

## 음정 교정을 위한 실시간 Pitch Tracer의 개발

정 영 철, 최 두 일, 조 우 연  
공주대학교 전기전자정보공학과  
전화 : 041-850-8607 / 핸드폰 : 011-9807-2263

### Development of Real Time Pitch Tracer for Training of Musical Tune

Young-Chul Jung, Doo-Il Choi, Woo-Yeon Cho  
Dept. of Electrical, Electronic, Information & Communication Engineering  
Graduate School, Kongju National University  
E-mail : jyc123@imusicsoft.com

#### Abstract

This research treated development of real time pitch tracer for training of musical tune of speech signal and pre-processing and post-processing technics were proposed to get higher accuracy in extraction of pitch.

Autocorrelation Function was used to get pitch frequency from 64Hz to 980Hz in real time.

Half Rectifier method and Envelop extraction method as a pre-processing was used to get higher accuracy in pitch detection, and improved results were obtained on noised speech signal.

Post-processing method using periodicity of Autocorrelation was proposed to get higher accuracy in the high frequency region.

요소로서 인간이 음정을 인지하기위한 방법에는 많은 경험과 훈련을 통해서 청음능력을 발달시키거나 기준이 되는 음정을 듣고 이를 비교하는 방법 등이 있다. 이러한 방법들은 일반인들에게는 쉽지 않은 방법들로써, 이에 손쉽게 실시간으로 음정을 인지할 수 있고, 이를 이용하여 음정을 교정할 수 있는 Pitch Tracer를 개발하였다.

본 논문에서 Pitch Tracer는 음성신호로부터 실시간으로 Pitch를 추출하여 음정으로 변환하는 것으로, 음성신호에서 64Hz에서 980Hz 사이의 주파수를 가지는 Pitch를 추출하는 것을 목표로 설계되었고 또한 음악의 Allegro의 빠르기에서 32분 음표에 해당하는 시간의 음정을 추출할 수 있는 시간특성을 갖도록 설계하였다. Pitch 추출을 위한 방법은 Autocorrelation Function을 이용하였고, Pitch 추출의 정확성을 높이기 위해서 음성신호의 전처리의 구현과 추출된 Pitch의 후처리 방법을 제안하였다.

#### I. 서론

인간의 음성은 가장 자연스러운 의사소통의 수단으로 음성에는 Pitch, 에너지 등의 정보가 포함되어 있다. 특히 Pitch는 음성신호의 주기적인 특징을 나타내는 것으로 음정을 결정하는 요인이 된다. Pitch를 통해서 결정되는 음정은 특히 음악분야에서 매우 중요한

#### II. Pitch 추출의 이론적 고찰

##### 2.1 음성신호의 기본이론

###### (1) Pitch

음성신호에서 Pitch는 기본 주파수라는 말과 같다. 기본 주파수는 음성신호의 시간축에서 커다랗게 나타

나는 Peak들의 주파수를 의미하며 성대의 주기적인 떨림에 의해서 생성된다. 이러한 Pitch는 인간의 청각에 매우 민감하게 반응하는 파라미터로써 음성의 높낮이를 결정하고 무성음과 유성음을 구분하는데 사용한다.

(2) Frame Window Size

Frame Window Size는 Sampling된 음성신호를 처리하는 단위로써 Sampling Rate, 그리고 시간특성과 관련을 가진다. 표 1은 Sampling Rate에 대한 Frame Window Size의 시간관계를 나타내었다.

표 1. Sampling Rate와 Frame Window Size의 시간관계

Sampling rate \ Frame window size	8 kHz	11 kHz	22 kHz
256 sample	32ms	23.22ms	11.61ms
512 sample	64ms	46.44ms	23.22ms
1024 sample	128ms	92.88ms	46.44ms
2048 sample	256ms	183.04ms	92.88ms

주파수의 해상도를 좋게 하기 위해서는 Sampling Rate와 Frame Window Size가 클수록 좋다. 그러나 Frame Window Size가 크면 시간 특성이 나빠지고 처리시간이 길어지게 된다. 따라서 음성신호에서 Pitch를 추출하기 위해서는 얻고자 하는 주파수의 대역과 시간 특성을 고려하여 적당한 Sampling Rate와 Frame Window Size를 결정하여야 한다. 본 논문에서는 음악의 빠르기표 Allegro의 빠르기에서 32분 음표에 해당하는 빠르기의 음정의 추출과 64Hz에서 980Hz 사이의 음정을 추출을 목표로 22050Hz의 Sampling Rate와 2048의 Frame Window Size, 그리고 50%의 Frame Window 중첩율을 사용하였다.

(3) 음정과 주파수의 관계

음성신호의 주파수는 음정을 결정하는 요소로써 주파수가 커지면 음정은 높아지고 주파수가 작아지면 음정은 낮아진다. 따라서 각 음정마다 다른 주파수 특성을 갖는데 이것이 소리의 높낮이를 구별할 수 있는 근거가 된다. 또한 어느 특정 주파수를 기준으로 두 배 높은 주파수를 가진 음을 옥타브라고 부른다.

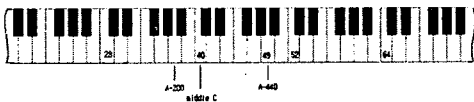


그림 1. 피아노 건반

음정은 피아노 건반을 기준으로 나타낼 수 있다. 한 옥타브에 12개의 건반이 있고 한 옥타브에서 이웃되는 건반은 상수 주파수 비를 유지한다. 따라서 이 비율을 이용하여 한 건반의 주파수만 알면 모든 건반의 주파수를 계산할 수 있다. 그림 1의 A-440이라고 불리는 건반의 주파수는 440Hz이다. A-440은 건반번호 49이고 다른 건반(음정)의 주파수는 식(1)로 계산할 수 있다.[1] 식(1)에서 N은 건반번호이다.

$$f = 440 \times 2^{(N - 49)/12} \tag{1}$$

2.2 Pitch 추출방법

Pitch추출 방법에는 시간영역에서의 추출방법과 주파수 영역에서의 추출방법이 있다. 시간영역에서의 추출방법은 백색잡음과 위상의 왜곡에 있어 강하고 계산이 간단하다. 주파수 영역에서의 추출 방법은 주기적인 음성신호가 주파수 스펙트럼 상에서 임펄스와 그것이 주기적으로 나타나는 점을 이용하여 Pitch를 추출하는 방법으로, 연산의 양이 많고 복잡하다. 본 논문에서는 Pitch추출 방법으로 시간영역의 방법인 Autocorrelation Function을 사용하였다.

Autocorrelation은 어떤 시간에서의 신호 값과 다른 시간에서의 신호 값과의 상관성을 나타내는 것으로 Pitch추출을 위해 적용된 Autocorrelation Function은 식 (2)과 같다.

$$\varphi(k) = \frac{1}{N} \sum_{m=0}^{N-1-k} x(m)x(m+k) \tag{2}$$

Autocorrelation Function에서 k가 Pitch의 주기와 일치하면 다른 영역에서 보다 큰 크기를 갖게 된다. 또한 주기가 P인 주기신호의 Autocorrelation은 역시 주기 P인 함수가 된다.

Autocorrelation을 이용한 Pitch추출은 k가 0이 아닌 부분에서 나타나는 Peak를 이용, Pitch를 추출한다. Pitch를 나타내는 Peak를 알아내면 Peak를 가지고 있는 k가 Pitch의 주기 시간이 된다.

III. Pitch Tracer 시스템 구현

3.1 실시간 Pitch Tracer의 개요

Pitch Tracer의 개발환경은 MicroSoft사의 Visual C++6.0을 사용하였고 Window OS 환경에서 동작할 수 있도록 개발되었다.

Pitch Tracer 시스템 블록도는 그림 2와 같다.

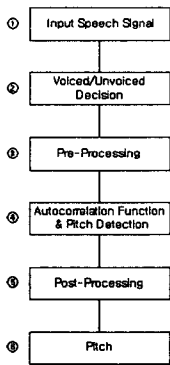


그림 2. 실시간 Pitch Tracer 시스템 블록도

### 3.2 Input Speech Signal

음성 입력은 MicroSoft 사의 DirectX SDK Sound Capture 인터페이스를 사용하여 구현하였고 음성 입력의 파라미터 값은 표 2와 같다.

표 2. 음성입력 Parameter

Parameter	Sampling rate	Sample당 bit	Frame Window Size	중첩율
값	22050Hz	16bit	2048 sample	50%

### 3.3 Voiced/Unvoiced Decision

음성유무의 판단은 에너지와 영교차율을 이용하여 음성유무를 판단하였다. 각각 Frame별 단구간 에너지와 영교차율의 값을 각각의 문턱값과 비교함으로써 판단하였다. 각각의 문턱값은 실험값으로 120000과 0.5를 사용하였다.

### 3.4 Pre-Processing

Pitch추출을 위한 전처리는 음성신호를 조작함으로써 Pitch의 윤곽을 더욱 뚜렷하게 하고 잡음이 있는 신호에서 Pitch 추출의 에러를 낮추는 것을 목적으로 Low Pass Filter, Half Rectifier기법, Envelop 추출기법을 적용하였다. Low Pass Filter는 FIR필터로 980Hz의 차단주파수를 가지는 89차의 필터를 설계하였다.[2]

Envelop 추출기법은 음성신호의 Peak값들을 2차 선형보간법을 이용하여 연결함으로써 음색에 해당하는

부분을 제거하여 음성신호를 단순화 시켜 Pitch를 강조하였다. 이 경우 음성신호가 정현파에 가까운 경우에는 신호가 심각하게 왜곡되므로 Half Rectifier 기법과 같이 사용하였고, Peak의 거리가 검출하고자 하는 최대 주파수의 거리보다 짧은 경우에 Peak를 연결하였다.

### 3.5 Autocorrelation Function & Pitch Detection

Autocorrelation은 식 (2)를 통해서 얻었다.

음정추출의 초기화 단계로 음성신호로부터 찾고자 하는 각각의 음정과 Autocorrelation Sample과의 관계를 나타내는 테이블을 작성하였다. Autocorrelation 결과에서 찾아낸 Pitch에 해당하는 Peak를 나타내는 Sample을 찾아내 테이블과 비교하여 Pitch에 해당하는 음정을 찾아내는데 사용하였다. Pitch추출은 실제 추출하고자 하는 음정의 범위를 옥타브를 기준으로 나눈 4개의 구간과 Pitch추출 오류를 방지하기 위한 1개의 구간으로 음정추출 구간을 나누어 각각의 구간별로 최대 Peak값을 구하고 이를 비교함으로써 Pitch를 추출하였다.

### 3.6 Post-Processing

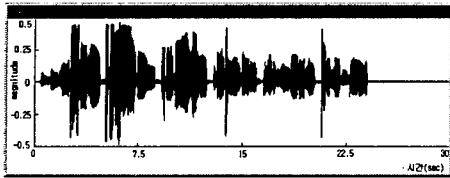
Pitch추출시 고주파 부분에서 주파수의 해상도가 떨어진다.[3] 이를 보완하기 위해서 추출된 음정의 후처리를 제안하였다. 후처리는 Autocorrelation은 주기성을 가지고 있다는 것과 저주파 부분에서 주파수의 해상도가 높다는 점을 이용하였다. 고주파 부분에서 음정을 추출하였을 경우 추출된 음정이 옥타브 배인 저주파 부분의 음정과 이 음정의 한음 높은음과 한음 낮은음의 Autocorrelation Sample의 구간에 있는 Peak값을 비교하여 정확한 음정의 값을 찾아내는 것이다.

## IV. 실험결과 및 고찰

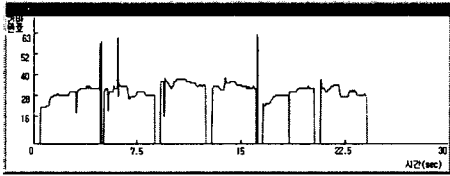
실험을 위하여 남성과 여성의 음성(노래)을 녹음한 30초의 시간을 가진 wave파일을 이용하여 음정을 추출하여 시각적으로 나타내었다. 또한 잡음과 전처리의 관계를 알아보기 위하여 원음성과 신호 대 잡음비가 20dB, 10dB, 6dB인 잡음이 포함된 신호를 만들어 전처리를 다르게 하여 실험하여 그 결과를 비교하였고, 각각 잡음이 섞이지 않은 음성에 Low Pass Filter의 전처리를 적용하여 제안된 후처리를 적용한 결과와 적용하지 않은 결과를 비교하였다.

남성과 여성의 음성 wave파일과 음정추출 결과는

그림 3, 4와 같다.

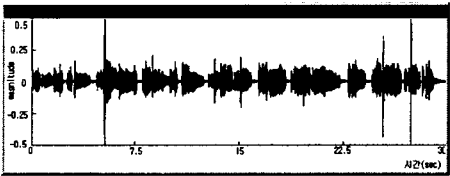


(a) wave 파형

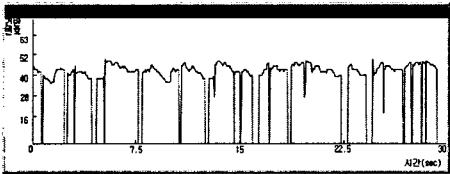


(b) 음정추출결과

그림 3. 남성의 음성 wave파형과 음정추출결과



(a) wave 파형



(b) 음정추출결과

그림 4. 여성의 음성 wave파형과 음정추출결과

그림 3, 4의 음정추출결과는 잡음이 없는 음성을 이용하여 전처리로 Low Pass Filter를 적용한 결과이다. 음정추출의 에러가 많이 발생하는 부분은 음정이 크게 변하는 부분과 음성에너지가 작은 부분에서 일어났다.

남성과 여성의 음성의 음정추출의 정확도는 표 3, 4와 같다.

표 3. 남성음성의 음정추출 정확도

신호 대 잡음비 전처리	원음성 (%)	20dB (%)	10dB (%)	6dB (%)
No	94.42	91.82	83.66	78.03
LPF	94.80	89.03	89.45	89.24
Half Rectifier	93.26	92.74	88.97	82.82
Envelop	93.46	91.88	92.86	91.09

표 4. 여성음성의 음정추출 정확도

신호 대 잡음비 전처리	원음성 (%)	20dB (%)	10dB (%)	6dB (%)
No	95.19	95.15	91.90	86.11
LPF	96.27	95.06	88.07	88.32
Half Rectifier	95.50	95.32	93.18	87.93
Envelop	94.72	94.32	94.84	94.38

잡음이 없는 음성에 대한 음정 추출결과 전처리를 하지 않은 경우와 전처리로 Low Pass Filter만을 적용했을 경우에 가장 높은 정확도를 보였다. 하지만 잡음이 증가할수록 전처리를 적용한 것이 높은 정확도를 보였다.

후처리를 적용하지 않았을 경우 주로 저음인 남성음성의 음정추출정확도는 94.12%로 후처리를 적용하였을 때와 별 차이를 보이지 않았고 상대적으로 고음인 여성음성의 경우에는 89.14%로 후처리를 적용했을 때와 큰 차이를 보였다.

## V. 결론

본 논문에서는 사람의 음성에서 음정을 추출하는 것을 목표로 하여 음정교정을 위한 실시간 Pitch Tracer를 개발하였다. Pitch추출의 방법은 Autocorrelation Function을 이용하였고, Pitch추출의 정확성을 높이기 위하여 전처리를 구현하였고, 후처리방법을 제안하였다. 음정추출결과 94%이상의 정확도를 보였다. 잡음이 있는 신호에 대하여 전처리를 적용하여 향상된 결과를 얻었으며 후처리를 적용하여 고음이 많은 여성의 음성의 경우 높은 정확도를 얻을 수 있었다.

향후 음정이 바뀌는 부분과 에너지가 작은 부분에서 생기는 음정추출의 오류를 개선시키고 음정추출의 범위를 넓히면 악기의 조율에도 사용할 수 있을 것이다.

## 참고문헌(또는 Reference)

- [1] McClellan, Schafer, Yoder, "DSP FIRST", Prentice Hall, 1998
- [2] Alan V. Oppenheim, Alan S. Willsky, "Signal Systems", Prentice Hall, 1997
- [3] Patricio de la Cuadra, Aaron Master, Craig Sapp, "Efficient Pitch Detection Techniques for Interactive Music", ICMC 2001, 2001