 공통의 특성으로 분류해내는 것은 어려운 작업이다. 음악 또는 악기의 인식은 이와 같은 데이터를 이해하는 과정이다. 주파수영역에서 음악인식의 가능성을 확인한 실험으로 D. Perrot 과 R. O. Gjerdige이 행한 연구에서 인간의 놀랄만한 음악인식 능력을 보여줬다[2, 3]. 클래식, 블루스, 컨트리, 댄스, 재즈, 라틴, 팝, 알앤비, 랩, 록의 10개의 장르를 사용하였고 장르마다 8곡을 사용했다. 음반제작사가 분류한 체계를 결과와 비교하였다. 4곡은 사람 목소리가 들어간 음악이고 나머지 4곡은 악기연주곡이다. 3000ms, 475ms, 400ms, 325ms, 250ms길이인 5개의 표본을 각각의 곡으로부터 추출하였다. 실험대상은 52명의 심리학과 학생이었다.

참여자들의 평균음악감상 시간은 주당 24시간 이었다. 학생들에게 음악 표본 중 무작위로 고른 음악을 들려 주고 10개의 장르중 1개의 장르를 선택하라고 지시하였다. 3000ms 음악표본의 분류정확도는 70%였다. 장르의 구분이 모호하고 일부 전문가들도 장르구분에 의견충돌이 있는 것을 감안하면 비교적 높은 성공률을 보였다. 250ms 표본의 성공률은 40%였다. 동일한 시간길이의 표본일 경우 순수한연주곡의 판단정확도가 약간 높았다. 연구결과는 장르판단이 고차원적인 추론 없이도 가능하다는 점에서 흥미로웠다. 250ms길이의 샘플은 리듬, 멜로디, 구조 등을 파악하기에는 매우 짧은 시간이다. 따라서 주파수, 음색분석으로도 장르의 판단이 가능함을 보여주는 실험이었다.

관련연구 분야들은 서로 직간접적으로 관련되어 있을 뿐 아니라 많은 문제, 아이디어, 이슈 등을 공유하고 있다. 음악과 음성이 혼합된 신호의 경우 음성인식시스템 또는 음악분류시스템으로 인식하기 어렵다. 따라서

음악과 음성분류 시스템은 어떤 시스템의 전처리 단계에 유용하게 사용될 수 있다. 최근의 연구는 Foote[4], Zhang과 Kuo[5]의 연구에 나타나 있다. 비디오 콘텐츠 분석에서 오디오부분은 중요한 역할을 한다. 오랫동안 중요성이 간과되어 왔으나 최근에 와서 그 중요성이 인정받고 있다[6].

비디오 내용을 분류함에 있어 시각적인 분류보다 청각적인 분류가 유용할 때가 있다. 울음소리, 총소리등을 인식함으로써 폭력적인 장면을 자동으로 분류할 수도 있다[7]. Liu[8]은 광고, 농구게임, 축구게임, 뉴스, 날씨예보의 5가지 프로그램을 오디오 특성값을 이용하여 75%의 정확도로 분류하였다. 오디오 데이터베이스의 출현으로 새로운 조회시스템이 필요해졌다. 고전적인 방법으로 특정음악의 멜로디를 흥얼거리면 시스템은 비슷한 멜로디의 음악을 검색한다[9]. 기본적인 개념은 인접하고 있는 음들의 음고차를 이용하여 멜로디 구하고 이를 이용하여 패턴매칭을 수행한다. 그러나 미디(MIDI)와 같은 표준화된 데이터를 이용하여야 하는 단점이 있다. Subramanya는 웨이브파일을 이용하여 이러한 단점을 극복하였다[10]. 각각의 파일을 작은 블록으로 나눈 후 이산코사인변환(DCT; Discrete Cosine Transform)을 행하여 일렬의 상관계수를 얻는다. 이 계수를 이용하여 입력데이터와 유사한 매칭파일을 검색한다. 자동악보 작성의 목적은 입력된 오디오파일 데이터를 이용하여 음악악보를 작성하는 것이다. 단성음악 입력일 경우 음의 높이 고저만 고려하면 되므로 쉽게 시스템을 구현할 수 있다. 그러나 다성 음악인 경우 시스템은 복수의 음높이를 고려해야 하므로 시스템은 복잡해진다. 더욱이 명확한 시퀀스가 존재하지 않으므로 이전의 연주되던 음이 끝나기 전에 다른 음이 연주될 경위 상황은 더욱 복잡해진다[11]. 현재 진행되고 있는 대부분의 연구는 다성의 오디오 스트림을 단성음악으로 나눈 후 처리하는 방법을 따르고 있다.

Roger J. McNab은 사용자의 노래소리를 인식하여 악보로 변환하는 시

시스템을 개발하였다[12]. 다성음악 악보작성에 대한 연구는 Keith D. Martin의 논문에 기술 되어있다[13]. 현재 연구되고 있는 컴퓨터 소리인식 시스템은 신호의 적절한 부분을 필터링하기 위해 청각장분석(ASA; Auditory Scene Analysis)를 사용하고 있다. 현재 구현되어 있는 청각장분석 시스템은 입력된 데이터를 처리하는 방식이다. 그러나 인간의 복합적인 소리인식 능력은 상황유추 능력에 강력하게 기반 하므로 Ellis는 예측기반 청각장분석시스템을 사용하여 복합적인 소리에 대응한 시스템을 구현하였다[14].

음향심리학(psycho-acoustics)은 음향신호, 청각의 물리적측면, 소리인식의 상관관계를 연구하는 분야이다. 음악심리학(music psychology)는 작곡가와 연주자가 음악을 연주하고 청취자가 음악적인 구조로 이를 변환하고 인식된 음악적인 구조가 어떤 영향을 미치는가에 대한 연구이다. 음악신호를 처리하고 이로부터 특징적인 값을 추출하고 응용분야에 적용하는 것을 음악신호처리(musical signal processing)라 한다.

현재 소리인식분야의 컴퓨팅 모델은 신호처리에 기반을 두고 있다. 특히 음악분야에서는 고립된 데이터를 다루고 있으며 제한된 그룹의 데이터들로 구성되어 있다. 서양음악 분석에서 사용되는 방법들을 이용하여 한국전통악기의 특징을 구분할 수 있는 컴퓨터모델을 구현하고자 하며 최적화된 방법을 찾고자 한다. 음악분류에 있어 불확실성을 제거해나가는 개념을 Bobick(1987)과 Minsky(1987)가 보여주고 있다[34]. 일련의 과정을 통해 가장 추상적인 단계인 소리의 원천으로부터 구체적인 단계로의 분류 과정을 보여준다.

음향심리학의 역사는 100여년이 넘는다. 19세기 초창기 연구는 소리의 크기와 음고에 대한 것이었다. 과학적인 연구이전에도 철학자들은 소리의 본질에 대해 고민하였는데 그 역사는 고대 그리스시대까지 거슬러 올라간



다. 피타고라스는 현의 길이가 정수배인 현들을 동시에 연주하면 좋은 소리가 들린다는 것을 알았다. 20세기 Stevens, Wegel, Fletcher 이후 현대 음향심리학이 성립되었다. 단순한 특성(음고 또는 음량), 시간-주파수관계에 따른 가림(masking)효과를 이용한 강인하고 효과적인 모델들이 구현되었다[35]. 최근의 연구는 청각장 분석 분야와 음향심리학의 새로운 분야로서 신호의 기저차원의 분석을 하는 주파수분포 패턴분석이 행해지고 있다.

음고이론(pitch theory)에서 음고는 단일음의 주파수와 관계된 지각적인 특성이다. 최근의 연구는 인간의 음고 결정과정을 모사한 시스템을 구현하는데 중점을 두고 있다. 위치음고 모델(place models of pitch)에서 음고는 주파수의 분포를 분석한 결과이다. 내이(cochlea)는 주파수분석기의 역할을 한다. 내이는 일련의 주파수성분의 고점을 중앙처리장치로 보내고 중앙처리장치는 고점의 위치관계를 해석하여 음고를 결정한다. 전통적인 모델로서 골드슈타인 최적처리(Goldstein optimum processor) 모델을 들 수 있다. 신호의 주파수성분의 고점이 추출되고 최대우도처리기(maximum likely-hood processor)를 이용하여 최적의 음고를 찾아낸다. 이 모델은 엄격한 수학적 바탕위에 구현되었다[36].

시간음고 모델(temporal models of pitch)은 내이 채널에서의 주기로 표현된다. 내이는 주파수성분 분석기로서 역할을 하는데 주파수성분의 고점을 추출하기보다는 시간영역에서 내이로 부터 대역통과된 신호를 측정한다. 최초의 시간음고 모델은 Licklider(1951)에 의해 제안되었다. 네트워크지연(network delay line)과 2차원 동시 감지(coincidence detector)에 기반한 분석 기법을 제안하였다. 각 채널의 상관관계를 계산하고 이렇게 취합된 정보는 주된 음고를 얻는데 사용한다[37]. 이 모델은 Slaney(1990), Meddis and Hewitt(1991)에 의해 다시 소개되었고 이후 음고분석의 상관관계법이라 불린다. 오늘날 많이 사용되는 기법으로서 Slaney는 변조스펙트로그램

(modulation spectrogram), 자동상관도(true auto-correlogram)등의 연구결과를 발표했다[38]. Tzanetakis[42], Kuo[6]는 음악인식에 있어서 푸리에 변환처리 후 음악의 질감을 수치화 할 수 있는 함수들을 이용하여 시스템을 구현하였다.

과거 음성관련연구가 활발하게 이루어졌고 음악과 관련된 분류 및 검색에 관한 연구는 최초 음성관련연구 결과들을 차용하여 이루어졌다. 이와 관련된 연구는 2000년 이후 본격적으로 이루어졌다. 지금까지 서양음악을 대상으로 연구가 이루어졌으며 동양음악, 한국전통음악을 대상으로 한 연구는 없었다.

본 논문에서는 한국전통음악의 분류라는 최종 목적을 세우고 한국음악에 널리 사용되는 리듬특성을 구분하기 위해 평균피크값을 측정할 수 있는 특성함수를 제안하였고 실효성을 확인하였다. 3장에서는 현재 음악인식분야에 사용되는 분류알고리즘을 성능을 비교하여 벡터거리법에 기반을 둔 분류알고리즘이 국악분류에 우수함을 입증하였다[44]. 특성함수중에서 롤-오프포인트 함수의 최적값을 찾아 실험에 적용하였다. 롤-오프는 저음프레임과 고음프레임의 상대적인 비를 나타낸다. 최적의 롤-오프 포인트를 찾기 위하여 롤-오프 포인트를 0.05에서 0.9까지 0.05간격으로 증가시키며 반복실험 하였다. 실험결과 롤-오프 포인트 0.85에서 가장 높은 분류성공률 85%를 나타냈다. 제4장에서 국악기분류를 실험하였다. 최적화된 실험환경에서 대금, 가야금, 바이올린을 94.44%의 성공률을 나타냈다[43]. 최종실험으로서 앞장에서의 실험결과를 종합하여 한국전통음악을 4개의 장르로 분류하여 94.5%의 성공률을 나타냈다[45]. 5장에서 본 논문의 실험결과들을 중심으로 결론을 맺었다.

## 제2장 음악인식 시스템

D. Perrot 과 R. O. Gjerdige이 행한 연구에서 보여주듯이 수초이내의 짧은 시간에도 음악장르 판단의 가능성을 보여주었다. 컴퓨터 처리능력의 발달로 푸리에변환, 주파수성분 함수처리를 빠르게 수행 할 수 있으므로 시스템구현은 충분히 가능하다. 어떤 미디어파일의 일부분을 잘라내고 이를 기반으로 푸리에변환을 행하여 주파수성분을 모두 구한 후 주파수성분의 분포관계 즉, 주파수분포평균, 가청 주파수 이내의 고주파와 저주파의 비, 통계적특성, 인접 주파수와의 상관관계를 함수화한 후 특정 데이터의 특징을 추출하여 시스템을 구현한다. 신뢰성 있는 자동인식, 분류시스템이 구현된다면 학술, 문화, 예술 등 다양한 분야에 적용가능하다.

### 2.1 음악인식 시스템의 개요

음악인식 시스템을 구성하는 전체적인 과정, 특성값 추출의 개념, 신호 처리의 배경 그리고 음악분석에 사용되는 특성들에 대하여 논의한다. 음악장르 분석의 핵심은 음악형식을 구분하는 특질을 찾아내는 것이다. 직접적인 비교가 불가능한 두 음악의 동질성과 이질성을 찾아내는 것이 문제의 핵심이다. 비교를 위해서 음악에서 추출된 데이터를 변환하고 본 논문에서 특성값 추출과정을 거친다.

특성값 추출의 목적은 입력데이터로부터 핵심적인 정보를 얻기 위함이다. 분류의 목적은 추출한 특성값을 처리하여 특정범주로 입력신호를 분류하는 것이다. 특성값 추출과 분류는 명확히 구분되는 과정으로 특성값 추출의 결과 데이터가 분류단계로 전달된다. 그림 2.1은 가장 단순한 형태

의 음악인식시스템을 나타낸다.

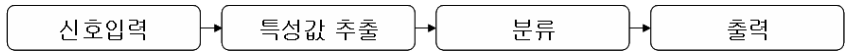


그림 2.1 음악인식 시스템의 과정.

Fig. 2.1 Schematic chart of music recognition system.

특성값 추출은 분류기법에 사용되는 전처리 단계로서 특성값은 음악과 일로부터 추출한 원데이터(raw data)를 변환함으로 서 얻어진다. 원데이터를 직접 사용하여 음악을 비교할 수 도 있으나 데이터셋의 차원이 매우 높고 중복되는 부분이 많기 때문에 흔히 사용되는 방법은 아니다.

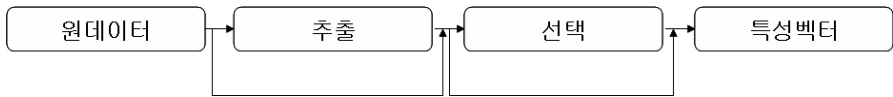


그림 2.2 입력된 데이터집합으로부터 특성값 벡터의 생성.

Fig. 2.2 Generating feature vector from input dataset.

## 2.2 음악신호 처리

특성값 벡터  $x$ 는 데이터처리에 사용되는  $d$ 차원의 한 개의 벡터이다:  $x=(x_1,...,x_n)$ . 각각의 각각의 특성값벡터는 특성공간의 한 점으로 볼 수 있다.  $n$ 개의 구성요소로 이루어진 패턴집합은 아래와 같이 표기한다.

$$X = \{x_1, \dots, x_n\} \quad (2.1)$$

일반적으로 패턴집합은  $n \times d$  차원의 행렬로 이루어져 있다. 아래 사항에 유의하여 특성값을 선택하여야 한다. 첫째, 유사한 객체는 특성공간상에서 인접하여 위치해야 한다. 판단경계에 의해 구분되는 영역간의 거리는 가능한 멀어야 한다[15]. 그러나 근접성이 필수불가결한 것은 아니다. 분류기에 따라 하나의 영역이 다른 영역과 충분히 구분될 정도이면 된다. 둘째, 특성값은 데이터에 담긴 중요한 정보들을 내재하고 있어야 한다. 데이터는 손실없이 처리되어야 하고 특성값을 통해 원신호를 회복할 수 있어야 한다. 변환처리에 사용되는 푸리에 변환이 좋은 예로서 가역적인 특성을 가지고 있다.

그림 2.3는 특성값 선택의 중요성을 보여준다. 2차원 공간의 점들로 구성된 데이터는 원점에서 비슷한 거리를 유지하고 있다. 그림 2.3(a)의 직교좌표를 선택한다면 분류의 판단경계를 찾기가 어려워진다. 반면 그림 2.3(b)의 극좌표를 사용한다면 특성공간에서의 판단영역 분리는 쉬워진다. 이는 단순한 처리도 결과의 선택에 큰 영향을 미칠 수 있다는 점을 보여준다. 선의 특성값을 찾는 것은 어려운 일이며 무수한 반복이 필요한 일이다. 일반적으로 계산의 효율성과 성공적인 수행을 위해 입력데이터의 차원을 낮추어야 한다.

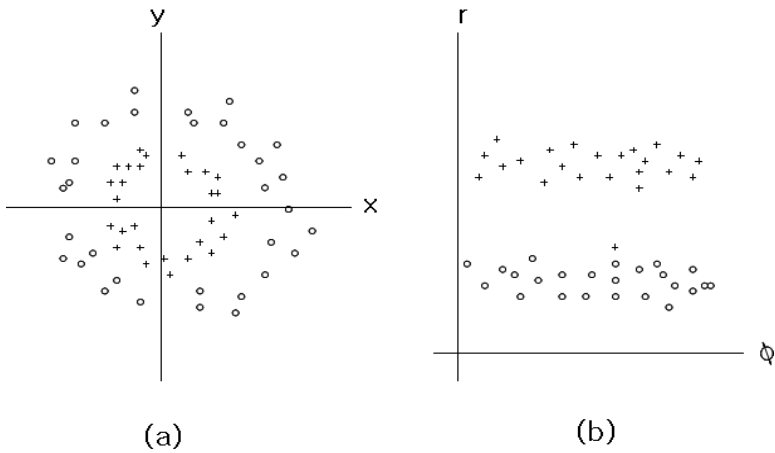


그림 2.3 (a) 직교좌표 (b) 극좌표  
**Fig. 2.3** (a) Cartesian coordinates (b) Polar coordinates.

### 2.2.1 음악신호의 푸리에 변환

시간영역(시간 경과에 따른 크기의 변화)의 소리신호 파형은 소리분석에 있어 큰 효용이 없다. 신호를 구성하는 주파수성분의 진동수, 위상, 크기를 통해 소리의 특성을 파악할 수 있다. 인간의 청각시스템의 처리방식과 유사한 방법으로 주파수성분을 분석함으로써 음악신호에 담긴 정보를 파악할 수 있다. 그림 2.4의 두파형은 주파수가 다른 정현파이고 가장 아래는 두 정현파의 합을 나타낸다.

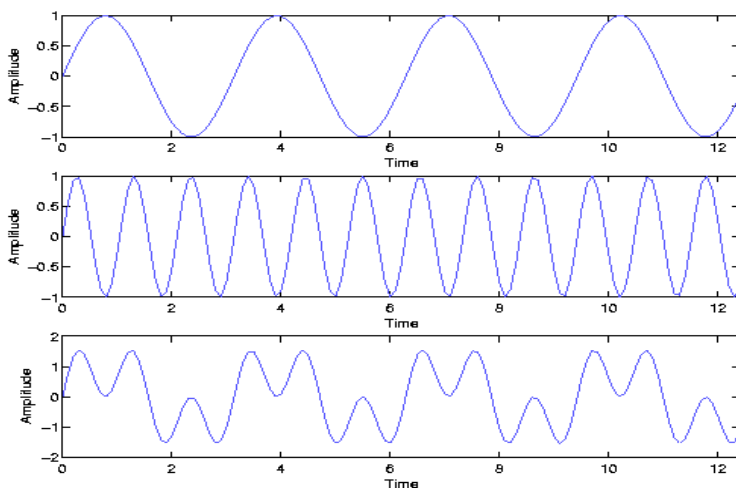


그림 2.4 정현파의 합성.  
**Fig. 2.4** Superposition of sine waves.

푸리에 합성의 역과정으로 푸리에 분석은 신호를 단순한 정현파로 나누는 것을 말한다. 푸리에 분석을 행함으로 서 어떤 신호의 정현파성분을 볼 수 있고 어떤 성분의 영향이 가장 큰지 파악할 수 있다. 푸리에 변환 수식은 아래와 같다.

$$X(f) = \int_{-\infty}^{\infty} x(t)e^{-j2\pi ft} dt \quad (2.2)$$

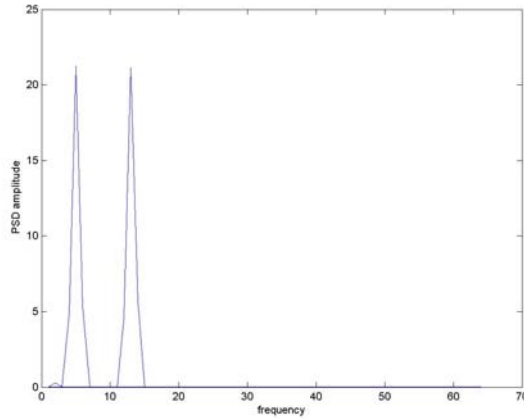
역변환은 아래와 같다.

$$x(t) = \int_{-\infty}^{\infty} X(f)e^{j2\pi ft} df \quad (2.3)$$

$x(t)$ 는 시간의 함수이고  $X(f)$ 는 주파수함수이다. 그리고,

$$e^{j\phi} = \cos \phi + j \sin \phi \quad (2.4)$$

푸리에 변환을 행함으로 서 시간영역의 신호를 주파수영역의 신호로 변환한다. 그림 2.4는 시간영역의 그래프인 그림 2.3의 주파수영역 그래프이다.



**그림 2.5** 그림 2.4 신호의 푸리에 분석.  
**Fig. 2.5** Fourier Analysis of Fig. 2.4 signal.

푸리에 변환은 이론적으로 무한길이의 신호에 적용할 수 있다. 전체 신호의 주파수성분을 보여준다. 그러나 비주기신호의 경우 다른 처리방법이 필요하다.

STFT(Short Time Fourier Transform)는 입력된 신호를 짧은 구간(frame)으로 나눈 뒤 각각에 대하여 푸리에 변환을 행한다. 그리고 시간에 따른 신호의 변화도 분석된다. STFT도 문제점을 안고 있는데 푸리에 변환은 입



력신호가 무한해야 한다. 각각의 짧은 구간은 무한길이의 신호로 확장되어야 한다. 구간경계에서의 갑작스런 신호의 중단은 원신호에 존재하지 않는 주파수를 발생시킬 수 있다.

이러한 문제를 처리하기 위해 창함수(window function)를 도입하여 구간을 구한다. 대표적인 창함수를 보면, 구형 창함수, Bartlett, Hanning, Hamming, Blackman 창함수 등이 있다[16,17]. 각각의 창함수는 식 (2.5)에서부터 식 (2.9)에 기술하였다. 여기서 대부분이 코사인열을 사용하여 최적화 되었음을 알 수 있다.

구형 창함수는

$$w(n) = \begin{cases} 1, & 0 \leq n \leq M \\ 0, & \text{others} \end{cases} \quad (2.5)$$

이고, Bartlett 창함수는

$$w(n) = \begin{cases} 2n/M, & 0 \leq n \leq M/2 \\ 2 - 2n/M, & M/2 < n \leq M \\ 0, & \text{others} \end{cases} \quad (2.6)$$

이며, Hanning 창함수는

$$w(n) = \begin{cases} 0.5 - 0.5\cos(2\pi n/M), & 0 \leq n \leq M \\ 0, & \text{others} \end{cases} \quad (2.7)$$

이다. 또한, Hamming 창함수는

$$w(n) = \begin{cases} 0.5 - 0.46\cos(2\pi n/M), & 0 \leq n \leq M \\ 0, & \text{others} \end{cases} \quad (2.8)$$

이고, Blackman 창함수는

$$w(n) = \begin{cases} 0.42 - 0.5\cos(2\pi n/M) + 0.08\cos(4\pi n/M), & 0 \leq n \leq M \\ 0, & \text{others} \end{cases} \quad (2.9)$$

으로 각각 정의된다. 창함수의 적용은 구간의 양쪽 끝에서 정보의 손실을 가져온다. 따라서 창함수를 중복(overlap)하여 적용한다. STFT는 아래와 같이 계산한다.

$$X(f, t') = \int_{-\infty}^{\infty} h(t' - t)x(t)e^{-j2\pi ft} dt \quad (2.10)$$

지금까지 논의한 푸리에 변환법은 연속적인 신호에 적용한다. 디지털 영역에서의 변환은 이산푸리에변환(DFT; Discrete Fourier Transform)을 사용한다. 이산함수  $f(t) \rightarrow f(t_k)$ 을 가정하면  $f_k = k\Delta, k = 0, \dots, N-1$ 이다. 주파수 간격(step)은

$$v_n = \frac{n}{N\Delta} \quad (2.11)$$

표준적인 DFT 수식은 아래와 같다.

$$F_n = \sum_{k=0}^{N-1} f_k e^{-j2\pi nk/N} \quad (2.12)$$

역변환 수식은 아래와 같다.

$$f_k = \frac{1}{N} \sum_{n=0}^{N-1} F_n e^{j2\pi nk/N} \quad (2.13)$$

다양한 알고리즘을 사용하는 DFT 계산법이 있다. 그중 가장 효율적이고 많이 사용되는 방법은 FFT(Fast Fourier Transform)이다.

### 2.2.2 물리적 특성과 인지적 특성

일반적으로 소리신호를 구분하는데 사용하는 두가지 특성은 물리적 특성과 인지적 특성이다. 물리적 특성은 소리신호의 수학적, 통계적 분석에 기반한다. 주파수, 에너지, 영교차율(zero-crossing)등을 물리적 특성값의 예로 들 수 있다. 인지적 특성은 인간이 소리를 인식하는 방법인 음고(pitch), 음색(timbre), 리듬에 기반한다.

인간의 소리인식은 물리적인 소리신호를 인식하는 것이기 때문에 물리적 특성과 인지적 특성은 어느 정도 연관성을 가지고 있다. 이중 명백한 연관성을 나타내는 것의 예를 들면 물리적인 신호의 크기(amplitude)는 인식되는 소리의 크기(loudness)에 대응된다. 물리적인 신호의 주파수는 인식되는 소리의 음고와 직결된다. 그러나 좀 더 자세히 살펴보면 이는 피상

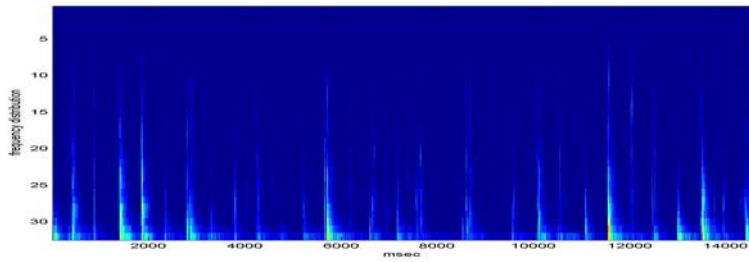
적인 관찰일 뿐이다. 예를 들면 인식상의 주관적인 소리의 크기는 신호의 주파수성분의 분포와 관련이 있다. 일반적으로 인지적인 특성들은 물리적인 특성의 복합적인 작용에 의해 결정되기 때문에 수학적으로 표현하는데 어려움이 있다. 현재로서는 두 가지 특성을 엄격히 구분하여 사용하는 것은 실익이 없다.

인지적인 특성을 구하기 위해 물리적인 특성들을 이용한다. 예를 들면 인지적 특성인 음고를 구하기 위해 음고의 인식에 중요하다고 판단되는 물리적 특성들을 동원하여 인지적 특성을 구현한다. 진정한 인지시스템을 구현하고자 한다면 특성값 추출에서부터 인지적모델을 사용하여야 할 것이다. 즉, 인간의 청각시스템이 소리를 인식하는 방식으로 시스템을 구현해야 할 것이다. 이러한 접근방법은 소리신호 분석에 최적의 모델이 될 것이나 현재 인간의 인지과정에 대한 이해는 이러한 시스템을 구현하기에는 아직 부족하다.

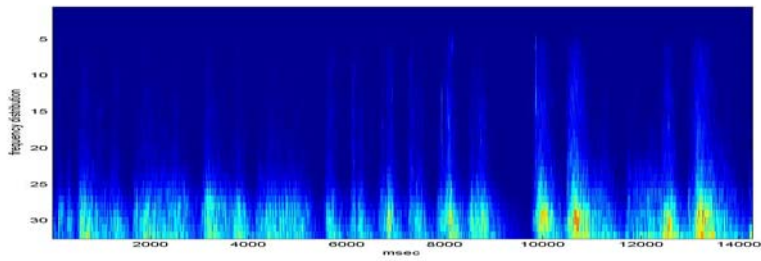
## 2.3 음악장르 구분에 사용되는 특성

푸리에 변환을 통해 주파수분포를 구하듯이 인간과 척추동물의 청각시스템은 유사한 처리를 거쳐 소리를 인식한다. 이는 소리의 주요정보들이 주파수를 통해 전달된다는 것을 의미한다. 음악장르 인식에 있어 주파수분포는 여러가지 특성들을 결정짓는 중요한 요소이다.

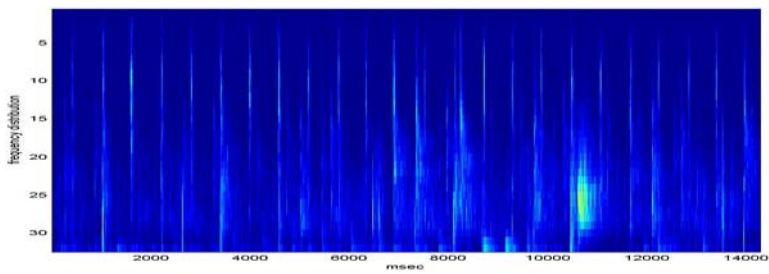
스펙트로그램(spectrogram)은 FFT 결과값을 시각적인 영역으로 변환하여 작업하는 직관적인 방법이다. 스펙트로그램은 FFT 결과값을 시간-주파수 영역으로 시각화한다. x축은 시간이고 y축은 주파수축이다. 한 점 (t, f)의 색(color)는 신호의 세기를 나타낸다. 그레이 스케일일 경우 흰색보다 검은색이 강한 신호임을 표지한다. 그림 2.6는 음악 장르별 스펙트로그램의 예이다. x축은 시간이며 y축은 주파수(위에서 아래로 증가)를 나타낸다. 그래프의 밝기는 강도를 나타낸다.



(a) 국악



(b) 클래식



(c) 재즈

그림 2.6 세가지 장르의 스펙트로그램.

**Fig. 2.6** Spectrograms for music samples from three genres.

그림 2.6은 각각의 장르음악의 10초 길이 표본을 시각화한 것이다. 모든 표본은 44100Hz로 샘플링되었으며 선택된 음악들은 해당 장르를 대표할 만한 표준적인 연주를 선택했다. (a)국악은 서울새울 가야금 삼중주단의 연주중에서 경복궁타령의 일부이다. (b)클래식은 아바도가 지휘한 말러 교향곡 2번중 5악장의 일부이다. (c)재즈는 자끄루시에르의 피아노트리오 연주이다. 가야금과 장고의 연주인 국악의 스펙트로그램은 불규칙한 리듬을 보여준다. 교향악단의 연주인 클래식은 리듬패턴을 관측하기 어려우며 에너지가 일정 주파수대역에 집중되어 있는 것을 볼 수 있다. 피아노, 베이스, 드럼으로 구성된 재즈는 규칙적인 리듬패턴을 볼 수 있으며 전주파수대역에 걸쳐 에너지가 분포하고 있다.

스펙트로그램은 음악장르인식을 이미지 분석 측면에서 접근한 것이다. 스탠포드대학의 연구자들은 시간영역의 신호(wave form)를 시각영역(스펙트로그램)으로 변환하여 시스템을 구현하였다. 이미지의 에지탐색(edge detection) 알고리즘을 적용하여 록, 클래식, 재즈를 75% 성공률로 분류하였다[15].

### 2.3.1 멜-주파수 쉐스트럴 계수(Mel-Frequency Cepstral Coefficients)

멜-주파수 쉐스트럴 계수(MFCCs)는 음성인식 시스템에 주로 사용되어 왔으며 주파수 분포 중에서 인간의 청각시스템에 민감한 부분을 포착한다. 소리신호 분류 시스템에 적용되어 왔다[18]. 인간 청각시스템의 소리 영역 주파수 인식은 선형적이지 않고 로그분포를 따른다. MFCCs는 청각의 주파수인식시스템을 모델로 하여 주파수를 인식 모델을 구현하였다.

처리절차는 아래와 같다[19].

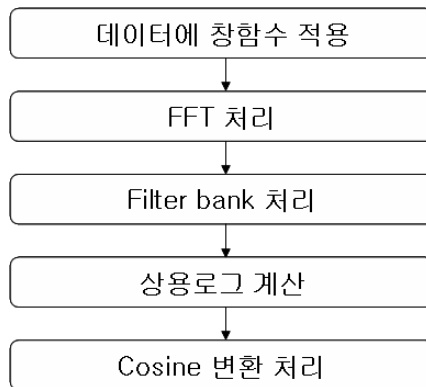


그림 2.7 MFCCs 처리순서

**Fig. 2.7** MFCCs processing sequence

필터집합(filter bank)는 MFCCs의 핵심으로 서 균등하게 배열된 13개의 필터와 로그간격으로 배열된 27개의 필터로 구성되어 있다. 이는 인간의 청각모델을 기반으로 하여 구성된 것이다. 균등배열된 필터들의 중심주파수 사이 간격은 133.33Hz이고 로그간격으로 배열된 필터는 주파수상에서 1.071 팩터(factor)로 분리되어 있다. 마지막 단계인 코사인변환은 출력의 복잡도(차원)을 경감시키기 위하여 시행한다. 부가적으로 신호의 세기를 계산하여 차원  $d=13$ 인 특성값 벡터( $13 \times 2911$ )를 출력한다. MFCCs 그래프는 스펙트로그램 처럼 직관적으로 이해하기 쉽지 않지만 컴퓨터를 이용한 소리의 분석에 있어 유용한 도구이다. 그림 2.8은 MFCC 계산시 데이터를 100Hz로 표본화하였다. x축에 2900개의 표본과 y축에 13개의 MFCC 특성값(위부터  $C_0$ - $C_{12}$ )이 출력되었다.



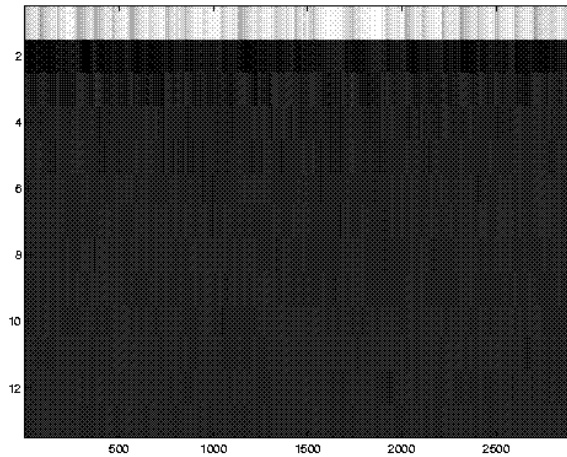


그림 2.8 국악(그림 2.6(a))의 MFCCs 플롯.  
**Fig. 2.8** Plot of MFCCs for the sample from Fig 2.6(a).

### 2.3.2 영교차율(average zero-crossing rate)

영교차율은 이산신호가 서로 다른 값을 갖을 때 발생한다. 따라서 영교차율의 빈도는 신호의 기본주파수(fundamental frequency)와 밀접한 관련이 있다. 정현파와 같은 단순한 신호의 경우 영교차율값은 주파수를 나타낸다. 정현파는 한주기당 영점축을 두 번 지나가므로 영교차율값을 2로 나누면 주파수가 된다. 그러나 모든 경우에 적용되지는 않는다. 그림 2.9(b)에서 보듯이 부분적으로 기본주파수보다 높은 주파수로 진동하는 신호의 경우엔 한주기당 영점축을 관통하는 횟수가 많아진다[18].

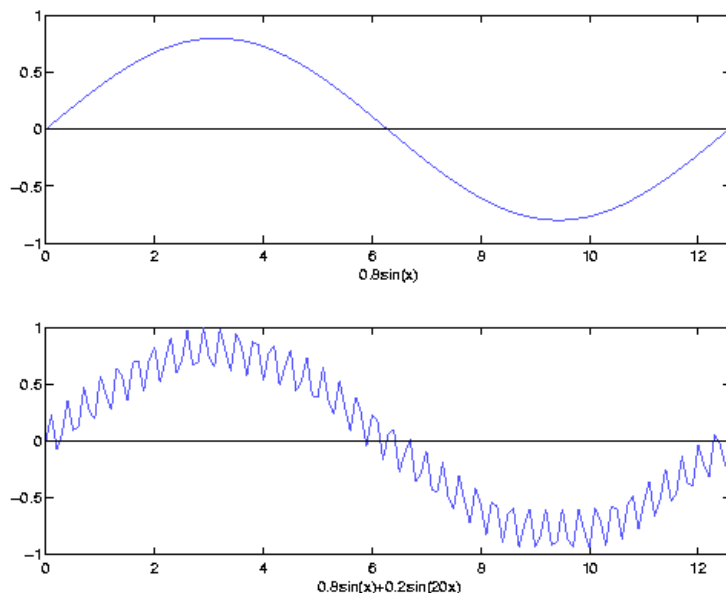


그림 2.9 영교차율을 이용한 기본주파수의 계산.  
**Fig. 2.9** Calculation of fundamental frequency using ZCR.

그림 2.9는 영교차율을 이용하여 기본주파수를 얻을 수 없는 경우의 예로서 (a)신호는 영교차율/2가 기본주파수가 되지만 (b)신호는 한주기당 여러 차례 x축을 관통하므로 영교차율을 이용하여 기본주파수를 구할 수 없다.

기본주파수를 찾는 일은 악보작성시스템과 멜로디인식시스템에 매우 중요하다. 그러나 음악장르 시스템에서 기본주파수는 큰 중요성을 갖지 않는다. 여러 가지 악기와 목소리, 효과음으로 구성된 음악에서 기본주파수를 찾는 일은 매우 어렵다. 실제 찾는다 해도 큰 의미를 부여하기는 어렵다. 만일 기본주파수를 찾아서 멜로디라인을 구한다 해도 음악장르 구분에는 도움을 주지 못한다. 동일한 멜로디를 연주하는 록그룹과 재즈밴드

를 연상하면 쉽게 이해할 수 있다.

기본주파수를 이용한 특성값을 사용하여 비디오 조각 및 분류 시스템을 구현한 연구가 보고되었다[19]. 이 분야는 복잡한 비음성 소리신호를 다룬다는 점에서 음악인식 시스템과 유사하다. 영교차율은 시간영역의 특성값으로 처리시간이 빠르고 시간에 따른 영교차율값의 변화를 분석하고 통계적 특성을 이용하여 음성(voice) 신호와 비음성 신호를 구분하는데 주로 사용되었다. 음성신호는 비음성신호보다 작은 영교차율값을 갖는다. T. Zhang과 C. Kuo는 영교차율특성을 사용하여 신호의 빈도, 주기성, 크기에 기반한 환경소음분류 시스템을 구현하였다[6]. 영교차율 특성은 아래의 수식으로 계산한다.

$$Z_n = \sum_m |sgn[x(m)] - sgn[x(m-1)]|w(n-m) \quad (2.14)$$

여기서

$$sgn[x(n)] = \begin{cases} 1, & x(n) \geq 0 \\ -1, & x(n) < 0 \end{cases} \quad (2.15)$$

그리고

$$w(n) = \begin{cases} 1/2, & 0 \leq n \leq N-1 \\ 0, & otherwise \end{cases} \quad (2.16)$$

m은 창함수의 윈도우크기이다. 음악은 일정시간 동안 안정적인 경향이 있기 때문에 음성신호보다 작은 영교차율 특성 분산을 갖는다. 발자국 소리, 새의 지저귐 소리도 음성신호와 유사한 특성을 갖으며 같은 악기중

에서도 피아노와 같은 단속적인 소리는 현악기보다 음성과 유사한 특성을 보인다.

### 2.3.3 단시간 에너지(short-time energy)

소리신호의 단시간 에너지는 아래와 같이 정의 된다.

$$E_n = \frac{1}{N} \sum_m [x(m)w(n-m)]^2 \quad (2.17)$$

여기서  $x(m)$ 은 이산시간 소리신호이며  $n$ 은 단시간 에너지의 시간지표(index)이다.  $w(m)$ 은 구형(rectangle) 창함수이다.

$$w(n) = \begin{cases} 1/2, & 0 \leq n \leq N-1 \\ 0, & otherwise \end{cases} \quad (2.18)$$

단시간 에너지 함수는 시간에 따른 세기(amplitude)의 변화를 보여준다. 에너지값은 100개의 표본마다 계산되어 지며 무음(silence)를 측정하기 위해 사용한다. 음성인식분야에서는 음성신호와 비음성 신호를 구분하기 위하여 사용한다. 비음성신호의 경우  $E_n$  값은 매우 작다[6].

### 2.3.4 리듬과 비트(rhythm and beat)

인간에게는 흥겨운 음악에 맞춰 박수를 치는 일이 어렵지 않다. 그러나 이러한 시스템을 구현하는 것은 매우 어렵다. 비트를 추적하고 리듬을 파악하기 위한 연구가 활발히 진행 중이다. 비트추적에 대한 개요는 Dixon[20]을 참조한다. 실시간으로 음악의 비트를 추적하는 시스템이 연구되고 있다[21].

비트추적을 어렵게 만드는 몇가지 장애물이 있다. 어떤 음악의 템포(tempo)나 구조에 대한 사전정보 없이도 인간은 쉽게 비트를 파악할 수 있다. 현재 구현된 상당수의 비트 추적 시스템은 이러한 사전정보에 의존하고 있다. 음악의 템포가 변하면 인간은 수초내에 이를 파악할 수 있으나 현재 까지 구현된 시스템은 이에 대처하기가 매우 어렵다.

다행스럽게도 정확한 비트검출은 음악인식에 있어서 큰 중요성을 갖지 못한다. 음악인식에서 관심을 갖는 것은 템포와 비트의 세기와 빈도이다. 템포는 일정한 소리가 일정한 시간간격으로 반복되는 것을 말한다. 여기서 시간간격이란 250ms~2s 범위를 갖는다[22]. 비트의 세기는 비트의 강도로서 신호에서 주기적으로 반복되는 강한 신호를 말한다. 신호의 평균세기와 비교하여 상대적인 비트의 세기를 구한다.

### 2.3.5 음색(timbre)

음색은 "다른 종류의 음원으로부터 발생하는 동일한 레벨의 소리를 구별하게 해주는 특성"으로 정의 한다. 이는 주파수의 분포, 음압, 순시특성과 밀접한 관계가 있다. 특별히 음악에서 음색은 악기의 수, 악기의 발음

강도에 의해 결정된다. 음색은 주관적인 특성으로서 아직까지 만족스럽게 음색특성을 이용한 시스템은 구현되지 못하였다. 스펙트로그램은 음색정보를 파악하기 위해 사용하는 일반적인 도구이다. 더욱 발전되고 지각적인 음색모델의 구현은 소리인식분야에 큰 발전을 가져올 것이다.

음악인식에 필요한 핵심적인 특성들에 대하여 연구하였다. 주파수영역에서의 음악신호는 시간영역에서 파악할 수 없었던 많은 정보들을 보여준다. 소리신호 주파수영역의 표현부터 시작하여 스펙트로그램을 이용하여 대표적인 장르의 주파수분포를 살펴보았다. 주파수영역에서의 장르별 특징은 어떤 특성값들을 사용하여 음악인식에 접근하여야 하는지 분명히 보여주었다. 국악의 경우 불규칙한 리듬과 비트의 세기가 일정하지 않았다. 클래식의 경우 일정 주파수영역에 에너지가 집중되어 있었고 리듬패턴이 일정치 않았다. 재즈의 경우 전대역에 걸쳐 고른 주파수분포를 보였고 리듬패턴이 매우 일정하였다.

특정주파수대에 민감한 청각시스템을 모델링한 MFCCs, 음성과 비음성 신호 구분에 유용한 영교차율, 주파수 에너지분포를 계산하는 단시간 에너지등의 특성값들은 스펙트로그램이 보여준 장르별 특징을 파악할 수 있는 함수들로서 음악인식시스템에 유용하게 적용될 수 있다.

## 2.4 청각 시스템

인간의 내이(inner ear)는 정교한 기관으로 작은 뼈, 근육, 달팽이관, 반고리관으로 이루어져 있다. 청각기관의 임무는 주위의 소리를 처리하여 그 결과를 뇌에 전달한다. 모든 감각기관은 주위의 신호를 감지하여 미세한 전기신호로 변환하고 다시 전기신호의 변화를 일종의 디지털 코드로 변환하여 뇌로 전달한다. 인간의 청각기관은 신경신호(neural code)로 변환하기 이전에 전처리 과정을 수행한다. 내이는 주위 소리가 담고 있는 개별 주파수들의 에너지의 세기를 결정한다. 달팽이관은 여러개의 채널로 구성되어 있으며 각각의 채널은 특정 밴드의 주파수에 민감하게 반응한다. 이러한 신호의 맵핑(mapping)은 입력된 정보를 코딩하는데 사용하는 일종의 알고리즘이다. 주위소리의 주파수분석은 내이에서 이루어진다.

인간의 청각모델(auditory model)의 구현은 인식기반 시스템을 구현하는데 목적이 있다. 이 중 음악인식에 유용한 모델을 선택하여 시스템에 적용한다. 달팽이관이 변환한 소리신호는 물리적인 소리의 패턴과는 다른 모습을 보인다. 예를 들면, 선형적인 주파수분포도인 스펙트로그램과 실제의 뉴로그램(neurogram)은 매우 다른 모습을 형태를 하고 있다. 최근 들어 신경의학의 연구에 의해 기저막, 내유모세포(inner haircell)과 청신경(auditory nerve)에 대한 연구가 활발히 진행되고 있다.

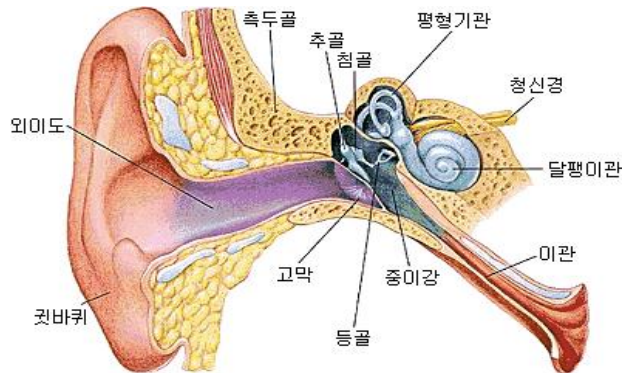


그림 2.10 청각기관의 구조.

Fig. 2.10 Structure of auditory system.

소리가 전달되는 매질이 달라지면 에너지중 일부는 매질의 경계에서 반사된다. 이러한 반사를 줄이고 공기로부터 림프액로 충전된 내이의 에너지 전달시에 손실을 줄이기 위해 포유동물의 외이는 소리를 모으고 중이는 이를 기계적으로 증폭하는 역할을 한다. 외이는 귀의 외이도 방향으로 소리를 반사하고 중이는 음파가 고막(ear drum)과 수직이 되도록 정렬시킨다. 주파수의 세기가 다르므로 주파수성분들의 반사각도는 서로 다르다. 외이의 모양에 따른 소리의 변이는 일종의 외곡으로 뇌는 이에 적응한다. 익숙한 소리의 주파수 분포는 음원의 방향을 인지하는데 사용한다. 방향에 따른 미세한 음파의 변이는 방향을 탐지하는데 사용된다. 중이의 뼈는 고막으로부터 전달된 소리를 난원창(oval window)를 통해 내이로 전달한다. 고막은 난원창보다 단면적이 큰데 이러한 면적의 감소는 소리의 압력을 증가 시킨다. 중이는 달팽이관 입구에서의 소리압력을 증대시키는 기계적인 증폭기의 역할을 한다. 이러한 모든 과정은 내이의 림프액에 전달되는 소리의 에너지를 최대화하기 위한 과정이다. 주위 환경과의 압력 평형은 귀와 코를 잇는 이관(eustachian tube)를 통해 이루어진다.



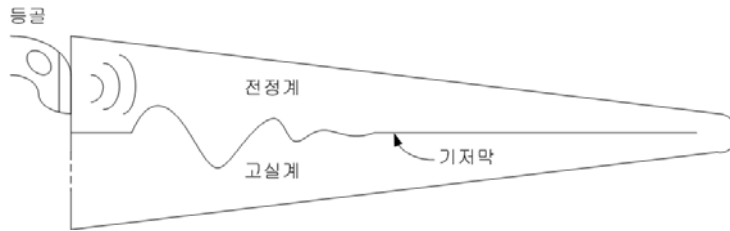


그림 2.11 달팽이관의 구조.

Fig. 2.11 Structure of cochlea.

내이는 평형감각부와 청각부로 나누어져 있다. 두개골의 중심부에 위치하고 있으며 딱딱한 뼈로 둘러싸여 있다. 내이의 청각부는 개별 주파수의 에너지 세기를 감지하여 뇌에 전달한다. 평형감각을 담당하는 반고리관과 청각을 담당하는 달팽이관으로 구성된 내이는 약 3cm 관으로 그 속에 림프가 충만해 있고 난원창을 통해 등골과 연결되어 있다. 달팽이관의 속은 막에 의하여 전정계, 달팽이세관, 고실계로 구분되어 있다. 전정계와 고실계는 달팽이관의 한쪽 끝에서 서로 통하고 있다. 달팽이세관 내에서 경계가 되는 나선형의 막을 기저막이라 한다. 달팽이관속에는 외림프가, 달팽이세관속에는 내림프가 가득 차있다. 이 두 림프의 조성은 서로 다르므로 인해 청각세포에 화학적인 에너지를 준다. 기저막에는 코르티기관이라고 하는 돌출된 구조가 있는데 여기에 청각을 담당하는 청세포가 자리 잡고 있다. 청세포에는 섬모가 표면에 존재하는데 이 섬모(내유모세포, inner hair cell)가 기저막의 진동에 따라 움직인다. 이 섬모의 움직임이 청세포를 자극하여 전기적 자극을 일으키고, 이 흥분이 청신경에 의해 대뇌 피질에 전달되어 소리를 감각하게 된다. 이 내유모세포를 통해 대뇌 피질과 정보를 주고 받는다. 기저막은 주파수분석을 수행하는 일련의 대역통과 필터의 집합으로 볼 수 있다. 섬모는 일종의 고유공진 주파수를 가지고 있어서 입력되는 대역통과된 소리의 특정 주파수에 공진된다.

## 2.5 청각응답 모델링

청각모델의 핵심은 달팽이관의 주파수분해에 있다. 음성인식분야 뿐만 아니라 음악인식분야에서도 주파수분해에 기초한 모델을 사용한다. 달팽이관내의 기저막은 대역통과 필터로 모델링할 수 있다(von Békésy, 1960)[23]. 본 논문에서 사용한 대역통과 필터모델은 Patterson이 제안한 것이다. 이는 신경외과, 심리학적으로 모델링 된 것이다(Patterson & Holdsworth, 1990; Patterson & Moor, 1986)[24,25]. Slaney가 이 모델을 도입한 이후 음악인식분야에선 대부분 Patterson모델을 기반으로 연구가 진행되고 있다[38]. 본 논문에서의 음악인식과정은 푸리에 변환과정을 거침으로서 달팽이관의 주파수분해모델을 모사한다.

음악의 인식은 인간의 판단결과이지 물리적신호자체로서의 음악신호는 전체의 일부일 뿐이다. 따라서 인간의 인식과정에 초점을 맞춘 시스템은 실제 우리가 느끼고 판단하는 음악에 대해 많은걸 설명해줄 수 있다. 그림 2.12은 청각처리의 핵심부분인 내이의 처리과정이다.

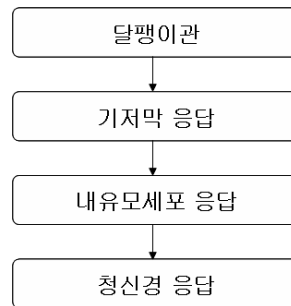


그림 2.12 내이의 신호처리 과정.

Fig. 2.12 Signal process of inner ear.

귀에 입력된 소리압력(sound pressure)은 외이도를 거쳐 고막으로 전달된다. 고막의 진동은 난원창을 지나 내이에 전달된다. 림프로 가득찬 달팽이관의 기저막에서 주파수성분분해가 발생하고 대역으로 분해된 주파수는 고유공진 주파수를 갖는 섬모에 모듈레이션 된다. 이렇게 섬모에 주어진 자극은 내유모세포의 준위(potential)을 변화시켜 변환된 신호는 청신경으로 전달되어 대뇌피질에 도달한다.

### 2.5.1 단일 채널 기저막 모델

달팽이관 내의 기저막은 다수의 채널로 모델링되는데 이해를 위해 단일 채널모델부터 기술한다. 그림 2.13는 기저막에 가해진 임펄스 신호를 나타낸다.

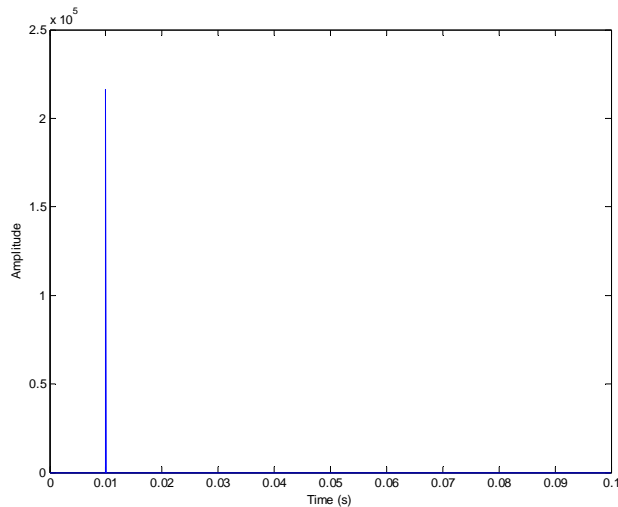
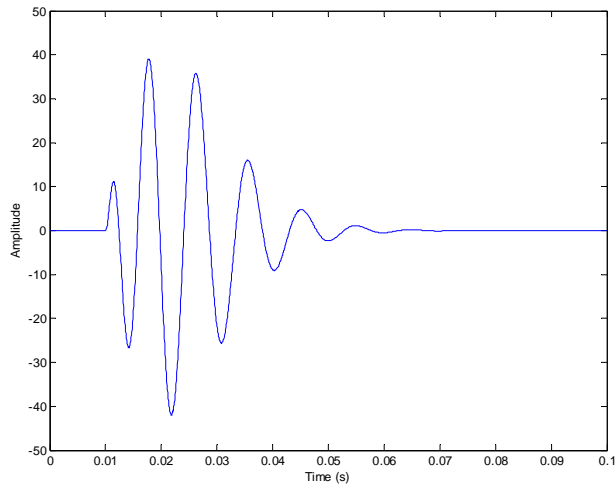


그림 2.13 기저막에 입력된 임펄스 신호.

**Fig. 2.13** Impulse signal applied to Basila membrane.

그림 2.14는 기저막의 응답으로서 임펄스응답에 대하여 양과자극이 발생하는 모습을 예시하였다. 이는 단일채널응답을 그래프화 한 것으로 인간의 청각은 멀티채널로 구성되어 있으므로 각각의 채널에서 그림 2.14와 같은 응답이 주파수대역별로 이루어진다.



**그림 2.14** 기저막의 단일 채널 임펄스 응답.

**Fig. 2.14** Single channel impulse response of Basilar memabrane.

기저막에서 행한 주파수 분석은 섬모에 전달되어 전기적인 준위로 변환된다. 한 개의 내유모세포는 청신경의 여러개의 돌기와 연결되어 있다. 인간의 경우 약 열 개정도의 신경돌기가 한 개의 내유모세포에 연결되어 있다. 각각의 신경돌기는 임계값을 가지고 있어서 이 값을 넘는 신호는 청신경신호로 변환(coding)된다. 청신경 모델에서는 자극확률을 입력신호로 받는다. 그리고 난수발생기를 이용하여 일련의 자극신호로 변환한다.

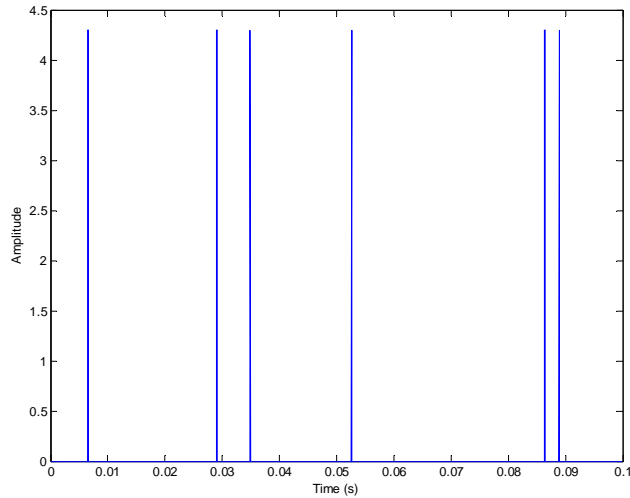


그림 2.15 청신경에 전달되는 일련의 자극열.  
**Fig. 2.15** Spike train transferred to auditory nerve.

## 2.5.2 다중채널 응답 모델

기저막은 내유모세포를 포함한다. 내유모세포는 기저막의 기계적 진동을 전기적신호로 변환하는 신호변환기(transducer)의 역할을 한다. 내유모세포에는 작은 섬모가 달려 있어서 기저막의 진동에 따라 움직인다. 섬모의 구부러짐 정도에 따라 내유모세포는 전기적자극을 발생 한다. 섬모는 한쪽방향으로 만 구부러짐으로 서 입력신호에 대한 일종의 반과정류를 행 한다. 길이가 짧은 섬모는 높은 주파수에 반응하고 길이가 긴 섬모는 낮은 주파수에 반응하여 주파수분해를 수행한다. 그리고 낮은 주파수에서 특정 위상의 신호에 만 반응하는데 이를 위상고정(phase locking)이라 한

다. 위상고정현상은 1.5Khz부터 사라지기 시작하고 5Khz가 넘으면 완전히 사라진다.

본 논문에서 사용하는 달팽이관 필터 모델은 1991년 Hewitt & Meddis[26]에 의해 개발되었으며 포유동물 내유모세포의 비선형적인 특성을 충실히 반영한 모델이다. 단순하면서도 필터사이의 상대적인 에너지 레벨을 잘 보존한다. 그림 16, 17에서 임펄스 응답은 주파수 대역별로 응답감과 응답시간의 차이를 보인다. 높은 주파수일수록 응답감이 좋고 응답시간이 빠른 것을 확인할 수 있다. 청각의 이러한 비선형성은 로그 스케일에 준한다고 알려져 있다. 달팽이관의 응답은 주파수대역별로 응답감과 응답시간이 비선형적이지만 본 논문에서는 이러한 비선형성은 고려하지 않았다. 푸리에변환 결과를 그대로 특성함수에 적용하여 데이터를 구하였다. [25]. 주파수대역별로 가중치를 주어 이러한 비선형성을 적용한 결과 실험결과가 고려하지 않을 때 보다 좋지 않았다. 청각시스템에 대한 의학적인 응답과정이 더욱 자세히 밝혀진다면 더욱 발전된 시스템을 구현할 수 있을 것이다. 그림 2.16은 기저막의 멀티채널 임펄스 응답으로 임펄스 신호를 달팽이관 모델에 입력했을 때 응답이다. 75개의 채널로 모의실험 했다. 아래쪽 채널 일수록 대역폭이 넓고 응답이 느리며 감도도 대역별로 달라짐을 확인할 수 있다. 그림 2.17은 반파정류된 내유모세포 임펄스응답으로 내유모세포가 한쪽방향으로 만 반응함으로서 반파정류된 못바를 확인 할 수 있다. 그림 2.18은 청신경에 전달되는 일련의 자극열을 이미지화 한것이다. 채널수는 75개이며 각각의 피크는 일종의 전기적신호로서 청신경에 난수적으로 맵핑된다.

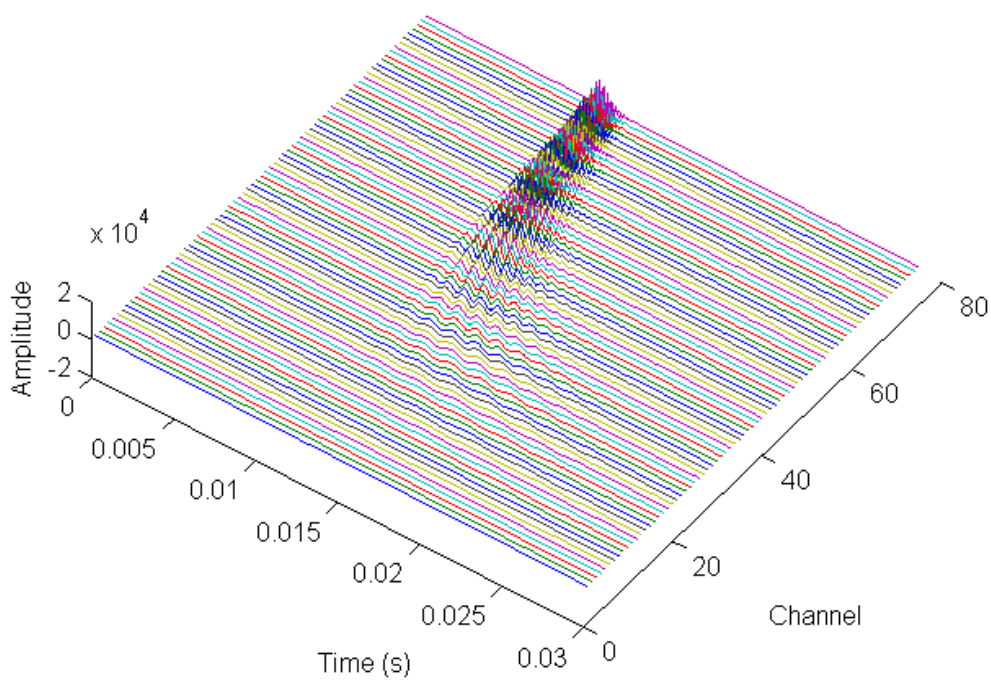


그림 2.16 기저막의 멀티채널 임펄스 응답.

**Fig. 2.16** Impulse response of multi channel Basilar membrane using impulse signal.

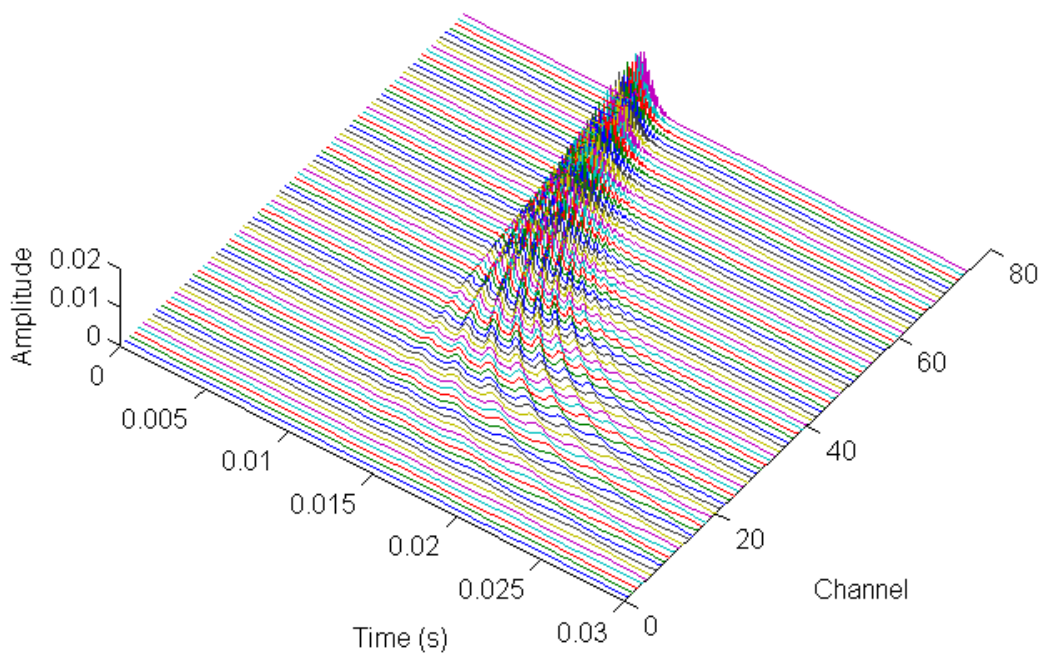


그림 2.17 반파정류된 내유모세포 임펄스응답.

**Fig. 2.17** Impulse response of Inner hair cell shows half-wave rectified signal.



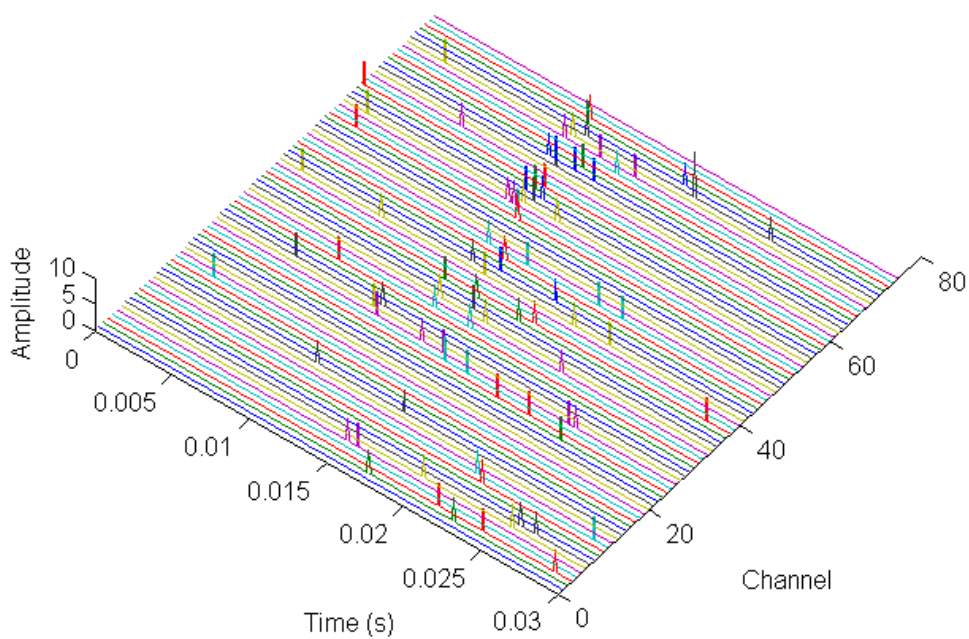


그림 2.18 청신경에 전달되는 일련의 자극열.  
**Fig. 2.18** Spike train transferred to Auditory nerve.

### 2.5.3 악기 대금의 다중채널 청각응답

실제 음악분석에 있어서 단순 FFT 결과를 사용하는 것 보다 청각모델을 적용하여 주파수 성분을 분해하면 한 차원 높은 인식시스템을 구현할 수 있다. 앞절에서 논의된 청각모델을 적용한 분석방법을 실제 악기신호에 응용하였다. 그림 2.20의 기저막 멀티채널분석과 그림 2.21의 악기의 내유모세포 응답모델은 실제 인식시스템에 직접적인 데이터로 사용할 수 있다. 음악신호의 기본주파수, 영교차율, 평균, 분산 등을 구하는데 있어 청각모델을 적용한 데이터를 사용한다.

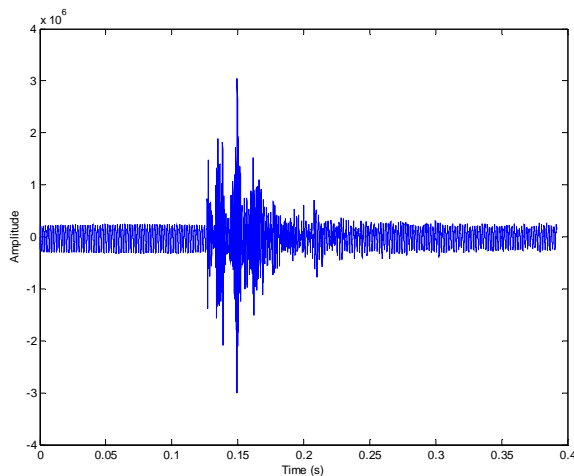


그림 2.19 대금산조의 오디오신호.

Fig. 2.19 Audio signal of Taegum.

그림 2.19의 대금산조파형을 청각모델에 입력하여 단계적 응답그래프를 표시하였다. 그림2.19는 대금산조의 오디오신호로서 0.4초 정도의 신호이며 크기가 평탄한 부분은 대금연주이며 피크부분은 북편소리이다. 그림

2.20은 대금산조의 기저막 멀티채널 응답으로 그림 2.19의 대금악기 신호를 달팽이관 모델에 입력했을 때 응답이다. 75개의 채널로 모의실험 했다. 그림 2.21은 반파정류된 대금의 내유모세포 응답으로 내유모세포가 한쪽 방향으로 만 반응함으로서 반파정류된 모습을 확인할 수 있다. 그림 2.22은 대금산조 연주에 의해 청신경에 전달되는 일련의 자극열이다. 일종의 전기적신호로서 청신경에 난수적으로 맵핑된다.

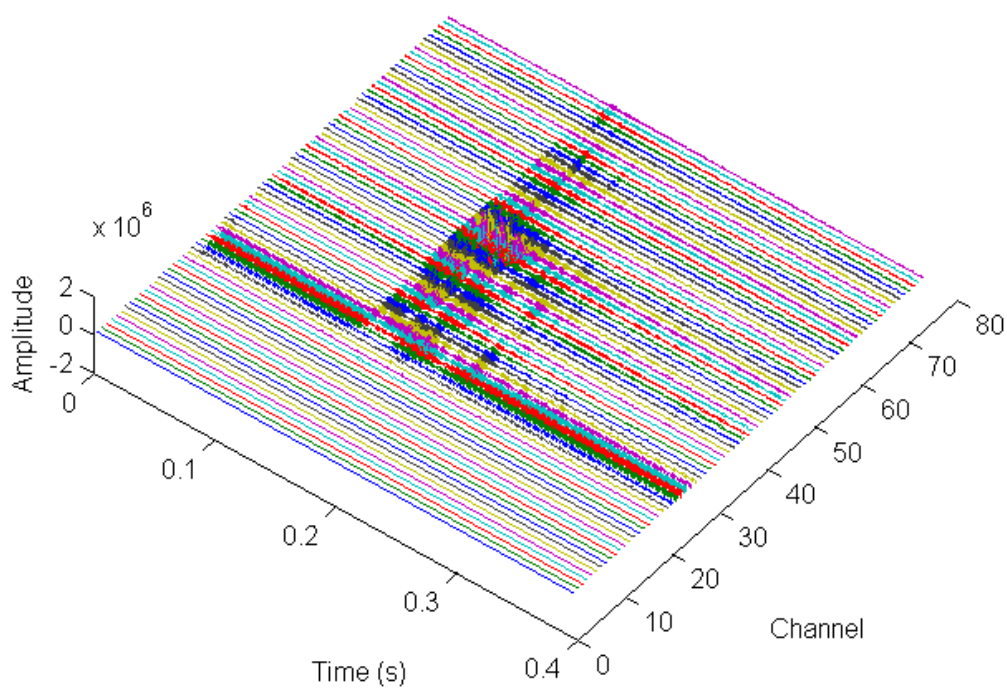


그림 2.20 대금산조의 기저막 멀티채널 응답.

**Fig. 2.20** Response of multi channel Basilar membrane using Taegum Sanjo;  
Korean traditional music.

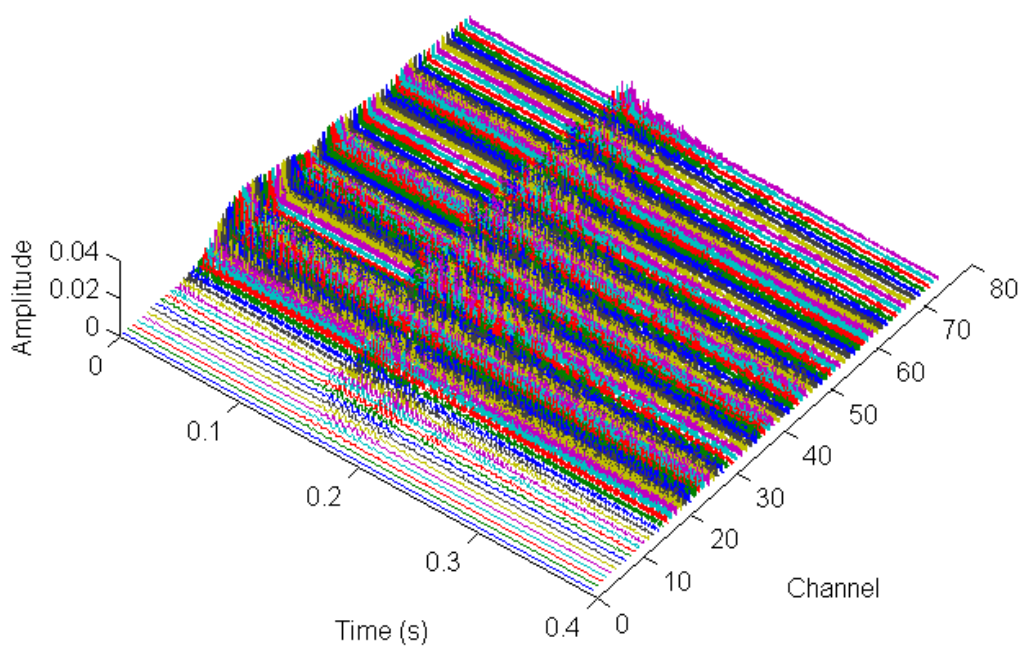


그림 2.21 반과정류된 대금의 내유모세포 응답.

**Fig. 2.21** Response of Inner hair cell shows half-wave rectified signal using Taegum Sanjo playing.

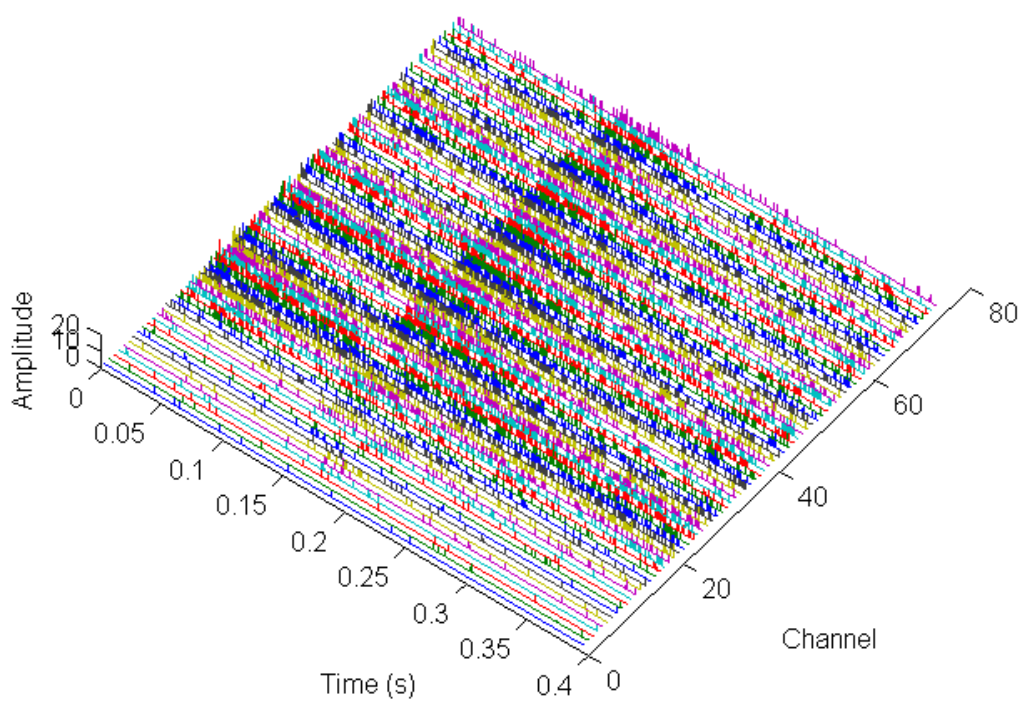


그림 2.22 대금산조 연주에 의해 청신경에 전달되는 일련의 자극열.  
**Fig. 2.22** Spike train caused by Taegum Sanjo transferred to auditory nerve.

기존의 음악인식 시스템은 선형적인 푸리에 변환의 결과를 이용하였다. Tzanetakis[27], Kuo[5] 등은 음악인식에 있어 푸리에 변환된 결과를 이용하여 음악의 질감을 수치화 할 수 있는 함수들을 이용하여 영교차율, PSD(Power Spectral Density), 리듬, 에너지등의 특성값을 추출하여 시스템을 구현하였다. 이러한 특성값을 구하는데 있어 사용한 데이터는 선형적인 푸리에 변환값을 사용하였다. 인간의 감각은 비선형적이다. 음악인식의 주체인 인간의 판단기준을 시스템에 도입하려면 인간의 인식특성을 고려한 비선형적인 데이터의 사용이 필요하다. 본 논문에서는 이러한 비선형적인 특성을 고려하기 위해 청각모델을 도입하였다.

푸리에 변환시에 기저막의 주파수분해 특성과 내유모세포의 응답 특성을 적용하여 음악인식시스템을 구현하고자 한다. 그림 2.16의 기저막 멀티 채널 임펄스응답을 보면 분해된 주파수 대역별 특성이 다르다. 기저막의 아래쪽 채널은 낮은 주파수분해를 담당하고 있으며 상대적으로 응답특성이 느리다. 반면 위쪽 채널의 경우 높은 주파수분해를 담당하고 있으며 낮은 주파수에 비해 응답특성이 빠르다. 이와 같이 알려진 응답특성 뿐만 아니라 차후 의학계에서 밝혀질 청각모델을 도입한다면 차원 높은 인식 시스템의 구현이 가능하다.

## 제3장 음악데이터 분류 알고리즘과 특성함수

분류성공률이 높은 인식시스템을 구현하기 위해서는 음악특성에 맞는 효율적인 분류알고리즘과 특성함수의 선택이 필수적이다. 3.1절에서는 현재 음악 분류에 가장 많이 사용되는 분류 알고리즘인 베이즈법(Bayes rule), 최근접이웃법(nearest neighborhood), k-최근접이웃법(k-nearest neighborhood)의 성능을 국악기와 양악기를 사용한 실험을 통하여 비교한다. 3.2절에서는 특성값중에서 롤-오프포인트의 최적값을 찾는 실험을 할 것이다.

특성함수로부터 얻어지는 특성값(feature vector)은 대상을 특정한 범주로 분류하는데 있어서 데이터로 사용된다. 음악인식시스템에서 분류단계는 특성추출 다음 단계에 해당한다. 다양한 분류기법이 존재하고 있으며 본 논문에 적절하다고 판단되는 기법에 대한 개념, 수학적 이론에 대하여 기술하였다.

### 3.1 분류시스템의 구성요소

분류이론은 판단이론(decision theory)의 하위 분야이다. 일정한 패턴이 특정한 군(class)에 포함된다는 가정에 기반 한다. 개별적인 패턴의 차이에도 불구하고 패턴으로부터 추출한 일련의 특성들이 일정한 범주에 포함된다. 이러한 특성들을 판단하여 범주에 대한 소속 여부를 결정 한다. 음악의 경우 같은 장르에 포함된 곡이라 할지라도 많은 부분에서 차이를 보인다. 이런 상황에서도 인간은 쉽게 장르 판단을 할 수 있는데 이는 동일 장르는 공유할 수 있는 기본특성을 가지고 있다는 것을 증명한다.

분류는 기하학적인 접근을 통하여 접근하는 것이 용이하다. 전술한 바와 같이 특성값은 특성공간의 점들로 위치한다. 분류시스템의 목적은 특정범주에



속한다고 판단할 수 있는 특성공간상의 판단경계선을 찾는 것이다. 그러므로 입력된 데이터는 공간상의 위치를 판단하여 범주를 결정할 수 있다[28,29].

특성값 추출 시스템은 특정한 상황에 적합하도록 설계되어야 한다. 기본적으로 영역(domain)에 종속적이다. 반면 분류기법은 영역으로부터 독립적인 작업이다. 이는 특성값 추출은 일종의 압축단계로서 영역에 종속적인 데이터를 일반적인 수치표현으로 옮기는 작업으로 이해할 수 있다. 이렇게 변환된 일반적인 수치는 분류시스템의 작업대상이 된다.

특성값추출은 음악, 음향심리학, 신호처리등 다양한 분야에 대한 이해가 필수적이다. 음성인식의 경우 기존 연구가 많이 축적되어 있어 음악인식에 필요한 다양한 도구를 제공할 수 있다.

## ● 설계요소

분류시스템의 설계에 있어서 고려해야 할 요소는 특성값의 선택, 특성값의 수, 데이터거리 측정, 시스템의 신뢰성 등이다. 분류시스템은 가공되지 않은 데이터(raw data)를 다루지 않는다. 특성값만을 분류대상으로 사용하므로 적절한 특성값의 추출은 분류시스템 설계에 있어 필수적이다.

## ● 해결해야할 과제

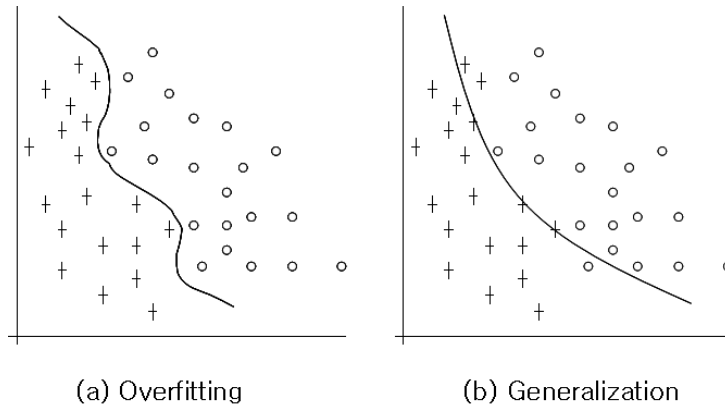
두 가지 요인이 같은 범주에 속하는 패턴의 차이를 발생 시킨다. 첫 번째, 시스템에 사용된 모델에서 기인할 경우이다. 모델이 단순할 경우 정제되지 않은 데이터를 출력하고 이 모델에 대한 정보가 없는 사용자는 큰 혼란을 겪게 된다. 두 번째, 의도하지 않은 잡음에 의한 차이발생으로 어떤 원인에 의하여 차이가 발생했는지 안다면 이에 대처할 수 있다.

## ● 시스템의 훈련 및 학습

분류시스템의 형태의 결정과 학습을 통한 파라미터값의 설정은 특정분류시스템 설계의 핵심이다. 여기서 학습이라 함은 표본데이터를 이용하여 미지의 파라미터값을 찾는 과정이다. 분류는 감독학습(supervised learning)으로 볼 수 있다. 상태를 알고 있는 특성값을 사용하여 분류시스템을 학습시킨다. 이렇게 학습된 분류시스템은 미지의 특성값을 입력 받아 분류결과를 출력한다. 비감독학습(unsupervised learning; clustering)의 경우, 미리 정해진 상태를 사용하지 않고 시스템 자신이 결정한 기준으로 분류한다.

## ● 판단경계의 일반화

학습 과정에 있어서 판단경계의 일반화 적절한 타협점을 찾는 과정이다. 만일 분류시스템이 트레이닝 데이터에 최적화된다면 미지의 데이터에 대한 분류 정확도는 감소할 것이다. 적절한 타협점을 찾음으로써 시스템의 성능을 높일 수 있다. 학습 데이터에 의해 결정된 경계는 적절한 타협점으로 일반화되어야 한다.



**그림 3.1** (a) 학습 데이터에 최적화된 판단경계. (b) 미지의 데이터를 위해 타협점 설정.

**Fig. 3.1** (a) Decision boundary fitted too closely to the training data. (b) General decision boundary for the unknown data.

그림 3.1(a)는 학습 데이터의 경우 완벽한 성공률을 보이지만 미지의 데이터의 경우 실패율이 높아진다. 그림 3.1(b)는 학습 데이터에 최적화 되지는 않았지만 전체적인 성공률은 높아진다.

분류시스템의 목표는 판단성공률을 높이는 것이다. 다음의 분류알리즘들은 이러한 접근방법으로 분류알고리즘의 종류와 개념을 수학적인 해석을 중심으로 설명하였다.

### 3.1.1 베이지언 추정이론

베이지언(Bayesian) 추정이론 확률이론에 근거하여 타협점(trade-off)을 정량화한다.  $x$ 는 개체의 상태이며  $P(x)$ 는 개체의 상태가  $x$ 일 확률이다.  $\omega_i$ 는 군(class)으로서 전체군을 몇가지 상태로 구분한 것이다. 예를 들면  $\omega_1$ =재즈,  $\omega_2$ =국악,  $\omega_3$ =클래식 이라 하면  $P(\omega_1)+P(\omega_2)+P(\omega_3)=1$ 이다.  $P(\omega_i|x)$ 는  $x$ 상태일 때  $\omega_i$ 일 조건부확률이며  $P(x|\omega_i)$ 는  $\omega_i$  군에 속할 때  $x$  상태일 확률이다.

$$P(x|\omega_i) = P(x, \omega_i) / P(\omega_i) \quad (3.1)$$

여기서  $P(x, \omega_i)$ 는 두 사건 동시에 일어날 확률이다.

$$P(\omega_i|x) = P(\omega_i, x) / P(x) \quad (3.2)$$

$P(\omega_i, x) = P(x, \omega_i)$ 이므로 식(3.2)를 식(3.3)으로 정리한다.

$$P(x, \omega_i) = P(\omega_i|x)P(x) \quad (3.3)$$

식(3.3)을 식(3.1)에 대입하면

$$P(x|\omega_i) = P(\omega_i|x)P(x) / P(\omega_i) \quad (3.4)$$

식(3.4)를 사후확률(posterior probability)이라 한다.

전체 군을  $\omega_1, \omega_2, \omega_3$  으로 나누었다면  $P(w_1|x) + P(w_2|x) + P(w_3|x) = 1$ 이 된다. 예를 들어  $P(w_1|x) = 0.8, P(w_2|x) = 0.1$  그리고  $P(w_3|x) = 0.1$ 이면 미지의  $x$ 를 만난다면 시스템은  $x$ 를  $w_1$  으로 분류하게 된다. 따라서 어떤 개체가  $x$ 라는 상태에 있으면  $P(w_i|x)$  값이 가장 큰  $w_i$  에  $x$ 가 속할 것이라고 예측한다.

$$P(w_1|x) < P(w_2|x) \quad (3.5)$$

라 가정하면 어떤 객체가  $x$ 라는 상태일 때  $\omega_2$  군에 속한다고 예측한다. 식 (3.5)의 좌우변을 모두 치환하면

$$P(x|w_1)P(w_1)/P(x) < P(x|w_2)P(w_2)/P(x) \quad (3.6)$$

정리하면

$$P(x|w_1)/P(x|w_2) < P(w_2)/P(w_1) \quad (3.7)$$

식(3.7)을 베이즈 추정법(Bayes' decision rule) 이라 한다. 식 (3.5, 3.6, 3.7)에서 왼쪽항이 더 크다면 어떤 객체  $x$ 는  $\omega_1$  에 속할 것이고 오른쪽항이 더 크다면 객체  $x$ 는  $\omega_2$  에 속할 것이다. 양쪽항 모두 학습을 통해 알 수 있는 값이다. 왼쪽항은 우도비(likelihood ratio), 오른쪽항은 문턱값(threshold)라 한다.

## ● 판별함수

판별함수(discriminant function)는  $n$ 개의  $g_i(x)$  판별함수를 계산하여 가장 큰 값으로 분류하는 네트워크로 표현할 수 있다.  $g_i(x) > g_j(x) (i \neq j)$  를 만족하면

특성값  $x$ 를  $w_i$  군으로 분류한다. 그림 3.2는  $d$ 개의 입력  $x_i$ 와  $n$ 개의 판별함수  $g_i(x)$ 로 구성된 분류시스템의 구조로서 각각의 판별함수는 범주  $w_i$ 와 관계되고  $g_i(x)$ 가 최대가 되는 범주로 분류된다.

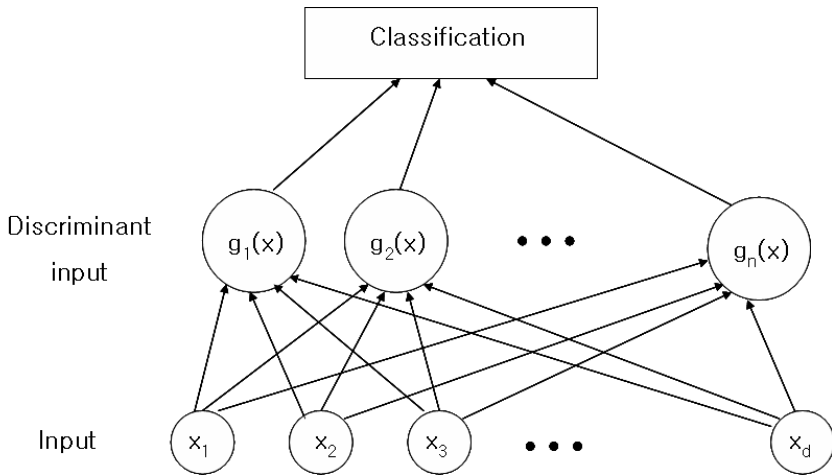


그림 3.2  $d$ 개의 입력  $x_i$ 와  $n$ 개의 판별함수  $g_i(x)$ 로 구성된 분류시스템의 구조.  
**Fig. 3.2** Structure of a classifier with  $d$  inputs  $x_i$  and  $c$  discriminant functions  $g_i(x)$ .

베이즈 분류시스템이 최소 에러비(error ratio)를 가질 경우, 판별함수의 최대값은 최대사후확률에 해당한다. 따라서,  $g_i(x) = P(w_i|x)$ 이며 판별함수를 식 (3.8)과 같이 나타낼 수 있다.

$$g_i(x) = P(x|w_i)P(w_i) \quad (3.8)$$

● 정규밀도 분포

베이지 분류시스템은 사전확률  $P(w_i)$ 와 조건부 확률밀도 함수  $P(x|w_i)$ 의해 결정된다. 정규밀도분포는 가장 일반적인 밀도분포이다. 정규밀도분포함수는  $w_j$  군의 특성벡터  $x$ 가 정형(prototype) 벡터  $\mu_j$ 에 기반 할 경우를 모델링한 것이다. d차원의 정규밀도 함수는 식(3.9)과 같이 표현할 수 있다.

$$P(x) = \frac{1}{(2\pi)^{\frac{d}{2}} |\sigma|^{\frac{1}{2}}} e^{-\frac{1}{2}(x-\mu)^T \sigma^{-1}(x-\mu)} \quad (3.9)$$

여기서  $x$ 는 d 차원의 열벡터이고  $\mu$ 는 d 차원의 평균(mean) 벡터이다.  $\sigma$ 는 d-by-d 공분산(covariance) 행렬이다. 식(3.9)의 단순화된 표현은 식(3.10)와 같다.

$$P(x) \sim N(\mu, \sigma) \quad (3.10)$$

기대값  $\epsilon$ 은 식(3.11)와 같다.

$$\epsilon[x] = \int x p(x) dx \equiv \mu \quad (3.11)$$

공분산(covariance) 행렬은 식(3.12)와 같다.

$$\epsilon[(x-\mu)(x-\mu)^T] = \int (x-\mu)(x-\mu)^T P(x) dx \equiv \sigma \quad (3.12)$$

만일 데이터가 정규분포를 이루고 평균과 공분산  $\mu, \sigma$  값을 알고 있으면 군판별함수의 조건부밀도함수인  $P(x|w_i)$ 은 식(3.9)에 정의된 정규밀도함수로 대체할 수 있다. 정규밀도함수는 강력한 도구로서 분류객체가 정규분포성질을 가지고 있을 때에만 적용할 수 있다. 그러나 자연상태에서 정규분포는 그리 흔한 경우는 아니므로 실제 분류시스템에서의 전형적인 문제는 다음과 같다.

완전한 확률분포구조에 대하여 알 수 없으므로 정규밀도와 같은 표준적인 분포모형을 따른다고 가정한다. 이런 경우, 평균벡터  $\mu$ 와 같이 조건부확률밀도함수의 계수(parameter)들을 결정해야 한다. 그리고 실제 데이터의 구조는 미리 알 수 없으므로 분류시스템은 학습 데이터를 통해서 정보를 얻어야 한다.

대부분의 분류상황에서 확률의 밀도분포는 알 수 없으므로 일반적인 분포형태에 대한 가정을 해야 한다. 일반적인 형태의 분포를 결정하는 것은 함수의 계수만 결정하면 되므로 분류시스템구현이 용이해진다. 미지의 확률분포는 샘플 데이터를 이용한 학습을 통해 구해진다. 예를 들면,  $P(x|w_i)$ 를 정규분포라 가정하면 평균  $\mu$ 와 공분산  $\sigma$  값만 찾아내면 된다.

## ● 최대우도 계수예측법

일반적인 계수예측기법 중 하나인 최대우도 계수예측법(maximum-likelihood parameter estimation)은 계수를 미지의 고정된 값으로 간주한다. 샘플로부터 구한 평균을  $\hat{\mu}$ 로 표시하면

$$\hat{\mu} = \frac{1}{n} \sum_{k=1}^n x_k \quad (3.13)$$

식(3.13)는 학습 샘플의 산술평균이므로 샘플의 수가 많아 질수록 정확한



값을 얻을 수 있다. 샘플데이터의 특성값은 특성공간의 점으로 표시되므로 샘플의 평균값은 그 점들의 중심으로 볼 수 있다.

## ● 비-계수법

계수예측은 조건부확률밀도의 구조에 대한 가정을 기반으로 한다. 그러나 다음 2가지 경우엔 이를 적용할 수 없다. 확률밀도함수의 구조를 전혀 파악할 수 없을 경우와 확률밀도가 다수의 최대값을 가질 경우(multi-modal), 앞서 논의된 정규분포는 평균  $\mu$  위치에서 한 개의 최대값(uni-modal)을 갖는다(uni-modal).

이 경우 최대우도 계수예측법은 적용할 수 없다. 데이터 구조에 따른 몇 가지 대안이 있다. 예를 들면, 만일 미지의 확률 분포가 밀도함수분포를 알고 있는 여러개의 uni-modal 분포의 합이라면 기대최대화(expectation-maximization)[4] 기법을 이용하여 계수들을 추측할 수 있다. 만일, 데이터에 대한 정보가 전혀 없다면 분류시스템은 학습샘플 정보에만 의지하여 구현된다. 이러한 비-계수 예측기법은 두부류로 나눌 수 있다. 첫째, 밀도함수예측  $P(x|w_i)$  두 번째, 사후 확률  $P(w_i|x)$ 예측이다.

### 3.1.2 은닉 마코브모델

은닉 마코브모델(Hidden Markov Models; HMMs) 은 일련의 판단을 내리는 문제에 대한 해법을 제시한다[30]. 분류시스템에 있어서 일련의 상태를 거치는 데 시간  $t$ 의 상태는 이전 상태의 영향을 받은 결과이다. HMMs은 시간  $t$ 의 상태는 시간( $t-1$ )의 영향만 받는 다고 가정한다. 음성은 연속적인 음의 발생이라는 측면에서 HMMs의 주요 응용분야중 하나이다. 마코브 모델의 주요개념은 전

이 확률(transition probability)이다. 전이 확률  $a_{ij} = P(w_j(t+1)|w_i(t))$  은 시간  $t$  의 상태  $w_i$  일때 시간  $(t+1)$ 의 상태  $w_j$ 일 확률이다. 전이 확률은 대칭적일 필요는 없고 같은 상태를 한번이상 거쳐 갈 수 있다. 상태  $w(t)$  에 접근할 수 없을 경우 심볼만이 관측가능하다. 상태  $w(t)$  에 있는 시스템은 심볼  $v(t)$  를 방출한다. 확률  $b_{jk} = P(v_k(t)|w_j(t))$  는 상태  $w_j(t)$  에 있는 시스템이 심볼  $v_k(t)$  를 방출 할 확률이다. HMMs에서 의미하는 은닉은 직접 접근할 수 없고 숨겨져 있다는 의미이다.

HMMs의 구조(상태와 심볼의 개수)와 학습샘플의 개수를 알면 확률  $a_{ij}$ 와  $b_{jk}$ 를 알 수 있다. 이를 마코브 모델 트레이닝(Markov model training) 이라 한다 [31].

### 3.1.3 최근접이웃법

사후확률을 예측하는데 가장 많이 사용하는 기법은 최근접이웃법(nearest-neighbour rule; NN)이다. 어떤  $x$ 를 분류하고자 할 때 인접한 기준점(prototype point)  $x'$ 가 군  $D = (x_1, \dots, x_n)$ 에 속한다면  $x$ 도 군  $D$ 에 속한다고 판단한다. k-최근접이웃법(k-nearest neighbor rule; k-NN)은 최근접이웃법의 확장판이다. 어떤  $x$ 의 k개수 인접 점들의 레이블(label)을 조사한다. 특정 레이블의 상대적인 빈도는  $P(w_i|x)$ 로 표시한다.  $P(w_i|x)$  값이 가장 큰 값일 때  $x$ 는  $w_i$ 로 분류된다. 여기서 최적의 k값을 찾는 것이 핵심이다. 예측에 사용된 인접점들은  $P(w_i|x') \approx P(w_i|x)$ 를 만족할 만큼 가까워야한다. k는 신뢰성있는 값을 보장할 수 있을 만큼 커야 한다.

최근접이웃법과 같은 알고리즘은 특성공간 상에서 특성벡터 사이의 거리를 측정한다. 일반적인 거리측정은 민코프스키 거리식(Minkowski metric)(3.14)을 사용한다.

$$d_p(x_i, x_j) = \sqrt[p]{\sum_{k=1}^d (|x_{i,k} - x_{j,k}|)^p} = \|x_i - x_j\|_p \quad (3.14)$$

연속적인 값의 거리를 측정하는 데 가장 많이 쓰이는 것은 유클리디안 거리 (Euclidean distance)이다. 이는 2차원 또는 3차원 상의 거리를 측정하는데 흔히 사용된다. 민코프스키 거리식에서  $p=2$ 인 경우가 유클리디안 거리이다.

$$d_2(x_i, x_j) = \sqrt{\sum_{k=1}^d (|x_{i,k} - x_{j,k}|)^2} = \|x_i - x_j\|_2 \quad (3.15)$$

### 3.1.4 k-최근접이웃법

k-최근접이웃법은 주어진 데이터 셋  $(x_i, y_i)$ 을 이용하여  $x$ 에 대한  $y$ 의 값을 예측한다.  $x$ 에 가까운  $x_i$ 의 k값을 이용하여  $y$ 값을 예측한다. 여기서 k는 일종의 계수로서 가장 영향이 큰 요인에 가중치를 두어 예측할 수 있다. 예측과정은 다음과 같다. 학습 데이터 셋은 분류대상의 분류에 사용될 계수를 생성하는데 사용한다. 즉, 분류시스템에는 분류대상(국악, jazz, classical)이 있고 이를 예측하기 위한 각 음악장르의 특징(음의 질감, 에너지, 리듬 등) 들이 존재한다. 분류대상과 가장 가까운 학습 데이터 셋의 k 계수를 구한다. 그리고 유클리디안 거리 측정법을 사용하여 분류대상과 학습 데이터의 거리를 구한다. 이러한 과정을 나머지 분류대상 데이터에 대하여 반복한다. k 계수가 많아질수록 계산시간은 길어지지만 비이상적인 데이터의 악영향을 감소시킬 수 있다. 실제로는 수십개 이하의 k 값을 사용한다.

NFL(Nearest Feature Line)은 최근접이웃법과 유사한 비계수 분류기법이다.

최근접이웃법은 개별적으로  $x$ 에 인접한 기준점  $x'$ 와 비교하지만 NFL은 한번에 여러 개의 인접 기준점 정보를 이용한다. NFL은 군내의 두개의 기준점  $x_i, x_j$ 를 통과하는 선이다. 각각의 군내에 가능한 기준점 쌍을 통과하는 NFL이 생성된다. 따라서 점  $x$ 와 NFL의 최소거리를 사용하여 분류작업을 한다[32].

### 3.2 악기 음원을 이용한 분류 알고리즘 성능비교 실험

같은 데이터그룹을 이용하여 베이지법, k-최근접이웃, 최근접이웃의 분류 성공률을 비교하고 자 한다. 3.2.1절의 실험데이터로는 국악기인 대금과 가야금의 표본을 사용하였다. 표본은 44.1k, 각각의 악기마다 15초 길이의 60개 데이터를 준비했다. 한 개의 표본마다 5개의 특성값을 추출하였다. 3.2.2절의 실험은 바이올린과 첼로음악을 이용하여 알고리즘 성능비교를 시행하였다. 베이지법, k-최근접이웃, 최근접이웃 알고리즘에 따라 데이터를 적용하였다. 분류 시스템은 10진 교차유효화(ten-fold cross-validation)을 사용하여 훈련하였다. 교차유효화(cross-validation)은 패턴인식에서 사용하는 일반적인 훈련기법으로 데이터집합은 같은 크기의  $n$ 부분으로 나뉘어지고  $(n-1)$ 부분은 분류시스템을 학습시키기 위해 사용한다. 나머지  $n$ 번째 부분은 학습된 시스템을 시험하기 위해 사용한다.

3.2.2 절의 실험데이터로는 바로크 바이올린, 바로크 첼로, 바이올린 표본을 사용하였다. 바로크 악기는 18세기 이전에 사용되던 악기로서 현대악기의 전신이다. 바로크 바이올린과 현대 바이올린의 차이점은 현의 재질이 현대악기는 금속이고 바로크악기는 동물의 내장을 꼬아서 만든 것이다. 외관상구조적인 모습은 유사하다. 바로크악기는 현대악기보다 음량이 작지만 섬세한 음색을 발성한다. 표본은 44.1k, 15초 길이의 96개 데이터를 준비했다. 각각의 악기

표본은 32개로 구성되어 있다. 한 개의 표본마다 5개의 특성값을 추출하였다. Bayes rule, k-최근접이웃, 최근접이웃 알고리즘에 따라 데이터를 적용하였다. 3.2.1절의 실험과 마찬가지로, 분류시스템은 10진 교차유효화(ten-fold cross-validation)을 사용하여 훈련하였다.

### 3.2.1 국악기를 이용한 알고리즘 성능비교

아래 나열된 특성값은 국악기와 서양악기에서의 분류 알고리즘 성능비교에 사용된 특성값들이다.

- **영교차율**

영교차율은 시간영역의 특성값으로 처리시간이 빠르고 시간에 따른 영교차율 변화를 분석하고 통계적 특성을 이용하여 음성(voice)신호와 비음성 신호를 구분하는데 주로 사용되었다. 음성신호는 비음성신호보다 작은 영교차율 값을 갖는다.

- **평균, 분산**

신호의 평균 및 분산값으로 일반적인 인식시스템에서는 사용하지 않는다. 그러나 본 실험에서는 한정된 데이터집합을 이용해 분류시스템을 하는 것이므로 유용하게 사용된다.

- **평균피크값(Average Peak level)**

신호의 피크레벨의 평균값이다. 일정한 임계값을 초과하는 신호는 피크로 간주하고 피크의 빈도와 세기를 측정한다. 음압 또는 음량에 따른 악기의 특성을 수치화한다.

$$AV_{peaklevel} = \sum \frac{x(t)}{f_p} \quad (3.16)$$

$x(t)$ 는 측정된 피크의 크기이고  $f_p$ 는 피크의 빈도이다.

- **군(Class)**

학습데이터가 소속된 범주를 말한다. 본 분류시스템은 감독학습(supervised learning)의 일종으로 데이터의 군을 이용하여 학습한다.

표 3.1 추출된 특성값의 예  
Table 3.1 Example of feature values

평균	분산	영교차율	평균피크	군
-0.001545,	0.000253,	7121.000000,	0.026573,	taegum
-0.002432,	0.000938,	20071.000000,	0.058750,	taegum
-0.002159,	0.000727,	19570.000000,	0.046796,	taegum
.	.	.	.	.
-0.000590,	0.000593,	9800.000000,	0.051839,	ajang
-0.000470,	0.000772,	12630.000000,	0.053978,	ajang
-0.000492,	0.000481,	7899.000000,	0.039112,	ajang
.	.	.	.	.
0.000096,	0.000070,	913.000000,	0.010253,	geomungo
0.000091,	0.000112,	1435.000000,	0.012191,	geomungo
0.000118,	0.000426,	2815.000000,	0.023039,	geomungo
.	.	.	.	.

**표 3.2** Bayes rule 실험결과, 분류성공률 96.67%

**Table 3.2** Experiment result of Bayes rule, Success rate 96.67%

구분	대금 출력	아쟁 출력	거문고 출력
대금	19	1	0
아쟁	1	19	0
거문고	0	0	20

**표 3.3** NN 실험결과, 분류성공률 100%

**Table 3.3** Experiment result of NN, Success rate 100%

구분	대금 출력	아쟁 출력	거문고 출력
대금	20	0	0
아쟁	0	20	0
거문고	0	0	20

**표 3.4** k-NN 시험결과, 분류성공률

**Table 3.4** Experiment result of k-NN, Success rate 85%

구분	대금 출력	아쟁 출력	거문고 출력
대금	13	7	0
아쟁	2	18	0
거문고	0	0	20



TP rate(True Positive)는 특정군의 요소로서 해당 군으로 분류된 것을 말한다. FP rate(False Positive) 특정군의 요소가 아니면서 해당군으로 분류된 것을 말한다. FN rate(False Negative)는 특정군의 요소가 아니면서 해당군이 아닌 다른 군으로 분류된 것을 나타낸다. 정밀도(Precision)은 특정클래스로 분류된 모든 요소들 중에서 정확히 분류된 요소들의 비를 나타낸다. 리콜(Recall)은 특정군에 속하는 모든 요소들 중에서 정확히 분류된 요소들의 비를 나타낸다.

$$Precision = \frac{|TP|}{|TP| + |FP|} \quad (3.17)$$

$$Recall = \frac{|TP|}{|TP| + |FN|} \quad (3.18)$$

F-값(F-measure)는 Precision과 Recall을 통합한 파라미터로서 Recall과 Precision 이 모두 필요할 때 사용한다.

$$F = \frac{2 \cdot Recall \cdot Precision}{Recall + Precision} = \frac{2|TP|}{2|TP| + |FP| + |FN|} \quad (3.19)$$

**표 3.5** Bayes rule, 군별 세부 정확도

**Table 3.5** Bayes rule, Detailed accuracy by class

구분	TP rate	FP rate	Precision	Recall	F-Measure
대금	0.95	0.025	0.95	0.95	0.95
아쟁	0.95	0.025	0.95	0.95	0.95
거문고	1	0	1	1	1

**표 3.6** NN, 군별 세부 정확도

**Table 3.6** NN, Detailed accuracy by class

구분	TP rate	FP rate	Precision	Recall	F-Measure
대금	1	0	1	1	1
아쟁	1	0	1	1	1
거문고	1	0	1	1	1

**표 3.7** k-NN, 군별 세부 정확도

**Table 3.7** k-NN, Detailed accuracy by class

구분	TP rate	FP rate	Precision	Recall	F-Measure
대금	0.65	0.05	0.87	0.65	0.74
아쟁	0.9	0.17	0.72	0.9	0.8
거문고	1	0	1	1	1

카파통계(Kappa statistics)는 만일 입력데이터 각각의 군별 소속을 알고 있다면 분류결과의 출력물과 비교한 성공률을 나타낸다. Kappa=1이면 분류결과의 완벽한 일치율을 나타내고 Kappa=0이면 전혀 일치하지 않음을 나타낸다. 음의 Kappa값은 임의의 분류확률보다 낮게 일치하였음을 나타낸다. 일반적으로 카파통계값이 0.9 이상이면 우수하다고 판단하며 0.7이하 는 개선이 필요함을 나타낸다. 평균근오차(MSE; Mean Squared Error)는 실제 값과 예측값의 차이를 측정한다[7].

$$MSE = \frac{1}{n} \sum_{i=1}^n (a_i - p_i)^2 \quad (3.20)$$

**표 3.8** Bayes rule, 전체 정확도  
**Table 3.8** Bayes rule, Entire accuracy

Correctly Classified Instances	58
Incorrectly Classified Instances	2
Kappa statistic	0.95
Mean absolute error	0.029
Root mean squared error	0.15
Relative absolute error	6.55%
Root relative squared error	31.31%
Total Number of Instances	60

**표 3.9** NN, 전체 정확도  
**Table 3.9** NN, Entire accuracy

Correctly Classified Instances	60
Incorrectly Classified Instances	0
Kappa statistic	1
Mean absolute error	0.023
Root mean squared error	0.025
Relative absolute error	5.26%
Root relative squared error	5.26%
Total Number of Instances	60

**표 3.10** k-NN, 전체 정확도  
**Table 3.10** k-NN, Entire accuracy

Correctly Classified Instances	51
Incorrectly Classified Instances	9
Kappa statistic	0.77
Mean absolute error	0.11
Root mean squared error	0.30
Relative absolute error	25.51%
Root relative squared error	63.88%
Total Number of Instances	60

### 3.2.2 바이올린과 첼로를 이용한 알고리즘 성능비교

국악기 실험과 같이 평균, 분산, 영교차율, 평균피크, 군이 사용되었다. 표 3.11은 실험에 사용된 특성함수에서 추출된 특성값의 예이다.

**표 3.11** 추출된 특성값의 예  
**Table 3.11** Example of feature values

평균	분산	영교차율	평균피크	군
-0.000002,	0.003485,	23154.000000,	0.146585,	modern-violin
0.000000,	0.014886,	39109.000000,	0.298086,	modern-violin
-0.000000,	0.008804,	25762.000000,	0.194512,	modern-violin
.	.	.	.	.
-0.000021,	0.005889,	9645.000000,	0.156339,	baroque-cello
-0.000050,	0.001725,	11793.000000,	0.098258,	baroque-cello
-0.000027,	0.001890,	8662.000000,	0.085151,	baroque-cello
.	.	.	.	.
-0.000003,	0.001840,	17589.000000,	0.088823,	baroque-violin
0.000000,	0.005120,	22058.000000,	0.143670,	baroque-violin
0.007161,	0.004754,	20438.000000,	0.160210,	baroque-violin
.	.	.	.	.

**표 3.12** Bayes rule 실험결과, 분류성공률 77.08%

**Table 3.12** Experiment result of Bayes rule, Success rate 77.08%

구분	modern-violin 출력	baroque-cello 출력	baroque-violin출력
modern-violin	32	0	0
baroque-cello	0	31	1
baroque-violin	21	0	11

**표 3.13** NN 실험결과, 분류성공률 88.54%

**Table 3.13** Experiment result of NN, Success rate 88.54%

구분	modern-violin 출력	baroque-cello 출력	baroque-violin출력
modern-violin	26	0	6
baroque-cello	0	32	0
baroque-violin	3	2	27

**표 3.14** k-NN 시험결과, 분류성공률 89.58%, 2개의 표본데이터는 미분류

**Table 3.14** Experiment result of k-NN, Success rate 89.58%, 2 sample data set are unclassified

구분	modern-violin 출력	baroque-cello 출력	baroque-violin출력
modern-violin	27	0	5
baroque-cello	0	31	0
baroque-violin	3	0	28

**표 3.15** Bayes rule, 군별 세부 정확도

**Table 3.15** Bayes rule, Detailed accuracy by class

구분	TP rate	FP rate	Precision	Recall	F-Measure
modern-violin	1	0.33	0.60	1	0.75
baroque-cello	0.97	0	1	0.97	0.98
baroque-violin	0.34	0.016	0.92	0.34	0.5

**표 3.16** NN, 군별 세부 정확도

**Table 3.16** NN, Detailed accuracy by class

구분	TP rate	FP rate	Precision	Recall	F-Measure
modern-violin	0.81	0.047	0.90	0.81	0.85
baroque-cello	1	0.031	0.94	1	0.97
baroque-violin	0.84	0.094	0.82	0.84	0.83

**표 3.17** k-NN, 군별 세부 정확도

**Table 3.17** k-NN, Detailed accuracy by class

구분	TP rate	FP rate	Precision	Recall	F-Measure
modern-violin	0.84	0.048	0.9	0.84	0.87
baroque-cello	1	0	1	1	1
baroque-violin	0.90	0.079	0.85	0.90	0.87



**표 3.18** Bayes rule, 전체 정확도  
**Table 3.18** Bayes rule, Entire accuracy

Correctly Classified Instances	74
Incorrectly Classified Instances	22
Kappa statistic	0.66
Mean absolute error	0.14
Root mean squared error	0.33
Relative absolute error	31.08%
Root relative squared error	70.70%
Total Number of Instances	96

**표 3.19** NN, 전체 정확도  
**Table 3.19** NN, Entire accuracy

Correctly Classified Instances	85
Incorrectly Classified Instances	11
Kappa statistic	0.83
Mean absolute error	0.076
Root mean squared error	0.28
Relative absolute error	17.17%
Root relative squared error	58.59%
Total Number of Instances	96

**표 3.20 k-NN, 전체 정확도**  
**Table 3.20 k-NN, Entire accuracy**

Correctly Classified Instances	86
Incorrectly Classified Instances	8
Kappa statistic	0.87
Mean absolute error	0.059
Root mean squared error	0.23
Relative absolute error	13.47%
Root relative squared error	50.21%
Total Number of Instances	96

소리신호 분류에 가장 많이 사용되는 분류알고리즘을 비교하였다. 실험결과 국악기실험에서는 최근접이웃 알고리즘이 분류성공률 100%로서 음악인식에 있어 빠른 계산과 가장 좋은 성공률을 보였다. k-최근접이웃 알고리즘은 85% 성공률로 가장 낮은 결과를 보였다. 오류는 대금과 아쟁에서 발생하였다. 서양 악기 실험에서는 국악기와의 비교를 위하여 같은 방법으로 바이올린과 첼로에 대하여 실험하였다. k-최근접이웃 알고리즘이 분류성공률 89.58%로서 악기인식에 있어 빠른 계산과 가장 좋은 성공률을 보였다. 베イズ법은 77.08% 성공률로 가장 낮은 결과를 보였다.

### 3.3 바이올린 음원을 이용한 롤-오프 포인트 검출 실험

특성함수 중에서 저음과 고음비의 수치를 측정하는 스펙트랄 롤-오프 함수의 최적파라미터를 찾는 실험을 행하였다. 음성신호의 특성을 분석하는 다양한 처리함수들이 존재하는데 대부분이 음성처리 분야에서 차용한 것으로 본 실험에서는 음악인식에 적합한 롤-오프 포인트를 찾고자 하였다. 롤-오프 포인트의 변화를 명확히 측정하기 위해 특성값의 수를 4개로 제한 하였다. 사용된 특성함수는 롤-오프, 분산, 평균피크값, 군이다. 스펙트랄 롤-오프식은 식 (3.21)과 같다.

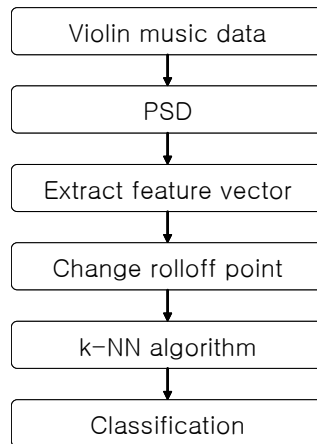
#### ● 스펙트랄 롤오프

PSD 최고값의 특정프레임(frame) 비율을 나타낸다. 고음역이 많은 악기의 경우 저음악기보다 더 작은 값을 나타낸다. 여기서 %비율은 목적에 맞게 임의로 조정할 수 있다.

$$\sum_{n=1}^R M_t[n] = r \cdot \sum_{n=1}^N M_t[n] \quad (3.21)$$

식(3.21)에서  $M_t[n]$ 은 프레임  $t$ 에서의 PSD 값이다.  $r$  값이 0.5이면 주파수분포의 중앙값을 나타낸다.  $R$ 값은 PSD의  $r\%$ 에 상응하는 값이다. 본 실험에서 분류성공률을 최대화 할 수 있는  $r$ 값을 0.05에서 0.95까지 0.05 간격으로 증가시키며 검출하였다.

데이터파일을 반복적으로 처리하여 분산, 스펙트랄 롤-오프, 평균피크값, 군의 분포벡터를 얻었다. 15초 길이의 현대바이올린 데이터 100개, 바로크바이올린 데이터 100개로 하여 한회의 실험을 하였고 롤-오프 포인트를 0.05에서 0.95 까지 변화시키며 총 19회의 실험을 행하였다. 실험순서는 그림 3.3과 같다.



**그림 3.3** 실험순서.

**Fig. 3.3** Flowchart of experiment.

데이터사이의 벡터거리를 측정하여 분류하는 k-최근접이웃분류 알고리즘을 적용하였다. 음악분류에서 k-최근접분류 알고리즘이 다른 알고리즘보다 우수한 성능을 보여 주었다. 분류시스템은 10분 교차 훈련법(ten-fold cross-validation)을 사용하여 하였다. 여기서 교차훈련(cross-validation)은 패턴인식에서 사용하는 일반적인 훈련기법으로 데이터집합은 같은 크기의  $n$  부분으로 나뉘어지고  $(n-1)$ 부분은 분류시스템을 학습시키기 위해 사용한다. 나머지  $n$ 번째 부분은 학습된 시스템을 시험하기 위해 사용한다.

표 3.21의 데이터는 바이올린음원파일로부터 얻은 롤-오프 포인트 0.4 일 때 특성값의 예이다. 오른쪽부터 분산, 스펙트랄 롤-오프, 평균피크값,

군을 나타낸다. 한 개의 악기음원 파일에서 추출한 4개의 특성값을 한 줄에 표시하였다. 특성값 중에서 군은 분류시스템의 훈련에 사용하기 위해 미리 현대바이올린과 바로크바이올린으로 입력한다. 따라서 각각의 롤-오프 포인트마다  $200 \times 4 = 800$ 개의 특성값이 사용되었다. 실험에서 설정된 롤-오프 포인트는 19개 이므로 전체  $200 \times 19 \times 4 = 15200$ 개의 특성값이 최적화된 롤-오프 포인트를 찾는데 사용되었다.

**표 3.21** 추출된 특성값의 예  
**Table 3.21** Example of feature values

분산	롤-오프	평균피크값	군
0.004920,	0.156220,	1378.125000,	modern
0.004454,	0.159770,	1378.125000,	modern
0.007294,	0.206301,	1378.125000,	modern
0.003660,	0.132678,	1378.125000,	modern
0.004243,	0.144828,	1378.125000,	modern
		.	
		.	
		.	

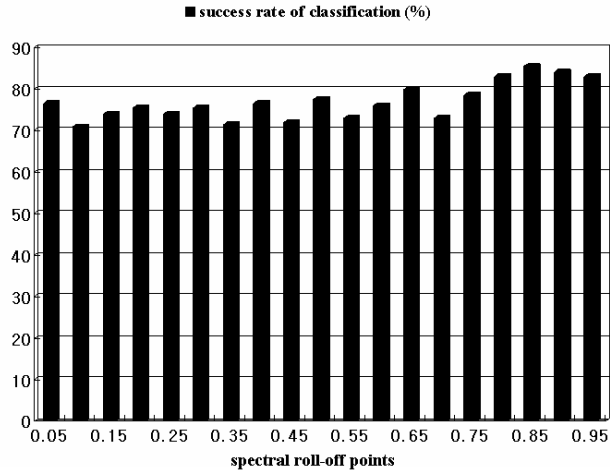


그림 3.4 롤-오프 포인트 변화에 따른 분류 성공률.

Fig. 3.4 Success rates of various roll-off point.

그림 3.4을 보면 스펙트랄 롤-오프 포인트 0.85일 때, 최대분류성공률 85.5%를 나타냈다. 표 3.22는 롤-오프 포인트 0.85에서 상세분류내용으로 200개의 입력데이터 중에서 171개의 데이터가 정확히 분류되었다. 표 3.22의 왼쪽구분은 실제 입력된 데이터의 종류를 나타내고 가로구분은 실험의 출력결과를 나타낸다.

**표 3.22** 스펙트럴 롤오프 포인트 0.85에서 분류결과  
**Table 3.22** Result of Success rate at spectral rolloff point 0.85

	output - modern violin	output - baroque violin
input - modern violin	87	13
input - baroque violin	16	84

**표 3.23** 전체 정확도  
**Table 3.23** All over accuracy

Correctly Classified Instances	171
Incorrectly Classified Instances	29
Kappa statistic	0.71
Mean absolute error	0.14
Root mean squared error	0.38
Relative absolute error	29%
Root relative squared error	76.16%
Total Number of Instances	200

현대바이올린과 바로크바이올린은 대역이 같고 음색이 유사하여 최적의 롤-오프 포인트를 찾는 데 유용할 것으로 판단하여 실험데이터로 사용하였다. 전체적인 분류성공률이 70%를 상회하는 것은 분산, 평균피크레벨과 군 특성값들의 유효성을 증명하는 것으로 롤-오프 포인트의 변화에 따라 최저 71%에서 최고 85.5%의 분류성공률은 나타났다. 저음의 비율이 높은 롤-오프 포인트 구간인 0.8 이후의 실험결과에서는 80%를 상회하는 분류성공률을 나타냄으로써 저음의 비를 상대적으로 높게 설정하는 것이 유효함을 알 수 있었다. 두 악기의 유사한 대역특성으로 인해 높거나 낮은 값의 최적 롤-오프 포인트를 찾을 것으로 기대했었다. 실험결과 롤-오프

프 포인트 0.85에서 최고의 분류성공률을 얻을 수 있었다.

### 3.4 실험결과 및 해석

3.1절에서는 소리신호 분류에 가장 많이 사용되는 분류알고리즘을 비교하였다. 실험결과 국악기실험에서는 최근접이웃 알고리즘이 분류성공률 100%로서 음악인식에 있어 빠른 계산과 가장 좋은 성공률을 보였다. k-최근접이웃 알고리즘은 85% 성공률로 가장 낮은 결과를 보였다. 오류는 대금과 아쟁에서 발생하였다. 대금과 아쟁의 음역대가 유사하고 일정시간동안 발성이 지속된다는 특징을 가지고 있다. 반면 거문고는 탄현악기로서 아쟁과 거문고에 비해 단속적인 특성을 갖는다. 음악 파일로부터 추출한 특성값은 2차원 벡터상의 위치이므로 순수한 확률과정법보다는 벡터거리를 적용한 최근접이웃 알고리즘의 분류결과가 우수한 것으로 확인되었다. 서양악기 실험에서는 k-최근접이웃 알고리즘이 분류성공률 89.58%로서 악기인식에 있어 빠른 계산과 가장 좋은 성공률을 보였다. 베이스법은 77.08% 성공률로 가장 낮은 결과를 보였다. 분류 오류는 바로크 바이올린과 현대 바이올린서 발생하였다. 두 악기는 바이올린으로서 음역대와 음색이 유사하여 오류가 발생한 것으로 판단된다. 반면 바로크 첼로는 음역대가 낮아 상대적으로 낮은 오류율을 보였다. 서양악기에서도 전체적으로 벡터거리 알고리즘을 사용한 k-최근접이웃법과 최근접이웃법이 좋은 성능을 보였다.

3.3절 최적의 톨-오프 포인트를 찾는 실험에서는 저음의 비율이 높은 톨-오프 포인트 구간인 0.8 이후의 실험결과에서는 80%를 상회하는 분류성공률을 나타냄으로 저음의 비를 상대적으로 높게 설정하는 것이 유효함을 알 수 있었다. 두 악기의 유사한 대역특성으로 인해 높거나 낮은



값의 최적 룰-오프 포인트를 찾을 것으로 기대했었다. 실험결과 룰-오프  
포인트 0.85에서 최고의 분류성공률을 얻을 수 있었다.

## 제4장 한국전통음악의 분류

방대한 양의 멀티미디어 데이터를 처리하는 것은 점점 중요해지고 있다. 모든 데이터를 일일이 확인하고 표시하는 것은 시간과 비용에 있어 적합하지 않다. 따라서 여러 가지 분류기법들이 소개되고 있다. 대부분은 음성처리 기법을 이용하고 있으며 데이터들의 유사한 특성을 지표로 삼아서 분류하고 있다. 유사한 특성값을 이용한 분류시스템이 Wold, Blum 등에 의해 구현되었다[33]. 이 시스템은 데이터의 인지적, 음향적 유사성을 이용하여 유사한 데이터를 분류할 수 있다. 본 논문은 음향데이터 분류 그 중에서도 한국 전통악기 분류에 초점을 맞추어 분류시스템을 구현하려고 한다.

음성신호의 분석 기법은 다양하게 연구되어 왔다. 그러나 음악을 포함한 오디오 신호의 스펙트럼 분석은 그 역사가 길지 못하고 음성신호의 연구결과물을 부분적으로 이용하고 있다. 음성신호 분석과 마찬가지로 음악 신호 역시 FFT 분석에 기반하여 처리한다. 처리 함수들은 주파수성분의 평균, 상관관계에 기반을 두고 있으며 음악의 질감을 수치적으로 표현해 준다. 따라서 분류하고자 하는 악기의 통계적 특성을 구해서 타 신호와의 임계점을 찾거나 통계적거리를 측정하여 임계점을 설정한다. 본 연구에서는 국악기를 주파수분포를 구하여 다양한 특성값을 추출한 후 분류알고리즘을 적용하여 악기들을 분류하였다. 이후의 연구 진행은 독주 뿐 아니라 사물놀이, 영산회상과 같은 합주도 주파수분석하여 장르 구분을 실험할 것이며 합주에는 다양한 악기가 포함 되므로 위에 언급한 여러 가지 주파수기법들을 포괄하여 주파수분석을 수행한다.

4.1절에서는 국악기의 스펙트럼분석을 통해 악기 간 자동분류시스템을 구현하는 것을 최종목표로 하여 이에 대한 선행 연구로서 가야금독주의

주파수 분석을 통해 그 가능성을 확인하였다. 4.2절에서는 국악장르를 악기구성과 연주형태를 기반으로 하여 4개의 장르로 묶고 본 논문의 연구결과들을 적용하여 장르분류실험을 하였다.

## 4.1 한국전통악기의 분류

데이터파일에서 추출할 수 있는 두 종류의 특성값이 있는데 첫 번째는 물리적인 특성값, 두 번째는 인지적인 특성값이 있다. 물리적인 특성값으로는 에너지함수, 주파수 분포, 기본 주파수(fundamental frequency) 등이 있으며 인지적인 특성값으로는 주관적인 요소로서 신호의 크기(loudness), 음고(pitch), 질감(timbre), 리듬(rhythm)등이 있다. 본 논문에서는 물리적인 특성값인 영교차율과 균형주파수를 사용하여 기본적인 분류시스템을 구현하였다. 주파수패턴이 복잡한 신호, 다양한 악기가 협연하는 협주곡의 경우엔 인지적인 특성값을 도입하여 분류시스템의 성능을 높여야 할 것이다.

### 4.1.1 악기신호의 분석

본 논문에서는 주파수영역의 데이터를 이용하여 시스템을 구성하였다. 시간영역에서의 시스템 구현은 데이터별 녹음크기가 다를 뿐 아니라 잡음에 큰 간섭을 받게 되므로 강인한 시스템을 구현이 어렵다. 이러한 이유로 음성인식 분야에서도 주파수영역의 분석을 기본으로 하고 있다.

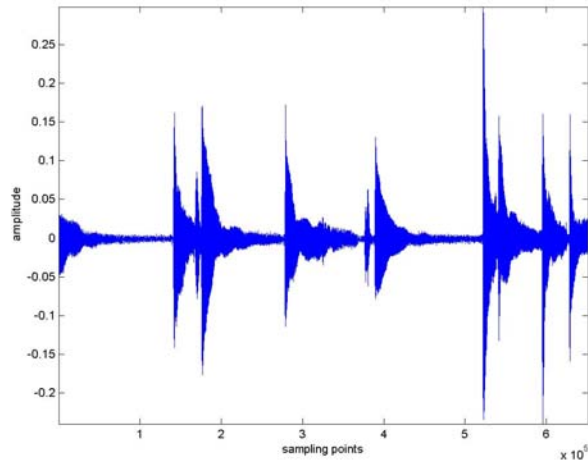
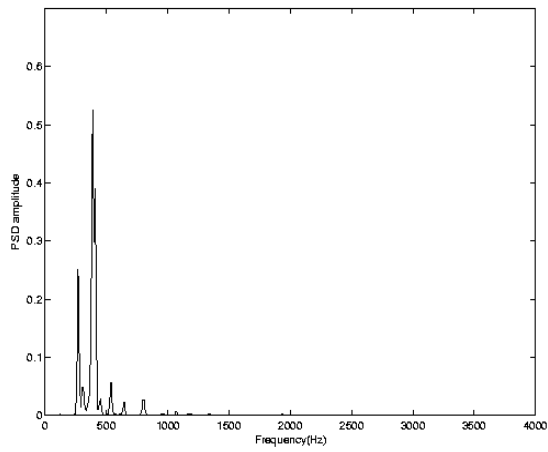


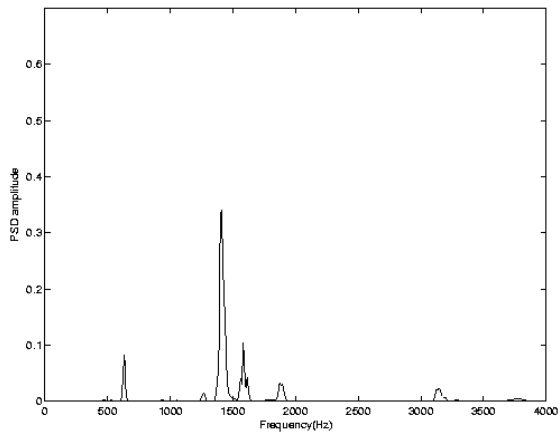
그림 4.1 시간영역에서의 가야금 신호.

Fig. 4.1 Kayagum signal in time domain.

각각의 데이터샘플(wav file)은 15초 정도의 길이를 가지고 있으며 CD, MP3 file로부터 추출하였다. 이를 FFT 분석하여 주파수 스펙트럼을 구한 후 이를 바탕으로 특성값들을 구한다. 음악신호의 FFT분석시에는 왜곡을 최소화 하기위해 샘플링 주파수를 충분히 크게 하였다. FFT분석의 안정성을 위해 직각 창함수(rectangular window) 대신 창함수 양끝의 포락선이 완만한 hanning 창함수를 사용하였다. DFT(Discrete Fourier Transform)알고리즘을 사용하는 FFT에서 원신호와의 차이를 줄이기 위해 창함수를 50% 중첩하였으며 256point의 FFT를 수행하여 충분한 해상도를 확보함으로써 원신호와의 유사성을 최대한 유지하고자 했다.



(a)



(b)

그림 4.2 (a) 가야금의 PSD 분포 (b) 해금의 PSD 분포.

**Fig. 4.2** (a) PSD distribution of kayagum (b) PSD distribution of Haegum.

그림 4.3(a)에서 가야금은 가청주파수의 1KHz 이하에 에너지가 집중되어 있고 그림 4.3(b) 해금의 경우 1KHz 이상의 주파수에 에너지가 집중하게 된다. 사용된 데이터는 44100Hz, 16bit 웨이브파일이다. 이와 같이 음역대가 다른 악기들 간의 구분은 중심주파수 값만으로 도 충분히 해낼 수 있다. 그러나 악기의 음역대가 비슷할 경우 또는 다양한 악기가 사용되는 합주곡일 경우에는 중심주파수 특성값만으로는 충분하지 않고 주파수특성에 적합한 함수를 추가적으로 적용해서 특성값을 추출해야 한다. 이 과정의 반복 수행을 통해 특성값의 임계값을 결정 한 뒤 이를 분류의 기준으로 삼는다. 장고는 낮은 음역대에 주파수성분이 분포되어 있고 피리는 넓은 주파수 분포를 보인다. 장고의 중심주파수는 579.6 Hz이고 피리는 3114.4 Hz이다. 예상과 같이 고음역의 악기인 피리의 중심주파수는 매우 높은 값을 보였다.

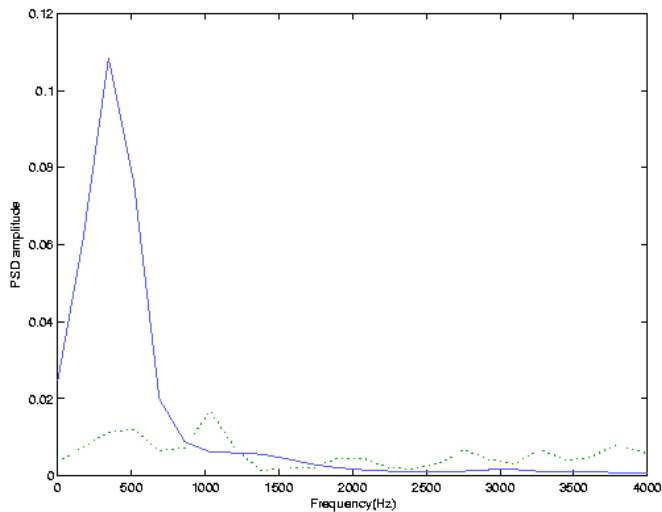


그림 4.3 장고와 피리의 PSD 분포 비교. 실선: 장고, 점선: 피리.  
**Fig. 4.3** Comparison of PSD distribution between Jango and Piri: solid line ; Jango, dotted line ; Piri.

#### 4.1.2 특성값과 분류알고리즘

음악신호의 특성을 분석하는 다양한 처리함수들이 존재한다. 대부분이 음성처리에서 차용한 것으로 음악신호에 적합한 처리함수들을 나열해 보면 다음과 같다.

- 주파수 밀도(PSD; Power Spectral Density)

Power spectral density(PSD)는 주파수별 응답의 밀도를 관찰하여 국악기의 주파수 분포 상태를 확인할 수 있다.

- 주파수 세기(RMS of frequency amplitude)

악기의 음량을 측정하기 위해 사용된다. 그러나 녹음환경, 장비의 특성에 따라 악기 소리의 크기는 가변적이므로 RMS 값을 규준화 하여 사용한다.

- 낮은에너지의 백분율(LEF; percentage of low energy)

PSD 최고값의 50% 이하인 프레임(frame) 비율을 나타낸다. 고음역이 많은 악기의 경우 저음악기보다 더 작은 값을 나타낸다. 여기서 %비율은 목적에 맞게 임의로 조정할 수 있다.

- **균형 주파수(spectral centroid)**

주파수분포의 균형값으로 식(4.1)로 계산한다. 악기의 음역대에 따라 특성값은 달라진다[39].

$$SC = \frac{\sum_k k A_k}{\sum_k A_k} \quad (4.1)$$

$A_k$ 는 각각의 주파수의 소리크기를 나타낸다.

- **스펙트랄 플럭스(SF; Spectral Flux)**

SF는 인접신호 프레임사이의 거리를 측정한 것으로 아래의 식으로 계산한다. 프레임간 주파수분포 변화를 계산한다.  $N_t$  값은 프레임  $t$ 에서 규준화 된 값이다.

$$SF_t = \sum_{n=1}^N (N_t[n] - N_{t-1}[n-1])^2 \quad (4.2)$$

- **영교차율**

위의 처리함수들은 크게 주파수성분의 크기와 분포에 기반을 두고 있다. 위 처리함수들을 크게 두부류로 PSD, RMS, LEF 와 균형주파수, SF, 영교차율 로 나눌 수 있다. 따라서 균형주파수와 PSD 함수만으로도 충분



히 악기를 분류해 낼 수 있었으나 더욱 다양한 악기와 합주곡 분류를 위해선 더 많은 함수들이 필요할 것으로 판단된다.

본 실험에서는 k-최근접이웃 분류알고리즘을 적용하여 분류하였다. 주어진 데이터 셋  $(x_i, y_i)$ 을 이용하여  $x$ 에 대한  $y$ 의 값을 예측한다.  $x$ 에 가까운  $x_i$ 의  $k$ 값을 이용하여  $y$ 값을 예측한다. 여기서  $k$ 는 일종의 계수로서 가장 영향이 큰 요인에 가중치를 두어 예측할 수 있다. 예측과정은 다음과 같다. 학습 데이터 셋은 분류대상의 분류에 사용될 계수를 생성하는데 사용한다. 즉, 분류시스템에는 분류대상이 있고 이를 예측하기 위한 각 음악장르의 특징(음의 질감, 에너지, 리듬 등) 들이 존재한다. 분류대상과 가장 가까운 학습 데이터 셋의  $k$  계수를 구한다. 그리고 유클리디안 거리(Euclidean distance) 측정법을 사용하여 분류대상과 학습 데이터의 거리를 구한다. 이러한 과정을 나머지 분류대상 데이터에 대하여 반복한다.  $k$  계수가 많아질수록 계산시간은 길어지지만 비이상적인 데이터의 악영향을 감소시킬 수 있다. 실제로는 수십개 이하의  $k$  값을 사용한다.

## 4.2 악기분류 실험

데이터파일을 반복적으로 처리하여 평균, 분산, 영점교차율, 중심주파수, 평균피크값, 군의 분포벡터를 얻었다. 각 악기의 데이터는 30개이며 총 90개의 데이터를 반복적으로 실행하여 특성값을 구하였다. 실험순서는 그림 4.4와 같다. 데이터로는 국악기인 대금, 가야금과 서양악기인 바이올린의 표본을 사용하였다. 한 개의 표본마다 6개의 특성값을 추출하였다. 데이터사이의 벡터거리를 측정하여 분류하는 k-최근접이웃분류 알고리즘을 적용하였다. 분류시스템은 10분 교차 훈련법(ten-fold cross-validation)을 사용하여 하였다. 여기서 교차훈련(cross-validation)은 패턴인식에서 사용하는 일반적인 훈련기법으로 데이터집합은 같은 크기의  $n$ 부분으로 나뉘어지고  $(n-1)$ 부분은 분류시스템을 학습시키기 위해 사용한다. 나머지  $n$ 번째 부분은 학습된 시스템을 시험하기 위해 사용한다. 표 4.1은 악기음원파일로부터 얻은 특성값의 예이다. 표 4.2는 분류 실험결과로서 분류성공률 94.44%의 결과를 얻었다. 세로항은 실제 입력된 악기이고 가로축은 분류결과를 나타낸다.

**표 4.1** 추출된 특성값의 예  
**Table 4.1** Example of feature values

평균	분산	영교차율	중심주파수	평균피크값	군
-0.001140,	0.000371,	4630.000000,	0.030745 ,	496.201280,	gayagum
-0.001312,	0.000351,	4154.000000,	0.030842 ,	545.832980,	gayagum
-0.001577,	0.000312,	6142.000000,	0.030981 ,	561.439203,	gayagum
			.		
			.		
0.000002,	0.000758,	8843.000000,	0.044950 ,	518.819700,	taegum
0.000003,	0.002055,	13460.000000,	0.076172 ,	733.871591,	taegum
0.000002,	0.001010,	12656.000000,	0.054974 ,	696.502076,	taegum
			.		
			.		
-0.000002,	0.003485,	23154.000000,	0.146585 ,	861.480532,	violin
0.000000,	0.014886,	39109.000000,	0.298086 ,	976.078632,	violin
-0.000000,	0.008804,	25762.000000,	0.194512 ,	875.031560,	violin
			.		
			.		

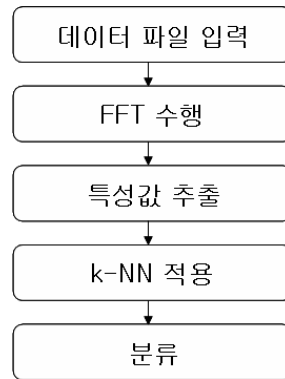


그림 4.4 실험순서.

Fig. 4.4 Flow-Chart of experiment.

표 4.2 분류 실험결과, 분류성공률 94.44%

Table 4.2 Experiment result of Classification, Success rate 94.44%

구분	가야금 출력	대금 출력	바이올린 출력
가야금	30	0	0
대금	0	26	4
바이올린	0	1	29

**표 4.3** 군별 세부 정확도  
**Table 4.3** Detailed accuracy by class

구분	TP rate	FP rate	Precision	Recall	F-Measure
가야금	1	0	1	1	1
대금	0.87	0.017	0.96	0.87	0.91
바이올린	0.97	0.067	0.88	0.97	0.92

**표 4.4** 전체 정확도  
**Table 4.4** Entire accuracy

Correctly Classified Instances	85
Incorrectly Classified Instances	5
Kappa statistic	0.92
Mean absolute error	0.052
Root mean squared error	0.19
Relative absolute error	11.61%
Root relative squared error	40.25 %
Total Number of Instances	90

실험의 객관성을 보장하고자 가급적 데이터 파일은 다양한 빠르기, 다양한 음역을 가지고 있는 연주파일들을 사용했으나 일부 음역대만 사용하는 극단적인 연주는 제외하였다. 90개의 데이터 중에서 85개의 데이터가 정확하게 분류되었으며 실험의 분류 성공률은 94.44% 로서 만족스런 결과를 얻었다. 잘못 분류된 5개의 데이터 중에서 4개는 대금을 가야금으로 분류한 경우이고 나머지 1개의 오류도 바이올린을 대금으로 분류한 것이다.

### 4.3 한국전통음악의 장르 분류 실험

장르는 종류 또는 계급을 뜻하는 라틴어 제누스에서 유래했다. 이는 현대 언어로 범위, 범주, 주제, 구조 등으로 번역될 수 있다. 그러나 장르의 의미는 객관적 사실에 근거하기 보다는 변화라는 사회, 문화의 과정에서 의미를 찾아야 할 것이다.

*장르는 반복되는 특정한 소통상황에서 개인, 사회, 기술의 복합으로 발생한 의사소통의 한 유형이다. 장르는 상호작용에 있어 어떤 유형에 대한 공통의 기대를 제 공함으로써 의사소통을 성립시킨다[1].*

장르의 구분은 개인 및 사회문화적 측면에서 모두 주관적이다. 본 논문에서는 일반적인 청취자가 음악을 청취함에 있어 다른 음악과 구별할 수 있는 공통의 특징을 구별해 낼 수 있는 특정음악의 범주로 정의한다. 특정장르는 해당 음악에 사용되는 악기와 리듬구조에 의해 구별할 수 있다. 그러나 이 외에도 인간은 다른 여러 가지 요소들을 사용하여 장르를 구분한다. 악기 또는 장르의 자동구분을 위해서는 이러한 요소의 파악이 필수적이다.

국악장르역시 단순한 악곡형식이나 구조로 장르구분하기 보다는 사회문화적인 요소를 고려해야 할 것이다. 국악은 계속변화 하고 있고 우리가 알고 있는 전통음악도 조선시대의 흔적이 가장 강하게 남아있다. 본 논문에서는 전통적인 장르 구분을 이용하여 유사한 음향특성을 갖는 경우 하나의 장르로 묶었다.

### 4.3.1 한국전통음악의 장르

한국의 전통음악은 크게 민속음악과 궁중음악으로 구분할 수 있다. 궁중음악은 연주의 목적에 따라 제례음악, 연희음악, 군대음악(취타)로 세분된다. 민속음악은 기악(정악, 산조, 농악)과 성악(판소리, 시조, 가사, 민요, 가곡)으로 세분된다. 또한 불교적인 전통이나 무속적인 전통과 같은 종교음악과 춤도 포함하고 있다. 음악양식의 일종인 시나위는 기악으로 연주되지만 목소리로 연주하는 구음시나위도 존재한다. 따라서 본 실험에서는 악기구성을 기준으로 장르를 구분하였다. 성악과 기악으로 나누고 기악은 독주/산조와 관현악으로 나누었다. 기악중에서도 농악과 사물놀이는 관악기와 타악기가 주로 사용되는 특징이 있으므로 따로 구분하였다.

- 판소리 : 사람의 소리꾼이 고수(鼓手)의 북 반주에 맞춰 극적(劇的)으로 구성된 긴 이야기를 '소리'와 '아니리'(말)와 '발림'(몸짓)을 통해 전달하는 우리나라 고유의 전통공연예술이다.
- 농악 : 풍물굿의 기원은 인류가 농경생활을 시작하면서부터 함께 했을 것이라고 본다. 풍물의 원형과 당시 사회적 역할을 고찰하는데 가장 보편적인 사료는 진수가 쓴 중국의 『삼국지』 위지동이전이다. 이 문헌에 의하면 마한에서는 "오월에 씨뿌리기를 마친 다음 귀신에게 제사하고 무리지어 노래하고 춤추고 술마시기를 밤낮으로 쉬지 않았다"고 하며 타악기 위주의 연주이다.
- 민요 : 인정, 풍속, 생활 감정 따위를 나타내어 민간에 전하여 오는 순박한 노래이다. 민요는 특별할 재주나 기교가 없이도 이 땅의 사람이 라면 누구라도 만들고 부를 수 있었으며, 그것을 즐기는데도 특별한

격식이 필요한 것이 아니었다.

- 무악(巫樂) : 전통적인 토속신앙은 굿이라는 의례를 통하여 구체화되는데, 종교와 음악의 결합체를 굿이라는 의식을 통해서 이를 확인할 수 있다. 이러한 현상은 멀리 상고시대의 가무전통에서부터 비롯되며, 그 맥을 잇고 있는 굿이라는 의식은 지금도 전국 각처에서 전승되고 있다.
- 산조 : 대금, 피리, 해금, 아쟁, 가야금, 거문고 등과 같이 주요 국악기는 모두 산조라는 연주 영역을 가지고 있다. 산조는 악장의 구분이 곧 장단의 구별과 상통한다. 즉, 장단의 변화에 따라 악장도 변한다. 산조에서의 장단 변화는, 느린 것에서 빠른 것으로, 즉 느린 장단에서 점차 빠른 장단으로의 변화하는 것을 말한다.
- 사물놀이 : 실내에서 한 자리에 붙박이로 앉아서 사물의 리듬을 정교하고 체계적으로 보여줌으로써 우리 민속 장단의 극단적인 아름다움과 신명을 느끼게 하는 놀이이다. 풍물굿은 가락에 있어서 맺고 푸는 가락이 반복하는 구조로 되어 있는 반면, 앞서 밝힌 바와 같이 사물놀이의 가락은 느린 것에서 빠른 것으로 이행하는 점층적 가속의 틀을 지니고 있다.
- 시조 : 초·중·종장의 사설을 3장 형식에 엮고 한 장단이 5박 또는 8박으로 되어 있는 장단에 맞춰 부르는 노래를 말한다. 18세기 중엽에는 시조를 시절가 또는 시절단가(詩節短歌)라고 불리었다.
- 줄풍류 : 줄(絲) 즉, 현악기 중에서도 거문고가 중심이 되는 풍류를 말한다. 가야금, 거문고 등 현악기가 편성 연주되는 점에서 현악영산회상



또는 줄풍류라 지칭하기도 하는데 이 현악영산회상은 율방을 중심으로 연주되던 소규모의 실내합주음악이며, 조선조 중인 선비계층 등 음악 애호가들이 즐기던 음악이기도 했다.

- 제례악 : 대표적인 제례악으로는 종묘제례악과 문묘제례악이 있다. 종묘제례악은 조선 역대 임금을 모신 사당인 종묘의 제향에 연주되는 음악이다. 세종 17년(1435) 중국계 고취악과 향악을 참작하여 제정, 세조 9년(1464)에 개작되어 종묘제향에 연주되어 왔다. 연주는 편경, 편종을 비롯한 팔음의 모든 악기가 동원된다. 문묘제례악은 문묘 제향에 연주되는 음악으로 문묘란 문선왕묘의 준말이다. 문선왕은 공자를 말하는 데, 문묘는 그의 제자인 안자, 증자, 지사, 맹자와 우리 나라에서 학덕이 높은 설총, 최치원과 같은 분을 모신 사당이다. 고려 예종 때 송나라에서 들어왔고, 세종 때 박연, 맹사성 등이 중국 역대 문헌을 참작하여 주 시대에 가까운 아악으로 복고, 개편했다.
- 여민락(與民樂) : 전체 125장의 <용비어천가>의 일부(1.2.3.4.125장)를 노래하기 위해 만들어졌지만 시대를 거치면서, 현재는 가사 없이 관현악곡으로 연주되고 있다.
- 수제천(壽齊天) : 악기는 당적, 대금, 향피리, 해금, 아쟁, 장구, 좌고, 박 등으로 편성되며, 처용무를 반주할 때는 삼현육각(향피리 2·대금 1·해금 1·장구 1·북 1 등으로 악기를 편성하는 방법으로 춤이나 굿노래의 반주 음악에 많이 사용)으로 연주하기도 한다. 속도가 느리고 장중하며, 우아함이 비길 데 없는 아악곡의 백미(白眉)로 손꼽힌다[40,41].

### 4.3.2 장르 구분 실험

본 실험에서는 유사한 악기구성을 사용하는 음악을 하나의 장르로 구분하였다. 연주형태에 따라 표 4.5과 같이 4개의 장르로 구분하여 실험하였다. 농악/사물놀이/무악은 타악기 중심의 음악으로 징, 꽁가리, 날나리, 북 등의 타악기가 주로 연주된다. 창/병창/시조는 사람의 목소리가 추가되는 성악으로 판소리와 시조의 경우 성악 독주 또는 고수의 북이 협연하는 형태이다. 병창은 가야금과 성악 또는 거문고와 성악과 같이 소수의 음율악기와 협연하는 형태이다. 독주/산조는 순수한 기악으로 서 가야금, 아쟁, 거문고, 해금 등의 독주가 널리 연주된다. 산조는 독주연주에 북과 같은 타악기가 협연하는 형태이다. 종묘제례악/실내합주/줄풍류는 현악기가 추가되는 기악협주곡이다.

**표 4.5** 연주형태에 따른 한국전통음악의 장르구분

**Table 4.5** Classification of Korean traditional music genre by way of playing style

구분	내용
타악	농악/사물놀이/무악
성악	판소리/병창/시조/민요
실내악	독주/산조
현악합주	제례악/관현악/줄풍류

**표 4.6** k-최근접이웃법 실험결과, 분류성공률 95.4%

**Table 4.6** Experiment result of k-NN, success rate 95.4%

구분	타악 출력	성악 출력	실내악 출력	현악합주 출력
타악 입력	189	4	4	3
성악 입력	7	184	5	4
실내악 입력	3	7	189	1
현악합주 입력	0	6	0	194

**표 4.7** k-NN, 군별 세부 정확도

**Table 4.7** NN, Detailed accuracy by class

구분	TP rate	FP rate	Precision	Recall	F-Measure
타악	0.94	0.017	0.95	0.94	0.95
성악	0.92	0.028	0.91	0.92	0.92
실내악	0.94	0.015	0.95	0.94	0.95
현악합주	0.97	0.013	0.96	0.97	0.96

**표 4.8** 전체 정확도  
**Table 4.8** Entire accuracy

구분	k-NN
Correctly Classified Instances	94.5%
Incorrectly Classified Instances	5.5%
Kappa statistic	0.93
Mean absolute error	0.03
Root mean squared error	0.16
Relative absolute error	7.84%
Root relative squared error	38.19%
Total Number of Instances	800

800개의 샘플데이터 중에서 잘못 분류된 44개의 데이터는 고르게 분포되어 있어 악기특성에 따른 장르분류는 유효하였다. 756개의 데이터가 정확하게 분류되어 실험의 분류 성공률은 94.5%이다.

## 4.4 실험결과 및 해석

4.2절의 국악기 분류실험에서 94.44%의 악기분류 성공률을 얻었다. 여기에서 가야금은 데이터전체가 정확히 분류되었고 대금과 바이올린에서 오류가 발생하였다. 탄현악기로서 음향적인 특성이 찰현악기인 바이올린, 취주악기인 대금과 구분되기 때문인 것으로 판단된다. 가야금의 음향은 단속적인 특성을 가지고 있으며 바이올린과 대금은 일정시간 음향이 지속되는 특성을 가짐으로서 바이올린과 대금에서 오류가 발생한 것으로 판단된다. 4.3절의 한국전통음악 장르분류 실험에서 사용된 음원파일의 표본주파수 44.1KHz이고 평균, 분산, 영교차율, 중심주파수, 평균피크값, 균형주파수, 스펙트랄 플릭스, 스펙트랄 롤-오프 8개의 특성값을 추출하여 분류 실험을 하였다. 성공률은 95.4%이고 오류는 골고루 분포되어 있다.

## 제5장 결 론

방대한 양의 멀티미디어 데이터를 처리하는 것은 점점 중요해지고 있다. 모든 데이터를 일일이 확인하고 인덱싱하는 것은 시간과 비용에 있어 적합하지 않다. 따라서 여러 가지 분류기법들이 소개되고 있다. 대부분은 음성처리 기법을 이용하고 있으며 데이터들의 유사한 특성을 지표로 삼아서 분류하고 있다. 데이터의 인지적, 음향적 유사성을 이용하여 유사한 데이터를 분류할 수 있다. 본 논문은 음향데이터 분류 그 중에서도 한국 전통음악 분류에 초점을 맞추어 분류실험을 하였다.

본 논문은 궁극적으로 인간의 소리인식방법과 유사한 시스템을 모델링하는데 목적이 있다. 복잡한 소리신호를 인지하고 특징적인 부분을 수치화하여 소리를 구분하는 시스템을 구현하고자 한다. 현재까지 이러한 연구는 인접연구분야인 음성인식분야의 연구성과들을 인용하여 전개되고 있다. 서양악기와 장르의 경우 다양한 연구결과들이 보고되고 있으나 동양의 전통악기 분야에선 연구가 미진하다. 한국 전통악기의 인식방법연구는 형태와 음역에서 유사한 중국, 일본악기의 연구에도 공통적으로 적용할 수 있을 것이라 기대한다.

미디어파일의 일부분을 잘라내고 이를 기반으로 푸리에변환을 행하여 주파수성분을 모두 구한 후 주파수성분의 분포관계 즉, 주파수분포평균, 가청 주파수 이내의 고주파와 저주파의 비, 통계적특성, 인접 주파수와의 상관관계를 함수화한 후 특정 데이터의 특징을 추출하여 시스템을 구현방법에 대하여 논하였으며 신뢰성 있는 자동인식, 분류시스템이 구현된다면 학술, 문화, 예술 등 다양한 분야에 적용가능하다.

제2장에서는 음악인식에 필요한 핵심적인 특성들에 대하여 연구하였다. 주파수영역에서의 음악신호는 시간영역에서 파악할 수 없었던 많은 정보

들을 보여준다. 소리신호의 주파수영역의 표현부터 시작하여 스펙트로그램을 이용하여 대표적인 장르의 주파수분포를 살펴보았다. 주파수영역에서의 장르별 특징은 어떤 특성값들을 사용하여 음악인식에 접근하여야 하는지 분명히 보여주었다. 국악의 경우 불규칙한 리듬과 비트의 세기가 일정하지 않았다. 클래식 경우 일정 주파수영역에 에너지가 집중되어 있었고 리듬패턴이 일정치 않았다. 재즈의 경우 전대역에 걸쳐 고른 주파수 분포를 보였고 리듬패턴이 매우 일정하였다.

특정주파수대에 민감한 청각시스템을 모델링한 MFCCs, 음성과 비음성 신호 구분에 유용한 ZCR, 주파수 에너지분포를 계산하는 short-time energy 등의 특성값들은 스펙트로그램이 보여준 장르별 특징을 파악할 수 있는 함수들로서 음악인식시스템에 유용하게 적용될 수 있다. 음악인식에 있어 푸리에 변환된 결과를 이용하여 음악의 질감을 수치화 할 수 있는 함수들을 이용하여 영교차율, PSD(Power Spectral Density), 리듬, 에너지 등의 특성값을 추출하여 시스템을 구현하였다. 이러한 특성값을 구하는데 있어 사용한 데이터는 선형적인 푸리에 변환값을 사용하였다.

푸리에 변환시에 기저막의 주파수분해 특성과 내유모세포의 응답 특성을 적용하여 음악인식시스템을 구현에 관하여 연구하였다. 기저막의 아래쪽 채널은 낮은 주파수분해를 담당하고 있으며 상대적으로 응답특성이 느리다. 반면 위쪽 채널의 경우 높은 주파수분해를 담당하고 있으며 낮은 주파수에 비해 응답특성이 빠르다. 기존의 특성값 추출을 위해서 사용하던 주파수분해 특성은 선형적이다. 본 논문에서는 특성값 추출에 사용되는 데이터추출에 있어서 비선형적인 청각모델을 적용하여 추출한다. 음악인식에 있어 핵심적인 데이터를 인간인식특성에 기반한 청각모델을 도입함으로써 한 차원 높은 인식 시스템의 구현이 가능하다.

제3장에서는 소리신호 분류에 가장 많이 사용되는 분류알고리즘을 비교하

였다. 실험결과 최근접이웃(NN) 알고리즘이 분류성공률 100%로서 음악인식에 있어 빠른 계산과 가장 좋은 성공률을 보였다. k-최근접이웃(k-NN) 알고리즘은 85% 성공률로 가장 낮은 결과를 보였다. 오류는 대금과 아쟁에서 발생하였다. 대금과 아쟁의 음역대가 유사하고 일정시간동안 발성이 지속된다는 특징을 가지고 있다. 반면 거문고는 탄현악기로서 아쟁과 거문고에 비해 단속적인 특성을 갖는다. 음악 파일로부터 추출한 특성값은 2차원 벡터상의 위치이므로 순수한 확률과정법보다는 벡터거리를 적용한 NN 알고리즘의 분류결과가 우수한 것으로 확인되었다. 서양악기 실험에서는 한국전통악기와의 비교를 위하여 같은 방법으로 바이올린과 첼로에 대하여 실험하였다. k-NN 알고리즘이 분류성공률 89.58%로서 악기인식에 있어 빠른 계산과 가장 좋은 성공률을 보였다. 베이스법은 77.08% 성공률로 가장 낮은 결과를 보였다. 서양악기에서도 전체적으로 벡터거리 알고리즘을 사용한 k-NN과 NN 알고리즘이 좋은 성능을 보였다.

특성값벡터를 추출하기 위해 롤-오프, 분산, 평균피크레벨을 사용하였다. 이중에서 롤-오프는 저음프레임과 고음프레임의 상대적인 비를 나타낸다. 최적의 롤-오프 포인트를 찾기 위하여 롤-오프 포인트를 0.05에서 0.9까지 0.05간격으로 증가 시키며 반복실험 하였다. 롤-오프 포인트를 증가시키며 분류성공률을 관찰하였다. 그리고 실험에 사용된 음원데이터는 바로크바이올린과 현대바이올린 연주이다. 두 종류의 악기는 모양과 주파수대역에 있어서 유사하지만 약간의 대역차와 질감의 차이를 가지고 있다. 이러한 특성이 최적의 롤-오프 포인트를 찾는 데 유용할 것으로 판단하였다. 실험결과 롤-오프 포인트 0.85에서 가장 높은 분류성공률 85%를 나타냈다.

제4장에서는 한국전통악기와 서양악기의 구분을 실험하기 위해 가야금, 대금, 바이올린으로 구성된 90개의 데이터를 대상으로 실험하여 85개의 데이터가 정확하게 분류되었으며 실험의 분류 성공률은 94.44% 로서 만족스



린 결과를 얻었다. 잘못 분류된 5개의 데이터 중에서 4개는 대금을 가야금으로 분류한 경우이고 나머지 1개의 오류도 바이올린을 대금으로 분류한 것이다. 가야금은 탄현악기로서 음향적인 특성이 찰현악기인 바이올린, 취주악기인 대금과 구분되기 때문인 것으로 판단된다. 가야금의 음향은 단속적인 특성을 가지고 있으며 바이올린과 대금은 일정시간 음향이 지속되는 특성을 가짐으로서 바이올린과 대금에서 오류가 발생한 것으로 판단된다.

본격적인 장르구분실험으로서 다양한 한국전통음악 장르 데이터를 대상으로 음악의 질감을 수치화 할 수 있는 함수들을 이용하여 영교차율, PSD, 리듬, 에너지등의 특성값을 추출하여 시스템을 구현하였고 87% 정도의 성공률을 보였다. 본 실험은 푸리에 변환시에 기저막의 주파수분해 특성과 내유모세포의 응답 특성을 적용하여 음악인식시스템을 구현하였다. 실험에 사용된 음원파일의 표본주파수 44.1KHz로 높이고 평균, 분산, 영교차율, 중심주파수, 평균피크값, 균형주파수, 스펙트랄 플렉스, 스펙트랄 롤-오프 8개의 특성값을 추출하여 분류시스템을 구현하였다. 800개의 샘플 데이터 중에서 잘못 분류된 44개의 데이터는 고르게 분포되어 있어 악기 특성에 따른 장르분류는 유효하였다. 756개의 데이터가 정확하게 분류되어 실험의 94.5%의 분류 성공률을 얻었다.

음악장르의 분류를 위해서는 음악을 구성하는 악기와 음악적인 차이점을 구분하는 것이 중요하다. 본 실험은 녹음이 비교적 잘된 데이터들을 이용하였으나 라디오, 공연장녹음등과 같은 노이즈간섭에 강한 분류시스템을 구축하는 것도 향후 과제이다. 음악의 내용을 파악하여 구분할 수 있는 특성함수를 개발하는 것도 정확하고 빠른 장르구분의 핵심임을 확인하였다.

## 참고문헌

- [1] T. Erickson, "Rhyme and punishment: The creation and enforcement of convention in an on-line participatory limerick genre," In Proceedings of the 32nd hawaii International Conference on System Sciences (HISS '99), 1999.
- [2] D. Perrot, and R. O. Gjerdingen, "Scanning the dial : An exploration of factors in the identification of musical style," In Proceedings of the 1999 Society for Music Perception and Cognition, 1999.
- [3] Jonathan foote, "An overview of audio information retrieval," In Multimedia Storage and Achieving Systems II, Proceedings of SPIE Corpora, pp.138-147, 1997.
- [4] Jonathan Foote, "A similarity measure for automatic audio classification," In Proc. AAAI 1997 Spring Symposium on Intelligent Integration and Use of Text, Image, Video, and Audio Corpora, March 1997.
- [5] T. Zhang, and C. Kuo, "Content-based classification and retrieval of audio," In SPIE's 43rd Annual Meeting - Conference of Advanced Signal Processing Algorithms, Architectures, and Implementations VIII, San Diego, July 1998.
- [6] T. Zhang, and C. Kuo, "Hierarchical system for content-based audio classification and retrieval," Technical report, University of Southern California, Los Angeles, CA, 1998.
- [7] Evritt BS, "Moments of the statistics kappa and the weighted kappa," British J. Math. Statist. Psych., 21, pp.97-103, 1968.

- [8] Z. Liu, Y. Wang, and T. Chen, "Audio feature extraction and analyze for scene segmentation and classification," *Journal of VLSI Signal Processing System*, 20:61-79, 1998.
- [9] Asif Ghias, Jonathan Logan, David Chamberlin, and Brian C. Smith, "Query by humming: Musical information retrieval in an audio database," In *ACM Multimedia*, pp.231-236, 1995.
- [10] S. R. Subramanya, Rahul Simba, Bhagirath Narahari, and Abdou Youssef, "Transform-based indexing of audio data for multimedia databases," In *international Conference on Multimedia Computing and Systems*, pp.211-218, 1997.
- [11] Jeremy Pickens, "A survey of feature selection techniques for music information retrieval," Technical report, Center for Intelligent Information Retrieval, University of Massachusetts, Amherst, MA, 2001.
- [12] R. McNab, L. Smith, and I. Witten, "Signal processing for melody transcription," In *Proc., 19th Australasian Computer Science Conf.*, pp.301-307, Melbourne, January 1996.
- [13] Keith D. Martin, "Automatic transcription of simple polyphonic music: Robust front-end processing," Technical report 399, Massachusetts Institute of Technology, The Media Laboratory, December 1996.
- [14] Daniel P. W. Ellis, "Prediction-Driven computational auditory scene analysis for dense sound mixtures," Technical report, International Computer Science Institute, Berkeley CA, 1996.
- [15] Hrishikesh Deshpande, Unjung Nam, and Rohit Singh, "Mugec : Automatic music genre classification," Technical report, Stanford University, June 2001.

- [16] Colin Campbell, Surface Acoustic Wave Devices and Their Signal Processing Applications, Academic Press, 1989.
- [17] V. Oppenheim, Ronald W. Schafer, Discrete-Time Signal Processing, Prentice-Hall International, Inc., 1989.
- [18] Curtis Roads, The Computer Music tutorial, MIT Press, 1996.
- [19] Z. Liu, Y. Wang, and T. Chen, "Audio feature extraction and analysis for scene segmentation and classification," Journal of VLSI Signal Processing System, 20:61-79, 1998.
- [20] Simon Dixon, "A beat tracking system for audio signals," In Proceedings of the Diderot Forum on Mathematics and Music, Austrian Computer Society, pp.101-110, 1999.
- [21] Paul E. Allen, and Roger B. Dannenberg, "Tracking musical beats in real time," In S. Arnold and G. Hair (editor), ICMC glasgow 1990 Proc., International Computer Music Association, pp.140-143 1990.
- [22] Keith D. Martin, Eric D. Scheirer, and Barry L. Vercoe, "Music content analysis through models of audition," in Proc. 1998 ACM Multimedia Workshop on content processing of music for Multimedia Applications, Bristol., UK, september 1998.
- [23] von Békésy. G, Experiments in Hearing, New York: McGraw Hill, 1960.
- [24] Patterson. R. D, Holdsworth. J, A functional model of neural activity patterns and auditory image, In W. A. Ainsworth(ed.) Advances in speech, hearing and language processing. London: JAI Press, 1990.

- [25] Patterson. R. D, Moore, and B. C. J, "Auditory filters and excitation patterns as representations of frequency resolution," In B. C. J. Moore (ed.) Frequency Selectivity in Hearing. London: Academic, 1986.
- [26] Hewitt. M. J, and Meddis. R, "An evaluation of eight computer models of mammalian inner hair-cell function," J. Acoust. Soc., Am 90(2), pp.904-917, 1991.
- [27] G. Tzanetakis, and P. Cook, "Automatic musical genre classification of audio signals," in Proc. Int. Symp. Music Information Retrieval (ISMIR), Oct 2001.
- [28] Richard O. Duda, Peter E. Hart, and David G. Stork, Pattern Classification. John Wiley & Sons, 1999.
- [29] Anil K. Jain, and Robert P.W. Duin, and Jianchang Mao, "Statistical pattern recognition: A review," IEEE Transactions On Pattern Analysis and Machine Intelligence., 22(1):4-37, 1999.
- [30] L. R. Rabiner, and B. H. Juang, "An introduction to hidden Markov models," IEEE ASSP Magazine., 3(1):4-16, January 1986.
- [31] Richards O. Duda, Peter E. Hart, and David G. Stork, Pattern classification, John Wiley & Sons, Inc., New York 2001.
- [32] Stan Z. Li, "Content-based classification and retrieval of audio using the nearest feature line method," IEEE Transactions on speech and audio processing, 9(5):619-615, September 2000.
- [33] E. Wold, T. Blum, D. Keislar, and J. Wheaton, "Content-based classification, search and retrieval of audio," IEEE Multimedia, vol.3, no.2, 1996.

- [34] Bobick. A, and Richards, W, "Classifying Objects from Visual Information," Massachusetts Institute of Technology A.I. Memo No.879, 1986.
- [35] Fletcher. H, Blackham. E. D, and Strraton. R, "Quality of piano tones," J. Acoust. Soc., Am. 34(6), pp.749-761, June 1962.
- [36] Therrien. C. W, Decision, Estimation, and Classification, New York: Wiley, 1989.
- [37] J.C.R Licklider, A duplex theory of pitch perception, in Psychological Acoustics, E.E. Schubert, Dowden, Hutchinson and Ross, Inc., Stroudsburg, PA, 1979.
- [38] Slaney, Malcom, and Richard F. Lyon, "Perceptual Pitch Detector," Proc., ICASSP 90(Albuquerque, New Mexico 1990).
- [39] G. Tzanetakis and P. Cooke, "Multifeature audio segmentation for browsing and annotation," in Proc., Workshop Application of Signal Processing to Audio and Acoustics (WASPAA), New Paltz, NY, 1999.
- [40] 전통소리문화 홈페이지, <http://www.sori.jeonbuk.kr>.
- [41] 김인평, 국악감상, 중앙대학교출판부, 1993.
- [42] G. Tzanetakis, and P. Cook, "Automatic musical genre classification of audio signals," in Proc., Int. Symp. Music Information Retrieval (ISMIR), Oct 2001.
- [43] 김재천, 김석호, 광경섭, “특성함수 및 k-최근접이웃 알고리즘을 이용한 국악기 분류,” 멀티미디어학회 논문지 제9권 3호, pp.279-286, 2006년, 3월.

- [44] 김재천, 광경섭, “ 바이올린과 첼로 연주 데이터를 이용한 분류 알고리즘의 성능 비교,” 한국통신학회 논문지 제30권, 50C호, pp.305-312, 2005년, 5월.
- [45] Jae Chun Kim, Kyung Sup Kwak, "Determining Key Features of Recognition Korean Traditional Music using Spectrogram," J. Acoustical Society of Korea, vol. 24, No.2e, pp.67-30, June 2005.