

2-채널(음성 및 EGG) 신호 분석에 의한 피치검출

Pitch Detection by the Analysis of Speech and EGG Signals

신 무 용*, 김 정 철**, 배 건 성**

(Mu Yong Shin*, Jeong Cheol Kim**, Kcun Sung Bae**)

*본 연구는 한국학술진흥재단의 '92 자유공모과제 연구비 지원에 의하여 수행되었습니다.

요 약

본 연구에서는 음성 및 EGG 신호를 이용한 2-채널 피치검출 알고리즘을 제안하였다. 성대의 떨림에 관한 정확한 정보를 얻을 수 있는 EGG 신호를 이용함으로써 음성신호로부터 피치를 검출하고자 할 때 수반되는 문제점들을 최소화 할 수 있으며, EGG 신호의 왜곡 및 불규칙한 변화는 음성신호의 분석을 통해 보완해 줌으로써 시간영역에서 음성신호에 동기된 정확한 피치 marker를 검출하였다. 2-채널 피치검출 알고리즘은 음성신호만을 이용한 일반적인 피치검출 알고리즘에 비해 보다 정확하고 개선된 피치제적을 얻을 수 있음을 실험을 통해 보였으며, 따라서 새로이 개발되는 피치검출 알고리즘의 객관적인 비교 및 평가에 이용될 수 있다.

ABSTRACT

We propose a two-channel(Speech & EGG) pitch detection algorithm. The EGG signal monitors the vibratory motion of vocal folds very well. Therefore, using the EGG signal as well as speech signal, we obtain a reliable and robust pitch detection algorithm that minimizes problems occurring in the pitch detection with speech only. The proposed algorithm gives precise pitch markers that are synchronized to the speech in the time domain. Experimental results demonstrate the superiority of the two-channel pitch detection algorithm over the conventional method, and it can be used in obtaining reference pitch for evaluation of other pitch detection algorithms.

I. 서 론

음성신호는 인간의 發聲기관에서 2개의 얇은막으로 구성된 聲帶(vocal folds)의 자발적인 운동에 의해 발생하는 공기의 흐름이 聲道(vocal tract)를 지나면서 변조되어 공기압의 波動형태로 나타나는 것이다. 이러한 음성신호는 성대를 통과한 공기흐름의 성질에 따라 크게 有聲음(voiced sound)과 無聲음(unvoiced sound)으로 나눌 수 있다. 보음과 같은 유성음을 발생할 경우 허파에서 방출되는 공기는 닫혀진 성대에 의해 공기압이 점차 증가하여 성대가 떨리기 시작하면서 좁은 공기통로를 형성하게 되는데 이를 聲門(glottis)이라고 하며, 성문을 통과하는 공기는 서로 다른 두 힘의 상호작용에 의해 성대가 규칙적으로 진동을 하도록 만든다.¹⁾ 유성음을 발생할 때의 성대의 단위시간당 진동횟수, 즉, 기본 주파수(fundamental

frequency) F_0 , 또는 반복되는 진동운동의 기본주기(fundamental period) T_0 를 피치(pitch)라고 하는데, 음성합성, 음성인식, 화자인식 등 음성신호처리 분야에 있어서 매우 중요한 파라미터 중의 하나이다.

피치는 여러가지 방법으로 정의되고 있는데, 피치의 정의 방법에 따라 동일한 음성신호에 대한 결과가 서로 다를 수 있게 된다. 입력 음성신호로부터 피치의 추정치를 찾아낼 때 피치검출에 직접적으로 사용되는 신호를 기준으로 하여 피치검출 방법은 크게 두 그룹으로 나눌 수 있다.²⁾ 하나는, 시간영역에서의 피치검출 방법으로 피치검출에 사용되는 신호는 입력 음성신호와 동일한 시간대(time base)를 가지게 되며 다음과 같이 T_0 를 정의한다.

" T_0 는 두 연속적인 성문 펄스(glottal pulse) 사이에 경과한 시간으로 정의된다. 측정시의 기준점은 한 주기 내에서의 일정한 한 순간을 임의로 설정할 수 있는데 일반적으로 성문이 닫히는 순간을 기준으로 하며 성문이 완전히 닫히지 않을 경우에는 열린 면적이 최소화 되는

*한국전자통신연구소

**경북대학교 전자·전기공학부

접수일자: 1994년 2월 15일

순간을 기준점으로 한다.”

다른 하나는, 짧은 구간에서의 분석에 의한 피치검출 방법으로 입력신호를 프레임 단위로 나누어 분석함으로써 다음과 같이 정의되는 피치에서 T_0 또는 F_0 를 구한다.

“ T_0 는 입력신호를 2개 내지 3개의 피치주기를 포함하는 프레임 단위로 나누어 그에 대해 여러가지 방법으로 얻어지는 주기의 평균값으로 정의된다. 프레임 단위로 하여 주파수영역에서 분석할 경우 그 프레임 내에서의 F_0 평균값을 얻게 된다.”

시간영역에서의 분석에 의한 피치는 EGG(electroglottograph) 신호 또는 음성신호를 역필터링(inverse filtering)하여 얻어지는 예측오차신호(residual signal) 등을 이용하여 구할 수 있으며, 짧은구간에서의 분석에 의한 피치검출은 음성신호에 내재된 주기성을 자기상관함수(auto-correlation function), 평균절대차함수(AMDF: average magnitude difference function) 또는 켈스트럼(cepstrum) 등을 이용해 찾아내게 된다.

음성신호를 이용한 피치검출 방법들은 피치를 검출할 때 수반되는 여러 문제점들을 만족스럽게 해결하지 못하고 모두 그들 나름대로의 한계를 가지고 있으며, 성능개선을 위한 연구가 꾸준히 진행되고 있다.¹³⁻¹⁶ 음성신호로부터 정확하고 신뢰성 있는 피치검출이 어려운 이유를 몇가지 나열하면 다음과 같다.

- 음성발생을 위한 勵氣신호가 완전한 주기성을 가지지 못하고 準주기성을 가진다. 특히, 유성음의 시작과 끝부분에서는 주기성이 약해지고 주기內에서도 불규칙한 신호의 변화를 나타낸다.
- 음성신호는 聲道와 여기신호의 상호작용에 의한 時變 신호이다. 따라서 성도의 低域 포먼트가 여기신호를 변화시킬 수 있으므로 실질적인 주기를 찾는 것이 어려울 수 있다.
- 음성신호를 분석하기 위해 녹음하거나 디지털化 할 때 주변잡음에 의해 유성음의 섬세한 구조가 왜곡되는 경향이 있다.
- 진폭의 크기가 작은 유성음 및 무성음의 구분이 어렵다.
- 유성음 구간에서 피치의 시작과 끝을 찾기가 어렵다.

특히, 음성신호를 이용한 피치검출 알고리즘의 경우 錄音환경 및 주변잡음의 영향이 피치검출 알고리즘의 성능에 많은 영향을 미치게 되므로 잡음환경에서의 신뢰성 있는 피치를 검출하기 위한 연구도 꾸준히 진행되고 있다.¹⁸⁻¹⁰

EGG 신호는 성대의 떨림을 임피던스 변화로 바꾸어 전기신호로 나타낸 것으로 laryngograph 신호라고도 하는데, 여기신호와 성도의 포먼트 사이의 상호작용에도 영향을 받지 않으면서 발성시의 주변잡음에도 관계없이 성대의 떨림에 관한 정확한 정보를 준다.¹¹⁻¹³ 따라서 EGG 신호는 喉頭기능의 이상 유무를 판별하기 위한 신호로서 의공학 분야에서 널리 연구되어 왔으며,¹⁴⁻¹⁶ 근

래에는 EGG 신호를 이용해서 음성분석시의 정확도를 높이려는 연구가 꾸준히 진행되어 왔다.¹⁷⁻²⁰ 또한, EGG 신호는 녹음환경 및 주변잡음의 영향을 받지 않고 성대의 떨림에 관한 정보를 정확하게 전달하므로 음성신호로부터 피치를 검출할때 수반되는 여러 문제점들을 해결해 줄 수 있으며, 파형의 단순성으로 피치검출 알고리즘의 성능 평가를 위한 기준 신호로 적합하다고 볼 수 있다.

EGG 신호의 단순성 및 특성이 피치검출에 적합하지만 EGG 신호는 화자에 따라 상당한 정도의 왜곡과 잡음을 수반하기도 하며 발성을 위한 경우가 아닌 때에도 상당한 크기의 신호변화를 보이기도 한다. 또한, 화자의 특성에 기인한 것인지 EGG 신호를 얻을 때의 artifact에 의한 것인지는 분명치 않지만, 음성신호는 뚜렷한 주기성을 나타내는데 EGG 신호는 그렇지 않다던지 그 반대의 경우도 존재한다. 이러한 문제점들은 피치검출時 적절한 방법으로 제거되어야 하는데, EGG 신호만으로는 해결하기가 어려우나 음성신호를 상호 보완적으로 이용함으로써 용이하게 해결될 수 있다. 따라서 본 연구에서는 음성신호와 그에 동기된 EGG 신호의 분석을 통해 음성신호의 분석만으로는 해결될 수 없는 피치검출 알고리즘의 문제점들을 해결하고 일반적인 피치검출 알고리즘에 대해 객관적인 비교 및 평가 기준이 될 수 있는 2-채널(음성 및 EGG 신호) 피치검출 알고리즘을 연구하였다.

본 논문의 구성은 다음과 같다. 먼저 2장에서 EGG 신호의 발생 원리 및 음성분석에서의 이용에 대해 간략하게 설명하고, 3장에서는 EGG 신호 및 음성신호를 이용하여 제안된 피치검출 알고리즘에 대해 설명한다. 제안된 피치검출 방법을 이용한 여러 실험결과 및 검토를 4장에서 다루며 전반적인 요약과 결론을 5장에서 언급한다.

II. ELECTROGLOTTOGRAPH(EGG) 신호

EGG 신호는 성대의 진동운동을 간접적으로 관측하기 위해 고안된 방법중의 하나로 성대의 떨림을 임피던스 변화로 바꾸어 전기신호로 나타낸 것이다. EGG 신호를 발생시키기 위한 시스템의 기능적인 블록선도 및 전형적인 EGG 신호의 파형은 그림 1과 같다.¹¹

성대가 위치한 후두 좌우에 약 3 cm²의 크기를 갖는 2개의 금속전극을 부착시키고 일정한 값을 갖는 약한 고주파 전류 또는 전압을 인가하면 성대의 진동에 따라 성대를 구성하는 얇은막을 통과하는 전류의 일부분이 변조되어 두 전극 사이의 임피던스가 변하게 되는데 그 변화를 전기신호로 검출한 것이 EGG 신호이다. 일반적으로 사용되는 고주파 신호의 주파수는 수 MHz 이며 신호의 크기는 약 0.5 V 내외의 값을 갖는다. 성대가 서로 떨어져 있을 때는 주로 열린 공간을 통해 전류의 통로가 형성되므로 두 전극 간의 임피던스가 큰 값이 되고, 성대가 서로 붙어있을 때는 얇은막을 통해 전류의 통로가 형성되므로 임피던스가 작은 값을 갖게 된다. 따라서 유성음을 발생

할 경우 성대의 연속적인 진동운동에 의해 그림 1에서와 같이 단조증가 및 단조감소 현상을 나타내는 단순한 형태의 EGG 신호를 얻게 된다. 많은 연구결과들에 의하면 EGG 신호에서 보여주는 임피던스 변화는 성문의 열린 정도에 의한 것이 아니고 성대의 측면 접촉면적의 변화를 나타내는 것으로 알려져 있다.^[12]

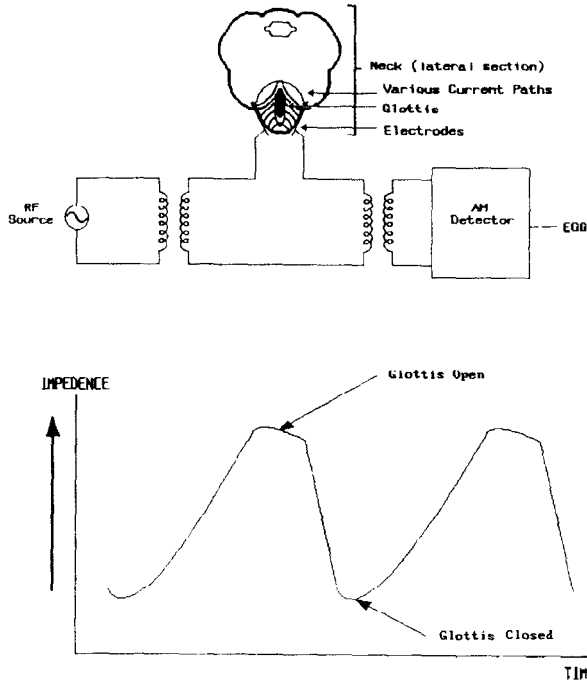


그림 1. EGG 신호의 검출원리 및 파형

EGG 방법은 원래 Farbe^[21]에 의해 설명되었는데, EGG 신호가 비록 발생되는 음성과 성대의 접촉면적과의 대응 관계를 명확하게 설명하기에는 충분하지 못했지만 성대의 진동운동을 간접적으로 쉽게 관측할 수 있다는 장점 때문에 음성병리학자 및 음성신호처리 연구자들에 의해 광범위하게 연구되어 왔다. Childers^[11-13] 등은 성대의 진동운동에 의한 성문의 열린 정도를 초고속 촬영하여 그 결과를 그와 동기된 EGG 신호와 비교함으로써 EGG 신호가 아래쪽으로 최대편향이 급속히 일어나는 순간, 즉 미분된 EGG 신호가 negative peak 값을 가지는 시점이 성대가 서로 가까워져 성문이 닫히는 순간과 일치한다는 것을 실험적으로 보였다.

음성신호에 있어서 피치정보는 음성분석 및 음성합성에서 아주 중요한 파라미터 중의 하나이다. 그러나 음성신호로부터의 정확한 피치검출은 앞에서 언급된 여러 문제점으로 인하여 제한적일 수 밖에 없다. 따라서 정확한 피치검출을 할 수 있는 보조신호로 EGG 신호를 이용해 음성분석을 하고자 하는 연구가 꾸준히 진행되어 왔다. 미분된 EGG 신호, 즉, DEGG 신호의 negative peak 값이 생기는 시점은 성대가 서로 접촉하여 성문이 최대로 닫

히는 순간과 일치하고 DEGG 신호의 positive peak 값이 생기는 시점은 성대가 서로 떨어져 성문이 최대로 열리는 순간과 대체적으로 일치한다는 특성을 이용하여 Childers^[17, 18] 등은 피치동기 분석(pitch synchronous analysis)과 더불어 성문이 닫혀있는 동안의 시간에 해당하는 음성신호만을 분석(closed-phase analysis)함으로써 성대의 진동운동에 의한 여기신호와 성도와와의 상호작용(source-tract analysis)을 배제하여 보다 정확한 성도의 파라미터들을 추출하고자 하였다. 또한 EGG 신호를 이용하여 음성신호를 유성음/무성음/묵음 구간으로 분류하였으며 그 결과를 포먼트 음성합성에 이용하여 합성음의 음질에 대한 실험을 하였다.^[20, 22]

III. 2-채널 피치검출 알고리즘

EGG 신호는 음성신호에 비해 파형이 단순하고 성도의 영향을 받지 않으므로 그의 주기성을 검출하여 피치를 구하는게 용이하다. 그러나 화자에 따라 EGG 신호의 base line이 많이 변하므로 영교차를 또는 레벨 값에 의한 주기성 검출에 어려움이 많다. 그에 비해 DEGG 신호는 EGG 신호의 base line 변화를 제거하여 주며 성대가 닫히는 순간에 negative peak 값을 나타내므로 peak-picking 알고리즘 등을 이용하여 DEGG 신호의 negative peak를 검출함으로써 피치를 구할 수 있다. 그림 2는 여성화자가 발성한 영어문장에서 안정된 유성음 구간에서의 음성신호, EGG 및 DEGG 신호, 피치 marker를 나타낸 것이다.

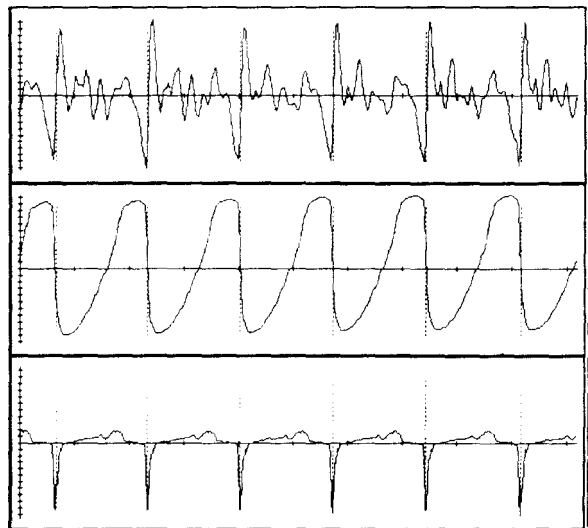


그림 2. 안정된 유성음 구간에서의 Speech/EGG/DEGG 신호 (점선은 피치 marker를 나타냄)

EGG 신호가 아래쪽으로 급격한 편향이 생기는 위치에 DEGG 신호의 negative peak이 존재하며, 시간영역에

시의 피치 marker가 바르게 표시되었음을 볼 수 있다. DEGG 신호의 negative peak 위치에 표시된 marker를 P_i ($i=1, 2, \dots$)라고 하면 그 위치에서의 순시적인 피치주기 (instantaneous pitch period)는 다음 식으로 정의할 수 있다.

$$\text{Instantaneous pitch period } T_i = P_{i+1} - P_i, \quad i=1, 2, \dots$$

EGG 신호는 발성을 위한 여기신호가 아닌 때에도 상당한 크기의 DEGG 신호의 negative peak을 보이는 경우가 많다. 또한, 화자의 특성에 기인한 것인지 또는 EGG 신호를 얻을 때의 artifact에 의한 것인지는 분명치 않지만 음성신호는 뚜렷한 주기성을 보이는데 EGG 신호는 그렇지 않다든지 그 반대의 경우도 있다. 이런 문제점들은 피치검출시 적절한 방법으로 제거되어야 한다. 특히, 유성자음과 같은 mixed sound인 경우에는 음성신호 또는 EGG 신호 하나만으로는 피치검출이 매우 어려울 수 있다. 따라서 본 연구에서는 2-채널 신호, 즉, DEGG 신호의 negative peak 위치와 그에 대응되는 음성신호의 분석을 통하여 시간영역에서 서로 다르게 정의되는 두가지 방법의 피치를 검출하고자 하였다. 하나는 피치에 동기된 분석을 통한 순시적인 피치값이며, 다른 하나는 순시적인 피치값으로 부터 일정한 크기를 갖는 분석 프레임 내에서 구해지는 평균적인 피치값이다.

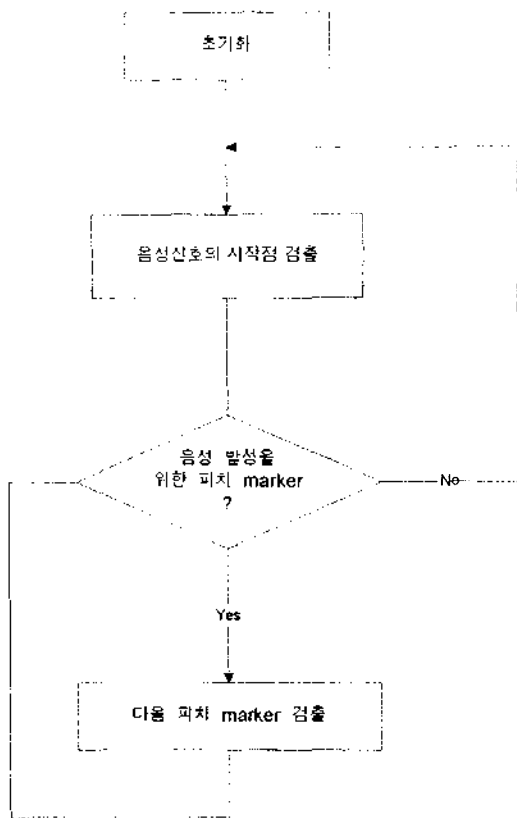


그림 3. 피치검출 알고리즘의 전체 흐름도

2-채널 신호를 이용하여 제안된 피치검출 알고리즘의 흐름도는 그림 3과 같다. 먼저 묵음 구간에서의 음성 및 DEGG 신호의 절대치 평균값과 음성신호의 영교차율을 이용하여 여기신호로 간주될 수 있는지를 음성 및 DEGG 신호의 분석을 통해 단계적인 방법으로 판정하게 된다. 이때 여기신호로 간주될 수 있다고 판정되면 같은 방법으로 그다음 negative peak을 찾게되며, 그렇지 않을 경우에는 음성신호의 끝부분으로 간주하여 그다음 음성신호의 시작점을 찾는 과정으로 되돌아 간다.

구현된 피치검출 알고리즘은 다음과 같다.

step 1: 초기화

음성신호 및 DEGG 신호에서 처음 30 ms 정도의 묵음 구간에 대해 각 신호의 절대치의 평균값 SPC_AVG 및 DEGG_AVG를 구한다. 이때, 음성신호의 영교차율 SPC_ZCR도 구한다.

step 2: 시작점 검출

정확한 시작점 검출을 통해 DEGG 신호의 첫번째 negative peak을 구하기 위하여 분석 프레임의 크기를 2 ms로 하여 이동시키면서 순차적으로 다음 과정을 수행한다.

1) $THR1 = SPC_AVG \times SF1$: ($SF1 = 2$)

음성신호의 절대치 평균값이 THR1 보다 적을 경우에는 묵음 또는 무성음 구간으로 간주한다.

2) $THR2 = DEGG_AVG \times SF2$: ($SF2 = 5$)

$THR3 = SPC_ZCR \times SF3$: ($SF3 = 3$)

검출된 DEGG 신호의 negative peak 절대치가 THR2 보다 작거나, 음성신호의 영교차율이 THR3 보다 클 경우에는 묵음 또는 무성음 구간으로 간주한다.

3) 음성신호에서 30ms 간격에 대해 AMDF를 적용하여 유성음 여부를 결정하고, 그에 따른 DEGG 신호의 첫번째 negative peak 위치를 찾는다.

step 3: 피치 marker 검출

피치 검출을 위한 분석 프레임의 크기를 20 ms로 하고, step 2에서 검출된 DEGG 신호의 negative peak이 피치신호, 즉, 음성 발성을 위한 여기신호인지의 판정을 위해 분석 프레임 내에서의 음성신호 절대치의 평균값 SPC_PIT 및 DEGG 신호의 첫번째 negative peak의 절대치인 DEGG_PIT을 구한다. 이때, DEGG_PIT은 분석 프레임내에 포함된 DEGG 신호의 negative peak들의 절대치 중에서 최대값의 1/3 이상이 되어야 한다. SPC_PIT 및 DEGG_PIT을 이용한 피치신호의 판정은 순차적으로 다음 과정을 통해 이루어진다.

- 1) SPC_PIT 값이 $(THR1 \times 2)$ 보다 적을 경우에는 묵음으로 간주한다.
- 2) DEGG_PIT 값이 $(THR2 \times 2)$ 이하이면서 SPC_PIT 값이 $(THR1 \times 2)$ 이하이면 묵음 또는 무성음 구간으로 간주한다.
- 3) DEGG_PIT의 위치가 그전까지에서 구해진 피치 평균값에서 크게 벗어나면 DEGG 신호의 negative peak이 올바르게 검출되지 못한것으로 간주하고 음성 신호에 AMDF를 적용하여 피치신호 輿呑를 결정한다.
- 4) DEGG_PIT 값이 $(THR2 \times 2)$ 이상이면 분석 프레임 내에서 구한 DEGG 신호의 첫번째 negative peak을 피치신호로 간주한다.
- 5) 음성신호에 AMDF를 적용하여 DEGG negative peak의 피치신호 輿呑를 결정한다.
- 6) 위의 모든 경우에서도 피치신호를 검출하지 못한 경우에는 음성신호의 끝부분으로 간주되어 새로운 시작점을 찾기위해 step 2로 되돌아 가고 그렇지 않을 경우에는 피치신호로 판정된 DEGG negative peak 위치에서 step 3의 과정을 반복한다.

여기서, 주어진 데이터에서 시간 및 화자에 따른 신호의 크기 변화를 고려해주기 위해서 처음 묵음구간에서 구한 SPC_AVG 및 DEGG_AVG의 값을 피치검출 과정을 거치면서 각 분석 구간에서 다음과 같이 새로운 값으로 update 해준다.

$$SPC_AVG = SPC_PIT/40, \text{ if } SPC_AVG < SPC_PIT/40$$

$$DEGG_AVG = | \text{Max Negative Peak} | /40,$$

$$\text{if } DEGG_AVG < | \text{Max Negative Peak} | /40$$

2-채널 피치검출 알고리즘에서는 EGG 신호의 왜곡 및 불규칙한 변화 등을 충분히 고려해주기 위하여 음성 및 EGG 신호에서 얻어진 파라미터들에 대해 여러 데이터에 대한 실험을 통하여 구해진 문턱값을 적용하여 순차적인 단계를 거쳐 피치 판정을 내리게 된다. 시간 영역에서의 순시적인 피치 및 분석 프레임내에서의 평균적인 피치를 구하는 방법은 그림 4와 같다. 그림 4(a)에서는 피치검출을 위한 분석 프레임의 이동 간격이 검출된 DEGG 신호의 negative peak 위치에 의해 결정되므로 음성신호의 피치에 동기된 피치 marker를 얻을 수 있고, 그림 4(b)에서는 분석 프레임의 이동간격이 일정하므로 분석 프레임내에서 구해진 여러 피치값들의 평균값을 얻게 된다. 이때, 분석 프레임의 이동 간격은 10 ms로 하였다.

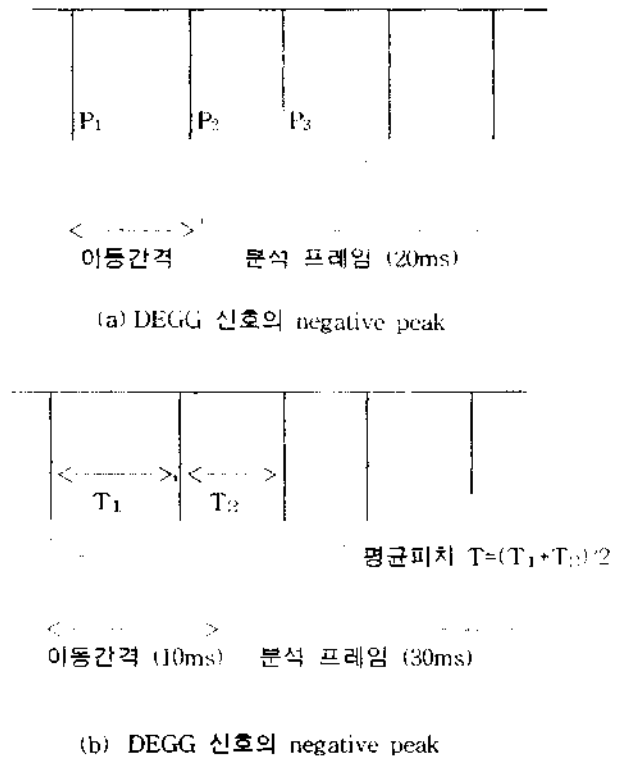


그림 4. DEGG negative peak 신호로부터의 피치검출 방법

IV. 실험 및 검토

실험에 사용된 음성 및 EGG 신호는 각각 10 kHz 샘플링에 16 bits/sample로 양자화 되었다. EGG 신호는 성문에서 검출되는데 비해 음성신호는 마이크에서 검출되므로 성문에서 입까지, 입에서 마이크까지 소리가 전달되는데 걸리는 시간만큼 음성신호는 EGG 신호보다 지연되어 나타난다. 따라서 음성신호에 동기된 EGG 신호를 얻기위해 EGG 데이터를 10 샘플 정도 지연시켜 사용하였다.

유성음의 시작과 끝부분에서는 성대의 진동운동의 주기성이 많이 깨어지고 불규칙하게 될 뿐만 아니라 음성 발생을 위해 성문을 통과하는 공기의 흐름(glottal volume velocity)이 적어 음성신호의 진폭이 작아지기 때문에 음성신호를 이용한 피치검출에서 제일 어려운 부분이다. 그림 5는 voice offset 구간에서의 Speech/EGG/DEGG/Pitch marker를 나타낸 것으로 EGG 신호로부터 성대의 진동운동에 많은 불규칙성을 보임을 볼 수 있다. 그림 5에서 보면 유성음이 끝나는 부분에서는 피치주기가 불규칙해지면서 두배 이상으로 길어지는 경우가 있는데, 이는 음성신호와 비교해 보면 잘 일치함을 알 수 있다. 그림 6은 여성화자가 발성한 영어문장 "We were away..."에서 "were"와 "away" 사이 부분에 대해 Speech/EGG/DEGG/Pitch marker를 보인 것이다. 피치 변화가 상당히 많음을 볼 수 있는데 표시된 피치 marker는 음성신호의 주기

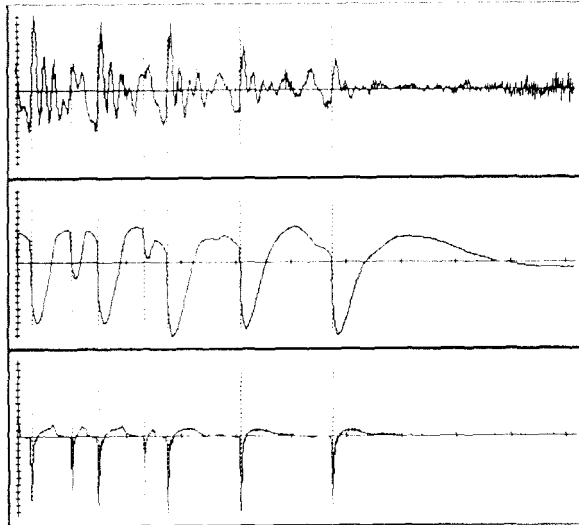


그림 5. Voice offset 구간에서의 Speech/EGG/DEGG 신호 (점선은 피치 marker를 나타냄)

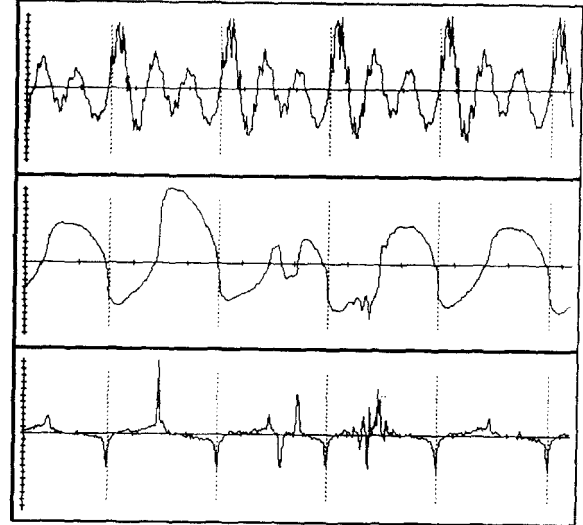


그림 7. 성대의 떨림이 불안정한 화자의 Speech/EGG/DEGG 신호 (점선은 피치 marker를 나타냄)

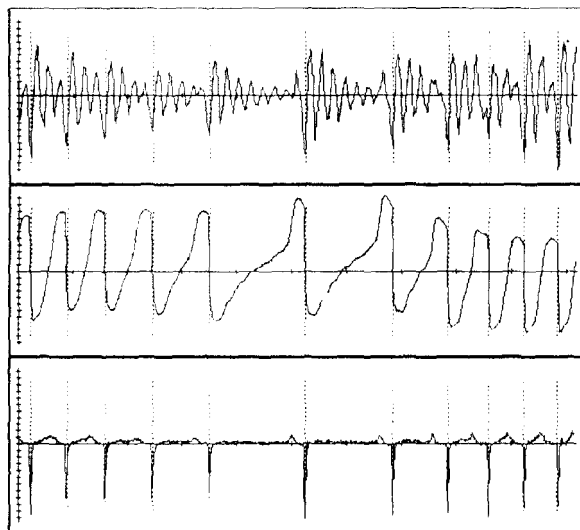


그림 6. 변이음 구간에서의 Speech/EGG/DEGG 신호 (점선은 피치 marker를 나타냄)

성과 잘 일치함을 볼 수 있다. 이런 경우 EGG 신호 등을 이용한 시간영역에서의 피치검출 방법을 사용하지 않고는 정확한 피치검출이 매우 어렵다.

그림 7은 성대의 떨림이 불안정한 화자의 Speech/EGG/DEGG/Pitch marker 파형을 보인 것이다. 음성신호는 주기성을 상당히 잘 나타내고 있는데 비해 EGG 신호에는 상당한 정도의 왜곡과 잡음 성분을 볼 수 있으며 그에 따라 DEGG 신호도 잡음성분을 많이 내재하고 있어 피치검출의 어려움을 알 수 있다. 그러나, 본 연구에서 제안된 2-채널 피치검출 알고리즘을 이용할 경우에는 음성신호에서 주어지는 뚜렷한 주기성을 이용할 수 있으므로 EGG 신호의 왜곡 또는 잡음에 의한 DEGG 신호의 부적

당한 negative peak을 잘 제거할 수 있음을 볼 수 있다.

제안된 2-채널 피치검출 알고리즘을 이용하여 정상적인 화자가 발성한 한국어 및 영어 문장에 적용하여 얻어진 피치 marker를 음성 및 EGG 신호 파형과 함께 직접 눈으로 관측하여 구해진 error rate를 실험에 사용된 문장과 함께 표 1에 나타내었다. 피치 추정 error는 voice onset 또는 offset 구간 및 유성음 구간에서 피치를 검출하지 못한 경우를 deletion error, 무성음이나 묵음구간에서 피치 marker를 만든 경우를 insertion error로 판정하여 관측된 전체 피치 주기에 대해 백분율로 나타내었다.

영어의 경우 남녀 각각 3명의 화자가 발성한 3개의 문장이 사용되었으며, 한국어의 경우 남녀 각각 4명의 화자가 발성한 5개의 문구가 사용되었다. 표 1에서 음성신호의 진폭이 작고 성대의 진동이 불규칙해지는 voice onset 및 offset 부분에서 생기기 쉬운 deletion error가 2-채널 피치검출 알고리즘에서는 거의 제거됨을 볼 수 있으며, 주요 error 원인으로서는 성대떨림의 불안정으로 인한 EGG

표 1. 2-채널 피치검출 알고리즘을 이용한 피치 추정 error (a) 한국어의 경우

	DELETION	INSERTION	TOTAL
남성 화자	0.0%	0.85%	0.85%
여성 화자	0.0%	0.09%	0.09%
평균	0.0%	0.47%	0.47%

문구 1: 바람과 햇님이...

문구 2: 힘이 더 세다...

문구 3: 사람의 특징이요...

문구 4: 우리 말은 우리 거예요...

문구 5: 문화 활동을 해오던 길...

(b) 영어문장의 경우

	DELETION	INSERTION	TOTAL
남성 화자	0.0%	0.11%	0.11%
여성 화자	0.08%	0.08%	0.16%
평균	0.04%	0.095%	0.135%

문장 1: We were a year ago.

문장 2: Should we chase those cowboys?

문장 3: We saw the ten pink fish.

신호의 왜곡과 이로 인한 insertion error가 실험에 사용된 한국어 남성화자에게서 주로 관측되었다. 실험결과를 종합해 볼때 제안된 2-채널 피치검출 알고리즘은 음성신호만으로는 피치검출이 어려운 voice onset 및 offset 부분에서 정확한 피치검출이 가능하며, 따라서 새로이 개발되는 피치검출 알고리즘에 대한 표준피치패턴 및 객관적인 자료의 제공에 이용될 수 있다고 하겠다.

그림 8은 남성화자가 발성한 한국어 문장 "바람과 햇님이 서로 힘이 더 세다"에서 "바람과 햇님이..."부분에 대해 음성신호만을 이용한 방법과 음성 및 EGG 신호를 이용한 분석 프레임내에서의 평균적인 피치궤적을 비교한 것이다. 음성신호만을 이용한 방법에서는 음성신호의 선형예측오차신호를 구한 후 AMDF를 적용하여 피치를 구하였는데, 2-채널 신호를 이용한 방법에서 훨씬 더 연속적이며 자연스런 피치궤적이 얻어짐을 볼 수 있다.

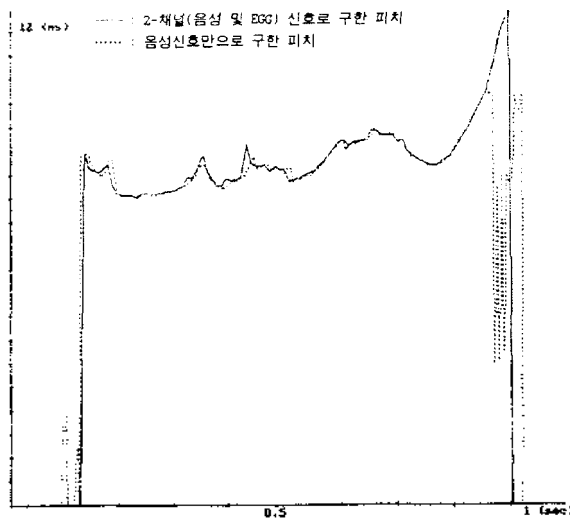


그림 8. 한국어 문장중 "바람과 햇님이" 부분에 대한 피치궤적

V. 결 론

음성신호의 피치는 음성합성, 음성인식 등 음성신호처리 분야에서 매우 중요한 파라미터중의 하나이다. 그러나 음성신호를 이용해서 현재까지 연구 개발된 많은 피

치검출 알고리즘들은 모두 그들 나름대로의 한계를 가지고 있으며, 성능 개선을 위한 연구가 꾸준히 진행되고 있다. 본 연구에서는 음성신호로 부터 피치를 검출하고자 할 때 수반되는 여러 문제점들을 최소화 시켜줄 수 있는 2-채널 신호 분석에 의한 피치검출 알고리즘을 제안하고 실험하였다.

유성음을 발생할 때의 성대의 진동운동에서 매 주기마다 성대가 서로 최대로 접촉되는 시점에 DEGG 신호가 negative peak 값을 가진다는 성질을 이용하고, EGG 신호의 왜곡 및 불규칙한 변화는 음성신호의 분석을 통해 보완해 줌으로써 시간영역에서 피치 동기된 순시적인 피치 marker를 정확히 검출 할 수 있는 알고리즘을 개발하였다. 2-채널 피치검출 알고리즘은 음성신호만을 이용한 일반적인 피치검출 알고리즘에 비해 월등히 개선된 피치 궤적을 얻을 수 있음을 실험을 통해 보였으며, 따라서 음성신호만을 이용하여 새로 개발되는 피치검출 알고리즘의 객관적인 비교 및 평가 기준으로 사용될 수 있고 보다 정확한 음성분석 과정에 유용하게 이용될 수 있으리라 기대된다. 또한, 2-채널 신호를 이용할 경우 음성신호의 유성음/무성음/묵음의 분류에도 효과적으로 이용될 수 있다.

참 고 문 헌

1. Berg, J. W. van den, "Myoelastic-aerodynamic theory of voice production," *Speech and Hear. Res.*, Vol. 1, pp. 227-244, 1958.
2. Wolfgang J. Hess, "Pitch and Voicing Determination," *Advanced in Speech Signal Processing* edited by Sadaoki Furui & M. Mohan Sondhi, Marcel Dekker, Inc. Part I, pp. 3-48, 1992.
3. D. T. L. Lee and M. Morf, "A Novel Innovations Based time-Domain Pitch Detector," *IEEE Proc. of Int. Conf. ASSP*, Vol. 1, pp. 40-44, 1980.
4. Peter Kroon and Bishnu S. Atal, "Pitch Predictors with High Temporal Resolution," *IEEE Proc. of Int. Conf. ASSP* pp. 661-664, 1990.
5. Jean-Paul Lefevre and Gang Feng, "Pitch Detection Based on Localization Signal," *Signal Processing V: Theories and Applications*, pp. 1159-1162, 1990.
6. Youv Medan, Eyal Yair, and Dan Chazan, "Super Resolution Pitch Determination of Speech Signals," *IEEE Trans. on Signal Processing*, Vol. 39, No. 1, Jan. 1991.
7. Etienne Barnard, Roland A. Cole, Mathew P. Vea, and Fileno A. Alleva, "Pitch detection with a Neural-Net Classifier," *IEEE Trans. on Signal Processing*, Vol. 39, No. 2, Feb. 1991.
8. Sanguoon Chung, and V. Ralph Algazi, "Improved pitch detection algorithm for noisy speech," *IEEE Proc. Int Conf. ASSP*, pp. 407-410, 198 5.
9. L. Hodgson, M. E. Jernigan, and B. L. Willis, "Nonlinear

- multiplicative cepstral analysis for pitch extraction in speech," *IEEE Proc. Int. Conf. ASSP*, pp. 257-260, 1990.
10. David A. Krubsack, and Russell J. Niederjohn, "An Autocorrelation Pitch Detector and Voicing Decision with Confidence Measure Developed for Noise-Corrupted Speech," *IEEE Trans. on Signal Processing*, Vol. 39, No. 2, Feb. 1991.
 11. D. G. Childers and A. K. Krishnamurthy, "A critical review of electroglottography," *CRC Crit. Rev. Bioeng.*, Vol. 12, No. 2, pp. 131-164, 1985.
 12. D. G. Childers, A. M. Smith, and G. P. Moore, "Relationships between electroglottography, speech and vocal cord contact," *Folia Phoniatrica*, Vol. 36, pp. 105-108, 1984.
 13. A. M. Smith and D. G. Childers, "Laryngeal evaluation using features from speech and the electrograph," *IEEE Trans. Biomed. Eng.*, Vol. 30, pp. 755-759, Nov. 1983.
 14. Fourcin, A. J., "Laryngographic assessment of phonatory function," *Proc. of the Conf. on the Assessment of Vocal Pathology, ASHA Report*, Edited by C. L. Ludlow and M. O. Hart, pp. 116-127, 1981.
 15. Fog-Pedersen, M., "Electroglottography compared with stroboscopy in normal persons," *Folia Phoniatrica*, Vol. 29, pp. 191-199, 1977.
 16. P. Kitzing, "Glottography, the electrophysiological investigation Phonatory biomechanics," *Acta Oto-Rhino-Laryngologica Belgica*, Tome40, Fasc. 6, 1986.
 17. Ashok K. Krishnamurthy, and Donald G. Childers, "Two-Channel Speech Analysis," *IEEE Trans. on ASSP*, Vol. 34, No. 4, Aug. 1986.
 18. J. N. Larar, Y. A. Alsaka, and D. G. Childers, "Variability in closed phase analysis of speech," *IEEE Proc. Int. Conf. ASSP*, pp. 1082-1092, Mar. 1985.
 19. Wolfgang Hess and Helge Indefrey, "Accurate Time-Domain Pitch Determination of Speech Signals by Means of a Laryngograph," *Speech Communication* 6 pp. 55-68, 1987.
 20. D. G. Childers, M. Harn, and J. N. Larar, "Silence and Voiced/Unvoiced/Mixed excitation(Four-way) classification of speech," *IEEE Trans. on ASSP*, Vol. ASSP-37, pp. 1771-1744, Nov. 1989.
 21. Fabre P., "Un procédé électrique percutané d'inscription de l'accolement glottique au cours de la phonation: Glottographie de haute fréquence; premiers résultats." *Bull. Acad. nat. Méd. (paris)*, S-4, pp. 66-69, 1957.
 22. Neal B. Pinto, Donald G. Childers, and Ajit L. Lalwani, "Formant Speech Synthesis: Improving Production Quality," *IEEE Trans. on ASSP*, Vol. 37, No. 12, Dec. 1989.

▲ 신 무 용(Mu Yong Shin)

1969년 4월 8일생



1991년 2월: 경북대학교 전자공학과 (공학사)

1994년 2월: 경북대학교 전자공학과 대학원(공학석사)

1994년 3월~현재: 한국전자통신연구소 이동멀티미디어연구실 연구원

※주관심 분야: 음성신호처리, 영상신호처리, 멀티미디어

▲ 김 정 철(Jeong Cheol Kim)

1972년 4월 7일생



1995년 8월: 경북대학교 전자공학과 (공학사)

1996년 2월~현재: 경북대학교 전자·전기공학부 대학원 석사과정

※주관심 분야: 디지털신호처리, 음성신호처리

▲ 배 건 성(Keun Sung Bac)

1953년 11월 9일생



1977년 2월: 서울대학교 전자공학과 (공학사)

1979년 2월: 한국과학기술원 전자공학과(공학석사)

1989년 2월: University of Florida(공학박사)

1989년 3월~현재: 경북대학교 전자·전기공학부 교수

※주관심 분야: 디지털신호처리, 음성신호처리, 디지털통신