DOI: 10.3745/KIPSTB.2009.16-B.2.141

SIDE를 이용한 자동 음악 채보 시스템

형 아 영[†] · 이 준 환[††]

요 약

본 논문에서는 사람의 노랫소리를 자동으로 채보할 수 있는 시스템을 제안한다. 먼저 입력된 음성으로부터 추출된 피치 정보를 안정화된 역확산 방정식(Stabilized Inverse Diffusion Equation : SIDE)을 이용하여 음절 단위로 분할한다. 이를 바탕으로 유전자 알고리즘에 기반한 클러스터링을 통해 음길이 인식을 수행하였다. 또한 시창자의 음 높이에 강인한 음정 인식을 위하여 상대 음정이라는 개념을 도입하였다. 그리고 휴지기 정보를 이용한 마디 추출 알고리즘을 적용하여 보다 정확한 노래의 채보를 가능하게 하였다. 제안된 시스템을 통하여 동요 16곡을 채보한 결과 마디 인식률은 91.5%였으며, DMOS 방법으로 측정한 악곡 전체 유사도는 3.82로써 시스템 성능의 유효성을 확인할 수 있었다.

키워드 : 안정화된 역 확산 방정식, 유전자 알고리즘, 마디 검출, 상대 음정

Automatic Music Transcription System Using SIDE

A-Young Hyoung[†] · Joon-whoan Lee[††]

ABSTRACT

This paper proposes a system that can automatically write singing voices to music notes. First, the system uses Stabilized Diffusion Equation(SIDE) to divide the song to a series of syllabic parts based on pitch detection. By the song segmentation, our method can recognize the sound length of each fragment through clustering based on genetic algorithm. Moreover, this study introduces a concept called 'Relative Interval' so as to recognize interval based on pitch of singer. And it also adopted measure extraction algorithm using pause data to implement the higher precision of song transcription. By the experiments using 16 nursery songs, it is shown that the measure recognition rate is 91.5% and DMOS score reaches 3.82. These findings demonstrate effectiveness of system performance.

Keywords : SIDE, Genetic Algorithm, Measure Detection, Relative Interval

1. 서 론

인류에 있어서 노래는 오래 전부터 존재해 왔던 하나의 문화 현상이며 개인과 사회 집단의 감정표현의 수단이자 유희의 도구였다. 노래는 음성의 범주에 속하며 발성 기관을 통해 표현되어 언어적인 모습을 지닌다는 점에서는 일반적인 말과 비슷하지만 음고(音高), 음량(音量), 음가(音價), 음색(音色) 등의 음악적 속성을 추가적으로 가지므로 그 차이가 있다. 이와 같이 사람의 노래 및 악기로 연주된 음악의 음정과 박자, 혹은 가사를 인식하거나 기존의 노래 자료와 입력된 노래의 비교를 통하여 곡명을 인식하는 등의 연구인 곡조 인식(music recognition)은 음성 인지 분야 중 하나로 노래의 각 특징량을 사용하여 최종적으로 원하는 데이터의 형태로 나타내 주는 것을 말한다.

자동 채보 시스템과 같은 곡조 인식은 노래를 통한 시창 교육 분야 및 여가 활동을 위한 엔터테인먼트 분야에 사용될 수 있다. 뿐만 아니라 최근에 중요시 되고 있는 노래의 저작권을 보호하기 위한 표절 검사의 도구로 사용 될 수도 있다. 그러나 다양한 정보 수단에 대한 컴퓨터의 처리 능력이 발전하고 있음에도 곡조 인식에 관한 연구는 현재까지 알려진 결과가 미미하다. 기존의 음악에 익숙한 전문가가 직접 노래를 듣고 채보하는 방법에 비하여 자동 음악 채보 시스템은 시창자(始唱者)의 노래가 가진 음악적 특징을 시스템이 자동으로 인식하여 일반인도 쉽게 노래를 악보화 할 수 있도록 한 것이다.

본 논문에서는 시창자 개개인의 음성 발성에 강인한 자동 음악 채보 시스템을 제안한다. 이는 사람의 음성 정보는 매우 애매하여 발성자의 성별, 나이, 감정 및 신체 상태, 성량 등에 따라 다양하고 같은 사람이 동일한 말을 발음하더라도 완벽하게 일치 하지 않을 만큼 많은 애매성을 지니기 때문이다.

따라서 제안된 시스템에서는 사람의 노래를 인식하기 위

† 준 회 원 : 전북대학교 컴퓨터공학과 석사과정
†† 정 회 원 : 전북대학교 전자정보공학부 교수

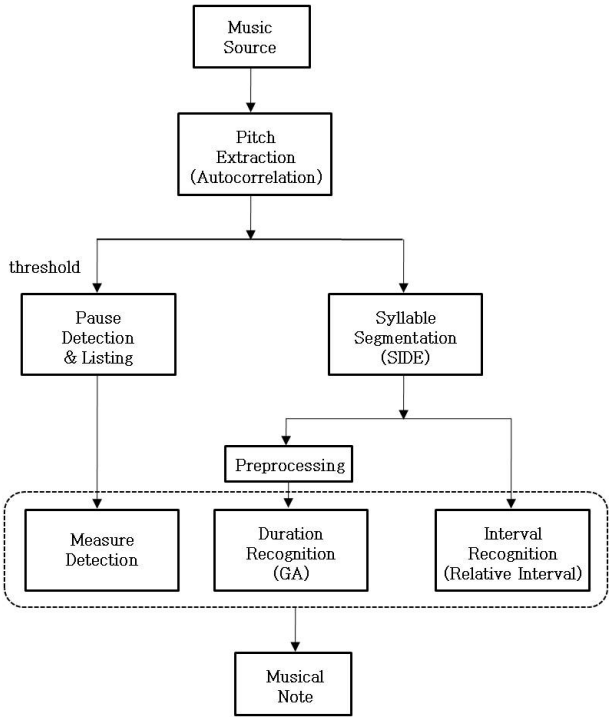
논문접수 : 2008년 12월 2일
수 정 일 : 1차 2009년 1월 15일, 2차 2009년 2월 10일
심사완료 : 2009년 2월 14일

하여 음정과, 박자의 특징을 사용한다. 노래의 가사도 인식을 위한 특징으로 사용할 수 있으나, 시창자 기억의 한계성으로 가사를 틀리거나 혹은 음만 부를 수도 있다는 점에서 가사는 인식 특징에서 제외된다. 또한, 기존의 시스템에서는 절대적인 기준치에 의거하여 음성을 인식하였으나 제안된 시스템은 이를 반영하고 해결하기 위하여 다양한 방법을 적용하였다.

먼저 각 시창자의 음길이(duration) 발성에 적합한 클러스터링을 위하여 쉽게 전역 최적 값에 도달 할 수 있는 유전자 알고리즘을 이용하였고, 연속적인 음성의 음절 분할을 위하여 잡음 제거와 객체 분할에 효율적인 SIDE를 적용하였다. 또한 상대 음정(relative interval)이라는 개별적인 기준치에 의거 하여 피치(pitch)을 매핑하므로, 시창자 개개인의 음 높이에 크게 영향을 받지 않는 음정 인식이 가능하게 되었다. 보다 나은 채보 시스템을 위하여 마디 추출을 수행하였으며, 이를 통하여 오분류된 음표의 후처리 가능성을 제시하였다.

2. 자동 음악 채보 시스템

본 장에서는 먼저 SIDE를 이용한 자동 음악 채보 시스템의 전체적인 구조와 각 단계별로 해당 알고리즘이 수행 되는 과정에 대하여 설명한다.



(그림 1) SIDE를 이용한 자동 음악 채보 시스템의 구조

2.1 피치 검출

음성파의 유성음 구간에 있어서 가장 낮은 파형의 반복 주파수를 음성의 기본 주파수(fundamental frequency)라 부르는데, 이는 성대 진동수와 대응한다. 이는 음성의 높낮이를 나타내게 되므로 피치(pitch)라고도 부른다. 모든 사람에게는 후두 구조에 따라 제약되는 피치 범위가 있으며 남자는 보통 50~250Hz, 여자는 120~500Hz 인데 이는 최대 범위로 보통은 자연스럽게 말 할 때 평균적으로 사용하는 습관적인 레벨을 가지고 있다. 기본 주파수의 시간축에 대한 변화 패턴은, 음성에 포함되는 악센트(accent), 억양(intonation), 강세(stress) 등 운율적 특징이라 불리우는 특성이 음향학적 특성으로 반영된 것이다. 특히 노래에 있어서 발성된 음의 주파수 크기를 대표하는 것이 기본 주파수가 된다.[2,3]

음성신호에서 피치를 검출하는 방법은 시간 영역, 주파수 영역 시간과 주파수 영역을 혼용하여 검출하는 방법으로 나뉠 수 있다. 이 중에서도 시간 영역 피치 추출 방법은 시간 영역 상에서 직접 처리하기 때문에 분해능이 높은 특징이 있다. 시간 영역 피치 검출법에는 병렬 처리법, 면적 비교법, 자기상관함수 이용법, ADMF 이용법 등이 있는데 여기서는 가장 널리 사용 하고 있는 자기상관함수에 의한 피치 검출 기법을 사용하도록 한다.

2.1.1 자기상관함수를 이용한 피치 검출

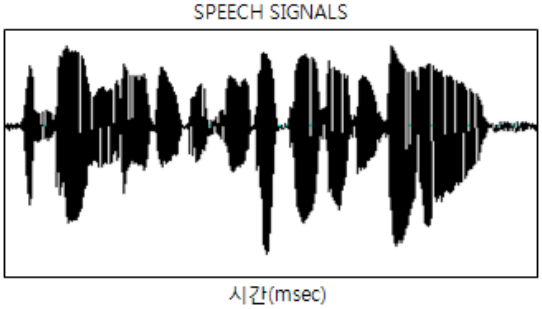
자기상관함수(Autocorrelation Function)는 어떤 시간에서의 신호 값과 다른 시간에서의 신호 값과의 상관성을 나타내는 것으로 연산결과로 나오는 신호는 신호의 주기적인 부분을 강조해 주어 피치를 측정할 수 있게 해 준다. 이 때 측정할 수 있는 피치는 계산에 포함 되는 평균 피치값이 된다.

$$\gamma_n(d) = \frac{1}{N}\sum_{n=0}^{N-d-1} x(n)x(n+d) \qquad (1)$$

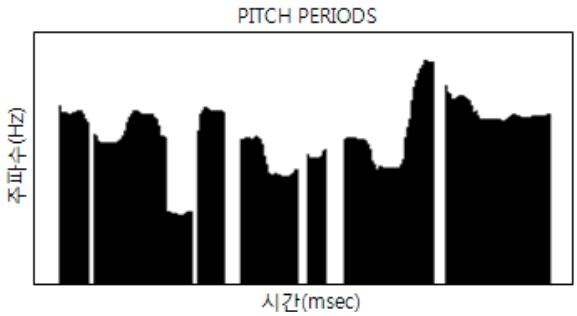
실제 음성 데이터에 적용되는 자기상관함수는 식 (1)과 같다. $\gamma_n(d)$은 지연 d 에서의 자기상관계수를 말하며, N은 프레임 크기이고 d 는 양수 지연, $x(n)$은 신호를 나타낸다. 따라서 자기상관계수는 어느 한 시점 n에서 표본 $x(n)$의 값과 그로부터 d만큼 떨어져 있는 표본의 값을 서로 곱한 것을 모든 n에 대하여 합한 것이라고 볼 수 있다.

신호와 점점 더 지연된 신호간의 상관을 계산함으로써 가장 높은 수준의 상관을 찾을 수 있다. 자기상관함수법에서의 최고는 신호의 주파수가 계산될 음높이 기본주기 길이의 배수에서 발생할 것이다.[4]

아래의 (그림 2)는 동요 "꼬마눈사람" 중 "한겨울에 밀짚



(그림 2) 음성신호, 동요 "꼬마눈사람"



(그림 3) 자기상관함수를 이용한 음성신호의 피치 추출

모자 꼬마눈사람 ”의 음성신호이다. 해당 음성 신호에 자기상관함수를 적용하여 피치 추출한 것을 (그림 3)에서 나타내었다.

검출된 피치는 자동 채보 시스템의 기본 정보가 되며 이를 통하여 음절 분할, 음정 인식 등의 일련의 단계가 수행된다.

2.2 음절 분할

자동 음악 채보 시스템에서 음절은 각각 하나의 음표를 나타내므로, 음표의 음길이와 음정을 인식하는 기본 단위가 된다. 그러나 곡조 인식에서 음절 분할 문제는 아직도 완전하게 해결되지 못한 문제로 남겨져 있다. 곡조인식에서의 분할이 곤란한 이유는 다음과 같다.

첫째, 한 음의 시작과 끝부분은 주파수의 변화가 심해 다른 음으로 분리시킬지 옆 음에 포함시킬지 애매하다.

둘째, 유성음만 발성되는 음의 경우 기본 주파수의 값은 앞 음과 붙어서 분할해야 할 경계를 구분하기 곤란하다.

셋째, 비음의 경우 기본주파수 값을 가지므로 무성음과과 유성음의 구별이 불명확해진다.

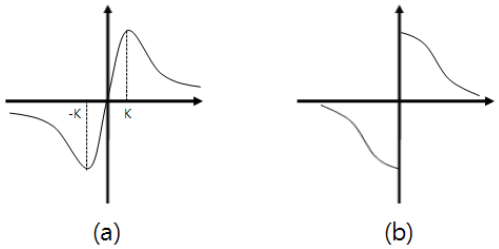
넷째, 실제 악보상의 음길이는 다음 음 발성 전까지로 되어 있지만 사람이 노래할 때는 중간에 숨을 쉬거나 발음의 변화, 무성음 발성 등의 이유로 단절이 생기게 된다. 이 때 단절된 구간이 길어 어디를 한음의 시작과 끝으로 할지가 문제가 된다.[1]

곡조 인식에서 음성 신호를 음절로 분할하는 과정은 필수적이나 분할의 오류가 인식 결과에 영향을 미치게 된다. 이러한 문제를 해결하기 위하여 본 연구에서는 잡음제거와 영역 분할에 효율적인 SIDE를 채택하였다.

2.2.1 안정화된 역 확산 방정식(SIDE)

일반적으로 영상처리 분야에서 객체 분할에 쓰이는 알고리즘인 안정화된 역 확산 방정식(SIDE : Stabilized Inverse Diffusion Equation)은 Perona와 Malik에 제안되어진 비선형 확산 방정식의 단점을 해결하기 위하여 I.Pollak가 고안한 것이다.[5] SIDE를 통한 분할은 잡음에 강인하며 안정된 분할 결과를 나타내므로 이를 곡조인식에서 음성신호를 음절 단위로 분할하는데 사용하였다.

초기 Perona-Malik 방정식에서 사용되는 F함수에서 K값을 0으로 설정하여 정의한 것이 SIDE이다. 이러한 SIDE에서 사용되는 F함수를 식(2)와 같이 정의 하며 이를 (그림 4)에 나타내었다.



(그림 4) F함수 : (a) Perona-Malik, (b) SIDE

$$\begin{aligned} &F'(v) \leqq 0 \ \ for \ v \neq 0. \\ &F(0^+) > 0, \\ &F(v_1) = F(v_2) \ \leftrightarrow \ \ v_1 = v_2 \end{aligned} \tag{2}$$

본 논문에서는 식(2)를 만족하는 식(3)과 같은 F함수를 사용하였다.

$$F(v) = \begin{cases} +e^{-\frac{v^2}{2\sigma^2}} & if \quad v > 0, \\ 0 & if \quad v = 0, \\ -e^{-\frac{v^2}{2\sigma^2}} & if \quad v < 0, \end{cases} \tag{3}$$

SIDE를 통한 다중 스케일 필터링은 분할을 목적으로 하여 영역 단위로 수행된다. 이를 위하여 식(3)의 F함수를 사용하여 식(4)와 같이 영역 값을 구하는 방정식이 사용된다.

$$\dot{u_i} = \frac{1}{m_i} \sum_{j \in A_i} F(u_j - u_i) p_{ij} \tag{4}$$

식 (4)에서 u_i는 영역 i의 값이고, 필터링이 진행될 때 스케일에 따른 변화율은 $\dot{u_i}$이다. 또한 A_i는 영역 i에 인접한 영역들의 집합이고 p_{ij}는 영역 i와 j간의 인접되어 있는 데이터 개수이다. 그리고 m_i는 영역 i의 면적이 된다.

SIDE에 기반을 둔 음절 분할은 다음과 같은 단계로 이루어진다. 초기 조건인 알고리즘의 반복 횟수(I)는 여러 번에 걸친 실험을 통하여 최적화된 값으로 설정하였다.

1단계 : 초기에 각각의 피치들을 독립된 영역으로 설정한다. 이때 각 영역의 면적(m_i)과 영역의 값(u_i), 분할된 영역의 수(N)를 계산한다.

2단계 : 인접한 영역 두 개 이상의 값이 같아질 때까지 각 영역의 값을 식 (4)에 의해 갱신한다.

3단계 : 동일한 값을 가진 인접 영역들을 병합한다. 이때 각 영역의 면적(m_i)과 영역의 값(u_i), 분할된 영역의 수(N)를 갱신한다.

4단계 : 2단계로 간다. 이때 반복 횟수(I)가 초기 설정값을 만족하면 알고리즘을 종료한다.

분할된 결과를 통하여 SIDE에 기반을 둔 영역 분할을 수행할 때 F함수에서 사용되는 σ의 값에 따라 그 수행 결과에 차이가 있는 것을 확인 할 수 있었다. 식(3)과 식 (4)에

서 알 수 있듯이 σ값이 클 경우 반복횟수 감소, 잡음제거, 에지 손실 등의 특징이 있는 반면, σ값이 작으면 반복횟수 증가, 잡음보존, 에지 보존 등의 특징을 지닌다.

따라서 본 연구에서는 SIDE를 통한 음절 분할 시 그 성능을 향상시키기 위하여 형아영 등이 제안한 SIDE의 수렴속도 향상을 위해 제안된 방법을 적용하였다.[6]

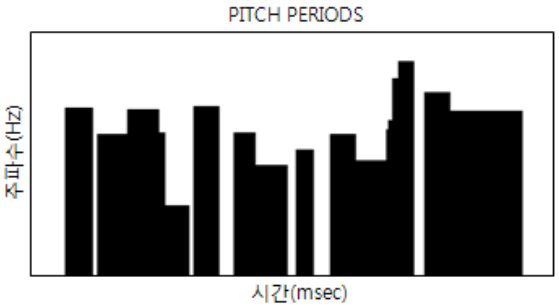
효율적인 분할을 위하여, 초기에 σ값을 크게 해서 잡음을 제거하는 동시에 수렴속도를 개선하고 분할이 진행 될수록 σ값을 작게 해서 그 성능을 높일 수 있는 방법으로 σ값을 설정해 주기 위하여 식 (5)와 같은 수식을 적용하였다.

$$\sigma_n = \sigma_{n-1}\left(\sqrt{\frac{N}{N_{init}}}+1\right) \quad (5)$$

식 (5)에서 σ_n은 현재 반복지점에서의 σ값을 나타내고, σ_{n-1} 은 이전 반복 지점에서의 σ값을 나타낸다. 또한 N은 현재 분할된 영역의 수를, N_{init}는 초기에 분할된 영역의 수를 나타내었다.

(그림 5)는 기존의 SIDE를 개선한 방법을 통하여 음절 분할을 수행한 결과를 나타낸 것이다. 음성 신호에서 모호한 음절 경계를 효과적으로 분할하고 병합하였으며 음정의 잡음을 제거함과 동시에 음정 인식에 필요한 음정의 대표값을 찾아낼 수 있다.

또한, 반복 횟수를 줄여 그 소요 시간을 줄이므로 그 성능을 높일 수 있었다. 분할 후에 일정 임계치(threshold)에 근접하지 못한 영역은 음표로서 유효하지 못하다고 보고 그 앞 음절에 포함하여 처리한다.



(그림 5) 개선된 SIDE를 통한 음절 분할 수행 결과

2.3 음길이 인식

음길이(duration) 인식은 각 음표(note)가 가진 유효한 음의 지속시간을 음악에서의 표준 음표로 매핑(mapping)하는 과정을 의미한다. 음표의 음길이는 음정(interval)과 함께 나타내어 한 마디 안에서 일정한 규칙에 따라 표현되며 곡조 인식에 있어서 핵심적인 정보가 된다.

기존의 음길이 인식 방법은 음표의 종류와 길이를 표준화한 데이터와 비교하여 이를 가장 가까운 음표로 근사하는 것으로서 시창자 개개인의 노래 특징을 고려하지 않게 되는 취약점이 있어 효율성이 높지 않다.

이를 위하여 본 연구에서는 쉽게 전역 최적 값에 도달하게 해주는 유전자 알고리즘에 기반한 음길이 클러스터링을 수행 하였으며, 그 결과는 노래를 부르는 개인의 발성 시간을 반영한 것이라고 볼 수 있다.

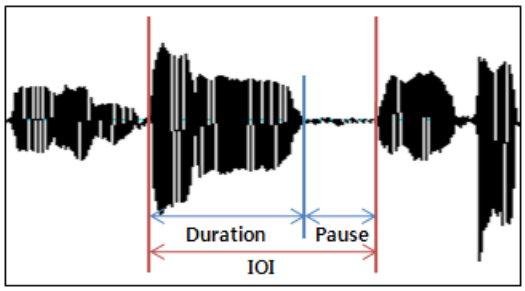
〈표 1〉 음표에 따른 표준 음길이 표

음표이름	음표	음길이(초)
온음표	𝅝	3.2
2분음표	𝅗𝅥	1.6
4분음표	♩	0.8
8분음표	♪	0.4
16분음표	𝅘𝅥𝅯	0.2
32분음표	𝅘𝅥𝅰	0.1

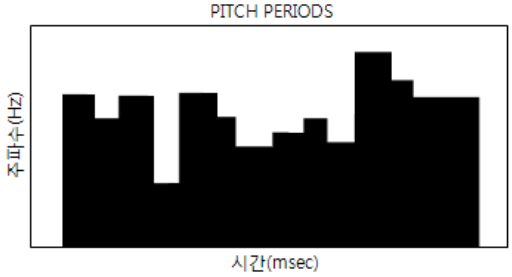
2.3.1 전처리(PreProcessing) 과정

음표의 음길이를 인식하기 위해 분할되어진 음절을 인식을 위한 처리 단위로 변환 하여야 한다. (그림 6)에서처럼 흔히 음성 신호는 음성이 발성되는 지속시간과 사람이 숨을 쉬거나 혹은 발음상의 이유로 인해 발생되는 휴지기(pause)로 이루어져 있다.

음절 분할에서 다루었던 것처럼 사람이 노래 할 때에도 음성의 단절이 발생하게 된다. 이때 발성 지속시간과 휴지기를 합친 것을 IOI(Inter-Onset-Interval)로 정의 할 수 있다. 여기서 발성 지속시간의 시작 시간과 휴지기의 종료 시간은 각 음표의 시작 시간과 다음 음표의 시작 시간의 차를 의미한다. 예비 실험에서 특징량으로 기존의 지속시간(duration) 데이터보다 IOI로 변형된 데이터를 사용한 경우가 좋은 성능을 얻었으므로 음절 데이터를 (그림 7)과 같이 IOI로 변환하는 전처리 과정을 거쳐 사용 한다.



(그림 6) 음성의 지속시간과 휴지기 및 IOI



(그림 7) IOI로 변형된 음절 길이 정보

2.3.2 유전자 알고리즘에 기반한 음길이 클러스터링

음길이 인식에 있어 시창자의 발성 시간은 개인차가 크며, 동일한 노래를 같은 시창자가 부르더라도 매번 그 차이가 있을 수 있다. 따라서 분할된 음절 데이터를 각각의 유효한 음표로 매핑하기 위한 방법이 요구되어진다. 본 논문에서는 이러한 문제를 해결하기 위하여 유전자 알고리즘(Genetic Algorithm)을 적용하였다.

유전자 알고리즘은 유전 정보의 교환에 의한 세대교체에 기반한 문제 해결 기법으로, 쉽게 전역 최적 값을 찾아 낼 수 있는 장점이 있다. 알고리즘의 진행이 반복적으로 수행될 때마다 사전에 결정된 개수의 우수한 유전자만이 다음 세대까지 생존하며, 나머지 열등 유전자들은 도태된다. 집단은 진화를 거듭함으로써 보다 우수한 상태로 빠르게 적응해 나가며, 이러한 집단의 진화 과정은 원래 문제의 목적에 최적화 상태로 수렴하게 된다.

제안된 알고리즘의 수행을 위하여 초기에 유전 정보를 표현하고 염색체를 구성한다. 이를 기반으로 음길이 정보를 클러스터링(clustering)하는 절차는 다음과 같다.

1단계 : 추출된 음절 데이터를 기반으로 염색체를 구성하고 모집단을 초기화 한다.

2단계 : 초기 클러스터의 센터 값을 바탕으로 각 객체 x_i, $\forall i \in 1,2,...,n$와 클러스터 센터 간의 거리를 최소화($d_{\min}$) 하는 클러스터링을 수행한다.

$$d_{\min}(x_i) = \min \| x_i - x_j \| \tag{6}$$

이 때, $j \in 1,2,...,n$이다.

3단계 : 위 결과의 적합도를 평가하기 위하여 클러스터간의 거리인 D_{inter}와 클러스터 내부 거리 D_{intra}를 식 (7), (8)로 정의한다.

$$D_{\text{inter}}\ (C_\alpha) = \max \| S_\alpha - S_\gamma \| * |C_\alpha| \tag{7}$$

$$D_{\text{intra}}\ (C_\alpha) = \sum_{\gamma=1}^{q} \left(\| S_\alpha - S_\gamma \| * \frac{|B_\gamma|}{|x|} \right) \tag{8}$$

이 때, $\alpha \in \{1,2,...,m'\}$이고 $B_\gamma \subset C_\alpha$이다.

C_α와 S_α는 현재 클러스터와 그 센터, B_γ와 S_γ는 나머지 클러스터와 각각의 센터를 말하며, 적합도 함수는 식 (9)과 같다.

$$F = \sum_{\alpha=1}^{m'} D_{\text{inter}}\ (C_\alpha) \Big/ \sum_{\alpha=1}^{m'} D_{\text{intra}}\ (C_\alpha) \tag{9}$$

4단계 : 계산되어진 적합도 함수의 결과를 룰렛 휠(Roulette wheel) 방법을 통하여 선택하고 이 결과를 통하여 구성된 염색체를 다시 일점 교차(one point crossover)한다. 최종적으로 주어진 돌연변이 계수를 이용하여 새로운 세대를 창출해낸다.

5단계 : 다시 2단계로 가서 알고리즘을 수행한다.
이때 설정되어진 반복 횟수(I)를 만족하면 알고리즘을 종료한다.

2.4 음정 인식

음정(interval)이란 순차적으로 또는 동시에 울리는 두 음 사이의 거리를 말하며 동시에 울리는 두 음의 관계를 화성적 음정이라 하고, 순차적으로 울리는 관계를 가락적 음정이라 한다. 두 개의 같은 음을 1도라 부르고, 음 사이의 간격이 한 자리씩 벌어짐에 따라 2도, 3도…등으로 부른다. 본 연구에서는 사람의 음정 인식을 위한 목적이므로 화성적 음정을 제외한 가락적 음정만을 다룬다.

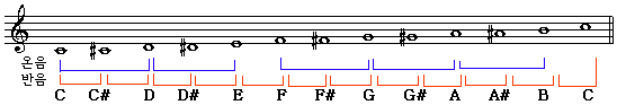
음악에 쓰이는 음높이를 통일시키기 위해 특별히 선정된 진동수를 표준 음고라 하고, 음악을 연주 할 때 모든 악기는 이 표준 진동수에 따라 음을 맞춘다. 국가에 따라 여러 가지 표준음을 쓰던 것을 1885년 비인회의의 결의에 따라 가(A)음을 435Hz로 정하였고, 최근에 연주 효과를 더욱 높이기 위하여 표준음고보다 약간 높은 440Hz로 조율하여 쓰게 되었다.

음악에서 사용할 수 있는 음계는 여러 가지가 있다. 어떤 음의 주파수에서 그 주파수의 2배음까지를 1옥타브(octave)라고 하는데, 음계는 1옥타브를 몇 개로 나누어 구성하느냐에 따라 달라진다. 한국 전통 음악에서는 '궁''상' '각' '치' '우'라 하여 5개음을 기준으로 음계를 구성해 사용하였고, 서양 음악에서 바하 시대 이후의 음계는 한 옥타브를 12개의 같은 간격의 반음으로 나눈 평균율을 적용하였다.[8] (그림 8)에서 12음계를 악보 상에 나타내었으며 본 논문에서는 음정의 결정시 이와 같은 12음계를 적용하였다.

〈표 2〉 옥타브 및 음계별 표준 주파수 표

(단위 : Hz)

옥타브 / 음계	1	2	3	4	5	6	7	8
C(도)	32.7032	65.4064	130.8128	261.6256	523.2511	1046.502	2093.005	4186.009
C#	34.6478	69.2957	138.5913	277.1826	554.3653	1108.731	2217.461	4434.922
D(레)	36.7081	73.4162	146.8324	293.6648	587.3295	1174.659	2349.318	4698.636
D#	38.8909	77.7817	155.5635	311.1270	622.2540	1244.508	2489.016	4978.032
E(미)	41.2034	82.4069	164.8138	329.6276	659.2551	1318.510	2637.020	5274.041
F(파)	43.6535	87.3071	174.6141	349.2282	698.4565	1396.913	2793.826	5587.652
F#	46.2493	92.4986	184.9972	369.9944	739.9888	1479.978	2959.955	5919.911
G(솔)	48.9994	97.9989	195.9977	391.9954	783.9909	1567.982	3135.963	6271.927
G#	51.9130	103.8262	207.6523	415.3047	830.6094	1661.219	3322.438	6644.875
A(라)	55.0000	110.0000	220.0000	440.0000	880.0000	1760.000	3520.000	7040.000
A#	58.2705	116.5409	233.0819	466.1638	932.3275	1864.655	3729.310	7458.620
B(시)	61.7354	123.4708	246.9417	493.8833	987.7666	1975.533	3951.066	7902.133



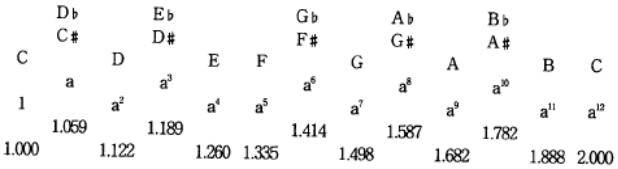
(그림 8) 본 논문에서 사용한 12음계

12음계 시스템은 인간 청각의 한계와 밀접한 관련이 있다. 12음계 내에서 주파수간의 관계는 배수 관계로 선형 비례가 아니다. 1도 차이 음의 주파수 비는 다음과 같은 식 (10)에 의하여 표현된다.

$$Y = \sqrt[12]{2}\,X \quad (\sqrt[12]{2} \approx 1.05946) \tag{10}$$

Y는 X에 대해 1도 위의 음

그러므로 옥타브가 12개의 반음으로 되어 있고, 모든 반음이 정확하게 같은 크기로 되어 있다면 C~C#의 음정은 C#~D의 음정과 같다. 이때의 음정 값을 a라 하고, C의 주파수를 1이라 하면 C#은 상대 주파수 a가 되고, D는 a를 또 곱한 a^2가 된다. 그렇다면 이와 같은 상대 주파수는 (그림 9)와 같은 평균율 음계로 나타낼 수 있다.[9]



(그림 9) 평균율 음계

2.4.1 절대 음정

절대 음고(absolute pitch)란 일반적으로 음을 들었을 때, 다른 기준 음과 비교 하지 않고도 즉각적으로 그 음의 이름을 판단할 수 있는 능력을 말한다. 절대 음고는 정확하게 고정된 음고들로 구성된 음계가 있다는 사실을 전제로 하는 것인데, 규정된 표준 음고(440Hz=A_4)를 사용하게 된 것은 최근 몇 백 년 사이의 일이므로 비교적 최근에 발달된 것임을 알 수 있다.

이를 바탕으로 하는 절대 음정은 표준 음고를 기준으로 음절의 대표 주파수를 가장 가까운 12음계 상의 대표음으로 근사하는 것으로서, 표준음고의 주파수와 가장 차이가 작은 음을 해당 음으로 간주한다. 이와 같은 방법은 기존에 연구되어진 곡조인식 분야에서 사용하는 것으로 표준화된 데이터를 기반으로 하여 음정을 인식하지만, 시창자의 음고 변화에 적응하지 못하는 모습을 보여준다.

2.4.2 상대 음정

절대 음고와 상반되는 개념으로서 상대 음고(relative pitch)가 있다. 이는 음을 지각할 시에 외부에서 주어지는 기준 음에 의존하여 음고를 판별 하는 방법이다. 상대 음고는 A_4만 들렸을 때는 어떤 음인지 모르지만, C_5음을 들려주고 이 음이 C_5음임을 알려준 다음 A_4음을 들려주면 이전에 들었던 C_5음에서부터 목표 음이 단 3도 아래임을 지각하고 C에서부터 C→B→$B^{\flat}$→A로 세 개 반음 내려와 A_4임을 알게 하는 능력이다.[9]

이와 같이 앞 음과의 상대적인 변화를 측정하여 변화정도로 음정을 결정하는 상대 음정(relative interval) 이용 방법은, 노래의 첫 음절을 기준 음이라 보고 평균율 음계를 이용하여 각 노래에 해당하는 상대 음정 주파수 표를 재구성하는 것이다. 만약 첫 음절이 표준 주파수 표에서의 "G(솔)"로 인식되고, 대표 피치 값에 해당하는 주파수가 200Hz라면 그 상대 음정 주파수는 <표 3>과 같다.

〈표 3〉 상대 음정 주파수 표

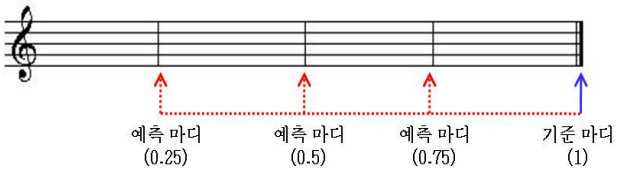
음 정	C	D	E	F	G	A	B	C
주파수	133.5	149.7	168.2	178.2	200	224.5	252	267
비 율	1.000	1.122	1.260	1.335	1.498	1.682	1.888	2.000

상대 음정을 이용할 경우 시창자마다 음역이 다름으로 인해 생기는 개인차를 고려하지 않아도 된다는 장점이 있다. 따라서 본 연구에서는 SIDE를 이용하여 분할된 음절 의 피치 대표 값을 음표의 음정을 인식할 수 있는 정보라고 생각하고 상대 음정을 이용하여 근사(approximation)하였다.

2.5 마디 검출

악곡에서 마디(measure)는 오선 위에 세로로 그은 세로줄(bar) 사이의 유효한 박자의 길이를 나타낸다. 음악의 시간적인 흐름을 구분하는 박자는 기본이 되는 음표의 종류와 1마디 안에 들어가는 음표의 수에 따라서 결정되므로, 마디 정보는 보다 정확한 악보를 구성하는 용도로 이용되어 진다.

한 노래에서 각 마디에는 해당 박자에 해당하는 음표들이 존재 하게 되며 이는 한 마디와 다른 마디의 발성 시간이 비슷한 대역에 분포한다는 것이다. 또한, 사람이 악보로 구성된 노래를 부를 때에는 대체로 곡의 중간 부분에 가장 긴 휴지기가 존재 하게 되는데 이것을 기준 마디라고 하고 나머지 마디 예상 지점을 찾아 낼 수 있다.



(그림 10) 기준 마디를 이용하여 찾은 예측 마디

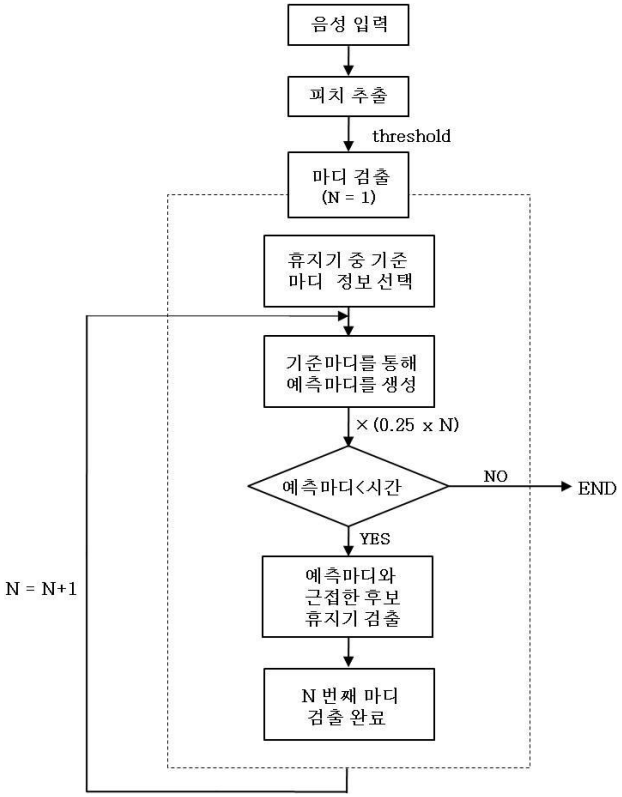
본 시스템에서 제안한 마디 검출을 위한 알고리즘은 (그림 11)과 같으며 크게 4단계에 걸쳐 수행되어진다.

1단계 : 추출된 피치 정보에 임계치(threshold)를 적용하여 음성/비음성 구간을 판단한다. 비음성 구간을 음성의 휴지기(pause)라 하며 악보 상에서 마디의 후보 위치라 가정한다.
이 때, $n = 1$로 정의 한다.

2단계 : 휴지기 데이터 분석을 통하여, 노래의 중심 휴지기를 찾아내고 이를 기준이 되는 마디(M_s) 정보로 사용한다.

3단계 : 기준 마디에 일정 배수(0.25, 0.5, 0.75…)를 곱해가며 마디의 예측 위치를 구한다. 예측 마디는 식 11에 의하여 구한다.

$$M_e = M_s(0.25 \times n) \tag{11}$$



(그림 11) 음성의 휴지기 정보를 통한 마디 검출 알고리즘

4단계 : 예측 마디(M_e)를 바탕으로 휴지기 데이터와 각 예상 위치와의 거리를 최소화($d_{\min}$)하는 근접한 후보 데이터를 마디 (M_n)로 검출한다.

$$d_{\min} = \min \| M_e - x_i \| \tag{12}$$

이 때, 휴지기 데이터는 $x_i,\ \forall i \in 1,2,...,n$이다.

다시 3단계로 가서 알고리즘을 수행한다.

$n = n + 1$로 재정의 된다.

검출된 마디를 이용하여 악보를 구성하면 보다 정확한 악곡의 채보가 가능하다. 또한 마디 정보를 이용하여 음표의 음길이 인식 시에 오분류된 음표의 후처리가 가능하다는 장점이 있다.

3. 실험 및 결과

본 장에서는 제안한 자동 음악 채보 시스템의 실험 결과 및 분석에 관하여 설명한다. 원곡 노래 파일을 제안된 시스템을 통해 인식한 악보의 생성 예를 보이고 그 성능을 DMOS(Degradational MOS)방법을 통하여 평가한 결과를 분석한다. 또한 마디 검출의 효율성과 그 후처리에 관하여 간략히 기술한다.

3.1 실험 방법

제안된 시스템의 성능을 평가하기 위하여 남녀 시창자 4인이 일반 실험실 환경에서 마이크를 통하여 음성(노래)를 녹음하였다. 또한 동일한 곡조 일지라도 부르는 사람에 따라 그 인식 결과가 같지 않음을 살펴보기 위하여 같은 곡을 다른 시창자가 불러 실험의 효율성을 고려하였으며,프로그램 상에서 음성을 녹음한 파일의 데이터 처리를 위하여 Sampling rate = 16000Hz, mono 타입으로 디지털 변환을 하였다. 실험에 사용된 12곡의 노래는 <표 4>와 같다.

〈표 4〉 실험에 사용한 노래 목록

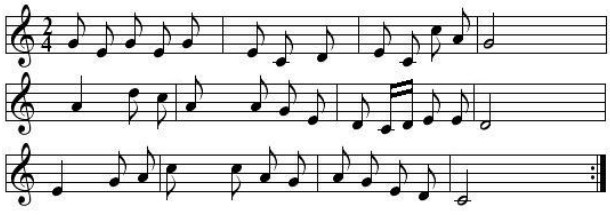
순 서	제 목	시 창 자
1	산토끼	남성1, 남성2
2	애국가	남성1, 남성2
3	눈꽃송이	남성1, 여성1
4	학교종	남성1, 여성1
5	꼬마눈사람	여성1
6	반달	여성1
7	비행기	여성2
8	곰세마리	여성1
9	새신	여성2
10	봄나들이	남성1
11	섬집아기	여성1
12	고드름	여성1

3.2 실험 결과 및 분석

각각의 wave 파일의 음성신호를 디지털 변환하여 제안된 시스템을 통해 인식하고 그 결과를 악보로 나타내었다. 마디 검출의 정확도는 원곡의 악보와 비교하여 그 수치를 계산하였으며, 노래를 부른 사람의 음악 파일과 채보된 악보를 연주한 파일의 비교를 위하여 DMOS를 적용하였다.

3.2.1 악보 생성 결과

구현된 인식 시스템에서 악곡의 채보 형태를 알아보기 위하여 음길이와 음정 정보 및 마디 데이터를 통하여 악보를 표현하였다. (그림 12)는 동요 "꼬마눈사람"의 원곡 악보를



(그림 12) 동요 "꼬마눈사람"의 원곡 악보

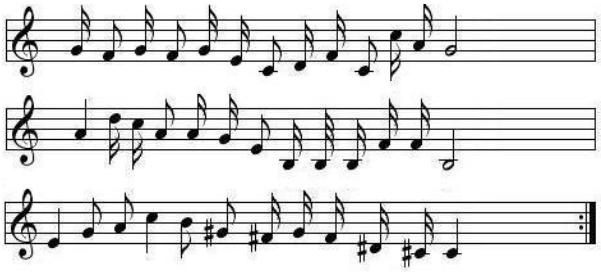
나타낸 것이다.

기존의 방법은 앞 장에서 언급했던 것처럼 표준화된 데이터를 이용하여 음정과 음길이를 인식하는 것으로 시창자 개개인의 노래 특징을 고려하지 않게 취약점이 있다. (그림 12)의 원곡 악보와 비교하여 (그림 13)에서 알 수 있듯이 지

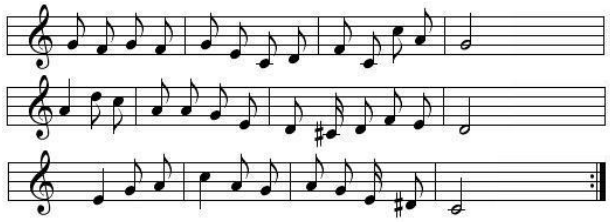
정된 곡을 빠르게 부른 시창자의 표준화된 데이터에 근거한 채보 결과는 음길이에 대한 인식 성능이 떨어지고 결과적으로 완전한 악보의 형태를 띠지 못하므로, 곡조 인식의 효율성 면에서 좋지 않다.

같은 샘플 데이터를 제안된 시스템으로 인식한 악보의 모습이 (그림 14)이다. 악곡의 음표의 길이와 음 높이의 인식 결과를 나타낸 것으로서 시창자의 발성 속도와 더불어 음의 높이까지 고려한 성능을 보여주고 있는 것을 알 수 있다. 또한 마디 추출을 가능하게 했으므로 (그림 15)와 같이 박자에 기준하여 음표의 오인식된 부분에 후처리가 가능할 것으로 예상된다.

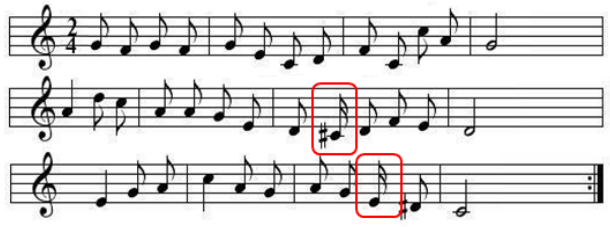
시창자가 원곡의 악보와 흡사하게 발성하여 노래를 부른 경우에 좋은 성능의 악보를 생성할 수 있으나, 상당 수 곡조에서 사람의 발성기관의 부정확함 및 인식 시스템의 한계로 인하여 음정과 음길이에 오차가 포함되어 나타내어진다.



(그림 13) 기존의 방법으로 인식한 악보



(그림 14) 제안된 시스템으로 인식한 악보



(그림 15) 마디 정보를 이용하여 검출된 오인식 음표

3.2.2 마디 검출 결과

제안된 마디 검출 알고리즘을 통하여 (그림 14)에서처럼 악보의 중요 정보인 마디를 표현 할 수 있다. 마디 검출의 성능 및 결과는 원곡의 악보에서 나타낸 마디 정보와 인식된 마디 정보를 일치도를 비교하여 실험하였다.

<표 5>는 <표 4>의 각기 다른 12곡에 대한 마디 검출

〈표 5〉 마디 검출 실험 결과

순 서	제 목	마디 개수	인식 개수	인식률(%)
1	산토끼(1)	7	7	100
2	산토끼(2)	7	6	85.7
3	애국가(1)	7	7	100
4	애국가(2)	7	6	85.7
5	눈꽃송이(1)	7	5	71.4
6	눈꽃송이(2)	7	6	85.7
7	학교종(1)	7	6	85.7
8	학교종(2)	7	6	85.7
9	꼬마눈사람	5	5	100
10	반달	7	7	100
11	비행기	7	6	85.7
12	곰세마리	5	5	100
13	새신	7	6	85.7
14	봄나들이	7	7	100
15	섬집아기	7	7	100
16	고드름	5	5	100

실험 결과를 나타낸 것이다. 16개의 악곡에서 전체 마디의 개수는 106개이고 인식된 마디의 개수는 총 97개 이므로 전체 마디 인식률은 평균 91.5%로 예측 할 수 있다.

마디 검출 알고리즘을 수행할 때 기준 마디 선택 시 한 곡에서 가장 긴 휴지기 부분이 시창자의 발성 상의 문제로 노래의 끝부분이나 혹은 마디 중간 부분에 위치하는 경우가 발생된다. 또한 마디의 위치에서 음절과 음절 사이를 연음으로 노래한 부분은 휴지기 정보가 없으므로 예측 마디를 통해 후보 마디를 찾을 수 없다. 이러한 문제로 인하여 마디 검출에서의 오인식이 발생하게 된다.

3.2.3 성능 평가 결과

본 논문의 제안된 시스템을 통해 인식된 채보 결과의 성능을 평가하기 위하여 20~30대의 남녀 성인 35명의 피실험자를 대상으로 원곡과 채보된 악보의 유사도를 측정 하였다.

유사도 측정 방법은 원곡 노래 파일을 듣고 채보하여 피아노 음으로 녹음된 파일을 들으면서 두 음악 사이의 유사도가 어느 정도 인지를 평가하는 방식으로 이루어졌다. 16곡의 노래의 빠르기와 총 시간은 모두 다르며, 이에 녹음된 파일을 반복 청취하여 실험에 응할 수 있게 하였다. 또한 본 연구에서 제안한 음길이 인식 방법과 음정 인식 방법의 성능 평가를 위하여 각 설문 항목을 '음길이 유사도', '음정 유사도', '악곡 전체 유사도'로 나누어 세밀한 관점에서의 설문이 가능하게 진행하였다.

본 연구의 목적은 사람마다 다른 발성 특징을 음성 채보 시에 얼마만큼 효율적으로 반영하느냐에 따라 그 성능의 효율성이 결정된다. 따라서 곡조 인식 실험에 따른 피실험자의 설문은 DMOS 방법을 채택하였다.

DMOS(Degradational MOS)는 MOS와 같이 주관적 음질 평가 척도 중 하나로 피실험자의 원래의 음성과 변환된 음성을 비교하여 측정함으로써, 원래의 음성에 대해 얼마만큼의 왜곡이 발생하였는지에 대해 등급을 부여한다.

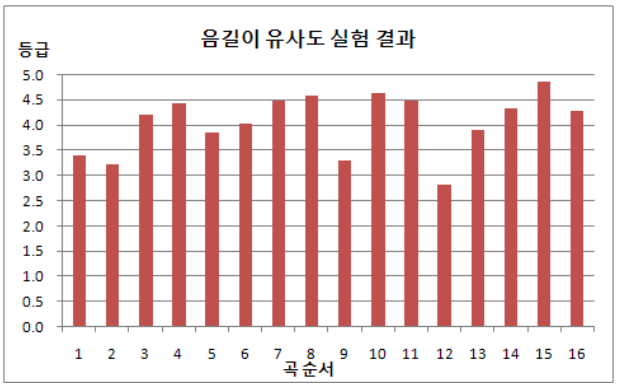
많이 사용되는 5점 오피니언 척도에서 5가지의 등급 5, 4, 3, 2, 1에 대한 구성은 <표 6>과 같다.

(그림 16)은 음길이 유사도에 대한 평균 분석 값을 그래프 형태로 표시한 것이다. 총 16곡의 노래에 대한 5점 척도의 평균값으로 변환하여 나타내었다.

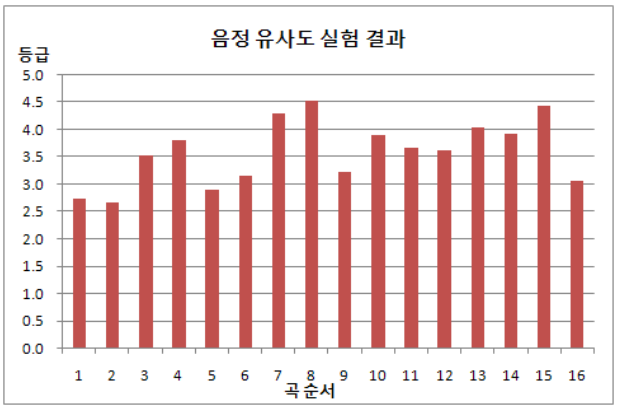
제안된 시스템을 통하여 인식된 노래의 음길이 유사도의 총 평균은 4.05로 DMOS 등급 표에서 “유사함”에 가까운 결과를 나타내며 노래의 15번곡 “섬집아기“는 그 평균값이 4.9로 매우 좋은 결과를 얻을 수 있었다. 이는 음길이 인식 시에 유전자 알고리즘을 통하여 클러스터의 중앙값을 찾고 이

〈표 6〉 5점 척도에서의 등급 표

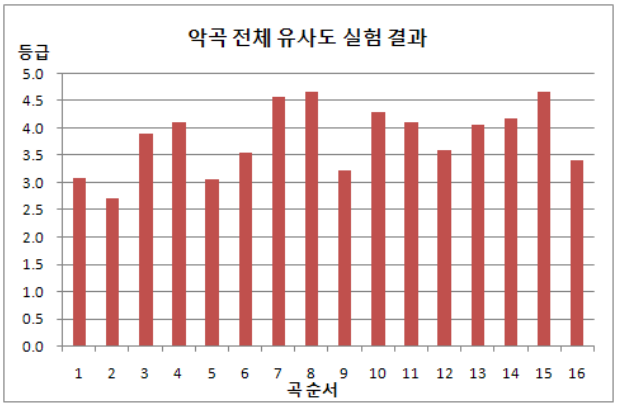
등 급	DMOS
5	매우 유사함
4	유사함
3	보통임
2	유사하지 않음
1	매우 유사하지 않음



(그림 16) 음길이 유사도의 평균 값



(그림 17) 음정 유사도의 평균 값



(그림 18) 악곡 전체 유사도의 평균 값

를 각 음표에 해당하는 클러스터로 각각의 악곡마다 재정립할 수 있게 클러스터링 하는 알고리즘을 수행함으로써 얻어진 것이라 할 수 있다.

따라서 음길이 유사도의 측정 결과는 본 시스템이 사람의 발성 속도에 강인한 음길이 인식 방법을 제안하였음을 보여준다. 결과적으로 기존의 시스템과 인식 방법에 있어 차별화를 두었을 때 그 유효성을 확인 할 수 있음을 나타낸 것이다.

위의 (그림 17)은 음정 유사도에 대한 평균 분석 값을 나타낸 것으로, 음정 유사도에 대한 피실험자 35명의 총 평균값은 3.58이다. 보통 사람이 한 곡의 노래를 부를 때에 앞부분에 비하여 뒷부분의 음정이 불안하고 많이 낮아지는 현상을 찾아볼 수가 있다. 본 연구에서는 기준 음정을 첫 음절로 사용 하였는데, 악곡의 뒤로 갈수록 음정이 약간씩 낮아지므로 이러한 부분에서 오인식된 음표가 발생된 것이다. 또한 대체로 남성 시창자들의 음정 유사도가 미흡하였는데 이는 여성 시창자에 비하여 고음 부분의 음역대가 낮고 이를 정확하게 발성하지 못하여 생긴 문제로 사료된다.

(그림 18)은 악곡 전체 유사도의 평균값을 나타낸 것이다. 악곡 전체에 있어서 유사도를 측정하는 기준은 인식된 음정과 음길이의 어울림과 곡 전체의 분위기에 해당한다고 할 수 있다. 실험 데이터 16곡에 대한 총 평균은 3.82였으며 이것은 “유사함”에 가까운 DMOS 등급을 나타낸다. 노래마다 그 차이는 있었지만 대체로 3~4등급에 해당하는 유사도를 지녔으며, 시스템의 유효성을 확인 할 수 있는 근거가 된다.

4. 결 론

본 논문에서는 시창자 발성의 특징을 반영할 수 있는 자동 음악 채보 시스템을 위하여 SIDE 외 다양한 곡조 인식 기법을 활용하였다.

먼저 SIDE를 이용한 음절 분할을 통하여 잡음 제거 및 효과적인 영역 분할을 수행하였으며, 분할된 음절은 악보에서의 한 음표를 나타내기 위한 기본 단위로 사용하였다. 또한 분할된 후에는 음절의 주파수 대표치를 알 수 있으므로 이를 통하여 음정 인식의 대표 주파수 값을 나타내었다.

그리고 음길이 인식을 위하여 유전자 알고리즘을 이용하였다. 음길이의 중심 값을 찾아 비슷한 발성 시간을 가진 음표를 클러스터링 함으로써 같은 노래라 하더라도 부르는 사람에 따라 각각의 노래 시간이 같지 않음을 반영한 것이다.

또한 음절 대표 주파수로부터 얻어진 음정의 기본 정보를 상대 음정을 통하여 매핑(mapping)하였다. 절대 음정은 표준 데이터를 사용하므로 인식 결과의 편차가 크다. 이러한 문제점을 해결하기 위하여 개개인의 상대 음정 주파수 표를 재구성하는 방법을 제안하였고, DMOS 방법을 통한 유사도 측정 실험 결과로 성능을 확인할 수 있었다.

기존의 채보 시스템에서는 음표의 음정과 음길이 정보만을 이용하여 악보를 구성하였다. 그러나 제안된 시스템에서는 마디 검출을 통하여 보다 정확한 악보의 채보를 가능하게 하였으며 박자 정보 등을 바탕으로 한 오인식된 음표의 후처리 가능성을 제시하였다.

본 논문에서 제안한 방법을 이용하여 휴대폰 상에서의 음성을 통한 자동 작곡 시스템과 같은 응용이 가능하며, 더 나아가서는 사용자의 음성 특징량을 분석한 감성 검색 시스템으로의 적용이 가능할 것으로 기대된다.

참 고 문 헌

[1] 장준영, “퍼지적분을 이용한 곡조 인식 시스템의 설계와 구현”, KAIST 석사학위 논문, 1996.

[2] W. Hess, “Pitch Determination of Speech Signals”, Springer-Verlag, NewYork, 1983.

[3] 오영환, “패턴인식론”, 정익사, 서울, 1991.

[4] 형아영, 이준환 “유전자 알고리즘을 이용한 음표의 음 길이 인식”, 제21회 신호처리 합동 학술대회 논문집, pp.176, 2008.

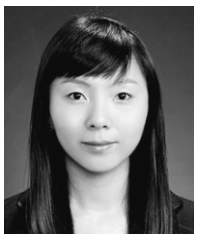
[5] Ilya Pollak, Alan S. Willsky and Hamid Krim, “Image Segmentation and Edge Enhancement with Stabilied Inverse Diffusion Equations”, IEEE Trans. on Image Processing. Vol.9, No.2, pp.256-266, 2000.

[6] 형아영, 이희신, 이준환, “안정화된 역 확산 방정식의 수렴 속도 향상”, IT-CONVERGENCE 학술대회 논문집, pp.78-80, 2008.

[7] Martin, D.W., “Musical Scales since pythagoras”, Sound, Vol.1, No.3, pp.22-24, 1962.

[8] 지정규, “오디오 데이터 베이스의 효율적 검색을 위한 선율 질의 처리기”, 숭실대학교 박사학위 논문, 1998.

[9] 이석원, “음악의 지각과 인지”, 한국음악지각인지학회, 서울, Vol.1, pp.31-51, 2005.



형 아 영

e-mail : aquashake@chonbuk.ac.kr

2007년 전북대학교 전자정보공학부 컴퓨터공학과(공학사)

2007년~현 재 전북대학교 컴퓨터공학과 석사과정

관심분야 : 음성처리, 인공지능, 패턴인식 등

이 준 환

e-mail : chlee@chonbuk.ac.kr

1980년 한양대학교 전자공학과(공학사)

1982년 한국과학기술원 전자공학과(공학석사)

1982년 전북대학교 전자공학과 조교

1985년 전북대학교 전자공학과 전임강사

1990년 미국 미주리대학 전산학과(공학박사)

1990년~현 재 전북대학교 전자정보공학부 교수

관심분야 : 영상처리, 컴퓨터 비젼, 인공지능