

양성후두 질환의 지속모음을 대상으로 한 기존 피치 추정 방법들의 성능 비교 분석*

Comparative Analysis of Performance of Established Pitch Estimation Methods in Sustained Vowel of Benign Vocal Fold Lesions

장 승 진** · 김 효 민** · 최 성 희*** · 박 영 철**** · 최 홍 식***** · 윤 영 로**
Seung-Jin Jang · Hyo-Min Kim · Seong-Hee Choi · Young-Cheol Park
Hong-Shik Choi · Young-Ro Yoon

ABSTRACT

In voice pathology, various measurements calculated from pitch values are proposed to show voice quality. However, those measurements frequently seem to be inaccurate and unreliable because they are based on some wrong pitch values determined from pathological voice data. In order to solve the problem, we compared several pitch estimation methods to propose a better one in pathological voices. From the database of 99 pathological voice and 30 normal voice data, errors derived from pitch estimation were analyzed and compared between pathological and normal voice data or among the vowels produced by patients with benign vocal fold lesions.

Results showed that gross pitch errors were observed in the cases of pathological voice data. From the types of pathological voices classified by the degree of aperiodicity in the speech signals, we found that pitch errors were closely related to the number of aperiodic segments. Also, the autocorrelation approach was found to be the most robust pitch estimation in the pathological voice data. It is desirable to conduct further research on the more severely pathological voice data in order to reduce pitch estimation errors.

Keyword: pitch estimation, pathological voice, pitch error, benign vocal fold lesions

1. 서 론

피치 추정(Pitch Estimation)은 오랫동안 음성처리 관련 분야에서 광범위하게 개발되고 적용되고 있는 연구 분야이다. 피치는 음성과 음악 분야에서는 인지를 위한 기본적인 청각적 특성이다. 악기

* 본 연구는 보건복지부 보건의료기술진흥사업의 지원에 의하여 이루어진 것임.(A020602).

** 연세대학교 의공학과, 교신저자

*** 위스콘신대학교 이비인-두경부과

**** 연세대학교 컴퓨터정보통신공학부

***** 연세대학교 이비인후과

소리 및 음악과 관련한 자동 전사(Automatic Transcription)(Klapuri, 2006) 기능 구현 또는 음성분야와 관련된 음성인식(Speech Recognition), 화자 인식(Speaker Recognition) 및 TTS(Text-to-Speech)와 같은 음성 합성(Speech Synthesis)(Gold, 1999) 기능 구현에서 피치 추정에는 필수적이고 중요한 음성처리를 위한 분석 방법으로 사용되어 왔다. 뿐만 아니라, 언어 장애(Dysphonia) 연구 분야에서도 피치 추정은 진폭변동(Shimmer), 주파수변동(Jitter)과 같은 음성 장애의 정도를 평가할 수 있는 분석 방법들의 기본적인 전처리 방법으로 활용되고 있다(Klingholz, 1985; Feijoo, 1990).

하지만, 오랜 연구기간 수많은 피치 추정 방법들의 개발에도 불구하고 여전히 피치 추정에 대한 많은 문제점 및 신뢰성들이 문제시 되고 있다. Bielamowicz(1996)의 연구결과에 의하면, cSpeech(© cSpeech), MDVP™(CSL™), SoundScope(GW Instruments, Inc.)와 같은 상용 음성 분석 시스템 및 소프트웨어와 수작업을 통한 음성 분석 방법을 대상으로 음성질환자 남성 29명과 여성 21명의 지속 모음에 대한 피치, 진폭 변동, 주파수 변동, HNR, SNR 검출 결과 동일한 음성에 대하여 각각의 동요 측정치 간에 상관관계가 좋지 않은 결과를 보였다. 특히 각 시스템에 존재하는 여러 피치 추정 방법들에 대한 비교 결과 각 알고리즘 적용마다 상이한 결과를 보임으로 인해 각 상용 시스템들 간의 측정 결과에 대한 호환성 및 정확성에 대한 신뢰성 문제를 제기하게 되었다. 또한 이 연구에서는 음성질환자와 같은 비주기적인 음성 데이터에 대한 피치 측정치와 피치로부터 파생되어 얻어진 동요 측정치들의 신뢰성에 대한 강한 의구심을 보고하고 있다. 그 외에도 Karnell (1991,1995)의 연구 결과에서 각기 다른 음성 분석 시스템으로부터 산출된 피치로부터 파생된 동요 측정치들에 대한 불일치성과 그 결과 값에 대한 의구심이 언어 장애 연구 분야에서 심각한 문제점으로 제기되어왔다.

그리하여 본 연구에서는 양성후두 질환(Benign Vocal Fold Lesions: BVFL)에 의해 성대의 구조적 변화가 발생하여 비주기적이고 비정상적인 패턴을 보이는 음성 신호를 대상으로 기존의 성능이 우수하다고 평가된 여러 피치 추정 방법들에 대하여 정확한 피치 검출에 대한 성능 평가를 수행하였다.

2. 피치 검출의 난해성

2.1 피치 검출의 어려움

앞에서 언급한 바와 같이 정확한 피치 추정은 여전히 어려운 일로 평가되고 있으며 다음과 같은 이유들로 피치 검출의 어려움을 설명할 수 있다.

첫째, 분석 도메인에 종속적인 피치 추정 방법의 모델링에 대한 문제가 존재한다. 현재까지 많은 수의 피치 추정 방법들이 개발되어 왔지만, 여러 분야에 걸쳐 적용될 수 있는 피치 추정 방법은 흔하지 않다. 예를 들어 음악과 관련한 자동 전사에 사용되는 피치 추정 방법은 다양한 악기들의 피치를 검출하기 위하여 광범위한 주파수 영역에 대하여 검출을 시도한다. 반면에 인간의 음성에 대하여 피치를 검출하는 경우 주로 저 대역의 주파수 영역에 대한 분석을 하게 된다. 이러한 경우 인간 음성의 피치를 잘 검출해내는 피치 추정방법이 자동 전사 기능에 적용될 경우 피치의 검출 결과에 대한 신뢰성은 문제를 발생할 여지가 많다. 이렇듯 다양한 도메인에서 사용되는 여러 가지 피치 추정 방법과 피치 검출 성능을 높이려는 특정한 파라미터의 조정 및 제한으로 인해 어느 한 도메인

에서 우수한 성능을 보인 피치 추정 방법이 적절한 변경 없이 다른 도메인에 적용되는 경우 좋지 않은 결과를 초래할 수 있게 된다.

둘째, 비록 인간의 음성에만 국한되어 피치를 추정하는 경우 일지라도, 피치 영역의 예상치 못한 동적 변화로 인해 문제가 발생할 수 있다. Fant(1956)의 연구결과 유럽언어들의 일상적인 대화에 사용되는 평균 피치 값은 남성의 경우 120 Hz, 여성의 경우 200 Hz, 아동의 경우 330 Hz이며, 각 개인 화자의 발음에 대한 피치 변동 폭은 옥타브 범주에 국한된다고 보고되고 있다. 또한, 다른 연구결과 일반적인 대화에 있어 기본 주파수의 최대 변동 폭은 평균적으로 남성의 경우 50-250 Hz이고, 여성의 경우 120-480 Hz로 조사되었다(Laver, 1994). 하지만 이러한 일반적인 피치 범주와 달리 남성의 경우 50 Hz보다 낮고, 아동의 경우 600 Hz보다 큰 기본 주파수를 나타내는 음성 신호들이 존재할 수 있다. 이러한 경우 일반적인 피치 검출 범주를 벗어나는 음성들은 잘못된 피치 검출 값을 발생하게 된다.

셋째, 인간의 음성이 현존하는 피치 추정 방법들의 능력보다 훨씬 더 복잡성을 나타낸다는 것이다. 일반적으로 피치 주기는 성문 개폐에 따른 주기에 종속적이다. 이로 인해, 대부분의 피치 추정 방법들이 가정하고 있는 분석 구간(일반적으로 15-30 ms)에서 음성신호는 안정적이고 주기적이라는 가정을 위배하게 되며, 이러한 결과로 인해 잘못된 피치 추정의 중요한 요인으로 작용하게 된다. 또한 기본 주파수에 잘못된 영향을 줄 수 있는 부하모닉의 존재 또한 잘못된 피치 검출의 요인으로서 작용한다. 이러한 두 가지 피치 오류의 발생 요인에 대한 대표적인 경우가 이 연구에서 다루고 있는 음성질환자의 음성 신호들에서 매우 빈번하게 발생하게 된다. 그러므로 음성질환자의 음성 신호 특성을 살펴보는 것이 필요하다.

2.2 음성질환자의 음성 신호 특성

음성질환자의 음성신호는 정상인의 음성신호에 비하여 뚜렷하게 구별되는 특성을 갖는다. 특히, 양성후두 질환에 의한 성대의 구조적 변화는 피치주기 또는 성문 여기파형의 패턴을 비정상적으로 만들어 내게 되어 다음과 같은 현상들을 빈번하게 발생시킨다.

- 발성 중 정 위치에서 성문접촉이 일어나지 않는다.
- 정상 음성신호보다 큰 진폭(Shimmer)과 피치(Jitter)의 변동이 발생한다.
- 성문의 물리적 변화의 영향으로 F0의 실제 값보다 반감되거나 배가 되는 피치 반감(pitch halving error) 또는 배가(pitch doubling error)와 같은 옥타브 에러를 발생시킨다.

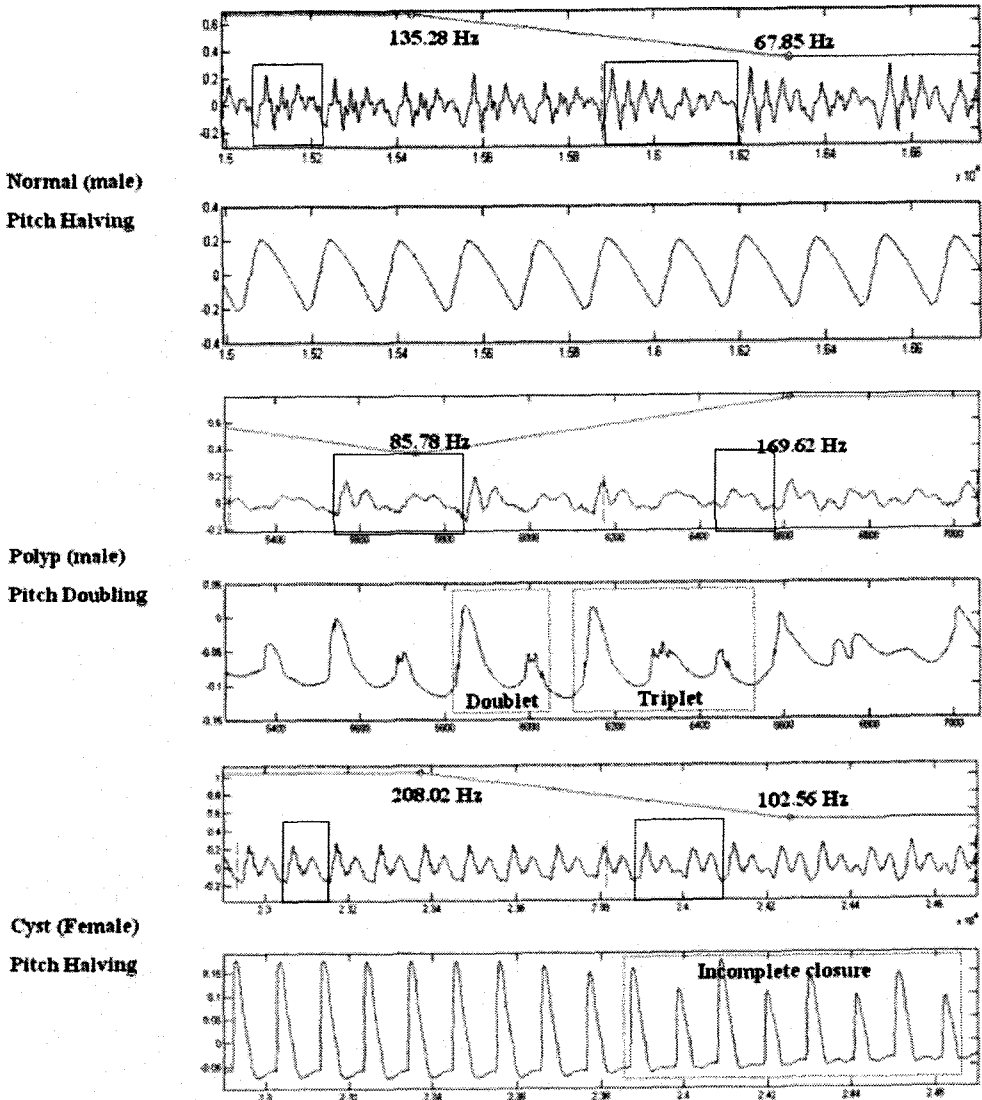


그림 1. 피치 반감, 배가 검출 사례: 각 그림의 위는 음성신호, 아래는 성문파형 신호, (상단) 정상인 남성 피치 반감, (중단) 성대 용종 환자 남성 피치 배가, (하단) 성대 낭종 여성 환자 피치 반감

이외에도 복잡하고 다양한 여러 가지 피치 추정 오류 요인들과 현상들이 존재하지만, 특히 옥타브 에러와 같은 경우 정상인, 음성질환자의 음성에 상관없이 종종 발생하게 된다. 예를 들어 <그림 1(상단)>에서, Autocorrelation(AC) 피치 추정 방법을 통하여 정상 남성의 음성신호를 분석하는 경우 실제 음성신호의 피치는 성문파형(Electroglottograph: EGG)과 일치해야 하는데(작은 박스), 중간 부분 이후 발생할 때 일정하지 못한 호기 량의 변동으로 인해 갑자기 일부분의 음성신호의 진폭이 줄어들게 되어 (큰 박스) 마치 2개의 피치 구간을 하나의 피치 구간으로 잘못 인식하는 피치 반

감에러를 발생하게 된다.

다른 예로 성대 용종(Vocal Polyp) 남성의 경우 실제 개폐되는 성문 파형이 아닌 성대 용종의 흔들림으로 인해 진짜 성대의 개폐와 유사한 파형을 만들어 내게 되는데, 이를 실제 피치 주기 구간에서 몇 번 발생하였는가를 따져서 실제 성문 개폐를 포함하여 2번 발생한 경우 Doublet으로, 3번 발생한 경우 Triplet으로 정의한다. <그림 1(중간)>에서 볼 수 있듯이 이전까지는 Doublet이었던 EGG파형이 중간 지점이후 Triplet으로 변동되면서 유사성이 줄어들게 되어 각각의 Wavelet에 대한 음성신호를 하나의 피치 구간으로 잘못 인식하는 피치 배가에러를 발생하게 된다. 이러한 Doublet과 Triplet 사이의 변동 또한 잘못된 피치 추정의 요인으로 작용하게 된다.

성대 낭종(Vocal Cyst)에서 성문의 완벽하지 못한 개폐 또한 잘못된 피치 추정의 요인으로 작용할 수 있다. <그림 1(하단)>의 EGG 파형에서 볼 수 있듯이 중간까지는 완벽하던 성문의 개폐가 중간이후에는 불안정하게 개폐되면서 이러한 현상이 음성신호에도 영향을 주게 되어 불안정한 성문 개폐에 해당하는 음성 신호의 진폭이 줄어들게 되는 것을 확인할 수 있다. 결국 하나의 피치 구간으로 보아야할 AC의 유사성보다 완벽한 성문 개폐구간과 완벽하지 못한 성문 개폐구간을 쌍으로 묶어서 2 개의 피치 구간을 하나의 피치 구간으로 잘못 인식하는 피치 반감에러가 발생하게 된다.

3. 기존의 피치 추정 방법의 고찰

대부분의 피치 추정 방법들은 음성신호가 분석 시간(주로 짧은 시간 구간) 내에서 안정적이고 주기적이라는 가정을 기반으로 피치를 검출하게 된다. 그러나 실제 음성신호는 단시간 구간에서 일지라도 비안정적이고 유사 주기적인(Pseudo-Periodic) 형태를 때때로 보이게 되며, 이것은 검출 에러에 큰 영향을 준다. 일반적으로, 피치 추정 방법들은 병적 음성의 모든 영역에서 사용되도록 최적화되어 설계되지 않았으며, 정확한 피치 추정 결과를 얻기 위해서는 각각의 피치 추정 방법에서 요구되는 신호의 주기성에 최소한의 충족 조건을 만족시켜야 하는 양상을 보인다.

이번 장에서는 고전적, 현대적인 알고리즘을 사용하는 성능이 우수하다고 밝혀지고 널리 사용되는 7 개의 피치 추정 방법들을 대상으로 음성질환자의 음성신호에 대한 정확한 피치 추정 능력을 평가하기 위하여 간략하게 피치 추정 방법들에 대한 소개 및 특성을 파악하려고 한다. 본 연구에서 분석한 피치 추정 방법은 시간 도메인, 주파수 도메인, 대안적인(Alternative) 분석방법과 같은 3가지 타입의 분석 영역에 의해 구분되고, 음성질환 음성신호에 대한 피치 검출 성능을 향상시키기 위해 코드의 일부분을 수정하여 테스트하였다. 그리고 일반적으로 피치 검출 후 리프팅 평활화(Lifting Smoothing) 과정을 통하여 차이가 큰 피치 검출 값을 보정하게 되는데, 이러한 과정은 피치 과대 에러를 줄일 수는 있으나 전체적인 피치 추정 결과 값에 변동을 주게 되어 순수한 피치 추정 알고리즘의 성능 분석에 장애를 발생시키기 때문에 본 연구에서 배제하고 분석하였다.

3.1 시간 도메인 피치 추정 방법

시간 영역 기반의 방법들로 Autocorrelation(AC)(Sondhi, 1968; Rabiner, 1977), Average Magnitude Difference Function(AMDF)(Deller, 1993), 그리고 YIN(Cheveigne, 2002) 알고리즘을 선택했다.

3.1.1 Autocorrelation (AC)

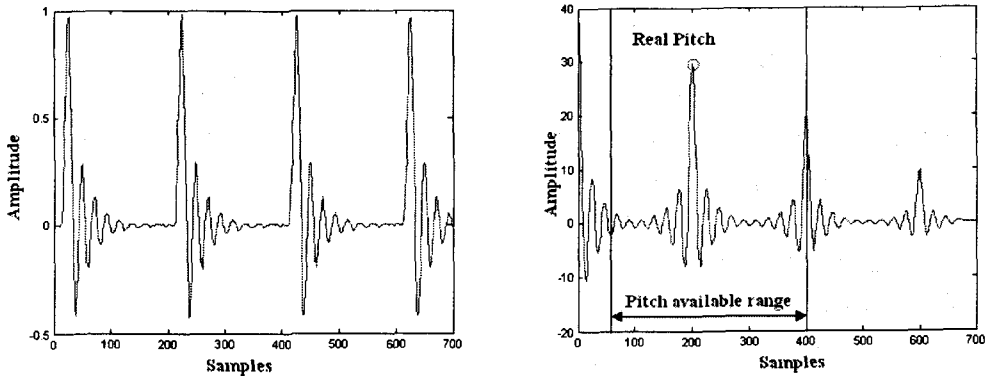


그림 2. 분석에 사용된 음성신호(좌측)와 AC 피치 추정 방법의 피치 검출 방법(우측)

대부분의 기본적인 피치 추정 방법들은 시간에 따라 변하는 공기압의 변화를 나타내는 음성신호의 진폭과 주기를 분석하여 피치를 검출하게 된다. 이러한 원리를 가장 잘 나타내는 피치 추정 방법이 바로 AC이다. 식 (1)과 같이 AC의 분석 방법은 어떤 신호와 그 신호가 이동한 자기 신호와의 유사성을 살펴보는 것이다. 이러한 상관성의 결과는 두 신호를 time lag의 함수에 따른 유사성의 패턴으로 분석 도메인을 취해서 분석해 볼 수 있다. 신호가 어떠한 일정한 주기를 갖고 있다면 그 주기만큼 해당하는 구간에서 자기상관 함수의 값이 커지는 패턴을 보일 것이라는 가정을 갖고 있다. <그림 2>에서 살펴보듯이, 좌측의 음성신호를 자기상관 함수를 통해 나타내면 우측과 같은 결과를 보이게 되며, 주기적인 신호의 경우 해당 주기마다 주기위치에 해당하는 값이 커지게 되는 것을 확인할 수 있다. 또한 자기상관 함수의 특성(식 (2) 참조)에 의하여, 자기 상관함수는 대칭 함수이며, $R(0)$ 에서 가장 큰 값을 갖게 되고, 신호가 완벽히 주기적이며 제한된 분석구간에서 분석 샘플이 phase를 벗어나지 않은 가정 하에 주기, N_p 마다 동일한 값을 나타내게 된다. 하지만, 실제 분석할 때는 이동구간에 따른 지연 시간에 해당하는 샘플이 존재하지 않기 때문에 분석 도메인에서 자기상관 함수는 점차 감소하게 된다. AC에서 일반적인 피치 검출 방법은 분석 도메인의 피치 검출구간 (본 연구에서는 50-350 Hz)에서 가장 큰 피크 값에 해당하는 주파수를 기본 주파수로 산출하게 된다.

$$R(k) = \sum_{m=-N}^N x(m)x(m+k) \quad 0 \leq k \leq N, \text{ 입력: } 1, \dots, N-1 \tag{1}$$

- i) $R(k) = R(-k)$
- ii) $R(0) \geq |R(k)|$
- iii) $R(k) = R(k + N_p)$ 만약 N_p 이 주기적이라면

3.1.2 Average Magnitude Difference Function (AMDF)

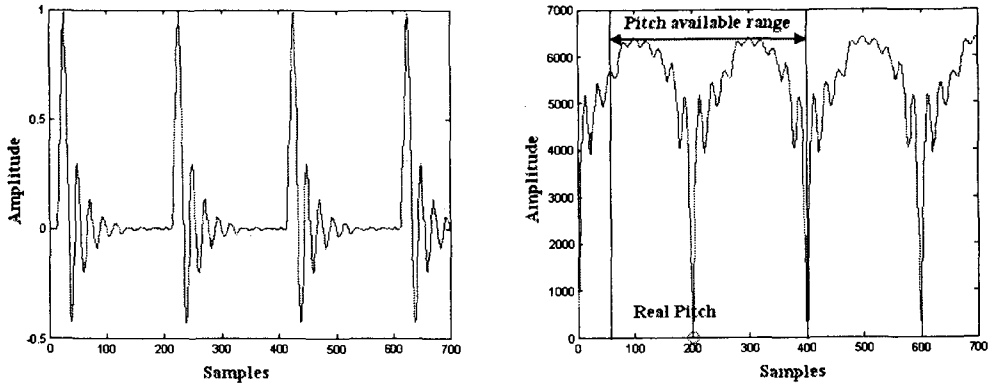


그림 3. 분석에 사용된 음성신호(좌측)와 AMDF 피치 추정 방법의 피치 검출 방법(우측)

AMDF분석은 AC 분석에서 변화, 수정된 분석 방법이다. 즉, AC가 자기 신호의 자기 상관 함수에 의하여 유사성을 파악하는 대신, AMDF는 자기 신호에서 이동된 신호와의 임의의 지점에서의 절대 값 진폭을 합한 값으로 유사성을 파악하게 된다. 좀 더 간단히 말하면 곱하기 더하기 연산으로 구성된 AC 방법대신, 식(3)과 같이 단지 + 연산(- 연산 또한 + (-)연산으로 표현 가능)만으로 구현 가능한 알고리즘을 이용하게 된다. 이와 같은 적은 계산 량의 장점 때문에 AMDF분석은 DSP 칩 기반의 하드웨어와 같은 응용에서 주로 적용되는 실시간 구현 알고리즘으로 널리 이용된다.

$$F(k) = \frac{1}{N} \sum_{m=1}^N |x(m) - x(m+k)| \quad 0 \leq k \leq N-1, \text{ 입력: } 1, \dots, N \quad (3)$$

신호와 k 만큼 이동한 자기신호에 대한 진폭 차에 절대 값 취한 후 샘플 구간에 대한 값을 모두 합하게 되면, <그림 3>의 우측과 같은 분석 도메인 함수를 얻을 수 있다. 또한, 식(4)의 특성 상 신호가 완벽하게 주기적일 경우 항상 거리 차(lag)가 0일 경우 $F(0)$ 값은 0이 된다. 그리고 N_p , 주기에 가까워질수록 결과 값은 $F(0)$ 에 가까운 값을 갖게 된다. 그러므로 피치 검출 방법은 AC 방법과는 달리 피치 검출 구간에서 가장 작은 피크 값에 해당하는 주파수를 기본 주파수로 산출하게 된다.

$$\begin{aligned} &i) F(0)=0 \quad \text{완벽히 주기적이라면} \\ &ii) F(0)=F(N_p) \quad \text{만약 } N_p \text{ 이 주기적이라면} \end{aligned} \quad (4)$$

3.1.2 YIN

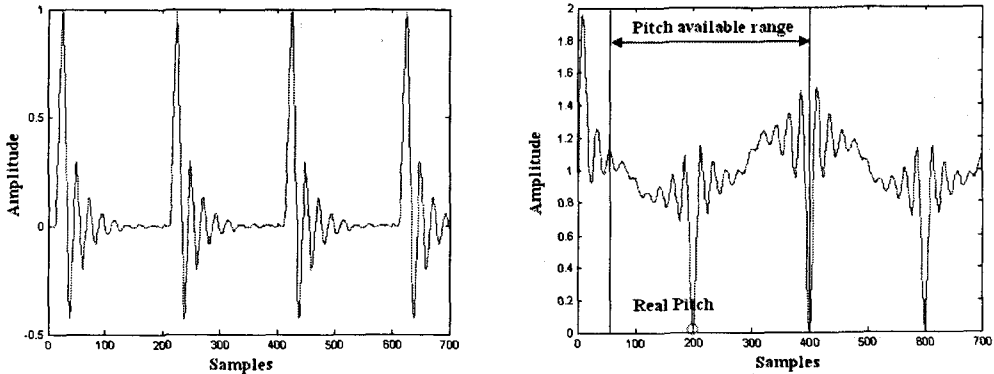


그림 4. 분석에 사용된 음성신호(좌측)와 YIN 피치 추정 방법의 피치 검출 방법(우측)

YIN 방법은 알고리즘 내부에 존재하는 자기상관 알고리즘과 부하모닉 오류를 소거하기 위한 알고리즘 사이의 균형을 조절해서 피치 추정을 수행하는 방법이다. YIN이라는 이름은 동양의 전통적이고 철학적인 균형에 관한 사상인 '음양철학'으로부터 유래되었다. 앞서 <그림 1>의 예에서 살펴 보았듯이, AC 방법의 단점으로 부하모닉의 크기가 주하모닉의 피치 주기 검출에 영향을 끼칠 만큼 커지게 되면 부하모닉 에러가 발생하게 되는데, YIN 방법은 여러 가지 절차적인 방식을 사용하여 이러한 문제들을 해결하게 된다. YIN은 자기상관 함수의 곱셈에 해당하는 수식을 사용하는 대신 식(5)와 같이 AMDF와 같은 방식으로 신호와 이동 신호 사이의 차를 곱한 값에 합을 가지고 분석 도메인의 함수를 산출하게 된다.

$$d(k) = \sum_{m=1}^N (x(m) - x(m+k))^2 \quad 0 \leq k \leq N, \text{ 입력: } 1, \dots, N-1 \quad (5)$$

$$i) \sum_{m=1}^N (x(m) - x(m+N_p))^2 = 0 \quad \text{만약 } x(m) - x(m+N_p) = 0, \forall m.$$

$$ii) d'(k) = \begin{cases} 1, & k=0 \text{의 경우} \\ \frac{d(k)}{\frac{1}{k} \sum_{m=1}^k d(m)} & \text{이외의 경우} \end{cases} \quad (6)$$

이는 식(6)의 특성과 같이 신호를 완벽한 주기, N_p 에 대한 함수로 모델링 할 경우, 신호 $x(m)$ 과 $x(m+N_p)$ 사이의 차는 매 주기 N_p 마다 0이 되기 때문에, 그 곱 또한 0이 되어 결과 값은 0이 된다. 이를 역으로 따져보면 식(5)에 대해 피치 검색이 가능한 구간에서 0과 같거나 0에 가까운 최소값의 함수 결과 값에 해당하는 k 를 검색하여 피치 주기 값을 산출하게 되며, 이러한 결과 <그림 4>의 우측과 같은 그래프를 얻을 수 있다. YIN 방법에 사용되는 절차적인 알고리즘에는 거리 차 함수에서 보다 높은 거리 차에 발생하는 값을 덜 강조하기 위한 누적 평균 함수(Cumulative mean function)를 이용하여 부하모닉 오류를 줄이고 있다. 또한, 피치 검출에 사용되는 윈도우 크기의 요

인에 피치 검출이 이루어지지 않을 경우 에러를 감소시키는 효과를 위하여 국소 최소값(local minima)들에 대한 포물선 보간법(Parabolic Interpolation)을 사용하여 에러 율을 감소시키는 알고리즘도 존재한다.

3.2 주파수 도메인 피치 추정 방법

주파수 영역 기반의 방법들로 Cepstrum(CEP)(Jindong, 2003)과 Simple Inverse Filter Tracking (SIFT)(Market, 1972) 알고리즘을 조사하였다.

3.2.1 Cepstrum (CEP)

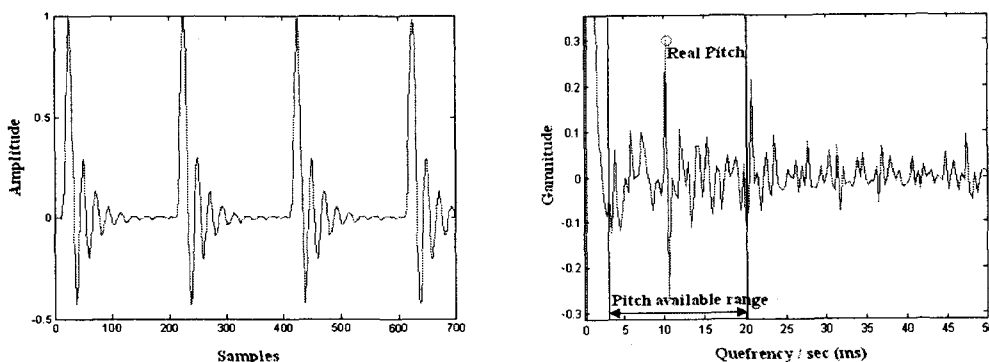


그림 5. 분석에 사용된 음성신호(좌측)와 CEP 피치 추정 방법의 피치 검출 방법(우측)

캡스트럼 피치 추정 방법은 사람의 음성신호로부터 이득(gain) 정보만을 추출하기 위하여 신호를 변환한 후 피치를 검출하는 방법이다. 일반적으로 유성음의 경우 에너지 소스는 성문의 개폐에 의한 임펄스 트레인으로 간주된다. 그러므로 유성음을 이러한 성문에서 발생한 주기적인 여기신호와 모음을 생성하는 성도를 표현하는 전달함수로 구분하여 모델링할 수 있다. 즉, $e(n)$ 을 성문 여기 시퀀스라고 가정하고, $q(n)$ 을 성도를 모델링한 임펄스 응답이라고 가정할 때, 유성음, $s(n)$ 은 $e(n)$ 과 $q(n)$ 의 컨벌루션으로 모델링할 수 있다. 이러한 음성신호를 식 (7)과 같이 FFT를 통하여 주파수 도메인으로 변환시키면 음성신호의 컨벌루션 연산은 곱셈연산으로 변화한다. 이후, $\log AB = \log A + \log B$ 와 같은 log함수의 특성에 의해 곱셈은 덧셈으로 변환될 수 있으며, 이것을 다시 IFFT를 통하여 변환시키게 되면 캡스트럼 계수를 구할 수 있게 된다. 이러한 파워 캡스트럼(Real Cepstrum)의 특성상 성문에서 발생된 주기적인 여기신호가 성도를 거치면서 공명, 반사되는 물리적 현상을 거쳐 시간 지연 및 원신호크기의 감소 등의 형태로 변화되어 측정되었을 지라도 원신호의 복원 및 지연 시간을 추측할 수 있게 해준다. <그림 5>와 같이 피치 검출은 Quefreny 분석 도메인의 피치 검출이 유효한 구간에서 Gamnitude가 가장 큰 값에 해당하는 $1/\text{Quefreny}$ 를 기본 주파수로 선정하게 된다.

하지만, 캡스트럼 시퀀스상의 인덱스 q 에 대한 캡스트럼의 계수 값이 $1/q$ 만큼 감소되기 때문에, 매우 잡음이 심한 음성의 경우 평탄한 잡음 시퀀스의 중간 지점에 위치한 피치 펄스들은 구별이 힘들기 때문에 기본 주파수의 복원이 쉽지 않다. 게다가 파워 캡스트럼의 계산에 필요한 STFT(Short Time Fourier Transform)와 STFT에 필요한 적당한 윈도우 크기의 선택은 피치 검출의 결과에 큰

영향을 주므로 신중히 고려되어야 한다.

$$c(m) = F_{FFT}^{-1}(\log|F_{FFT}(x(m))|) = \frac{1}{2\pi} \int_{-\pi}^{\pi} \log|S(\omega)|e^{j\omega m} d\omega \quad (7)$$

$$S(\omega) = \sum_{m=-\omega}^{\omega} x(m)e^{-j\omega m} \quad \text{입력: } 1, \dots, N-1 \quad (8)$$

3.2.1 Simplified Inverse Filtering Techniques (SIFT)

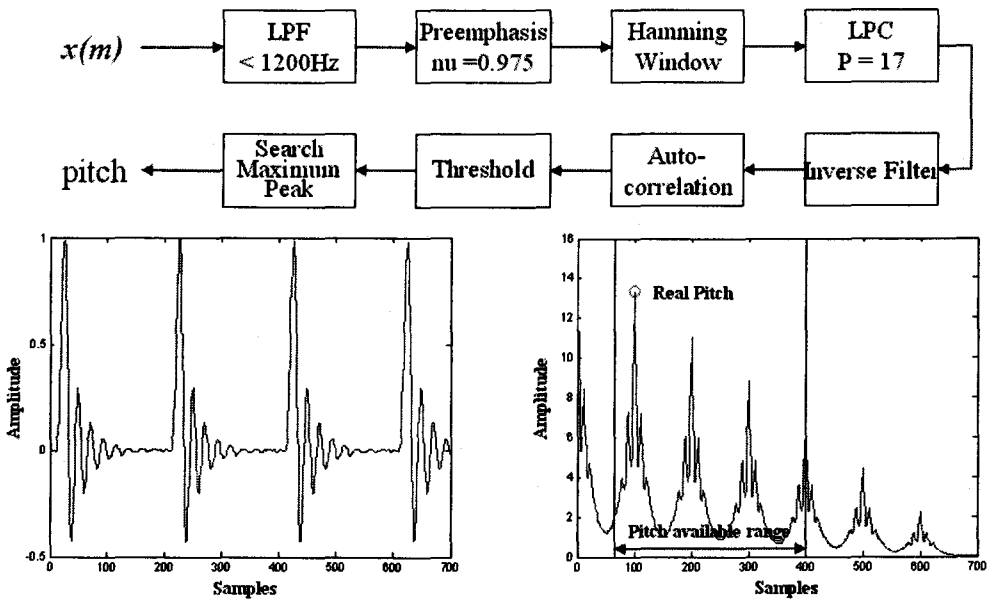


그림 6. SIFT 전체 알고리즘 흐름도(상단) 및 분석에 사용된 음성신호(좌측)와 SIFT 피치 추정 방법의 피치 검출 방법(우측)

SIFT 피치 추정 방법은 선형 예측(Linear Prediction: LP) 알고리즘을 기반으로 데이터를 분석하는 알고리즘이며, 주요 원리는 역 필터(Engin, 2004)를 이용하여 성도의 임펄스 응답을 필터링 시키게 되면 남아 있는 것은 주기적인 성문 여기신호이고, 이를 앞서 언급한 AC 방법을 이용하여 피치를 검출한다는 것이다. 본 연구에서는 일반적인 SIFT에서 사용되는 파라미터를 일부 수정하여 병적인 음성들에 대하여 성능이 향상되도록 조정하였다. 먼저, 저역 통과 필터는 1200 Hz 이하로 설정하여 계산 량을 줄임과 동시에 피치 검출 주파수 대역을 음성 주파수 대역으로 맞추었으며, LP 분석의 특성상 입을 통해 방사되는 음성신호의 점차 감소되는 성도 임펄스 응답을 보정하기 위하여 Preemphasis의 $\nu=0.975$ 로 설정한 후 분석구간에 대하여 해밍 윈도우를 취하고 LP 분석을 통하여 계수를 구하였다. 일반적으로 사용되는 $p=4$ 와 달리 $p=17$ 을 사용하였다. 이 차수 ($p=17$)는 병적인 음성 데이터를 대상으로 차수를 결정하기 위한 MDL(Minimum Description Length) 분석을

통하여 얻어 낸 결과이다. 이렇게 역 필터를 취한 데이터는 주기적인 여기신호를 갖는 유성음의 경우 일정한 패턴을 가지고 있기 때문에 앞에서 언급한 AC 방법을 통하여 주기의 유사성을 파악한 후 <그림 6>의 예와 같이 피치 검색 구간에 대하여 가장 피크 값이 큰 위치에 해당하는 주파수를 검출하면 기본 주파수 값을 산출 할 수 있다.

3.3 대안적 피치 추정 방법

시간, 주파수 도메인이 아닌 다른 분석 방법으로 웨이블릿을 이용한 피치 추정 방법(WAV) (Mallat, 1989; Yisong, 2000)과 시간 delay를 이용한 n 차의 상태 공간 거리 차 분석 (Phase Space: PS)(David, 1999; Dmitry, 2002) 피치 추정 방법을 선택했다.

3.3.1 Wavelet (WAV)

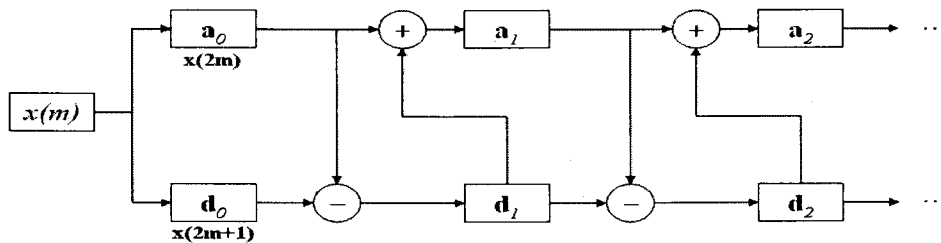


그림 7. 고속 리프팅 웨이블릿 변환(FLWT) 변환 다이어그램

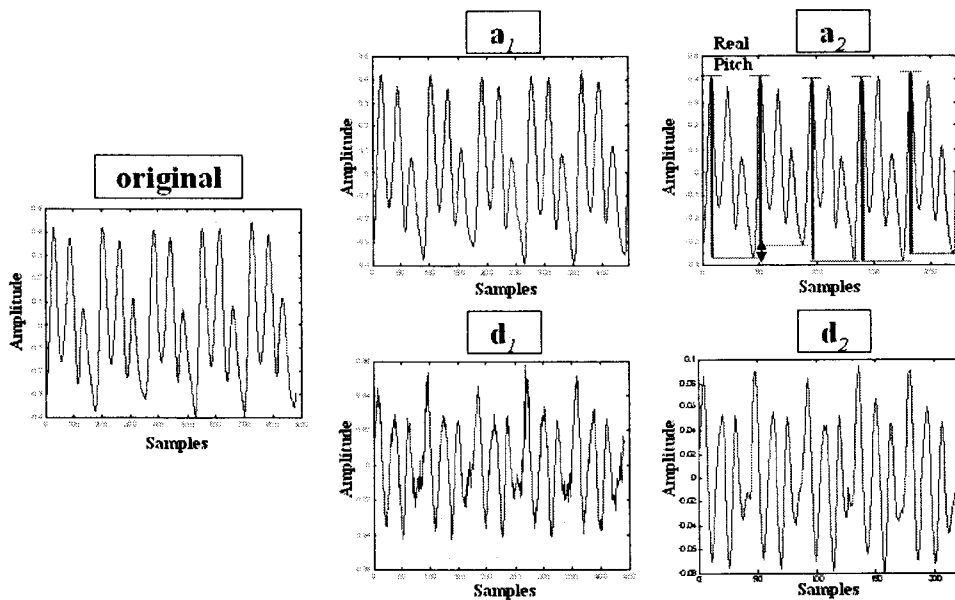


그림 8. 분석에 사용된 음성신호(좌측)와 FLWT에 의해서 단계적으로 분해된 음성 신호 흐름 및 피치 추정 방법의 피치 검출 방법(우측)

웨이블릿 변환은 비정상적인(Non-Stationary) 데이터를 다루는데 적합하고 잡음이 내재된 신호들의 시-주파수 분석에 용이한 분석 도구이다. 뿐만 아니라 알고리즘의 계산 량에 있어서도 상당한 절감이 가능하기 때문에 실시간 알고리즘에도 적용 가능하다는 장점이 있다. 그리고 이분(Dyadic) 웨이블릿의 경우 선형성(Linearity)과 변이 불변성(Shift Invariance)의 특성으로 인해 변이 적이고 진폭이 감소하는 사인 파형들의 선형적인 결합과 유사하게 모델링된 음성 신호들에 대해 좋은 결과를 보이는 분석 방법이다. 본 연구에서는 이산 웨이블릿 변환(Discrete Wavelet Transform: DWT) 중에서 알고리즘 계산량이 단순하고 빠른 고속 리프팅 웨이블릿 변환을(Fast Lifting Wavelet Transform: FLWT) 이용한 피치 검출 알고리즘을 구현하였다. <그림 7>과 같이 DWT 타입의 Harr 웨이블릿을 사용하는 FLWT은 원 신호를 근사(Approximation) 성분과 세부(Detail) 성분으로 분리한다. 이것은 수학적으로 저대역 통과 필터와 다운 샘플링을 산출하는 근사 성분과 고대역 통과 필터와 다운 샘플링을 산출하는 세부 성분으로 원 신호를 분리하게 된다. FLWT를 최대 4번 까지 수행한 이후 <그림 8>과 같이 각각의 최대값과 최소값들에 대한 피크 점들을 검출한 후 각 포인트들 간의 거리 차에 대한 평균값으로 해당 구간의 피치 구간을 설정하게 된다.

3.3.2 Phase Space (PS)

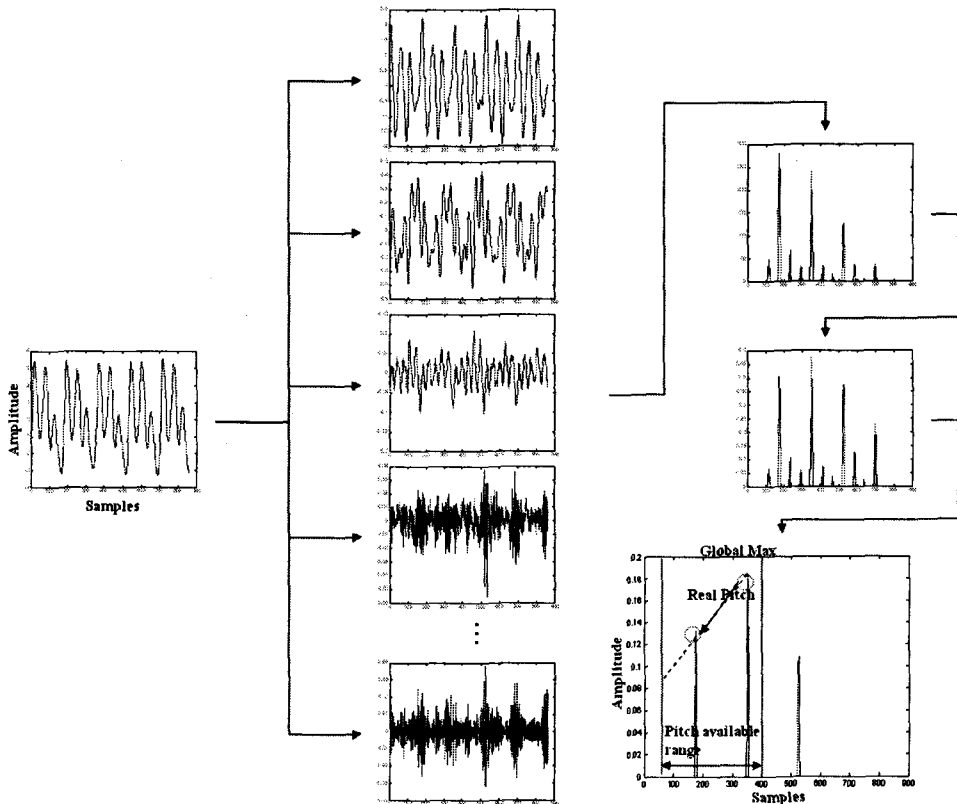


그림 9. 분석에 사용된 음성신호(좌측)와 PS에 의해서 단계적으로 분해된 음성 신호 흐름 및 피치 추정 방법의 피치 검출 방법(우측)

위상 공간(Phase Space:PS) 또는 상태 공간 삽입(State-Space Embedding)으로 더 알려진 PS 방법은 음성신호의 주기적인 사이클을 보다 명확하게 만드는 방식으로 파형의 단 시간적인 추세를 관측하는 방법이다. 기본적인 PS 방법의 분석 방식은 임의의 시간 t에 해당하는 위치에 대한 파형의 기울기를 디스플레이한 후 히스토그램에 의한 분석 결과 주기적인 피치를 검출하게 된다. 만약 신호가 주기적이라면 상태 공간 삽입에 의한 분석 도메인에 있어 동일한 값과 기울기를 갖는 지점들이 반복되는 사이클을 산출하게 된다. 예를 들어 n 차원 매립 차원(Embedding Dimension)으로 음성신호를 재 표현할 때, 신호는 n-1 개의 도함수와 값으로 재해석되게 된다. <그림 9>와 같이 이렇게 재 표현된 주기적인 신호의 히스토그램 분석을 통하여 피치 검출 분석 구간에서 문턱치 이상의 피크 값들 중에서 위치 값이 작은 피크에 해당하는 위치를 피치 주기로 결정한다. 본 연구에서는 병적인 음성 신호에 대한 피치 검출 능력을 향상시키기 위하여 매립 차원은 7, 시간 지연은 19의 계수를 가지고 분석하였다.

4. 피치 추정 평가를 위한 프로토콜

이번 장에서는 피치 추정 평가를 위해 사용되는 음성 신호 데이터베이스 소개 와 기준 피치 값 추출 방법, 음성 신호 분류 및 피치 추정 오류 측정 방법 설명을 하게 될 것이다.

4.1 음성 신호 데이터베이스

연세대학교 영동세브란스 병원 음성언어의학연구소에서 2003년부터 2007년 사이에 치료된 99명에 대한 양성후두 질환군의 치료 전 음성과 비교를 위한 정상군 30명의 음성을 대상으로 피치 추정 능력을 평가하였다. 음성질환은 후두에 외형적인 변형이 발생한 양성후두 질환군(용종, 낭종, 결절) 음성만을 대상으로 검사하였다. 이러한 이유는 양성후두 질환이 흔하게 발병할 수 있는 질환이며, 성대에 물리적인 변형으로 인해 비주기적이고 비정상적인 음성 신호를 자주 산출하는 음성이기 때문에 선택하였다. 성별에 상관없이 양성후두 질환군의 나이 분포는 18~71세이며, 평균 나이는 43.7세이고, 정상군의 경우 나이 분포는 21~68세이며, 평균 나이는 48.4세이다. 음성 data는 잡음이 없는 밀폐된 환경에서 컴퓨터 화된 음성녹음 장비를(Kay Elemetrics CSL) 이용하여 음성 신호와 EGG 신호를 동시에 수집하였다. EGG 신호는 차후에 얘기되겠지만 기준 피치 검출을 위하여 취득되어진다. 모든 음성신호와 EGG data는 22 kHz로 샘플링 되어졌고 16bit 분해능을 갖고 있으며, 실험 프로토콜에 의해 최소 2-3 초간 유지되는 모음(/a/, /e/, /i/, /o/, /u/) 샘플들을 각각 1 초 간격으로 2 회 발생시켜서 분석하였다. 신호 분석은 PC기반의 Matlab 7.1을 이용하여 각 피치 추정 알고리즘들을 구현한 후 평가하였다.

4.2 기준 피치 값 선정 및 신호 분류

4.2.1 기준 피치 값 선정

정확한 성문 폐쇄 순간(Glottal Closure Instance: GCI)(Ananthapadmanabha, 1979)은 음성 신호의 피치 주기와 관련된 EGG 신호로부터 취득할 수 있다. 음성신호가 복잡한 형태의 신호를 보일지

라도 EGG 신호는 좀 더 단순화된 패턴으로 음성신호의 주기를 반영하게 된다. 그러므로 거의 주기가 일정한 음성 신호와 관련된 EGG신호를 미분한(Differentiated) DEGG신호를 기반으로 기준(reference) 피치를 검출하였다.

하지만 성문에 외형적 변형으로 인해 EGG 신호에 잡음이 많이 내재되고 패턴이 불규칙하며 주기가 일정하지 않은 경우 전문가에 의해 육안을 통한 수동적인 피치 검출 방법을 통하여 기준 피치를 검출하였다. 이러한 경우 자동화된 알고리즘을 통하여 선정된 점을 기준으로 허용된 구간(± 25 샘플 범위) 안에서 진폭이 가장 큰 점을 피치 피크로 선정하고 이후 반복하여 검증하였다.

4.2.2 비주기성에 따른 음성 신호 분류

정상인 음성에 비하여 음성질환자의 음성은 신호의 패턴 및 주기가 상이하기 때문에 이를 단계적으로 분류하여 분석해 볼 필요가 있다. Titze(1995)의 음성 신호 분류 기준에 의하면 음성 신호는 다음과 같이 Type-1, 2, 3으로 분류될 수 있다. 각각의 Type 분류에 대한 정의를 살펴보면, Type-1의 음성 신호는 분석 구간에서 물리적인 변화가 거의 없이 주기적인 신호를 의미한다. Type-2의 음성 신호는 기본주파수의 에너지에 근접하는 변조된 주파수 또는 서브주파수를 갖는 신호에 의해 분석 구간에서 변화가 발생하는 신호로 정의된다. 마지막으로, Type-3신호는 전혀 주기적인 신호를 알 수 없는 혼돈된(Chaotic) 신호로 정의된다. 이러한 정의에 의해 Titze(1995)의 연구결과 오직 Type-1의 음성신호일 경우에 한해서 피치 검출 값의 유효성이 신뢰될 수 있다고 주장하였다.

본 연구에서는 Titze(1995)의 Type에 따른 음성신호 분류에 따라 각각의 음성신호를 Type-1, 2, 3로 구분하였으며, 구분기준은 비선형 동역학 기반의 Lyapunov Exponents(Kumar, 1996)를 이용하여 계수가 1 이상인 경우 Type-3로, 아닌 경우 Type-1,2로 구분하여 분류한 후, 다시 <그림 10>과 같이 스펙트로그램을 이용하여 분기가 발생한 음성신호의 경우를 Type-2로 아닌 경우 Type-1으로 분류하였다. 그리하여, 본 실험에서는 Lyapunov Exponents와 스펙트로그램 분석을 통하여 Type-3로 판명된 혼동성(Chaotic)이 강한 음성 신호는 제외하고 실험을 수행하였다.

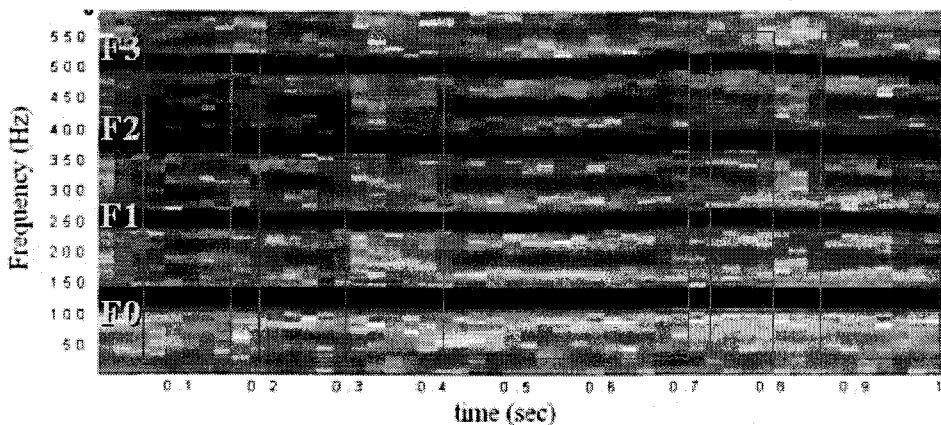


그림 10. Type-2 신호 EGG의 기본주파수(위), 스펙트럼 분석(아래) 스펙트럼 분석에서 0.1s, 0.2s, 0.4s, 0.7s, 0.85s에서 분기점을 볼 수 있다.

4.3 피치 추정 오류 측정 방법

피치 추정 알고리즘의 성능 평가를 위한 피치 추정 에러 측정 방법을 아래와 같이 7 가지 타입으로 정의하였다.

- 무성음 피치 오인식 에러 (Voiced/Unvoiced: V/U): 실제 무성음이지만 유성음으로 잘못 인식한 경우 발생한 에러이며, 아래와 같은 수식으로 표현될 수 있다.

$$V/U\text{에러}(\%) = \frac{\text{무성음 피치 오인식 발생 분석 구간 전체 숫자}}{\text{실제 무성음 구간의 전체 숫자}} \times 100(\%)$$

- 유성음 피치 오인식 에러 (Unvoiced/Voiced: U/V): 실제 유성음이지만 무성음으로 잘못 인식한 경우 발생한 에러이며, 아래와 같은 수식으로 표현 될 수 있다.

$$U/V\text{에러}(\%) = \frac{\text{유성음 피치 오인식 발생 분석 구간 전체 숫자}}{\text{실제 유성음 구간의 전체 숫자}} \times 100(\%)$$

- 피치 배가 에러 (Pitch Doubling Error: X2): 피치 추정 알고리즘에 의하여 계산된 피치 추정 값이 기준 피치 추정 값보다 20% 높게 추정 된 경우 발생하는 에러이며, 아래와 같은 수식으로 표현될 수 있다.

$$X2\text{ 검출 기준} = \frac{\text{피치 예측 방법에 의해 계산된 값} - \text{기준 피치 예측 값}}{\text{기준 피치 예측 값}} > 0.2$$

$$X2\text{에러}(\%) = \frac{X2\text{검출 기준에 해당하는 분석구간 발생 전체 숫자}}{\text{분석 구간의 전체 숫자}} \times 100(\%)$$

- 피치 반감 에러 (Pitch Halving Error: /2): 피치 추정 알고리즘에 의하여 계산된 피치 추정 값이 기준 피치 추정 값보다 20% 낮게 추정된 경우 발생하는 에러이며, 아래와 같은 수식으로 표현될 수 있다.

$$/2\text{ 검출 기준} = \frac{\text{피치 예측 방법에 의해 계산된 값} - \text{기준 피치 예측 값}}{\text{기준 피치 예측 값}} \leq -0.2$$

$$/2\text{에러}(\%) = \frac{/2\text{검출 기준에 해당하는 분석구간 발생 전체 숫자}}{\text{분석 구간의 전체 숫자}} \times 100(\%)$$

- 피치 과대 에러 (Pitch Gross Error: G): 위에서 언급한 피치 배가 에러와 피치 반감 에러를 합한 값으로서 기준 피치 추정 값보다 계산된 피치 추정 값이 절대치 20% 이상 차이가 발생한 경우를 의미한다.

$$G\text{ 에러}(\%) = X2\text{ 에러}(\%) + /2\text{ 에러}(\%)$$

- 피치 미세 에러 (Pitch Fine Error: F): 위에서 언급된 피치 에러들이 검출된 음성신호 분석 구간과는 배제된 유성음 구간에 한하여 계산된 피치 추정 값과 기준 피치 추정 값과의 차에 대한 절대 값을 기준 피치 추정 값에 대한 백분율로 계산한 값들의 평균으로 정의한다.

$$F \text{ 에러 } (\%) = \frac{|\text{피치 예측 방법에 의해 계산된 값} - \text{기준 피치 예측 값}|}{\text{기준 피치 예측 값}} \times 100 (\%)$$

∈ G 에러 발생 구간과 배제된 유성음 분석 구간

- 피치 분산 에러 (Pitch Standard Deviation Error: S) F 에러에 의해 정의된 분석구간과 계산 방법에 대한 표준 편차 값을 의미한다.

$$S \text{ 에러 } (\%) = \text{std} \left(\frac{|\text{피치 예측 방법에 의해 계산된 값} - \text{기준 피치 예측 값}|}{\text{기준 피치 예측 값}} \times 100 \right) (\%)$$

∈ G 에러 발생 구간과 배제된 유성음 분석 구간

5. 음성질환 및 정상인 대상 피치 추정 실험 결과

5.1 최적 윈도우 크기 결정

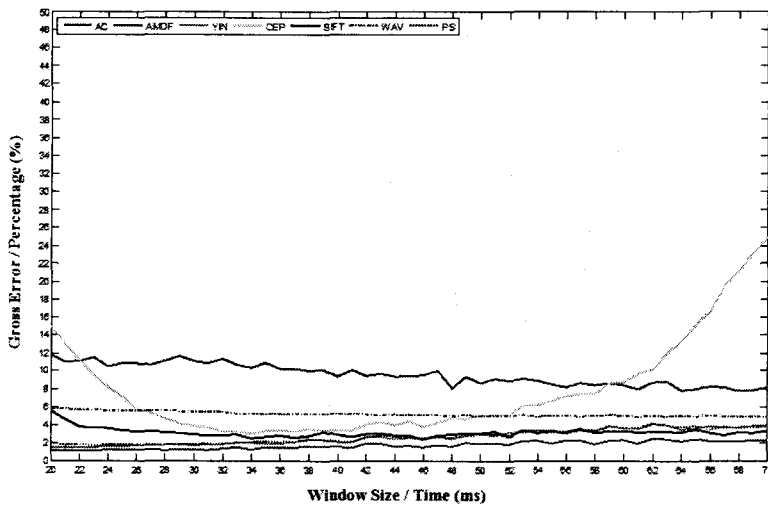


그림 11. 정상군을 대상으로 한 윈도우 크기에 따른 피치 과대 에러 함수 그래프

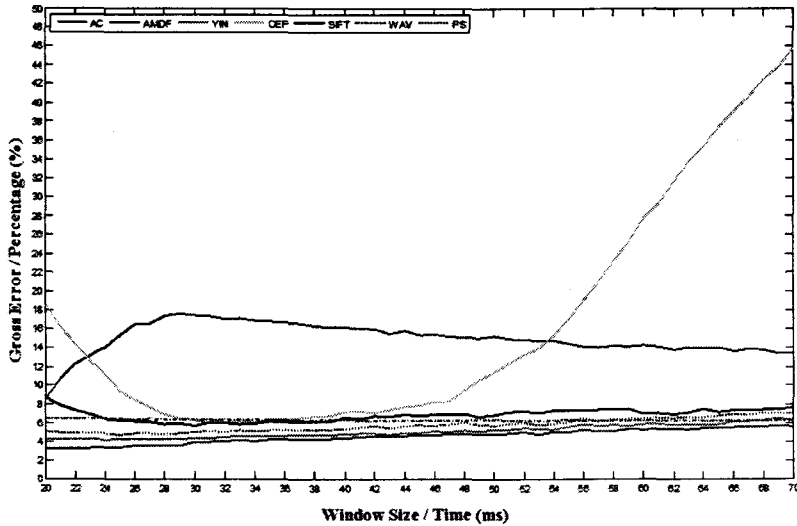


그림 12. BVFL군을 대상으로 한 윈도우 크기에 따른 피치 과대 에러 함수 그래프

각 피치 추정 방법들은 각자의 도메인에 따라 피치를 추정하기 때문에 공통적으로 최적화된 피치 검출 성능을 보이는 분석 구간 크기를 갖기 보다는 각각의 분석 구간의 크기에 따라 다른 피치 추정 성능이 보이게 된다. 그렇기 때문에 정상군과 BVFL군을 대상으로 최적의 성능을 보이는 윈도우 크기 (분석구간 크기)를 분석한 결과 <그림 11, 12>와 같은 결과를 도출하였다. 최적 윈도우 크기 결정에 피치 과대 에러가 사용된 이유는 이 피치 에러가 피치 추정 값 오류에 가장 큰 영향을 미치기 때문이다. 결과를 분석해 볼 때 정상군에서 최적 윈도우 크기는 AC(22), AMDF(68), YIN(20), CEP(34), SIFT(46), WAV(70), PS(23)이고, BVFL군에서는 AC(21), AMDF(20), YIN(24), CEP(31), SIFT(30), WAV(59), PS(25)였다. 윈도우 크기의 단위는 ms이다. 결과적으로 BVFL군에서도 각 알고리즘의 윈도우 크기 패턴을 그려 본 결과 정상군과 비슷하였고, AMDF, SIFT와 WAV를 제외하고는 두 그룹의 최적 윈도우 크기의 차이는 크지 않음을 알 수 있었다. 특이할 만한 사항은 정상군에 비해 BVFL군의 피치 과대 에러가 전반적으로 높아졌다는 것을 알 수 있었으며, AMDF, CEP 피치 추정 방법의 성능이 윈도우 크기에 매우 민감한 결과를 보였다. 이후 각 알고리즘의 최적화된 성능을 비교하기 위해서 각 피치 추정 방법에 대한 각자의 최적 윈도우 크기를 가지고 피치 추정 관련 성능을 평가하였다.

5.2 피치 에러 분석

5.2.1 양성 후두 질환과 정상 음성에 대한 피치 추정 에러 비교 분석

<표 1>은 BVFL군과 정상군의 음성신호에 대한 각 피치 추정 방법들의 성능 평가를 앞서 언급한 7가지 피치 추정 에러를 가지고 통계적으로 분석한 결과이다. 정상군 30 명과 BVFL 군 99 명의 /a/, /e/, /i/, /o/, /u/ 5 가지 지속 모음 중에서 Type-3 음성 신호를 제외한 정상군(NNormal=150)과 BVFL군(NBVFL=447)에 대한 각 피치 추정 방법의 성능을 평가 하였다. <표 1>에 사용되는 유의 확률, p 는 95%의 유의 수준을 갖는 Welch t-검정을 사용하여 정상군과 BVFL 군의 평균을 비교한 결과 이다.

그리고 <표 1>에서 음영 처리된 부분은 각 열에서 가장 성능이 좋은 알고리즘을 나타낸다. V/U에러는 각 피치 추정방법들에게 동일한 휴지 감지기(Silence Detector)를 적용하였기 때문에 비슷한 결과 값이 산출되었다. 다만, 지속 모음이 시작되거나 끝나는 말단 부분에서 실제 무성음 구간이나 유성음 구간으로 오인식하여 잘못 검출한 다소 차이가 발생함을 알 수 있었다. 반면에 U/V에러는 V/U 에러에 비하여 크게 발생함을 알 수 있었다. 특히 주파수 도메인 기반은 피치 추정 알고리즘인 CEP 또는 SIFT에서 상대적으로 크게 발생함을 알 수 있었다. 모든 항목의 피치 에러 분석에 있어 정상군의 음성신호에 비하여 BVFL군의 음성신호에 대한 피치 에러가 전반적으로 증가하고 통계적으로 유의하게 차이가 발생함을 알 수 있었으며, 이로 미루어 보아 정상인 음성에 비해 양성 후두 질환자의 음성에서 정확한 피치를 검출하는 것이 더 어렵다는 것을 알 수 있었다. 결과에서 살펴 볼 때 AC 방법이 피치 과대에러 와 피치 반감 에러에서 좋은 성능을 보여주고, WAV 방법은 피치 배가 에러와 피치 미세 에러 및 피치 분산 에러에서 강점을 보였다. 반면에 피치 과대 에러에서는 AMDF와 WAV가 좋지 않은 성능을 보였고, 피치 미세 에러에서는 YIN 방법이 좋지 않은 결과를 보였다. 비록 WAV가 우수한 피치 미세 에러를 보이지만 높은 피치 과대 에러를 산출하기 때문에 전반적으로 AC 알고리즘이 피치 과대 에러 및 피치 미세 에러에 대하여 안정적이고 강건한 피치 추정 방법으로 볼 수 있다.

표 1. 정상군과 BVFL군 지속모음을 대상으로 한 피치 추정 방법들의 성능 비교 결과

PE Methods Error Type	AC		AMDF		YIN		CEP		SIFT		WAV		PS		
	Normal	BVFL	Normal	BVFL	Normal	BVFL	Normal	BVFL	Normal	BVFL	Normal	BVFL	Normal	BVFL	
V/U	Mean	0.05	0.02	0.05	0.03	0.00	0.00	0.07	0.04	0.04	0.03	0.01	0.08	0.01	0.02
	S.D.	0.03	0.01	0.04	0.02	0.00	0.00	0.02	0.01	0.01	0.02	0.00	0.05	0.00	0.01
	p	.002		.007		0.997		0.045		0.294		0.007		0.068	
U/V	Mean	0.15	0.32	0.47	0.54	0.35	0.42	4.95	5.93	1.27	1.90	0.07	0.18	0.76	0.93
	S.D.	0.08	0.11	0.12	0.34	0.29	0.36	2.91	3.48	0.72	0.58	0.04	0.13	0.52	0.66
	p	<.001		.149		.118		.114		.201		.044		.159	
G	Mean	1.07	3.20	7.69	8.66	1.43	4.12	2.97	5.98	2.44	5.72	4.89	6.27	1.68	4.69
	S.D.	1.31	5.99	10.72	12.88	1.59	6.97	3.67	9.08	2.73	9.42	2.10	1.91	1.84	7.46
	p	<.001		.008		<.001		<.001		<.001		.015		<.001	
X2	Mean	0.13	0.41	0.15	1.06	0.53	0.66	1.30	2.48	0.66	0.97	0.01	0.01	0.63	1.27
	S.D.	0.45	2.27	0.81	6.14	0.82	2.02	3.58	6.98	2.06	3.14	0.07	0.09	1.50	4.64
	p	.009		.005		.085		.017		.181		.276		.031	
f2	Mean	0.94	2.80	7.55	7.60	0.90	3.46	1.67	3.51	1.78	4.75	4.88	6.25	1.05	3.42
	S.D.	10.22	5.45	10.77	11.85	1.37	6.29	1.83	4.51	1.98	8.10	2.10	1.91	1.40	5.47
	p	<.001		.017		<.001		<.001		<.001		.012		<.001	
F	Mean	1.31	2.14	6.35	7.53	14.01	14.75	3.49	5.56	4.83	6.44	0.39	0.83	1.90	3.93
	S.D.	0.82	3.69	7.63	9.06	0.54	3.52	5.56	11.89	2.40	8.69	0.76	1.40	10.88	6.57
	p	<.001		.002		<.001		.005		<.001		<.001		<.001	
S	Mean	2.61	4.48	11.17	11.35	2.63	4.56	7.90	10.73	6.34	8.87	0.44	0.58	5.45	8.15
	S.D.	3.91	7.22	10.58	10.16	2.32	5.88	10.93	13.17	9.07	11.85	0.82	0.97	6.27	8.73
	p	<.001		.668		<.001		.012		.071		<.001		.002	

5.2.2 비주기성 정도에 따른 음성신호에 대한 피치 추정 에러 비교 분석

이번 실험에서는 정상군과 BVFL군 상관없이 음성신호의 비주기성 정도에 대한 피치 추정 에러를 분석하였다. Type-3 신호를 제외한 Type-1 ($N_{\text{Type-1}}=419$)과 Type-2 ($N_{\text{Type-2}}=178$) 음성 신호를 대상으로 실험을 수행하였다. 우선 비주기성에 대한 분류는 앞에 언급한 바와 같이 Lyapunov Exponents 계수가 1이하 이고 스펙트로그램 상의 분기가 발생한 경우에는 Type-2로 아닌 경우는 Type-1 음성 신호로 분류하여 검사하였다. 즉, 이 분석은 피치 배가 또는 반감이 빈번하게 검출되는 음성 신호에 대한 각 피치 추정 방법의 강건성(Robustness)을 테스트하기 위하여 수행하였다. <표 2>의 결과에서 보듯이 피치 과대 에러의 경우 등분산이 가정되지 않은 평균에 대한 95%의 유의 수준을 갖는 Welch t-검정 결과 AMDF를 제외한 모든 피치 추정 방법들이 Type-2에서 더 높은 에러를 산출하였으며, 특히 피치 반감 에러로 인해 이러한 차이가 발생하였다. 반면에 피치 미세 에러의 경우 CEP, WAV, PS를 제외한 나머지 피치 추정 방법들은 Type-1과 Type-2의 통계적으로 유효할 만큼 크게 차이가 나지 않았다. 이러한 결과는 피치 분산 에러에서도 비슷한 경향을 보였다. 피치 과대 에러 측면에서 볼 때 BVFL군과 정상군 비교 실험에서와 마찬가지로 AC 알고리즘이 가장 우수한 성능을 보인 반면, AMDF가 좋지 않은 성능을 보였다. 피치 미세 에러 또한 WAV 방법이 가장 좋은 성능을 보인 반면, YIN 방법이 좋지 않은 성능을 나타내었다. 이로 미루어 보아 AC 알고리즘이 비주기성을 갖는 음성 신호 분석에서도 가장 성능이 우수한 피치 추정 방법으로 간주될 수 있다.

표 2. Type-1, 2 지속모음을 대상으로 한 피치 추정 방법들의 성능 비교 결과

PE Methods Error Type	AC		AMDF		YIN		CEP		SIFT		WAV		PS		
	Type-1	Type-2	Type-1	Type-2	Type-1	Type-2	Type-1	Type-2	Type-1	Type-2	Type-1	Type-2	Type-1	Type-2	
V/U	Mean	0.04	0.02	0.05	0.04	0.00	0.00	0.06	0.04	0.03	0.03	0.03	0.06	0.01	0.01
	S.D.	0.03	0.02	0.03	0.02	0.00	0.00	0.02	0.03	0.02	0.02	0.02	0.05	0.01	0.01
	p	.042		.034		.998		0.077		0.702		0.172		0.838	
U/V	Mean	0.24	0.27	0.39	0.44	0.33	0.45	4.95	4.93	0.97	1.74	0.09	0.16	0.56	1.23
	S.D.	0.11	0.10	0.22	0.34	0.27	0.30	10.9	2.34	1.25	0.71	0.05	0.11	0.42	0.97
	p	.354		.620		.094		.914		.032		.039		.009	
G	Mean	1.12	2.16	8.20	8.63	1.44	2.58	2.56	3.80	2.37	3.53	4.99	6.26	1.58	3.06
	S.D.	1.34	3.98	10.98	12.38	1.61	4.29	2.48	4.46	2.77	3.79	2.08	1.80	1.55	4.07
	p	<.001		.706		<.001		<.001		<.001		<.001		<.001	
X2	Mean	0.11	0.21	0.09	0.73	0.49	0.35	0.80	1.15	0.60	0.58	0.00	0.01	0.47	0.55
	S.D.	0.40	1.27	0.63	5.63	0.79	0.78	2.14	2.57	2.09	1.62	0.05	0.08	0.94	0.99
	p	.168		.029		.074		.124		.913		.440		.402	
/2	Mean	1.01	1.96	8.11	7.90	0.95	2.23	1.76	2.66	1.77	2.95	4.98	6.25	1.11	2.51
	S.D.	1.24	3.77	11.00	11.45	1.40	4.26	1.85	3.39	2.00	3.44	2.08	1.80	1.44	4.02
	p	<.001		.847		<.001		<.001		<.001		<.001		<.001	
F	Mean	1.32	1.55	6.68	7.42	14.01	14.10	2.73	3.33	4.75	4.66	0.41	0.73	1.70	2.59
	S.D.	0.81	2.34	7.79	8.61	0.55	2.07	3.22	4.95	2.25	2.75	0.79	1.29	1.55	3.72
	p	.099		.351		.470		.110		.702		.001		<.001	
S	Mean	2.56	3.34	11.54	11.29	2.60	3.38	6.51	7.99	5.57	6.39	0.44	0.57	5.01	6.59
	S.D.	3.91	5.57	10.41	10.22	2.37	3.63	8.93	10.66	7.87	7.12	0.85	0.99	5.73	7.04
	p	.075		.804		.005		.113		.284		.147		.010	

6. 결론 및 고찰

본 논문에서는 음성언어 장애 분석을 하기에 알맞은 피치 추정방법을 찾아내기 위하여, 기존의 잘 알려진 피치 추정 방법들(Autocorrelation, AMDF, YIN, Cepstrum, SIFT, Wavlet 및 Phase Space)에 대한 피치 검출 성능을 비교 평가하였다. 정상인(n=30)과 양성후두 질환자(n=99)의 지속 모음 /a/, /e/, /i/, /o/, /u/를 대상으로 하여 피치 과대, 반감, 배가, 미세, 분산 에러에 대한 평가를 수행하였다. 정상군 및 양성후두 질환군 모두에서 피치 과대 에러에 대해서는 Autocorrelation 방법이 우수한 성능을 보였다. 또한 피치 미세 에러에 대해서는 Wavelet 방법이 우수한 성능을 보였으나, 종합적으로 피치 과대 에러와 미세 에러를 고려할 때, Autocorrelation 방법이 우수한 피치 추정 방법으로 결론내릴 수 있다. 또한 Titze의 음성 분류 기준에 의한 비주기성 정도에 따른 음성 신호의 피치 추정 성능 분석 결과 음성 신호의 비주기성이 증가할수록 피치 에러들 또한 증가함을 알 수 있었다. 이 실험 결과에서도 AC 방법이 강건한 피치 추정 방법으로 선정되었다.

피치 추정은 음성의 기본 주파수를 검출함으로써 병적인 음성에 대한 정도를 파악하는데 유용하고 다양한 파라미터들을 산출해낸다. 우수한 피치 추정 방법이라면 몇 가지 가져야 할 특성이 있는데, 음성신호가 비주기적이고 비정상적일지라도 피치 과대 에러가 낮아야 하고, 상호 보완적인 피치 미세 에러 또한 낮아야 한다. 특히, 피치 추정 방법의 성능 개선 문제에 있어 이 두 가지 에러는 기회비용(trade-off)의 관계를 갖는다[13]. 즉, 한 쪽 에러를 낮추면 다른 쪽 에러가 높아지는 성질을 갖기 때문에 성능 향상을 위한 선택의 문제에 직면하게 된다. 본 연구에서는 피치 과대 에러를 중심으로 피치 추정 에러의 성능을 평가하였지만, 향후 연구에서는 이 두 가지 에러를 모두 고려한 효율적인 피치 추정 방법을 평가하는 것 또한 필요하리라 생각된다.

참 고 문 헌

- Ananthapadmanabha, T. & Yegnanarayana, B. 1979. "Epoch extraction from linear prediction residual for identification of closed glottis interval." *IEEE Trans. Acoust., Spch, and Sag. Proc.* 27, 309-319.
- Bielamowicz, S., Kreiman, J., Gerratt, B. R., Dauer, M. S. & Berke, G. S. 1996. "Comparison of voice analysis systems for perturbation measurements." *J. Speech Hear. Res.* 39, 126-134.
- Cheveigne, A. D. & Kawahara, H. 2002. "Yin: A fundamental frequency estimator for speech and music." *J. Acoust. Soc. Am.* 111, 1917-1930.
- David, G. 1999. "Audio visualization in phase space." In *Bridges, mathematical connections in art, music and science*, 137-144.
- Deller, J. R., Proakis, J. G. & Hansen, J. H. L. 1993. *Example short-term features and applications, discrete-time processing of speech signals*. New York: Macmillan.
- Dmitry, T. 2002. "Fundamental frequency estimation using signal embedding in state space." *J. Acoust. Soc. Am.*, 112, 2279.
- Dmitry, T. 2002. "Robust pitch determination using nonlinear state-space embedding." In

- International Conference on Acoustics, Speech and Signal Processing I*, 345-348.
- Engin, T. T. 2004. "Deconvolution and preequalization with best delay LS inverse filters." *Signal Processing* 84, 2207-2219.
- Fant, C. 1956. "On the predictability of formant levels and spectrum envelopes from formant frequencies." In M. Halle, H. Lunt, and H. MacLean (eds.), *For Roman Jakobson*. The Hague: Mouton.
- Feijoo, S. & Hernandez, C. 1990. "Short-term stability measures for the evaluation of vocal quality." *J. Speech Hear. Res.* 33, 324-334.
- Gold, B. & Morgan, N. 1999. *Speech and audio signal processing: processing and perception of speech and music*. New York: Wiley.
- Jindong, C., Kuldeep, K. P., Satoshi, N. 2003. "Cepstrum derived from differentiated power spectrum for robust speech recognition." *Speech Communication* 41, 469-484.
- Karnell, M. P., Scherer, R. S. & Fischer, L. 1991. "Comparison of acoustic voice perturbation measures among three independent voice laboratories." *J. Speech Hear. Res.* 34, 781-790.
- Karnell, M. P., Hall, K. D. & Landahl, K. 1995. "Comparison of fundamental frequency and perturbation measurements among three analysis systems." *J. Voice* 4, 383-393.
- Klapuri, A. & Davy, M. 2006. *Signal processing methods for music transcription*. New York: Springer-Verlag.
- Klingholz, F. & Martin, F. 1985. "Quantitative spectral evaluation of shimmer and jitter." *J. Speech Hear. Res.* 28, 169-174.
- Kumar, A. & Mullick, S. K. 1996. "Nonlinear dynamical analysis of speech" *J. Acoust. Soc. Am.*, 100, 615 - 629.
- Laver, J. 1994. *Principles of phonetics*. Cambridge: Cambridge University Press.
- Mallat, S. G. 1989. "A theory for multiresolution signal decomposition: the wavelet representation." *IEEE Trans Patt Anal Mach Intell* 11, 674 - 693.
- Market, J. D. 1972. "The SIFT algorithm for fundamental frequency estimation." *IEEE Trans. Audio Electroacoust* 20, 367-377.
- Rabiner, L. R. 1977. "On the use of autocorrelation analysis for pitch detection." *IEEE Trans. Acoust. Speech Signal Process.* 25, 24-33.
- Sondhi, M. M. 1968. "New methods of pitch determination." *IEEE Trans. Audio Electroacoust.* 16, 262-266.
- Titze, I. R. 1995. *Workshop on Acoustic Voice Analysis, Summary Statement*. National Center for Voice and Speech, Denver.
- Yisong, D. 2000. "The time-frequency analysis approach of electric noise based on the wavelet transform." *Solid-State Electronics* 44, 2147-2153.

접수일자: 2007. 10. 8

게재결정: 2007. 11. 28

▲ 장승진

강원도 원주시 흥업면 매지리
 첨단의료기기테크노타워 201호 (우: 220-842)
 Tel: +82-33-760-2809 Fax: +82-33-763-1953
 E-mail: highnoon@yonsei.ac.kr

▲ 김효민

강원도 원주시 흥업면 매지리
 첨단의료기기테크노타워 201호 (우: 220-842)
 Tel: +82-33-760-2809 Fax: +82-33-763-1953
 E-mail: highnoon@yonsei.ac.kr

▲ 최성희

Dept of Surgery, Division of Otolaryngology-Head & Neck Surgery,
 K4/719 Clinical Science Center, 600 Highland Avenue,
 University of Wisconsin School of Medicine and
 Public Health, Madison, WI, USA (우: 53792)
 E-mail: shgrace67@hanmail.net

▲ 박영철

강원도 원주시 흥업면 매지리
 연세대학교 원주캠퍼스 창조관 241호 (우: 220-842)
 Tel: +82-33-760-2744 Fax: +82-33-763-4323
 E-mail: young00@yonsei.ac.kr

▲ 최홍식

서울시 강남구 도곡동 146-92
 연세대학교 영동세브란스병원 음성언어의학연구소 (우: 135-720)
 Tel: +82-2-3497-3460 Fax: +82-2-3463-4750
 E-mail: highnoon@yonsei.ac.kr

▲ 윤영로

강원도 원주시 흥업면 매지리
 첨단의료기기테크노타워 201호 (우: 220-842)
 Tel: +82-33-760-2809 Fax: +82-33-763-1953
 E-mail: yoon@yonsei.ac.kr