배경잡음하에서 주파수영역 피치검출에 관한 연구 -스펙트럼 AMDF에 의한 제 1 포먼트 영향 제거법-

On the Frequency Domain Pitch Detection of Noise Corrupted Speech Signals -Minimizing the Effects of the F1 by the Spectral AMDF-

배 명 진* 박 찬 수* 안 수 길**

(Myungjin BAE, Chansou PARK, Souguil ANN)

본 연구는 한국과학재단의 기초연구 과제의 자금지원으로 이루어 졌음
※과제번호 : 911-0801-015-1※

요 약

음성 신호처리 분야에서 기본주파수를 정확히 검출하는 것이 아주 중요하다. 주파수 영역에서 피치검출 방법의 문제점은 대체로 배경잡음이나 제 1 포만트에 의하여 발생한다. 그러므로, 본 논문에서는 스펙트럼 AMDF 함수를 이용하여 잡음의 영향이나 제 1 포만트의 영향을 줄이는 주파수영역 피치검출 앨고리즘을 제안하였다. 여러가지 컴퓨터 시뮬레이션 결과 제안한 앨고리즘이 기본주파수 검출에 효과적으로 나타났다.

ABSTRACT

Detecting the fundamental frequency(Fo) of the speech signal is a problem in many speech applications. A problem of the pitch detection method in the frequency domain is occured by the first formant and the background noise. Thus, in this paper, we proposed a pitch detection algorithm in the frequency domain that reduces the effects of the first formant and the background noise by the spectral AMDF function. Several computer simulation results showed that the proposed algorithm was very effective for fundamental frequency detection.

I. 서 론

음성 신호처리 분야에 있어서 기본주파수를 정확히 검출하는 것은 매우 중요하다. 만일 음성신호의 기본주파수를 정확히 검출할 수만 있다면 음성인식시에 화자에 따른 영향을 줄일 수 있으므로 인식의 정확도를 높일 수 있고, 음성합성시에 자연성과 개성을 쉽게 변경하거나 유지할 수 있다. 또한 음성분석시에 피치에 동기시켜 분석하게 되면 성문(vocal cord)의 영향이 제거된 정확한 성도 파라미터를 얻을 수 있다.

기본주파수 검출에 관한 연구는 시간영역법, 주파수영역법, 시간주파수혼성법으로 구분할 수 있다. 주파수영역 피치검출 방법은 하모닉스를 파라미터로 하여 피치를 검출하는 방법으로써 harmonic product spectrum[6], peak valley detection[5,7], comb-filtering, spectrum autocorrelation[11] 등이 있다. 이 방법은 음성 스펙트럼상의 하모닉스 간격을 측정하여 유성음의 기본주파수를 검출하고 있다. 일반적으로 스펙

* 숭실대학교 전자공학과
** 서울대학교 전자공학과

프레임을 한 프레임(20-40ms) 단위로 분해시키므로 이 구간에서 음소의 천이나 변동이 있더라도 배경잡음이 발생하여도 평균화(averaging)하므로 이 영향을 적게 받는다. 그러나 처리과정을 주파수영역으로의 변환과정이 필요하므로 계산이 복잡하며 기본주파수의 정밀성을 높이기 위해 FFT의 포인트수를 늘리면 그 만큼 처리시간이 길어진다[13]. 그렇지만 최근에는 신호처리 전용칩인 DSP칩의 발달에 따라 영역변환에 따른 계산시간의 부담이 해결되었기 때문에 주파수 영역법이 연구 대상으로 다시 대두되고 있다.

Noll에 의해 제안된 하모닉스 합(곱)의 방법[6]은 다음 식으로 하모닉스를 강조시킨다.

$$P(f)=2\sum_{k=1}^{k}\log|X(f,e^{jwr})| \qquad (1\text{-}1)$$

이 식에 의해 하모닉스의 봉우리는 봉우리끼리, 골은 골끼리 더해져서 하모닉스가 강조된다. 이 방법은 음성이 전화선로를 통해올 때 기본 하모닉스가 감쇄되거나 잡음이 많은 경우에도 얼마간 성공적이지만 포먼트에 의해 평탄치 못한 스펙트럼의 영향을 많이 받아서 피치 결정시에 doubling이나 halving이 많이 나타난다.

Seneff에 의해 제안된 하모닉스의 peak-valley 측정법[5]은 1100 Hz 이하의 스펙트럼상 하모닉스에 대해 복잡한 봉우리 검출 앨고리즘(harmonic detector)을 적용하여 피치의 주기를 결정하고 있다. 비슷한 방법으로 Screenivas와 Rao는 선택된 하모닉스들의 최대공약수를 얻기 위한 peak-valley검출 앨고리즘[7]을 제안하였다. 이러한 방법들은 배경잡음이 없는 경우에만 우수한 결과가 얻어진다.

Lahat에 의해 제안된 스펙트럼 상관계수(Autocorrelation)법[11]은 스펙트럼을 한 신호로 취급하여 특정 리프터들에 통과시키는 방법이다. 이것은 하모닉스가 기본주파수의 정수배로 나타난다고 볼 수 있기 때문에 하모닉스를 특정 리프터에 통과시켜서 최대값이 찾아지는 위치가 기본주파수가 된다는 방법이다. 이러한 방법 역시 포먼트의 영향을 크게 받고 여성이나 어린이 화자인 경우에는 순수한 정이 현과 스펙트럼에서 처럼 다임 펄스형이 되기 때문에 어려운 방법으로서, 피치를 추정하는데 다소 어려움이 있다[4].

대표적인 주파수영역 피치검출 방법은 스펙트럼의 하모닉스를 강조한 후 간단한 검출기에 의해 피치를 검출하고 있다. 이들 방법은 스펙트럼을 파라미터로 사용하기 때문에 SNR이 0-dB인 경우에도 검출이 가능하다고 알려져 있지만 피치검출의 정확성이 떨어지게 되고, 포먼트의 영향을 많이 받아 검출에러가 증가된다. 주파수 영역법에서 피치검출의 정확성과 분해능을 악화시키는 원인을 분석하여 제거해 줄 수만 있다면 잡음에 강력하면서도 정확한 피치검출법이 제안될 수 있게 된다.

따라서 본 논문에서는 이 문제점들을 해결하기 위하여 스펙트럼상에서 AMDF를 취하는 SAMDF 함수를 제안 하였다. 이 함수를 적용하면 배경잡음의 영향이나 포먼트들의 영향을 최소화하면서 기본 하모닉스를 강조할 수 있게 된다.

본 논문의 진행 순서는 우선, 주파수영역의 스펙트럼 하모닉스 특성에 대해 II 장에서 알아본 다음에 III 장에서는 포먼트의 조락 특성을 그대로 유지하면서 스펙트럼의 최고점까지 AMDF를 취하는 SAMDF 함수를 제안한다. 다음으로 IV 장에서는 SAMDF 함수를 이용하여 기본 하모닉스를 강조하는 ISAMDF 함수의 제안과 그 결정논리를 다룬다. 그리고는 제안된 앨고리즘의 우수성을 보기 위하여 결과들을 비교 분석하고 결론을 짓는다.

II. 스펙트럼 하모닉스의 특성

시간영역에서 음성신호 s(t)의 성분 주파수가 f/2로 대역 제한되어 있고 주기 T를 갖는 주기신호로 이루어져 있다면 이에 대한 퓨리에 급수전개는 다음과 같다.

$$s(t)=\sum_{k=-N}^{N}c(k)e^{j2\pi kt/T} \qquad (2\text{-}1)$$

여기서, c(t)는 k번째 퓨리에 계수 값으로

$$c(k)=\int_{0}^{1}s(t)e^{-j2\pi kt}dt$$

가 된다. 따라서 음성신호 s(t)에 대한 주파수 영역의 표현을 S(f)라 하면 S(f)는

$$S(f)=\int_{-\infty}^{\infty} s(t)e^{-j2\pi ft}dt$$
$$=\int_{-\infty}^{\infty}\left(\sum_{k} c(k)e^{j2\pi k\frac{1}{T}t}\right)e^{-j2\pi ft}dt$$
$$=\sum_{k} c(k)\int_{-\infty}^{\infty} e^{-j2\pi\left(f-k\frac{1}{T}\right)t}dt$$
$$=\sum_{k} c(k)\delta\left(f-k\frac{1}{T}\right) \qquad (2\text{-}2)$$

과 같이 된다. 식 (2-2)에 나타내 보인 바와 같이 시간영역에서 나타나는 T주기 성분은 주파수 영역에서 1/T 간격을 갖는 하모닉스들로 반복되어 나타나게 된다. 이렇게 반복적으로 나타나는 하모닉스들의 간격을 기본주파수(F_0)라 한다.

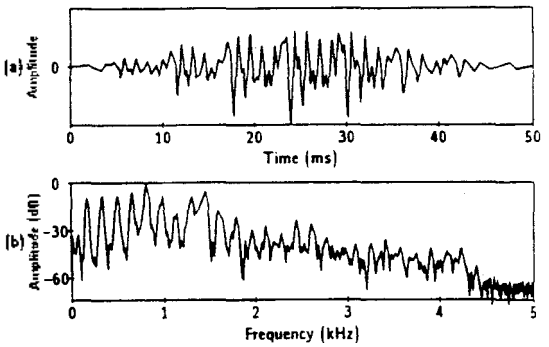

그림 1. 유성음에 대한 파형(a)과 스펙트럼(b)
Fig. 1 Speech waveform and the spectrum for an utterance /ah/.

그림 1의 (a)는 유성음 파형에 해밍창을 적용한 파형을 나타내며 (b)는 이에 대한 스펙트럼을 예시한 것이다. 그림 1에서 볼 수 있는 것과 같이 유성음은 성대가 진동해서 발생하는 것이기 때문에 성도의 공진특성의 영향을 받아 그림 1의 (b)와 같이 공진주파수의 봉우리 부분이 관찰된다. 대략 4KHz 이하에서는 2-4개가 관찰되는데 낮은 주파수로 부터 제 1 포먼트(F1), 제 2 포먼트(F2) 등으로 부른다. 제 1 포먼트(F1)는 다른 포먼트들에 비해 약 10dB 가량 높게 나타나며 0 주파수에서 제 1포먼트(F1)까지 에너지가 증가되는 봉우리 형태를 띠게 된다.

또한 하모닉스들 사이에 세세한 국부 봉우리(local peak)들은 잡음에 의해 나타나게 된다. 스펙트럼상에서 하모닉스를 검출하고자 할 때는 이러한 포먼트의 영향이나 배경잡음의 영향을 받게된다.

Ⅲ. 스펙트럼상의 AMDF(SAMDF)

이상에서 살펴 본 바와같이 유성음의 스펙트럼 하모닉스들은 주기성이 유지될 때 마다 골과 봉우리를 형성하며 0 주파수에서 첫번째 포먼트까지 기울기를 가지고 증가하는 형태로 나타난다. 이 봉우리 특성은 화자의 영향이나 음성이 변화함에 따라 폭과 형태가 제각기 다르게 나타나며 여기서 하모닉스의 구조를 찾으려면 사전에 제 1 포먼트까지의 봉우리 특성을 알아야만 하는데 이것은 또다른 별도의 과정이 필요하게 된다. 따라서 영교차(zero crossing) 방법으로 하모닉스의 간격을 측정하고자 할 때, 평탄치 못한 포먼트의 영향으로 인하여 기본주파수를 검출하는데 많은 에러가 발생할 수 있다. 또한 잡음이 인가되면 하모닉스상의 골과 골, 또는 봉우리와 봉우리 사이에는 삼음성 국부 봉우리들이 나타나게 되며, 이 경우 피크를 검출함에 있어서 자칫하면 이 국부 봉우리를 하모닉스의 피크치로 오인하게 되어 에러가 발생한다.

따라서 본 논문에서는 기본 하모닉스들의 주기성을 강조하면서도, 포먼트의 영향을 제거할 수 있는 스펙트럼 AMDF 함수식(SAMDF)을 다음과 같이 제안한다 :

$$SAMDF(d)=\sum_{k=1}^{N}|S_p(k)-S_p(k-d) \qquad (3\text{-}1)$$
$$(d=1,2,\cdots, Fm)$$

여기서 $S_p(\cdot)$는 음성신호의 진폭 스펙트럼이며, d는 주파수 단위로 지연시키는 값이다. 주파수 지연기 d의 계산 범위는 제 1 포먼트의 주파수인 Fm 주파수까지 수행한다.

이 식은 시간영역에서 피치를 검출하기 위해 자기상관함수(autocorrelation) 대신에 주기성을 강조하기 위해 적용했던 AMDF법과 유사한 함수식으로 우리는 스펙트럼상에서 수행하였다. AMDF식은 세세한(instantaneous) 변화와 함께 포락(envelope)이 변화

하고 있는 신호파형에 대하여 주기성 강조도 분석 [illegible] 나타낼 수 있는 함수이다. 식 (3-1)의 AMDF 또한 음성신호 파형의 [illegible] 주기성 분석을 위해 사용되어 왔지만 여기에서는 이 함수가 갖는 평균 포락정보를 이용하고자 한다.

이 함수는 포먼트에 의해 스펙트럼이 정단하지 않은 경우에도 유리하게 된다. 이 포먼트들의 공명에 의해 하모닉스들의 포락선이 거울이질 수록 식 (3-1)의 값에서 최대값과 억눌한 값의 차이는 더욱 두드러지게 된다. 이것은 주파수 지연 d에 따른 SAMDF 값이 제 1 포먼트 봉우리의 거울기를 잘 나타내게 됨을 의미한다. 또한 이 함수는 하모닉스의 정수배로 주파수 지연을 시킬 때 최소값이 되지만 d가 기본 하모닉스 간격의 1/2의 정수배일 때는 최대값이 된다.

Ⅳ. 스펙트럼상의 AMDF에 대한 결정논리

유성음 스펙트럼에 대해 식 (3-1)을 통과한 SAMDF(d)의 최소값은 그림 2(c)에서 처럼 하모닉스의 주기가 일치할 때 마다 골을 형성한다. SAMDF(d) 함수를 통과시킨 스펙트럼 값에 대해 우선 초기값인 0 스펙트럼에서 부터 첫 최고점 봉우리까지를 SAMDF(d)의 최대값으로 대치한다. 이것

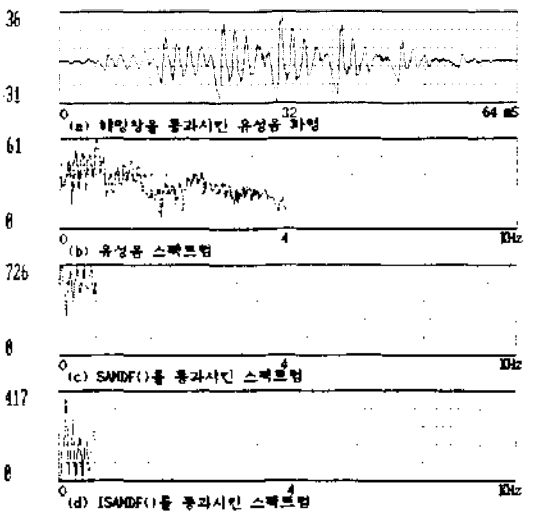

그림 2. 유성음 스펙트럼에 대한 기본하모닉스 강조
Fig. 2. Emphasizing the fundamental harmonic for utterance /ah/

[illegible] AMDF 함수값 중 d=0일 때는 자기 자신과의 상관관계로 최대값을 갖게되어 매번에 이 결과을 [illegible]

유성음의 스펙트럼이 제 1 포먼트까지 봉우리를 이루기 위해 그림 2(b)에서 처럼 하모닉스로 진동하면서 증가되는 형태를 나타낸다. 이 때문에 SAMDF를 통과한 값은 세세한 국부 봉우리들이 제거되며 주기성이 강조된 하모닉스들이 제 1 포먼트까지 포먼트 봉우리의 평균 기울기로 진폭을 이루면서 증가하게 된다. 이 때 SAMDF함수를 통과한 스펙트럼에 대해 그 역 스펙트럼(Inverse SAMDF)을 다음과 같이 구한다. :

$$ISAMDF(k)=P_{max}-SAMDF(d) \qquad (4\text{-}1)$$

여기서, P_{max}는 SAMDF를 통과한 스펙트럼의 최대값을 나타낸다.

이렇게 하면 그림 2(d)에서 처럼 스펙트럼의 하모닉스들 중에서 첫(fundamental) 하모닉스가 가장 강조되어 나타나게 된다. 따라서 식 (4-1)을 통과한 스펙트럼에 대해 최대 봉우리를 이루는 하모닉스를 찾으면 이 위치가 기본주파수가 된다. 이렇게 강조된 첫번째 봉우리까지의 간격은 쉽게 측정될 수 있으며, 이 값이 기본주파수 F_0가 된다.

Ⅴ. 실험 및 결과

이상의 과정을 처리하기 위해 마이크가 장치된 12bit A/D변환기를 IBM-PC/AT에 인터페이스 시키고, 음성시료 1로는 24세의 남성화자가 연속 발음한 "인수네 코마가 전제소년을 좋아한다"와 음성시료 2로는 27세의 여성화자가 연속 발성한 "감사합니다"를 8KHz의 샘플링 주파수로 양자화하여 시뮬레이션에 대한 시료로 사용하였다.

그림 3은 본 논문에서 제안한 처리과정을 블럭도로 나타낸 것이다. 각 음성샘플은 한 프레임을 512샘플로 설정하여 50%씩 겹치게 처리하였다. 그림 4에서 부터 그림 7까지는 제안한 알고리즘에 대한 결과들이다.

그림 4(a)는 음성시료 1의 "좋아한다"에 해당하는

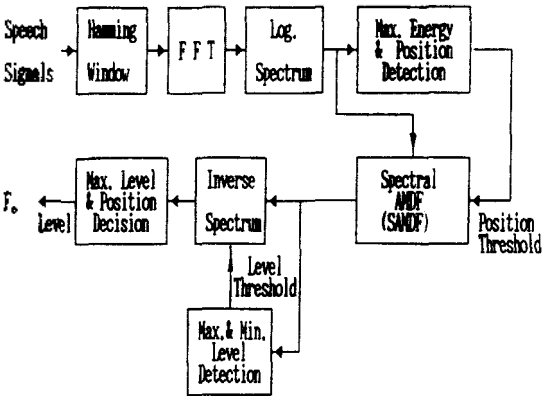

그림 3. 스펙트럼 AMDF를 이용한 기본 하모닉스 강조 및 피치 검출과정
Fig. 3. Processing the fundamental harmonic emphasis and pitch detection by using the spectral AMDF.

1초(8192샘플) 동안의 음성파형이다. 입력된 음성신호는 추정에러를 감소시키기 위하여 해밍 창함수(Hamming Window)를 적용하였고, 이 결과의 파형을 그림 4(b)에 나타내었다. 다음으로 512포인트 FFT를 수행함으로써 시간영역의 파형에 대한 스펙트럼을 그림 4(c)에 나타내었다.

다음은 스펙트럼상의 값들 중에서 최대 에너지 값에 해당하는 주파수 Fm값을 구한다. 주파수 지연값 d를 0에서 부터 Fm 주파수까지 증가시키면서 (3-1)식에 의해 스펙트럼상의 AMDF를 수행한다. 이렇게 얻어진 스펙트럼에 대해 0 주파수에서부터 첫번째 하모닉스 봉우리까지의 값을 그 최대값으로 대치시킨다. 이렇게 수행한 결과의 예를 그림 4(d)에 나타내었으며 스펙트럼에 깔려있던 세세한 국부 봉우리들이 제거되었음을 알 수 있다. 또한 SAMDF(·)를 통과한 스펙트럼은 하모닉스의 주기가 일치될 때 마다 깊은 골을 이루게 되고, 골의 변화는 제 1 포먼트의 봉우리 형태를 나타내게 된다.

그 다음에는 스펙트럼상의 AMDF를 통과한 값에 대해 역수를 취하여 골을 형성했던 스펙트럼을 봉우리가 되도록 식 (4-1)에 의해 역 스펙트럼 ISAMDF(·)를 구한다. 이러한 결과의 예로는 그림 4(e)와 같이 나타나며, 여기서 기본 하모닉스가 최대의 높이를 띠게 된다. 따라서 결정논리는 그림 4(e)의 스펙트럼에 대해 최대값을 갖는 하모닉스의 위치를 선택

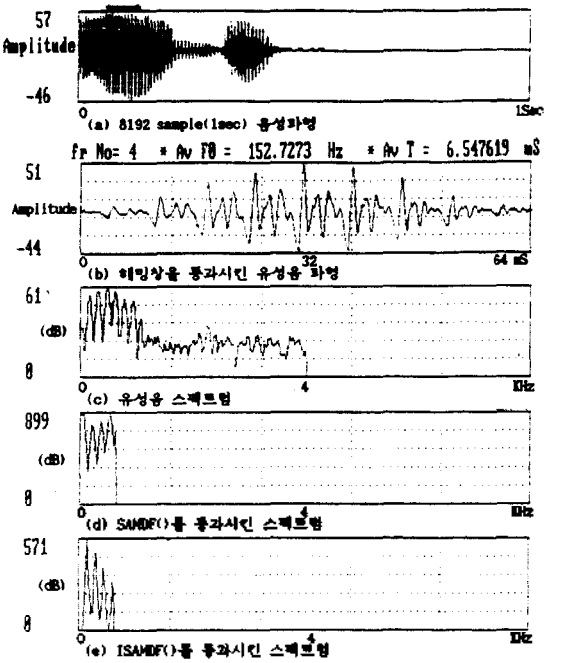

그림 4. 무잡음하에서 발성 1에 대한 결과
Fig. 4. Processing result for the utterance 1 with no noise.

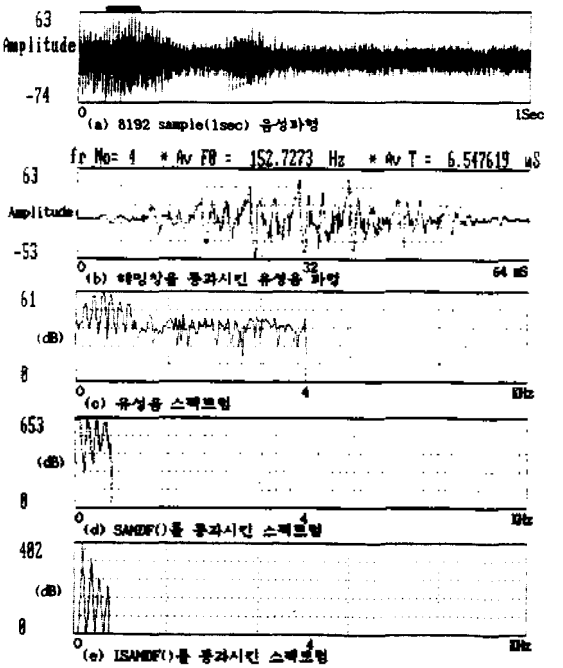

그림 5. 0-dB 가우샨 잡음하에서 발성 1에 대한 결과
Fig. 5. Processing result for the utterance 1 with 0-dB SNR.

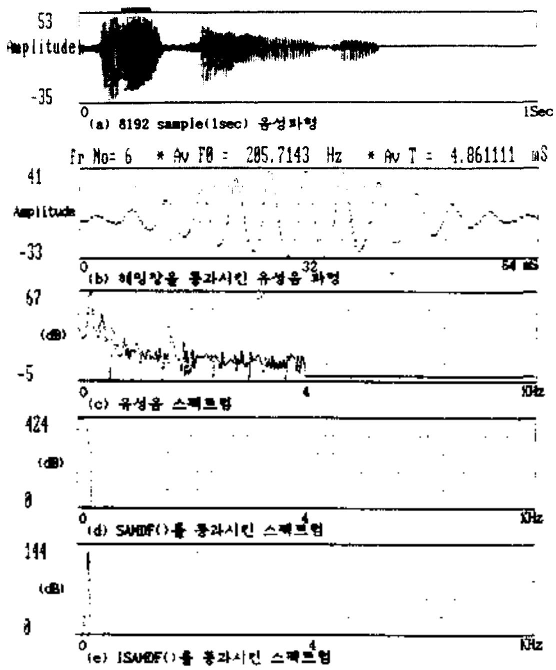

그림 6. 무잡음하에서 발성 2에 대한 결과
Fig. 6. Processing result for the utterance 2 with no noise.

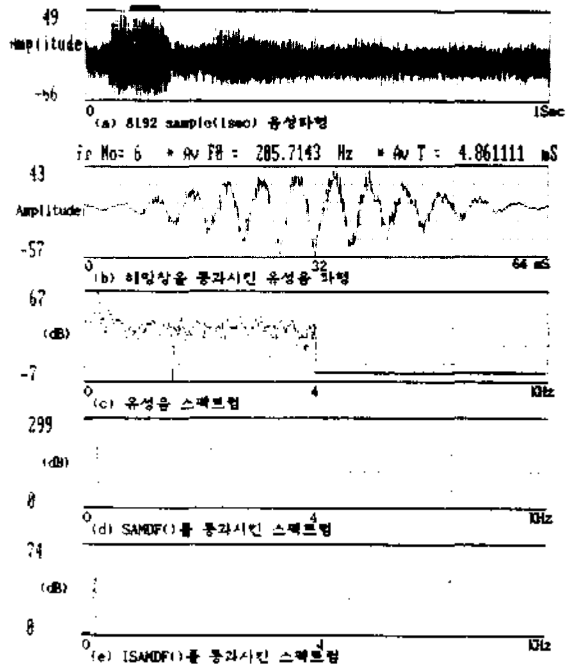

그림 7. O-dB 가우산 잡음하에서 발성 2에 대한 결과
Fig. 7. Processing result for the utterance 2 with O-dB SNR.

하면 기본주파수가 검출된다.

그림 4에서 처리한 음성시료를 배경 잡음이 있는 경우에 대해서도 스펙트럼의 기본 하모닉스를 잘 검출할 수 있는지를 보기 위해, 음성시료의 파형에 대해 실효치를 계산하고 그와 똑같은 값으로 가우산 잡음을 만들어 SNR이 O-dB가 되도록 섞었다. 이때의 음성 파형, 해밍창을 적용한 파형, 이의 스펙트럼, SAMDF(·)에 통과시킨 스펙트럼, 그리고 그의 ISAMDF 함수에 통과시킨 스펙트럼을 그림 5의 (a), (b), (c), (d), (e)에 결과를 각각 나타내었다. 잡음이 인가된 경우에도 잡음이 없는 경우와 마찬가지로 기본 하모닉스 검출이 잘 이루어졌다.

그림 6은 음성시료 2에 대한 처리결과이며, 그림 7은 이에대해 O-dB 가우산 잡음을 섞었을 경우의 결과이다. 이 결과에서 볼 수 있는 것과 같이 잡음이 인가된 경우, 또는 비음 구간이나 여성화자에 많이 나타나는 단일 펄스형 음성파형에 대해서도 포먼트의 영향에 무관하게 우수한 결과를 얻었다.

VI. 결 론

주파수 영역에서의 피치검출법은 다른 영역법에 비해 잡음에 강인하며 비교적 정확한 피치를 검출할 수 있다. 그렇지만 지금까지 제안된 방법들은 제1 포만트의 영향을 많이 받는데 특히 어린이나 여성화자와 같은 단일 펄스형 스펙트럼을 갖는 경우에는 검출에러가 증가하는 등의 단점이 있었다.

이상과 같은 단점을 보완하기 위해 본 논문에서는 스펙트럼상의 포락특성을 이용한 스펙트럼 AMDF 함수를 제안하였다. 이 SAMDF함수를 적용한 결과, 유성음 스펙트럼이 갖고 있는 포먼트와 포락특성은 그대로 유지되었고 이에 대해 역스펙트럼을 구하였기 때문에 하모닉스의 피크만을 강조할 수 있었다. 또한 배경잡음이 인가된 경우에도 평탄화 되어 나타나기 때문에 잡음성 국부봉우리들이 제거되었다. 이로써 하모닉스만이 강조되어 기본주파수 검출이 용이하였다.

향후 연구해야할 과제는 주파수영역에서 스펙트럼의 하모닉스의 간격을 측정하여 피치를 검출하면 분해능이 약화 되는데 이 분해능을 높이는 연구가 이루어져야 한다.

참 고 문 헌

1. L.R. Rabiner and R. W. Schafer, *Digital processing of Speech Signals*, Englewood Cliffs, Prentice-Hall, New Jersey, 1978.
2. S. E. Stearns & R. A. David, *Signal Processing Algorithms*, Prentice Hall, Inc., Englewood Cliffs, New Jersey, 1988.
3. P. E. Papamichalis, *Practical Speech Processing*, Prentice Hall, Inc., Englewood Cliffs, New Jersey, 1987.
4. M. BAE, and S. ANN, "Fundamental Frequency Estimation of Noise Corrupted Speech Signals Using the Spectrum Comparison", J. Acoust. Soc. Korea, Vol.8, No.3, pp.57-64, June 1989.
5. S. Seneff, "Real Time Harmonic Pitch Detector", IEEE Trans. Acoust., Speech, and Signal Processing, Vol. ASSP-26, pp.358-365, Aug. 1978.
6. A. M. Noll, "Pitch Determination of Human Speech by the Harmonic Product Spectrum, the Harmonic Sum Spectrum, and a Maximum Likelihood Estimate", Proc. Symp. Comput. Processing in Commun., pp.779-798, Apr. 1960.
7. T. V. Screenivas and P. V. S. Rao, "Pitch Extraction from Corrupted Harmonics of the Power Spectrum", J. Acoust. Soc. Amer., vol. 65, pp.223-228, Jan. 1979.
8. M. BAE and S. ANN, "On the Time-Frequency Hybrid Technique for Detecting the Pitch of Noise Corrupted Speech Signals(Time Domain Processing)", J. Acoust. Soc. Korea, Vol.9, No.1, pp.87-94, Feb. 1990.
9. L. R. Rabiner, "On the use of Autocorrelation Analysis for Pitch Detection", IEEE Trans. Acoust., Speech and Singal Proc., Vol. ASSP-25, pp.24-33, Feb. 1977.
10. M. Ross, H. Schafer, A. Cohen, R. Freuberg, and Manley "Average Magnitude Difference Function Pitch Extractor", IEEE Trans., Vol. ASSP-22, pp. 353-362, October 1974.
11. M. Lahat, R. J. Niederjohn, and D. A. Krubsack, "A Spectral Autocorrelation Method for Measurement of the Fundamental Frequency of Noise Courrupted Speech", IEEE Trans., Acoust., Speech, and Signal processing, Vol. ASSP-35, No.6, June 1987.

▲배 명 진(정회원)
: 현 호서대학교 전자공학과 조교수(9권 1호 참조)

▲안 수 길(정회원)
: 현 서울대학교 전자공학과 교수(9권 1호 참조)

▲박 찬 수(학생회원) 1965년 1월 15일생

1991년 2월 : 호서대학교 전자공학과 졸업(공학사)
1991년 3월~현재 : 호서대학교 대학원 전자공학과 석사과정