

배경 잡음하에서 스펙트럼 누설현상을 이용한 음성신호의 중심 피치 검출

On the Center Pitch Estimation by using the Spectrum Leakage Phenomenon for the Noise Corrupted Speech Signals

강 동 규*, 배 명 진**, 안 수 길***

(Dong Kyu Kang, Myung Jin Bae, Sou Guil Ann)

요 약

지금까지 제안된 피치 추정 알고리즘들은 나이노소에 무관하게 폭넓은 피치범위를 검출하기가 어렵다. 조음기관의 물리적 한계성 때문에 피치의 분포도는 일반적으로 중심 피치에 몰려있는 형태가 된다. 이 중심 피치를 본격적인 피치검출 과정에 적용한다면, 그 처리과정이 간단해지고 정확도가 개선될 수 있다. 본 논문에서는 스펙트럼 누설현상을 이용하여 중심피치를 정확하게 검출하는 알고리즘을 제안한다.

ABSTRACT

The pitch estimation algorithms which have proposed until now are difficult to detect wide range pitches regardless of age or sex. A little deviation are observed with reference to the center pitch in the distribution diagram of pitches, since pitches are characterized by a physical limitation of the coarticulation mechanism. If the center pitches are referred to the accurate pitch extraction procedure, the algorithms will be not only simplified in procedure but also improved in accuracy. In this paper, we proposed an algorithm that the center pitches are accurately detected by using the spectrum leakage phenomenon for the noise corrupted speech signals.

I. 서 론

현대사회가 정보사회로 발달하고, 디지털 신호 처리기술과 반도체기술의 급속한 발전에 따라, 인간

의 기본적인 의사전달 수단인 음성을 인간과 기계의 인터페이스(Man-Machine interface)로 이용하고자 하는 음성 신호처리에 관한 연구나 응용이 활발히 추진되고 있다.

음성 신호처리에 있어서 음성신호의 기본주파수를 정확히 검출하면 분석시 피치에 동기시켜 분석할 수 있어 성문의 영향이 제거된 정확한 성도 파라미

*전자통신연구소 신호처리연구실
**호서대학교 전자공학과
***서울대학교 전자공학과

더욱 얻을 수 있다. 또한, 합성시 소량의 네이마르 개성이나 자연음을 쉽게 변경하고 유지할 수 있게 된다.

지금까지 알려진 피치 검출법으로는 시간영역법, 주파수영역법, 시간-주파수 혼성법으로 나눌 수 있다[1-3] [7-12]. 이들의 보편적인 방법은 그림 1-1과 같이 음성신호의 주기성분을 강조한 후 결정논리에서 주기성을 검출하고 있다.

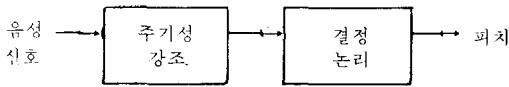


그림 1-1. 보편적인 피치 추출과정
Fig. 1-1. Typical processing for the pitch extraction.

기존의 피치검출법에서 주기성을 강조할 때, 남노소에 따라 널리 분포한 피치성분을 모두 강조하지 못하고 특정영역만 살리고 나머지는 오히려 약화시킬 수 있다. 결정논리는 주기성이 강조된 결과값으로 주기성 검출을 하기 때문에 주기성 강조시 발생한 오차와 함께 검출오차가 더해질 수 있다. 더욱이, 배경 잡음하에서는 더 많은 어려움이 따르게 된다. 이상에서 알 수 있듯이 모든 경우에 적용할 수 있는 강력한 피치 추출법은 아직까지도 어려운 과제로 남아 있다[1,3].

화자의 피치변화는 조음기관의 물리적 한계와 정신적인 안정감에 따라 제한된다. 피치는 준주기적인 성질을 갖고 있으므로 짧은 구간내에서 급격한 변화를 나타내지 않는다. 오랜 시간 동안 검출된 피치의 분포도를 보면 어떤 중심된 피치를 기준으로 약간의 이탈치를 갖는다. 이와같은 피치의 기본적인 성질을 피치 검출 알고리즘에 적용할 수 있다. 피치 검출시 주기성 강조와 동시에 몇개의 프레임에 대한 중심피치를 검출하여 결정논리에 참고한다면, 검출과정이 간단해지고 검출의 정확도를 높일 수 있다. 따라서, 본 연구에서는 스펙트럼 누설현상을 이용하여 비교적 안정된 구간내의 중심된 피치를 검출할 수 있는 새로운 중심피치 검출 알고리즘을 제안하였다.

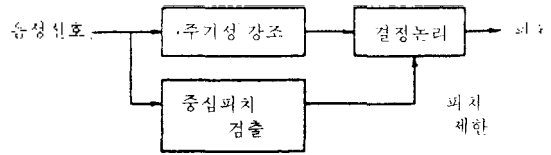


그림 1-2. 중심피치 적용
Fig. 1-2. Center pitch application.

본 논문의 진행 순서는 II장에서 음성신호의 일반적인 발생모델과 음성음에서의 피치에 대하여 서술하였다. III장에서는 본 논문에서 이용하고자 하는 누설현상에 대하여 설명하고, 이 누설현상이 음성신호에서는 어떻게 나타나는가에 대해 기술하였다. IV장에서는 음성에너지 계산시 창함수길이에 따른 음성에너지변화가 어떤 원인에 의해 발생하는지 살펴보고 V장에서는 창함수길이를 선행적으로 변화시키면서, 시간영역 컨볼루션(convolution)할 때 구해진 음성에너지에서 누설에너지를 검출하고, 창함수길이에 따른 평균누설에너지의 변화도에서 중심피치를 추출하는 방법에 대해서 논하였다. VI장에서는 실제 음성에 대한 처리 방법과 결과를 제시하고 VII장에서 결론을 맺는다.

II. 음성신호의 발생모델

음성신호는 발생음원에 따라 유성음, 무성음, 묵음으로 구분된다. 음성신호를 생성모델 측면에서 고찰하면 그림 2-1에서 처럼 무성음의 경우에는 불규칙잡음 발생기(random noise generator)가 그 생성원이므로 주기성은 나타나지 않는다. 그러므로, 무성음의 스펙트럼은 2500Hz 근방에서 주된 공명봉우리를 갖는 준색잡음의 형태를 띤다. 무성음은 공명 주파수가 높기 때문에 유성음에 비해 평균 영교차율(zero crossing rate)이 크다[2-4].

유성음은 준주기적인 성대의 진동펄스(glottal pulse)가 그 생성원이며 성대(vocal cord)의 진동에 따른 성대(vocal tract)의 영향이 강조되어 나타나 일반적으로 진폭이 크고 준주기적인 성질을 갖는다. 성대의 진동펄스가 성도를 여기하여 발생되는 유성음은 각 음소마다 성도에서 고유한 공명이 일어

난다. 이러한 공명봉우리를 포먼트(formant)라 하고 낮은쪽 주파수에서 부터 누드러진 포먼트들을 제 1, 제2, 제3 포먼트 등으로 부른다.

성대의 진동주기를 피치(pitch)라 부르며, 남성과 소 및 주변환경에 따라 다르지만 2.5-25msec 정도로 알려져 있다. 성도특성은 보편적으로 성별이나 연령에 관계없이 나타나지만, 성대특성은 화자에 따라 고유한 특성을 나타내고 있으므로 화자식별에 중요한 파라미터로 사용되고 있다. 또한, 성대특성은 감정상태에 따라 다르게 나타나므로 합성시 성대의 진동주기를 변경함으로써 감정상태가 조절되어 자연스런 음성을 합성할 수 있다.

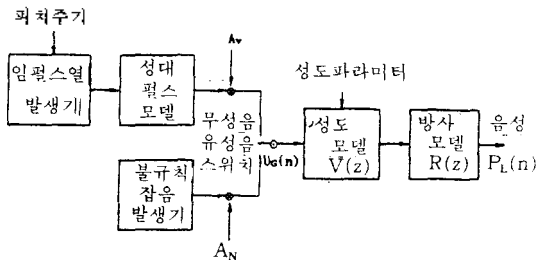


그림 2-1. 일반적인 음성생성에 대한 모델
Fig. 2-1. General model for speech production.

III. 누설현상

A. 누설현상의 정의

$x(t)$ 가 128 sps로 표본화된 32Hz 단위천목의 정현파라고 하면, 이에 대한 진폭스펙트럼은 그림 3-1 a)와 같이 나타난다. 이 스펙트럼은 128 데이터 포인트(data point)를 사용하였고 편의상 22-42Hz의 값만을 표시하였다. 그림에서 값이 존재하는 것은 32Hz뿐이다. 이것은 값이 존재하는 기본주파수 f_0 가 데이터 포인트 N과 표본화 간격 T에 의해 다음과 같이 결정되기 때문이다.

$$f_0 = 1/NT = 1 \text{ Hz} \quad (3-1)$$

$y(t)$ 는 $x(t)$ 와 약간 다른, 31.5Hz를 갖는 단일 정현파이고, $x(t)$ 와 같은 데이터 포인트 및 표본화 간격을 갖을 때 이에 대한 진폭스펙트럼은 그림

3-1 b)와 같이 나타난다. 그림 3-1 b)는 그림 3-1 a)와는 달리 31.5Hz 근방에 몇개의 스펙트럼 성분을 갖는다. 이러한 현상이 발생된 원인은 구간 NT가 31.5Hz 정현파 주기의 정수배가 아니기 때문이다 [5].

이와같이, 창함수 구간 NT내에 정수배의 주기를 갖지 않는 신호에 대한 스펙트럼은 원래의 신호주파수 근처에 대칭으로 분산되어 나타나는 현상을 스펙트럼 누설(leakage)현상이라고 한다. 이 현상은 DFT (Discrete Fourier Transform)의 기본적인 가정에 의해 발생된 것이다. 즉, DFT식은 구간 NT에 걸쳐 있는 모든 성분이 주기적으로 반복된다는 가정하에 푸리에 급수(Fourier Series)를 전개한 식이기 때문이다 [5].

B. 음성신호에서의 누설현상

함수 A와 B의 컨벌루션을 $A * B$ 로 표시하기로 한다. 유성음에서 성도특성을 $f(n)$, 성대의 진동특성을 $g(n)$ 이라 할때 발생된 음성신호 $s(\cdot)$ 은

$$s(n) = f(n) * g(n) \quad (3-2)$$

이 되고, 이에 대한 주파수 응답은

$$S(e^{j\omega T}) = F(e^{j\omega T}) \cdot G(e^{j\omega T}) \quad (3-3)$$

으로 표시된다. 성도특성은 시간영역에서 임펄스열 (impulse train)로 나타나므로 주파수영역에서도 임펄스열로 나타난다. 그러므로, 식(3-2)는 성도특성에 성대특성이 곱해진 라인(line)스펙트럼의 형태로 나타난다. 유성음에 대해 일정 위치에서 유한 구간 창함수를 적용하여 구한 스펙트럼은 창함수를 고려하지 않은 유성음의 스펙트럼에 창함수의 스펙트럼이 컨벌루션된 결과가 된다. $s'(\cdot)$, $S'(\cdot)$ 을 각각 음성신호에 창함수를 적용한 신호 및 이에 대한 스펙트럼이라 하면,

$$s'(n) = s(n) \cdot w(n-k) \quad (3-4)$$

$$S'(e^{j\omega T}) = S(e^{j\omega T}) * W(e^{j\omega T}) \quad (3-5)$$

으로 표현할 수 있다. 유성음은 피치주기 마다 반복되는 신호로 볼 수 있으므로, 만일 창함수 구간이 피치주기의 정수배로 되지 않으면 상기한 바와같이 누설현상이 발생하여 스펙트럼은 왜곡된다.

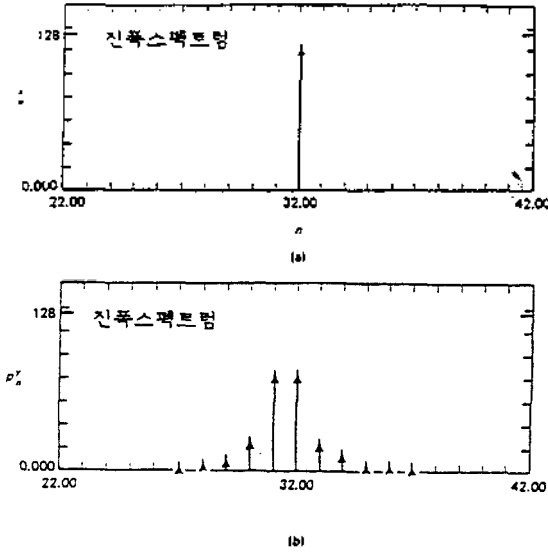


그림 3-1. 스펙트럼누설현상의 영향
Fig. 3-1. Effect of leakage problem on spectra.

IV. 음성에너지에서의 누설현상

A. 음성에너지 계산

유, 무성음의 에너지 레벨이 서로 다르기 때문에 음성에너지의 변화형태가 음소의 변화를 근사적으로 나타낸다. 음성신호 $s(\cdot)$ 의 단시간 에너지는, $w(\cdot)$ 이 창함수이고 $a(n)=[s(n)]^2$ 일 때 다음과 같이 나타낼 수 있다.

$$E(n) = \sum_{k=-\infty}^{\infty} a(k) \cdot w(n-k) = a(n) * w(n) \quad (4-1)$$

여기서 방형창함수가 적용될 경우, 이에 대한 주파수 응답은 다음과 같다.

$$E(e^{j\omega T}) = A(e^{j\omega T}) \cdot w(e^{j\omega T}) = A(e^{j\omega T}) \frac{\sin(\Omega NT/2)}{\sin(\Omega T/2)} e^{j\omega T(N-1)/2} \quad (4-2)$$

식 (4-2)에서 단시간 평균 에너지계산은 f_s 를 표본화 주파수라 할때 차단 주파수 $f_c=f_s/N$ 인 저역통과 필터에 제공된 음성신호 $a(\cdot)$ 을 통과시킨 것과 같다. 이상적인 저역통과 필터라면 차단주파수 이상의 성분들에 의한 영향이 완전히 제거될 수 있다. 그러나, 방형창함수일 경우에는 식(4-2)에 주어진 것처럼 주파수응답이 싱크함수(sinc function)형 저역통과 필터이기 때문에 차단 주파수 이상의 성분들이 싱크함수의 부분우리에 의해 살아 남게 된다. 창함수 종류에 따라 통과대역과 차단대역의 에너지비를 근사적으로 나타내면, (1) 방형창: 약 10dB (2) 삼각형창: 약 30dB (3) 해밍창: 약 40dB (4) 블랙맨창: 약 60dB 등이 된다[7].

B. 창함수길이에 따른 음성에너지 변화

식(4-2)에서 음성에너지 계산은 주파수영역에서 각 스펙트럼의 곱으로 표시됨을 표현하고 있다. 부분적으로 선택된 유성음에 대해, 표본화 주기(T)는 고정되어 있으므로 유성음의 스펙트럼은 창함수길이(N)에 따른 함수가 된다. N이 변화하므로써 창함수 스펙트럼은 주봉우리 및 부분우리의 간격이 변화하게 된다. 주봉우리의 대역폭이 유성음의 기본주파수보다 작으면 음성에너지의 변화도에서 국부봉우리(local peak)가 많이 나타나고, 커지면 스무딩 현상이 현저해져 음소의 변화특성을 잘 나타내지 못한다. 창함수의 길이에 따른 음성에너지의 변화를 그림 4-1에 예시하였다.

유성음은 스펙트럼상에서 기본주파수의 하모닉스 마다 골과 봉우리를 형성하고 있다. 창함수 스펙트럼의 골과 봉우리 간격이 선형적으로 변화하면서 유성음의 스펙트럼과 곱해지면, 각 스펙트럼의 봉우리와 골이 일치되거나 어긋나는 경우가 발생된다. 일치될 경우에는 차단주파수 이상의 성분이 통과대역의 성분과 더해지고, 어긋날 경우에는 차단주파수 이상의 성분이 상쇄되는 현상이 발생된다. 이 현상은 창함수길이가 신호주기의 관계가 누설현상과 유사하게 나타난다. 이와 같이, 창함수길이의 변화에 따라 음성에너지는 창함수 스펙트럼의 주봉우리 뿐 아니라 부분우리에 의해서도 많은 영향을 받게 된다.

창함수길이(N)에 따른 에너지변화는 부분우리가
 가상 한 방형창함수일 경우 두드러지게 나타난다.
 창함수길 적용할때 발생하는 누설현상과 창함수의
 길이를 선형적으로 변화시키면서 시간영역 컨벌루션
 할때 발생하는 유사한 누설현상이 곱해지면 음성에
 너지의 변화는 더욱 심화된다. 누설현상의 정의에
 따라, 창함수길이 커 피치의 정수배가 될때 누설현상
 이 최소가 되어 차단대역과 통과대역의 에너지비는
 가장 커지게 된다. 상기한 바와같은 현상으로 인해
 변동되는 에너지를 편의상 누설에너지라 규정하겠
 다.

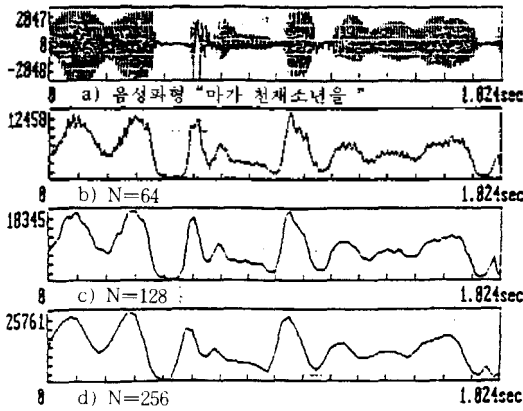


그림 4-1. 창함수길이에 따른 음성에너지
 Fig. 4-1. Speech energy for various window length.

V. 누설에너지 검출

A. 누설에너지 검출

누설에너지값을 $E_1(\cdot)$ 이라 할때, n-번째 샘플부
 터 창함수를 적용하여 얻어지는 에너지값, $E_1(\cdot)$
 은,

$$E_1(n) = E(n) + E_1(n) \tag{5-1}$$

이 된다. $E_1(n)$ 은 창함수의 위치와 창함수의 길이에
 따라 변동되는 누설에너지값이다. 일정한 창함수
 길이 N에 대하여 누설에너지를 측정하기 위해 방형
 창함수를 음성 샘플단위로 적용시키면서 에너지
 $E_1(n, N)$ 를 구하면 다음과 같이 표시할 수 있다.

$$E_1(n, N) = E(n, N) + E_1(n, N)$$

$$E_1(n-1, N) = E(n, N) + E_1(n-1, N)$$

$$\dots \dots \dots$$

$$E_1(n-N+1, N) = E(n, N) + E_1(n-N+1, N) \tag{5-2}$$

여기서, $E(n, N)$ 은 통과대역내에 있는 에너지값이
 로 시간지연에 무관한 에너지 값이다.

식 (5-2)에 의해 에너지 $E_1(n, N)$ 값을 구한후, 창함
 수길이 N에 대한 누설에너지의 값을 구할 수 있다.
 창함수 구간내에서 검출된 에너지 $E_1(n, N)$ 들의 최대
 및 최소값을 구하면,

$$E_{\max}(n, N) = \text{Max} [E_1(n, N), E_1(n-1, N), \dots, E_1(n-N+1, N)]$$

$$E_{\min}(n, N) = \text{Min} [E_1(n, N), E_1(n-1, N), \dots, E_1(n-N+1, N)] \tag{5-3}$$

이된다. 여기서, $E_{\max}(n, N)$ 와 $E_{\min}(n, N)$ 함수는 주어
 진 변수에서 최대값과 최소값을 각각 의미하는 함수
 라 정의한다. 누설에너지 $E_1(n, N)$ 의 변동분이 평균치
 들 중심으로 같다면, 평균 에너지값 $E(n, N)$ 은

$$E(n, N) = [E_1(n, N) + E_1(n-1, N) + \dots + E_1(n-N+1, N)] \tag{5-4}$$

이 되고, 근사적으로 다음과 같이 표시할 수 있다
 [7].

$$E(n, N) = [E_{\max}(n, N) + E_{\min}(n, N)] / 2 \tag{5-5}$$

식 (5-5)에서 누설에너지값은 근사적으로 다음식
 같이 표현할 수 있다.

$$E_1(n, N) = E_{\max}(n, N) - E_{\min}(n, N) \tag{5-6}$$

실제의 음성신호 "감사합니다."에 대하여 식(5-
 3)의 최대 및 최소값을 그림 5-1에 도시하였다. 그림
 에서 볼 수 있듯이 최대 및 최소치가 거의 같은
 폭으로 변화하고 있다. 이것은 식(5-4)와 (5-5)에서

이 가정의 합당하고 있음을 보이고 있다.

제1 프레임의 음성신호에 창함수길이 N으로 컨볼루션하면서 산출된 샘플단위별 누설에너지에 대한 평균 누설에너지를 구한다. 창함수길이 N에 대한 평균 누설에너지 $G_1(N)$ 는 다음과 같이 표시된다.

$$G_1(N) = \frac{1}{M-N+1} \sum_{n=0}^{N-1} E_1(n, N)$$

$$= \frac{1}{M-N+1} \sum_{n=0}^{N-1} [E_{\max}(n, N) - E_{\min}(n, N)] \quad (5-7)$$

여기서, $N=20, 21, \dots, 200$ 이다.

으로 근사할 수 있다. M은 프레임의 크기이고, 창함수길이 N은 포분화 주파수 F_0 가 8kHz일때 발생가능한 최저 피치인 20샘플(2.5msec)로 부터 최고 피치인 200샘플(25msec)까지이다. 그러므로 최고 피치까지 검출하기 위한 프레임구간은 적어도 M이 400샘플정도는 되어야 한다. N을 20에서 200까지 변화하면서 $G_1(N)$ 의 변화도를 구하면 피치길이의 정수배 마다 누설에너지가 최소로 되어 협곡점(null point)을 형성할 것이다. 이때 최초의 협곡점 위치에 해당하는 샘플수가 그 프레임내에서 가장 지배적인 피치가 된다.

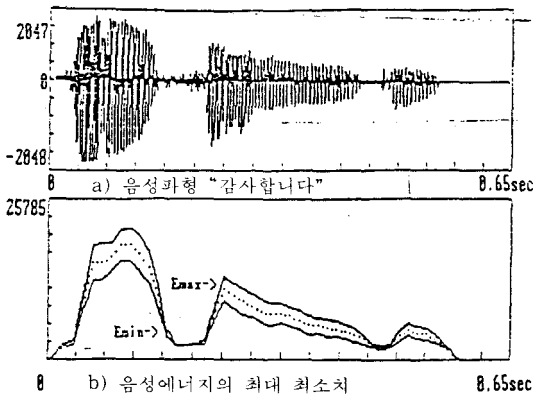


그림 5-1. $E_{\max}(\cdot)$ 와 $E_{\min}(\cdot)$ 의 변화도
Fig. 5-1. $E_{\max}(\cdot)$ and $E_{\min}(\cdot)$ Contour.

상기한 가정에 대해 실제와 일치하는가를 보기 위해 실제 음성신호를 적용하여 $G_1(N)$ 의 변화도를

구하여 그림 5-2에 도시하였다. 음성시료는 26세 여성화자가 발성한 숫자음 /삼/ 중에서 /사-/이며, 한 프레임을 512샘플로 했다. 그림 5-2의 변화도에서 두번째 이상의 협곡점은 피치의 정수배마다 나타나고 있음을 알 수 있고 협곡점의 예리한 정도는 그 프레임내에서 피치가 얼마나 안정되어 있는가에 따른다. 그림 5-2의 경우 최초의 협곡점이 창함수길이 29 샘플일 때 발생하여 중심피치는 3.625 msec로 검출되었다. 이상의 결과에서 볼 수 있듯이 제안된 시간영역 중심피치 검출법은 음성이 시작하거나 음소가 천이하는 구간에서도 그 프레임내의 중심되는 피치를 검출하고 있음을 알 수 있다.

B. 단축알고리즘

창함수길이 N에 대하여 식 5-1을 계산하려면 $E_1(\cdot)$ 를 평균진폭으로 구하여도 N이 200일때 샘플당 200번의 덧셈과 최대 및 최소값을 구하는데 2번의 비교연산이 소요된다. 연산량을 줄이기위해 $E_1(n-1)$ 을 평균진폭에 관한 식 $M_1(n-1, N)$ 로 다시 쓰면,

$$M_1(n-1, N) = |s(n-1)| + |s(n-2)| + \dots + |s(n-N+1)| + |s(n-N)|$$

$$= |s(n)| + |s(n-1)| + \dots + |s(n-N+1)| - |s(n)| + |s(n-N)|$$

$$= M_1(n, N) - |s(n)| + |s(n-N)| \quad (5-8)$$

이 되어 방금전의 $M_1(n, N)$ 을 안다면 반복적으로 $M_1(n-1, N)$ 을 계산할 수 있다. 따라서, 평균진폭 계산에서는 샘플당 덧셈과 뺄셈이 각각 한번씩 필요하게 된다.

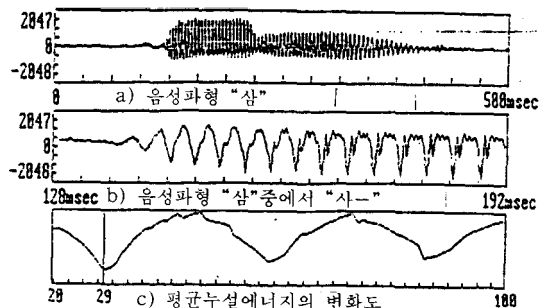


그림 5-2. 음성신호 /삼/중에서 /사-/에 대한 $G_1(\cdot)$ 의 변화도
Fig. 5-2. $G_1(\cdot)$ contour for utterance /sa-/ of /sam/.

VI. 실험 및 결과

A. 실험 장비 및 시료

식 (5-7)에 대한 컴퓨터 시뮬레이션(computer simulation)을 하기 위해 IBM-PC / AT에 마이크로 장치된 12-비트 A/D 변환기를 인터페이스 시켰다. 화자는 남성화자 2명으로 다음 음성을 각각 발생하도록 하고 8kHz로 표본화하면서 저장하였다.

발성1) 23세 남성화자: "인수네 꼬마는 천재소년을 좋아한다."

- 1-1) "인수네 꼬마"
- 1-2) "꼬마가 천재소년을"
- 1-3) "좋아 한다."

발성2) 32세 남성화자: "호서대 전자공학과 음성 신호처리 연구."

- 2-1) "호서대 전자 공"
- 2-2) "학과 음성"
- 2-3) "신호처리 연구"

B. 실험 과정 및 결과

그림 6-1에 실험과정을 흐름도로 나타내었다. 각 음성시료에 프레임길이를 512샘플로 하고 128샘플씩 중첩시키면서 방형창함수를 적용했다. 창함수의 길이(N)를 20부터 200까지 변화시키면서 컨벌루션을 수행하여 창함수의 길이에 따른 평균 누설에너지분 산출하였다. 검출된 평균 누설에너지분 기초로 변화도를 작성하고 최선의 항목점을 찾아 각 프레임내에 중첩된 피치로 하였다.

각 그림의 a)는 음성파형은, b)에서는 누설에너지분 검출된 피치(실선)와 목적으로 구한 피치(점선)를, c)에는 목적으로 구한 피치와 제안중인 후출범으로 구한 피치와의 편차를 도시하였다. 동일한 음성시료에 영데시벨(0dB)의 가우시안 잡음(Gaussian noise)을 섞어 실험과정을 반복하고 결과를 도시하였다.

음성신호의 시작과 끝에서 목적(eye check)으로 구한 피치와 제안중인 방법으로 구한 피치가 서로 차이를 보이고 있다. 이것은 프레임길이를 512 샘플

로 했기 때문에 프레임끝이 잘 서로 일치하지 않아 발생한 것이다. 또한 6) 1)과 1)2)에는 검출된 값은 음성이 없는 구간과 있는 구간과 대략적으로 일치하는 것으로 보아도 주기적의 편차 항목을 인식할 수 있다.

실질영역에서 피치값이 리운 즉, 음성의 시작부분 그리고 음성의 끝부분에서 편차의 변화가 심해도 제안중인 방법은 음성신호의 추가성만 유지되면 편차가 검출되었다. 제안된 검출법은 제어통과의 특징을 내포하고 있어 배경잡음에 강력한 것으로 나타났다. 또한, 편차가 변화하고 있어도 평균화되어 검출되었다. 무음이나 무성음일 경우 2.5msec 이하로 나타내므로 음, 무성음 구간도 판별할 수 있다.

VII. 결 론

화자의 피치변화는 조음기관의 물리적 한계와 정신적 안정감에 따라 제한된다. 조음기관의 물리적 한계성 때문에 피치는 몇개의 프레임내에서 급격한 변화는 보이지 않는다. 이와 같은 피치의 기본적인

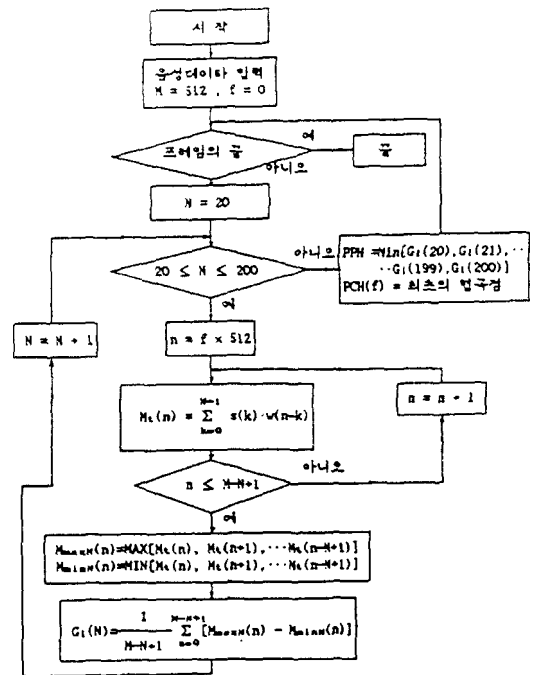


그림 6-1. 실험에 대한 흐름도
Fig. 6-1. Flow chart for computer simulation

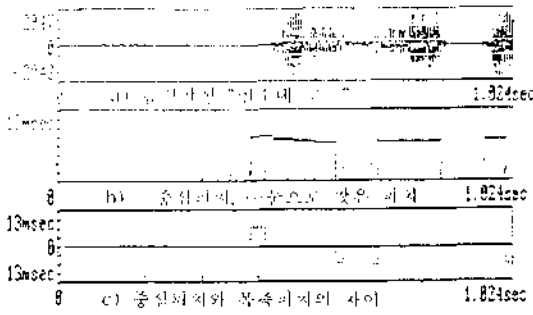


그림 6-2. 발성 1-1)에 대한 피치변화도
Fig. 6-2. Pitch contour for utterance 1-1).

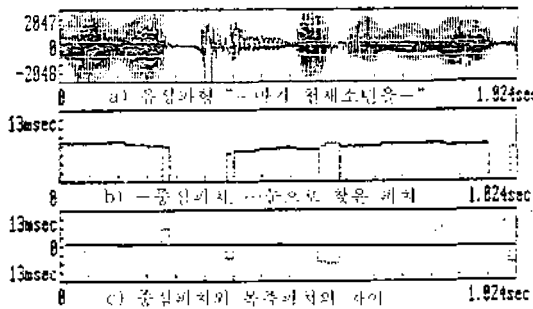


그림 6-3. 발성 1-2)에 대한 피치변화도
Fig. 6-3. Pitch contour for utterance 1-2).

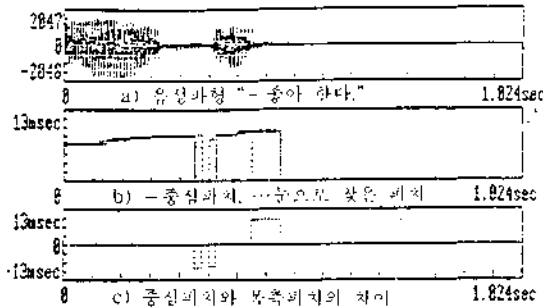


그림 6-4. 발성 1-3)에 대한 피치변화도
Fig. 6-4. Pitch contour for utterance 1-3).

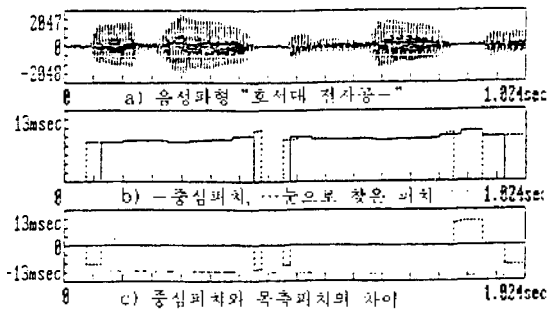


그림 6-5. 발성 2-1)에 대한 피치변화도
Fig. 6-5. Pitch contour for utterance 2-1).

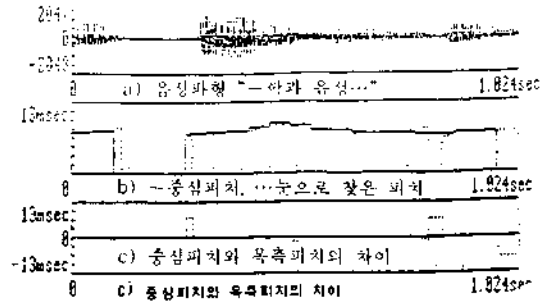


그림 6-6. 발성 2-2)에 대한 피치변화도
Fig. 6-6. Pitch contour for utterance 2-2).

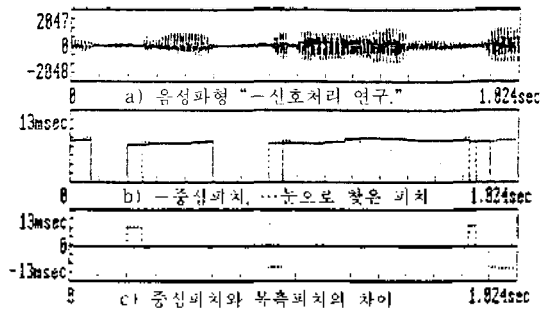


그림 6-7. 발성 2-3)에 대한 피치변화도
Fig. 6-7. Pitch contour for utterance 2-3).

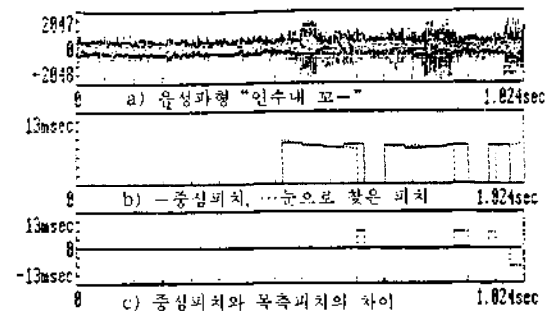


그림 6-8. 발성 1-1)에 대한 피치변화도
Fig. 6-8. Pitch contour for utterance 1-1).
(0dB Gaussian Noise)

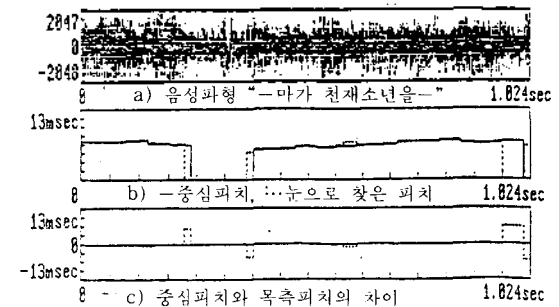


그림 6-9. 발성 1-2)에 대한 피치변화도
Fig. 6-9. Pitch contour for utterance 1-2).
(0dB Gaussian Noise)

87.

2. L.R. Rabiner / R.W. Schafer, *Digital Processing of Speech Signal*. Prentice-Hall inc., Englewood Cliffs, N.J., 1987.

3. Thomas W. Parsons, *Voice and Speech Processing*, McGraw-Hill, 1987.

4. Douglas O'Shaughnessy, *Speech Communication*, Addison Wesley, 1987.

5. Nasir Ahmed, T. Natarajan, *Discrete-Time Signal and System*, Prentice-Hall inc, 1983.

6. J.D. Markel, A.H. Gray, Jr. *Linear Prediction of Speech*. Spinger-Verlag, 1980.

7. 배명진, 안수길, "음성 에너지계산에서 창함수-길이 개선에 관한 연구", 한국음향학회, 제9권 2호, pp.34-41, 1990년 4월.

8. M. BAE, and S.ANN, "Fundamental Frequency Estimation of Noise Corrupted Speech Signal Using

the Spectrum Comparison", J., Acoust., Soc., Korea, vol.8, No.3, 1989.

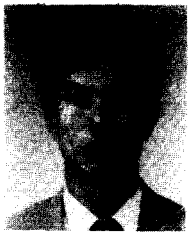
9. M.BAE, S.SHIN, S.ANN, "The pitch extraction of voiced speech by the comparison between the Original and Repeated Partial Waveform", J., Acoust., Soc., Korea, vol.7, No.5, 1988.

10. M.BAE, and S.ANN, "Inverse Rate Type Filtering for the Pitch Extraction", J., Acoust., Soc., Korea, Vol.5, No.3, 1986.

11. 배명진, 안수길, "면적 비교법에 의한 고속 피치추출", 대한전자공학회, Vol. 22, No. 2, pp. 13-17, March 1985.

12. H.Duifhuis, L.F. Willems, and R.J. Sluyter, "Measurement of pitch in speech: An implementation of Goldstein's theory of pitch perception", J., Acoust. Soc., Amer., vol. 71 pp.1568-1580, June, 1982.

▲강 동 규



1960년 3월 2일생
 1979년 : 금오공고 전자과 졸업
 1989년 : 호서대학교 전자공학과 졸업
 1991년 : 호서대학교 전자공학과 대학원 졸업
 현재 : 전자통신연구소 신호처리 연구실 연구원

▲배 명 진 9 권 5 호 참조

▲안 수 길 9 권 5 호 참조