ISSN 1975-8359(Print) / ISSN 2287-4364(Online)
The Transactions of the Korean Institute of Electrical Engineers Vol. 63, No. 1, pp. 156～161, 2014
http://dx.doi.org/10.5370/KIEE.2014.63.1.156

멜로디 비교 시각화를 통한 음정 교정 시스템 구현

Implementation of a Tone Correction System Through a Visualization of Melody Comparison

이 혜 인* · 박 주 현* · 이 석 필†
(Hye-In Lee · Ju-Hyun Park · Seok-Pil Lee)

Abstract – With the proliferation of digital music, public's interest in music and desire to sing well are increasing. This paper presents the implementation of a tone correction system through a visualization of comparison between music and humming data. For this we extract MIDI note from music and humming data and then design a matching engine using DTW algorithm which is for robust matching results against local timing variation and inaccurate tempo. This system is expected to correct the user's wrong tone by visualization and feedback from the result.

Key Words : Melody extraction, Humming, Tone matching, DTW

1. 서 론

인터넷이 보편화 되고 디지털 음원의 보급이 확산되면서 대중은 많고 다양한 음악을 쉽게 접할 수 있게 되었다. 음악뿐만 아니라 많은 미디어들이 디지털화 되어 대중은 다양한 매체를 통해 서로의 정보를 공유했고, 그 중 사용자가 직접 제작하는 동영상, 음악, 이미지 등을 일컫는 UCC는 큰 화제가 되었다. 그러면서 직접 노래를 부른 동영상을 인터넷 사이트에 올려 인기를 얻고 유명세를 타는 UCC가수가 생겨나기 시작했고 대중매체인 TV에서는 오디션 프로그램이 성행하게 되었다. 이러한 것들에 접하면서 사람들은 노래를 잘 부르고 싶어 하고 더 정확한 음정으로 노래를 부르는 것에 관심이 집중되고 있다. 노래를 잘 부른다는 것은 정확한 음정으로 부르는 것을 말한다. 여기서 음정이라는 것은 음악에서 보컬 부분의 멜로디이며 음악 신호에서 이러한 멜로디를 정확히 검출하기 위한 많은 선행연구들이 있다.

음악의 신호 레벨에서의 정보처리는 MIREX(Music Information Retrieval and Exchange)를 중심으로 많은 연구가 이루어지고 있다[1]. 음악에서 멜로디 추출의 정확도를 향상하는 기술은 멜로디 규칙 기반 방법과 신호 확률 모델 방법으로 분류된다[2]. 멜로디 규칙 기반 방법에는 Goto가 제안한 다중 트랙 구조를 사용하여 기본 주파수를 찾는 방법이 있다[3]. Gomez등은 앞의 두 프레임과 현재프레임, 그 다음 프레임 이렇게 4개로 구성된 프레임의 기본 주파수의 확률적인 값을 Goto의 트래킹 에이전트에 적용하여 이 4개의 프레임 중 가장 우세한 집단을 주요 멜로디로 결정하였다[4]. Yoon등은 검출된 멜로디 주파수의 연속성을 기반으로 프레임 단위로 검출된 주요 멜로디 주파수를 주파수 특성을 사용해 집단화 한다. 이 방식은 그룹간의 비율을 다르게 적용하여 좀 더 정확한 방법으로 병합하여 멜로디의 흐름을 더 매끄럽게 될 수 있도록 한다[5]. Han등은 음의 시작점과 피크점 그리고 일시적인 영역을 검출하며 MWAE(Modified Windowed Average Energy)를 이용하여 끝점을 찾은 뒤 각 윈도우 프레임에서 정확한 기본 주파수를 찾는다[6]. 신호 확률 모델 방법으로는 Vincent등이 제안한 HMM(Hidden Markov Model)을 사용하여 주요 멜로디 검출의 정확도를 향상 시킨 연구가 있다[7]. Vincent등이 주요 멜로디 검출 정확도를 향상시키기 위해 HMM을 적용한 반면 Ryynanen등은 주요 멜로디 검출 정확도 향상과 멜로디의 시작점과 끝점을 결정하기 위해 HMM을 사용하였다[8].

기존의 노래 평가 시스템들은 결과적으로 점수만 보여주어 사용자가 어느 부분에서 어떤 음을 틀렸는지 알 수 없다는 문제점이 있다. 따라서 본 논문에서는 멜로디 검출 방법을 응용하여 음정을 교정 할 수 있도록 도와주는 시스템을 그림 1과 같이 구현하였다.

음치인 사람은 자신이 내는 음이 어떤 음인지를 인지하지 못하며, 음을 제대로 발성하는 것에도 어려움을 겪으므로 노래 부른 결과를 시각화하여 피드백을 줌으로써 음치인 사람이 음정을 교정하는데 도움이 된다. 다성 음악에서 멜로디를 추출하여 MIDI note를 얻는 것은 [5]의 방법을 이용하여 DB를 구성하였으며, 사용자로부터 허밍 데이터를 받아 Peak picking을 통해 기본 주파수를 추출하고 이 주파수를 MIDI 값으로 바꾸는 Chroma과정을 거치고, 부정확한 피치를 보상시켜주는 Compensation을 적용하여 MIDI note를 도출해 낸다. 톤 매칭은 MIDI note값 끼리 비교 해 주며 허밍 데이터와 음악의 길이가 다른 것을 맞춰주기 위해 DTW 알고리즘을 사용하였다. 또한 결과 값을 시각화함으로써 사용자가 어느 위치에서 음정을 틀리게 불렀는지를 확인 할 수 있도록 한다.

† Corresponding Author : Dept. of Digital Media Technology, Sangmyung University, Korea
E-mail : esprit@smu.ac.kr
* Dept. of Digital Media Technology, Sangmyung University, Korea
Received : October 31, 2013; Accepted : November 30, 2013

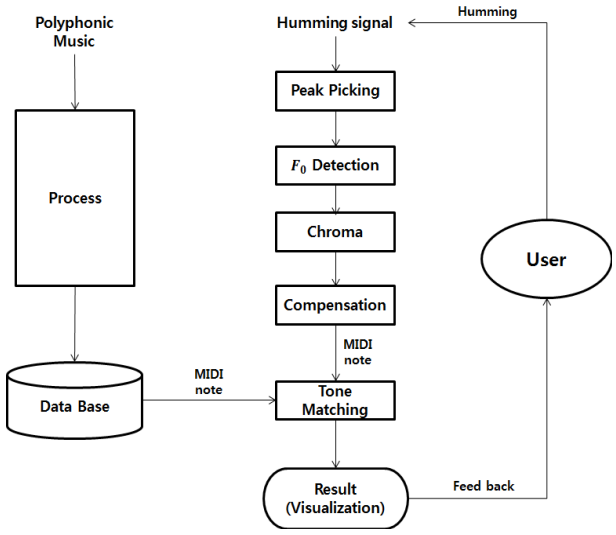

그림 1 음정 교정 시스템의 구조
Fig. 1 Architecture of Tone correction system

본 논문의 구성은 다음과 같다. 2장에서 멜로디를 추출하는 방법에 대한 내용을 설명하고 3장에서 추출된 멜로디를 패턴 매칭 하는 방법에 대해 설명하며, 4장에서 구현한 결과와 함께 실험 결과를 보여주고, 5장에서는 결론에 대해 논의한다.

2. 멜로디 추출 방법

허밍 데이터는 8kHz로 다운 샘플링 된 모노 채널의 데이터를 사용하며 주파수 영역에서의 처리를 위해 64ms 길이의 Hamming window를 적용하여 STFT(Short Time Fourier Transform)를 해 준다. 윈도우의 길이가 길수록 분해능이 높아져 주파수 검출의 정확도를 높여주므로 실질적으로 처리되는 데이터길이는 64ms이며 분해능을 높여주기 위해 프레임 길이를 128ms look-ahead를 적용한다[9].

2.1 Peak Picking

허밍 데이터는 단성음악이므로 간단하게 한 윈도우 안에서 신호의 크기가 가장 큰 값을 갖는 주파수를 기본 주파수로 검출하여 멜로디를 추출한다. 기본 주파수를 검출하기 위해 앞과 뒤 주파수의 신호 크기보다 큰 피크를 검출하는데 이때 의미 있는 피크를 찾기 위해 경계 값을 설정한다. 경계 값은 주파수의 포락선 모양에 따라 2kHz를 기준으로 저주파, 고주파 대역으로 나뉘어져 가변적으로 설정된다. 포락선의 모양은 Skewness(SK)를 구하여 알 수 있으며 SK는 식 (1)과 같이 정의된다. 여기서 $\overline{X}$는 $|X[k]|$의 평균값이다.

$$SK = \sum_{k=0}^{N/2} (|X[k]| - \overline{X})^3 \tag{1}$$

SK < 0 인 경우는 고주파대역에 더 많은 에너지가 분포되어있고, SK > 0 인 경우는 저주파대역에 더 많은 에너지가 분포 되어 있음을 의미하며 SK = 0 인 경우는 대칭적으로 분포 되어 있음을 의미한다. SK에 따른 경계 값은 표 1과 같다.

표 1 SK에 따른 경계값
Table 1 Threshold from SK

	PTH_{low}	PTH_{high}
$SK<0$	$\overline{X_{all}}+\sigma_{all}/2$	$\overline{X_{high}}+\sigma_{high}$
$SK=0$	$\overline{X_{all}}$	$\overline{X_{all}}$
$SK>0$	$\overline{X_{all}}+\sigma_{all}$	$\overline{X_{high}}+\sigma_{high}/2$

구해진 경계 값보다 신호의 크기가 큰 주파수가 피크로 검출되며, 검출된 피크 중 사람이 노래를 부를 때 낼 수 있는 음의 범위인 $100Hz \le F_0 \le 1500Hz$ 구간에 있는 주파수만을 기본 주파수 후보로 검출한다. 그리고 그 중 신호의 크기가 가장 큰 주파수를 기본 주파수로 설정한다. 그림 2는 검출된 피크의 주파수와 신호의 크기를 나타내고 있다.

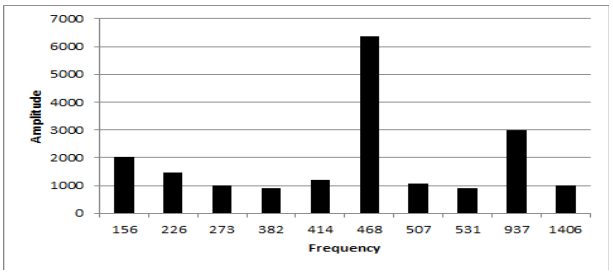

그림 2 검출된 기본 주파수 후보
Fig. 2 Extracted candidates of Fundamental frequency

2.2 Chroma

기본 주파수로 검출 된 멜로디를 한 음으로 표현하기 위해 주파수 값을 MIDI 값으로 변환한다. 식(2)를 통해 주파수 값을 MIDI note로 변환시킬 수 있다.

$$MIDI = 69 + 12 * \log_2 (\frac{f_0}{440}) \tag{2}$$

그림 3은 허밍파일에서 검출된 주파수를 MIDI값으로 변환하여 나타낸 그래프이다. 값이 0 인 부분은 멜로디가 존

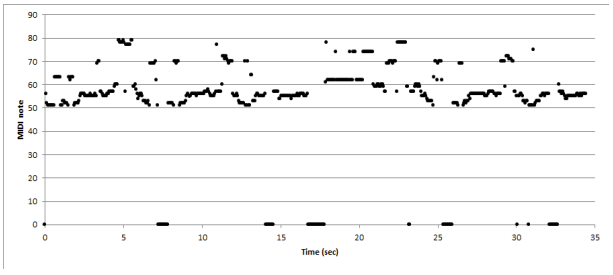

그림 3 허밍 파일에서 검출된 피치
Fig. 3 Pitch sequence extracted from Humming

재하지 않는 부분으로 전체 주파수의 에너지의 평균보다 에너지가 낮아서 멜로디로 인식되지 않는 부분이거나 멜로디가 없는 부분을 나타낸다.

반주가 없는 상태로 노래를 부르기 때문에 허밍의 조성이 실제 비교하는 음악의 조성과 다를 수 있다. 이것을 해결하기 위해 허밍의 MIDI 값을 한 옥타브 C1(24) 에서 C2(36) 사이로 정규화 시켜 만약 범위를 넘어서는 음이 있으면 한 옥타브 낮추어 정규화 범위 안에 들어갈 수 있도록 한다. 또한 정규화 시킨 MIDI note에서 위아래로 12음계와의 거리 값을 계산하여 가장 거리가 짧은 것으로 허밍데이터의 조성을 맞춰준다. 그림 4는 그림 3의 MIDI note를 정규화 시킨 그림을 나타낸다.

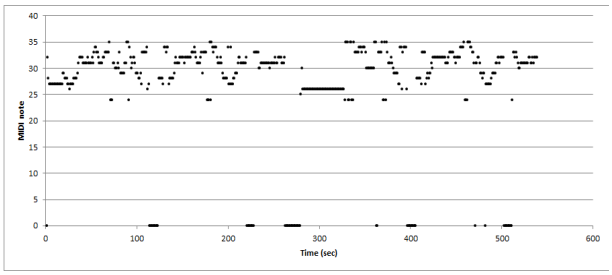

그림 4 검출된 피치 정규화

Fig. 4 Normalization of extracted pitch sequence

2.3 Compensation

허밍데이터의 멜로디 추출은 사람이 부르는 노래를 녹음한 데이터를 사용하므로 주변의 환경잡음이나 노래의 진행을 방해하는 큰 소리 등이 들어갈 수 있다. 또한 사람은 정확한 피치를 부르지 못하므로 허밍에서 검출된 피치는 정확하게 한 음으로 나타나지 않는다. 따라서 320ms를 그룹화하여 멜로디가 매끄럽게 이어질 수 있도록 보상을 해 준다[7]. 그림 5는 허밍으로부터 검출된 피치의 MIDI note의 그래프이고 그림 6은 검출된 피치의 MIDI note를 보상시켜주고 C1(24)에서 C2(36)사이로 정규화 시킨 결과를 나타낸다.

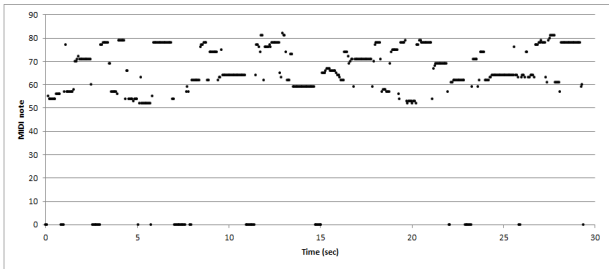

그림 5 허밍으로부터 검출된 피치

Fig. 5 Pitch sequence extracted from Humming

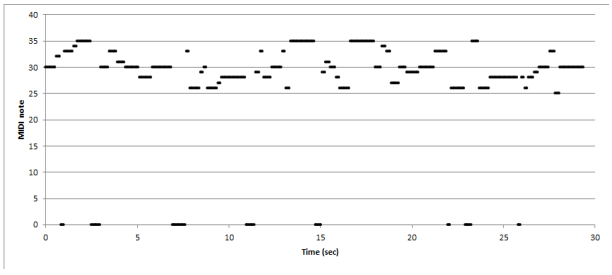

그림 6 검출된 피치 보상

Fig. 6 Compensation of extracted pitch sequence

3. 매칭 방법

DTW(Dynamic Time Warping) 알고리즘은 시간의 길이가 서로 달라도 패턴 매칭을 수행 할 수 있도록 하여 QbSH에서도 많이 사용되는 알고리즘으로 노래의 길이가 다르거나 박자가 맞지 않을 때 이를 보완하기 위해 쓰인다[10][11]. 본 연구에서 허밍 데이터는 반주가 없이 취득되기 때문에 실제 음악과 박자가 맞지 않을 수 있어 이를 해결하기 위해 DTW 알고리즘을 사용한다. 그림 7은 DTW 알고리즘을 나타낸다. 길이가 m, n으로 각각 다른 두 신호에 대해 더 긴 신호를 기준 신호로 잡은 뒤 짧은 신호와 첫 점, 끝 점을 각각 맞춰주고 기준 신호와 매칭 되어 있는 윈도우와 그 다음 윈도우 간의 유클리디안 거리를 측정하여 최솟값이 나오는 윈도우와 매칭 시켜 준다. 그림에서 $w1$, $w2$, …는 선택된 경로를 나타낸다.

본 논문에서는 하나의 윈도우가 하나의 MIDI note값을 나타내며 DTW를 수행하여 매칭 되는 note들의 차이 값이 저장되어 후에 사용자에게 시각화 하여 보여줄 결과 화면의 값으로 사용 된다.

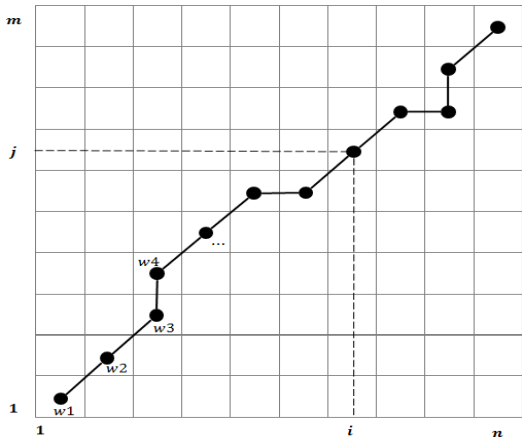

그림 7 DTW 알고리즘

Fig. 7 DTW Algorithm

4. 실험결과

본 실험에서는 2125개의 다성 음악으로부터 추출된 MIDI note값들을 DB로 사용하였고, 사용자로부터 받은 허밍 데이터를 실험 데이터로 사용하였다. 허밍 데이터는 샘플링 주파수 8kHz, 모노 채널인 데이터를 사용하였고, 실험자는 20대의 남녀 각각 15명으로 총 30명이 노래를 불렀다. 논문에는 두 명이 부른 두 개의 노래에 대한 결과를 첨부하였다. 결과화면은 초록색, 노란색, 빨간색 세 가지 색으로 나타난다. 초록색은 MIDI note가 일치하도록 제대로 불렀거나, 사

람이 부른 허밍과 음악의 음정 차가 3개 미만인 것을 의미하며, 노란색은 3개, 4개인 것을 의미하며, 빨간색은 그 이상 틀린 것을 의미한다. 또한 컬러를 지원하지 않는 기기를 위해 그레이 스케일로 결과화면을 나타내주는데, 초록색, 노란색, 빨간색을 각각 흰색, 회색, 검은색으로 나타낸다.

멜로디 추출 프로그램 검증을 위해 피치가 정확하게 나타나는 피아노에서의 멜로디를 추출하여 음악의 MIDI note와 톤 매칭을 수행한 결과이다. 그림 8은 음악 이승철의 "그런 사람 또 없습니다"의 MIDI note이다. 그림 9는 피아노에서 같은 음악을 연주하여 멜로디 추출을 수행한 결과 MIDI note이다. 그림 10은 두 MIDI note를 비교한 결과를 컬러 스케일로 시각화한 것이다. 피아노는 정확한 음을 낼 수 있기 때문에 음악에서 추출된 MIDI note와 톤 매칭을 수행하여 비교했을 때 결과 값이 전부 초록색으로 나타나 값의 차이가 없는 것을 확인할 수 있다.

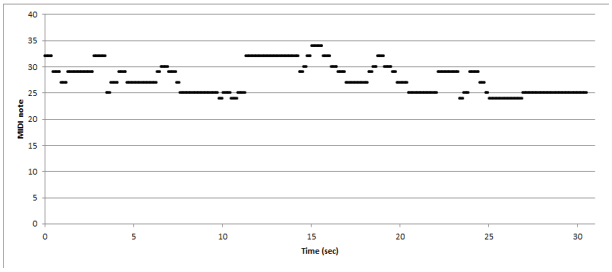

그림 8 이승철 "그런 사람 또 없습니다" 미디노트
Fig. 8 Lee Seung Cheol "No one else" MIDI note

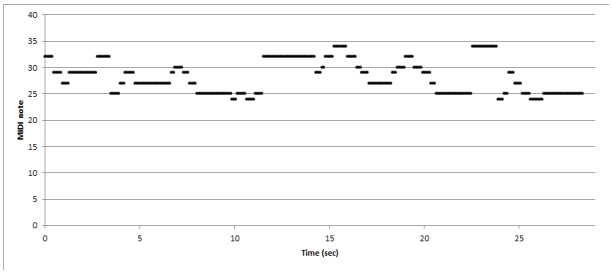

그림 9 이승철 "그런 사람 또 없습니다" 미디노트(피아노)
Fig. 9 Lee Seung Cheol "No one else" MIDI note (Piano)

그림 10 이승철 "그런 사람 또 없습니다"미디노트(피아노)
Fig. 10 Lee Seung Cheol "No one else" MIDI note(Piano)

그림 11은 이문세의 "가로수 그늘 아래 서면"음악의 MIDI note이다. 그림 12, 그림 13, 그림 14는 실험자 A의 멜로디 추출 결과 얻어진 MIDI note, 컬러 스케일의 결과화면, 그레이 스케일의 결과화면이다. 그림 15, 그림 16, 그림 17은 실험자 B의 MIDI note와 컬러 스케일의 결과화면, 그레이 스케일 결과화면이다. 실험자 A에 비해 음정이 정확하지 않은 실험자 B에게 결과를 보여준 뒤 음정을 교정할 수 있도록 한다. 그림 18, 그림 19, 그림 20은 실험자 B가 시각화 결과데이터를 보고 다시 부른 허밍의 MIDI note, 컬러 스케일 결과화면, 그레이 스케일 결과화면이다.

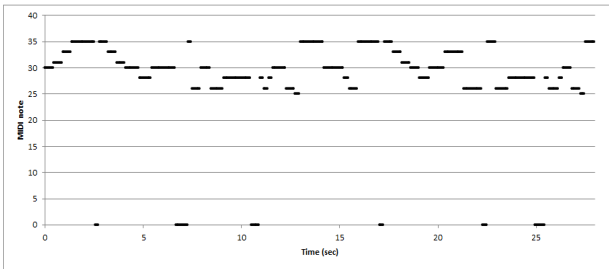

그림 11 이문세 "가로수 그늘 아래 서면" 미디노트
Fig. 11 Lee Mun Se "Underneath the shade roadside tree" MIDI note

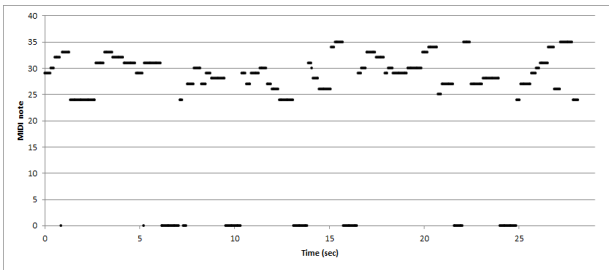

그림 12 실험자 A의 미디노트
Fig. 12 MIDI note of subject A

그림 13 실험자 A의 결과 시각화 (컬러 스케일)
Fig. 13 Visualization of Result subject A (color scale)

그림 14 실험자 A의 결과 시각화 (그레이 스케일)
Fig. 14 Visualization of Result subject A (gray scale)

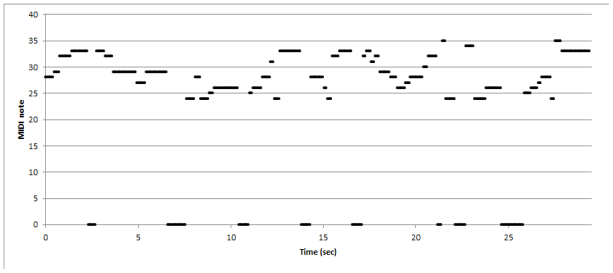

그림 15 실험자 B의 미디노트
Fig. 15 MIDI note of subejct B

그림 16 실험자 B의 결과 시각화 (컬러 스케일)
Fig. 16 Visualization of Result subject B (color scale)

그림 17 실험자 B의 결과 시각화 (그레이 스케일)
Fig. 17 Visualization of Result subject B (gray scale)

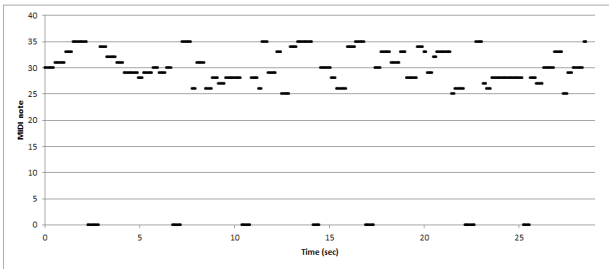

그림 18 실험자 B의 교정 후 미디노트
Fig. 18 MIDI note of subject B after correction

그림 19 실험자 B의 교정 후 결과 시각화 (컬러 스케일)
Fig. 19 Visualization of Result subject B after correction (color scale)

그림 20 실험자 B의 교정 후 결과 시각화 (그레이스케일)
Fig. 20 Visualization of Result subject B after correction (gray scale)

그림 21는 이승철의 "그런 사람 또 없습니다"의 음악 MIDI note이며, 그림 22, 그림 23, 그림 24는 실험자 A의 MIDI note와 컬러 스케일 결과화면, 그레이스케일 화면이고 그림 25, 그림 26, 그림 27은 실험자 B의 MIDI note와 컬러 스케일 결과화면, 그레이스케일 화면이다. 실험자 A에 비해 음정이 정확하지 않은 실험자 B에게 결과를 보여준 뒤 음정을 교정할 수 있도록 한다. 그림 28, 그림 29, 그림 30은 실험자 B가 역시 시각화 데이터를 보고 다시 부른 허밍의 MIDI note와 컬러 스케일 결과화면, 그레이 스케일 화면을 나타낸다.

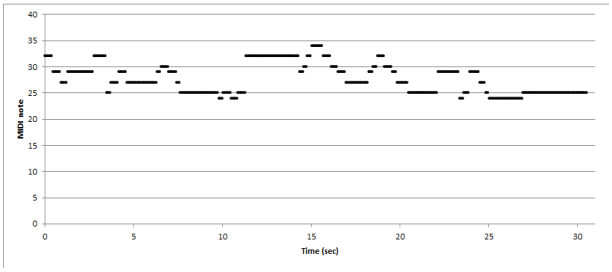

그림 21 이승철 "그런 사람 또 없습니다" 미디노트
Fig. 21 Lee Seung Cheol "No one else" MIDI Note

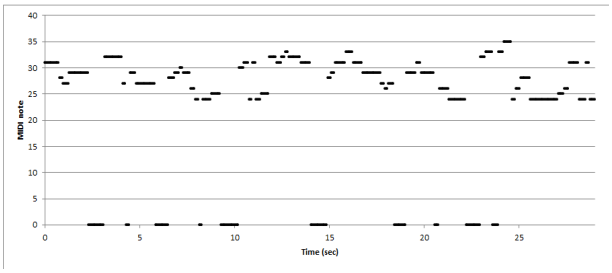

그림 22 실험자 A의 미디노트
Fig. 22 MIDI note of subject A

그림 23 실험자 A의 결과 시각화 (컬러 스케일)
Fig. 23 Visualization of Result subject A (color scale)

그림 24 실험자 A의 결과 시각화 (그레이 스케일)
Fig. 24 Visualization of Result subject A (gray scale)

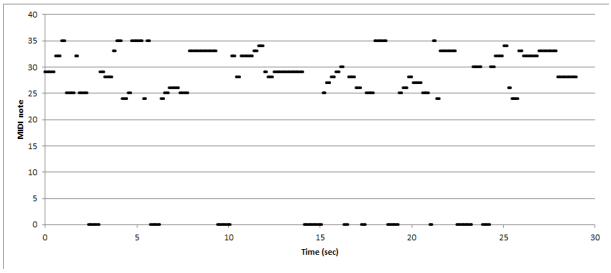

그림 25 실험자 B의 미디노트
Fig. 25 MIDI note of subject B

그림 26 실험자 B의 결과 시각화 (컬러 스케일)
Fig. 26 Visualization of Result subject B (color scale)

그림 27 실험자 B의 결과 시각화 (그레이 스케일)
Fig. 27 Visualization of Result subject B (gray scale)

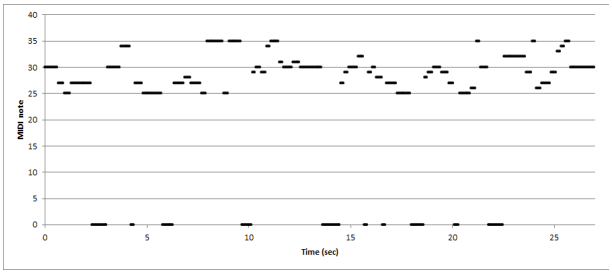

그림 28 실험자 B의 교정 후 미디노트
Fig. 28 MIDI note of subject B after correction

그림 29 실험자 B의 교정 후 결과 시각화 (컬러 스케일)
Fig. 29 Visualization of Result subject B after correction (color scale)

그림 30 실험자 B의 교정 후 결과 시각화 (그레이스케일)
Fig. 30 Visualization of Result subject B after correction (gray scale)

실험자들은 모두 중간의 고음 부분에서 음정의 정확도가 많이 떨어졌으며, 일정한 음을 불러야 하는 끝부분 에서도 동일한 음을 유지시키지 못하고 피치가 많이 흔들렸음을 시

각화 결과를 통해 알 수 있다. 실험자 A에 비해 음의 정확도가 떨어지는 B에 대해 음정을 교정할 수 있도록 결과를 보여준 뒤 다시 노래를 부르게 하였더니 결과 데이터에서 보는 바와 같이 교정 전보다 교정 후에 음의 정확도가 높아진 것을 결과를 통해 알 수 있다.

5. 결 론

본 연구에서는 음정 교정을 위해 음악의 멜로디와 허밍의 멜로디를 비교하여 매칭 결과를 시각적으로 보여주는 시스템을 구현하였다. 사용자가 부른 허밍 데이터를 멜로디 추출하여 주파수를 MIDI note로 바꾸고 반주가 없이 부르기 때문에 기존 DB에 있는 MIDI 파일과 조성이 다르다는 문제점을 해결하기 위해 한 옥타브 안에 정규화 시키는 Chroma과정을 수행하였다. 사람의 목소리가 정확한 피치를 부르지 못해 이를 보상해주기 위한 Compensation 과정을 수행하였다. MIDI note와 허밍 데이터의 시간차를 줄이기 위한 방법으로 DTW 알고리즘을 사용하여 톤 매칭을 하였으며 후에 비교 결과를 시각화하여 나타내었다. 사용자는 정확한 음정과 본인의 음정의 차이에 대한 결과를 즉시 알 수 있고 시각화된 결과를 피드백 받으므로 어느 부분에서 틀린 음을 불렀는지 알 수 있어 스스로 음정을 교정할 수 있다. 이는 단순히 엔터테인먼트적인 기능뿐만 아니라 반복적인 테스트를 거쳐서 음정을 교정할 수 있어 교육적 효과를 기대할 수 있다.

감사의 글

본 연구는 2013년도 상명대학교 교내연구비를 지원받아 수행하였음.

References

[1] S. P. Lee, S. I. Shin, “MIREX - The Status and Future of Music analysis Technology”, Trans. KOSBE, Vol.16, No.4, p75-86, 2011
[2] J. Y. Yoon, J. J. Song, S. P. Lee, H. J. Park, “Accurate Melody Extraction Technology using Polyphonic music”, Trans. KOSBE, Vol.16, No.4, p84-92, 2011
[3] M. Goto, “A real-time music scene description system: Predominant-F0 estimation for detecting melody and bass lines in real-world audio signals”, Speech Communication, vol.43, 2004
[4] E. Gomez, S. Streich, B. Ong, R. P. Paiva, S. Tappert, J.-M. Batke, G. Poliner, D. Ellis and J. P. Bello, “A quantitative comparison of different approaches for melody extraction from polyphonic audio recordings”, Technical Report MTG-TR-2006-01, Music Tech. Group, 2006
[5] J. Y. Yoon, S. J. Park, S. P. Lee, H. Park, “Extracting Predominant Melody of Polyphonic Music based using Harmonic Structure”, MIREX, 2011
[6] B. J. Han, S. M. Rho, E. J. Hwang, “A threshold Adaptation based Voice Query Transcription Scheme for Music Retrieval”, Trans. KIEE, Vol.59, No.2, 2010
[7] E. Vincent and M. Plumbley, “Predominant-f0 estimation using Bayesian harmonic waveform models”, MIREX, 2005
[8] M. Ryynanen and A. Klapuri, “Transcription of the singing melody in polyphonic music”, ICMIR, 2006
[9] H. S. Joo, S. H. Jo, C. D. Yoo, “Analysis of the Effect of Window Length on the Melody Pitch Extraction Using Harmonic Structure Model”, Trans. IEEK, Summer Conference, Vol.33, No.1, p8-11, 2010
[10] Eamonn J. Keog, Michael J. Pazzani, “Derivative Dynamic Time Warping”
[11] Nattha Phiwma and Parinya Sanguansat, "An Improved Melody Contour Feature Extraction for Query by Humming", International Journal of Computer Theory and Engineering, Vol.2, No.4, August, 2010

저 자 소 개

이 혜 인 (李 惠 仁)
2008년 ~ 현재 상명대학교 디지털미디어학과 학사과정
주관심분야 : 오디오 신호처리, 패턴인식

박 주 현 (朴 奏 絃)
2011년 ~ 현재 상명대학교 디지털미디어학과 학사과정, 2013년 ~ 현재 상명대학교 디지털미디어학과 학석사 연계과정
주관심분야 : 오디오 신호처리, 패턴인식

이 석 필 (李 錫 弼)
1990년 연세대학교 전기공학과 공학사, 1992년 연세대학교 전기공학과 공학석사, 1997년 연세대학교 전기공학과 공학박사, 1997년 ~ 2002년 대우전자 영상연구소 선임연구원, 2002년 ~ 2012년 KETI 디지털미디어연구센터 센터장, 2010년 ~ 2011년 미국 Georgia Tech. 방문 연구원, 2012년 ~ 현재 상명대학교 디지털미디어학과 교수
주관심분야 : 멀티미디어 검색, 방송통신 시스템, 인공지능