

고음질 음성합성을 위한 LSP를 이용한 피치검출 성능향상에 관한 연구

A Study on the Pitch Extraction Improvement Using LSP for the Synthesis of High Speech Quality

서 지 호*, 김 종 국*, 배 명 진*
(Ji-Ho Seo*, Jong-Kuk Kim*, Myung-Jin Bae*)

*숭실대학교 정보통신전자공학부
(접수일자: 2009년 11월 16일; 채택일자: 2009년 12월 16일)

본 논문에서는 스펙트럼 신호를 최대한 평탄화시킴으로써 포먼트의 영향을 제거하고 고조파 성분을 분리해 내어 이를 피치검출에 사용한다. 스펙트럼 신호로부터 포먼트의 영향과 천이진폭의 영향을 제거하기 위해 주파수 대역을 LSP(Line Spectrum Pair)를 기준으로 서브밴드로 나누고 각각의 서브밴드에서 기울기를 취한 후에 역기울기로 스펙트럼을 보상한다. 실험 결과 제안한 방법이 LPC법, Lifter법, Cepstrum법을 이용하여 평탄화시킨 때 보다 평탄화 정도가 좋아짐을 알 수 있다. 또한 제안한 방법 이외에 가장 양호한 성능을 나타낸 LPC법을 이용하여 피치를 구했을 때 제안한 방법의 조오율이 평균 1.30% 감소하였다. 또한 제안한 방법은 잡음을 부가한 음성의 경우에도 낮은 에러율을 보여 배경잡음에 강하다는 것을 알 수 있었다.

핵심용어: 피치, 피치검출, 스펙트럼 평탄화, LSP, 포먼트

투고분야: 음성처리 분야 (2,4)

In this paper, the pitch is detected after the elimination of formant ingredients by flattening the spectrum in frequency domain. In order to remove impact of formant and transition frequency in the signal spectrum, formant envelop is made by linear interpolation with any points each sub-band and the spectrum of speech signal is compensated by the reverse of the envelop interpolated linearly after we divide frequency band into several segment based on LSP and detect the points. The experimental result showed the proposed method appeared an outstanding performance in compared with LPC, Cepstrum, Lifter methods. The method reduced the gross error rate 1.30% than the LPC method which appeared a good performance except the proposed method. Also, the proposed method showed low error rate in noise environment.

Keywords: Pitch, Pitch Detection, Spectrum Flattening, LSP, Formant

ASK subject classification: Speech Signal Processing (2,4)

I. 서론

음성신호의 피치주기를 정확하고 신뢰성 있게 측정하는 것은 다음과 같은 여러 가지 이유 때문에 매우 어렵다. 첫 번째 이유는 성분의 여기파형이 완전히 주기적인 파형이 아니라는 것이다. 비록 주기적인 파형의 주기를 찾는 것이 간단하다고 할지라도 한 주기 내에 있는 파형의 세

부구조와 함께 변화하는 음성파형의 주기를 측정하는 것은 어려운 작업이다 [5].

피치주기를 측정하는데 있어서 두 번째 어려움은 성도와 성문간의 상호작용 때문이다. 어떤 경우에는 성도의 포먼트가 성문파형의 구조를 완전히 바꿀 수 있기 때문에 피치주기를 검출하기가 어렵게 된다. 일반적인 상호작용은 조음기관이 빠르게 변화할 때 일어나며 피치검출 시에는 방해요소로 나타나게 된다 [3].

세 번째 문제점은 유성음구간 동안 각 피치주기의 정확한 시작과 끝을 정의하기 어렵다는 것이다. 피치주기의

정확한 시작과 끝의 위치를 선택하는 것은 임의적일 수 있기 때문이다. 예를 들면 음성파형에 근거하여 각 피치 주기의 시작과 끝점을 정의하기 위한 후보로는 주기내의 영교차 간격을 들 수 있다. 이러한 측정에서 필요한 것은 각 피치주기의 시작과 끝점의 정확한 위치를 정의하기 위해 주기에서 주기까지 일관성이 있어야 한다는 것이다. 이러한 일관성이 부족하면 잘못된 피치주기를 측정하게 된다. 음성파형의 주기성분을 영교차 단위로 검출하면 포먼트 구조에 민감하게 된다 [1][5].

피치를 검출할 때 나타나는 네 번째 문제점은 무성음과 낮은 레벨의 유성음을 구별해야 하는 것이다. 많은 경우에 무성음 구간과 레벨이 낮은 유성음 구간의 변화는 별로 뚜렷하지 않기 때문에 정확한 위치를 찾기가 어렵다.

다섯 번째 문제점은 전화시스템을 통하여 전송되는 음성이 피치검출 문제에 직면할 때에 부가적인 복잡성이 발생하게 된다는 것이다. 음성에 대한 전화시스템의 영향은 선형, 비선형처리를 포함하면서 음성신호에 잡음이 추가되는 경우이다. 전화시스템을 선형여파기로 간주하면 기본주파수와 고조파들의 성분을 약하게 만들 수 있는 대역통과 여파기로 볼 수 있다. 이 때문에 전화음질에 대해 신호의 주기성을 검출하기가 더 어렵게 된다. 음성신호에 대한 전화시스템의 비선형특성은 전송시스템의 채널에 의해 좌우된다 [10].

본 논문에서는 스펙트럼 신호를 최대한 평탄화시킴으로써 포먼트의 영향을 제거하고 고조파 성분을 분리해 내어 이를 피치검출에 사용한다. LSP를 이용하여 적응적 밴드에서 평탄화를 시도하고 이를 피치 검출에 이용하였다.

II. 피치검출법

피치 검출방법은 시간영역, 주파수영역, 시간주파수 혼성영역 검출법으로 나눌 수 있다 [7].

시간영역 피치검출법은 시간영역에서 직접 처리하기 때문에 다른 영역으로의 변환과정이 불필요하며 합, 차, 비교논리 등에 의해서만 처리가 가능해진다. 피치의 범위는 보통 2.5에서 25 msec 로 알려져 있고 음성을 8 kHz 로 표본화하여도 그 범위가 20-200 표본 사이에 나타나기 때문에 시간영역에서의 검출은 분해능이 높다는 장점이 있다. 보통 음성의 분석은 프레임 단위로 처리되고 파열음의 경우, 음소의 변화가 한 프레임 안에서도 일어날 수 있다. 따라서 시간영역에서의 피치측정은 음소변화에

따른 특성을 검출할 수도 있다. 그렇지만, 음소의 종류나 결합위치에 따라 음소파형의 외곽선이 변화하는데, 이 변화를 측정하기 어렵기 때문에 이러한 파형부분에서 피치검출 오차가 많이 발생한다. 또한 특성전송 채널을 통과한 경우나 배경잡음이 부가된 경우에는 이의 영향이 커져서 피치검출 오차가 높아지는 단점이 있다 [11].

시간영역에서의 피치검출법들은 일반적으로 음성파형의 주기성을 강조하고 나서 결정논리에 의해 주기성을 검출하게 된다. 주기성 강조는 성도의 공명현상에 나타나는 포먼트들의 영향을 제거시키고 여기원의 피치만을 강조하는데 그 목적이 있다. 시간영역 피치검출법으로는 Parallel Processing, Autocorrelation, AMDF, Data Reduction, Harmonic Matching법 등이 종래의 기법으로 알려져 있다 [12].

주파수영역에서의 피치검출은 그 처리과정상 주파수 영역으로의 변환이 필요해서 계산시간이 많이 걸리고 복잡하지만, DSP칩의 시판으로 인해 계산시간의 부담이 제거되어 최근에 다시 관심의 대상이 되고 있다. 주파수영역 피치검출기는 보통 음성스펙트럼에서 고조파들의 간격을 측정하여 기본주파수를 구한다. 고조파를 파라미터로 하여 기본주파수를 검출하는 방법으로는 Harmonic Spectrum, Peak-Valley, Comb Filtering, Spectrum Autocorrelation법 등이 제안되어져 있다 [4][9].

피치 검출 시에는 각 영역별로 처리가 용이하고 특징적인 장점을 취할 수 있다. 시간-주파수 혼성법에서는 시간영역에서의 장점과 주파수영역에서의 특징을 변환특성을 이용하여 처리하는 방법이며, Cepstrum, SIFT, LPC inverse filtering and AMDF법 등이 제안되어져 있다 [2][8].

III. 제안한 방법

대부분의 음성분석에서는 신호가 시간적으로 변하는 성도 여파기성분과 여기성분으로 모델링될 수 있다고 가정한다. 여기 성분은 성문을 통과하여 나오는 준 주기적인 공기의 흐름과 성도를 스치며 발생하는 넓은 대역의 잡음으로 분류한다. 성도여파기의 응답은 일반적으로 천천히 변화하는데 주파수 영역에서 스펙트럼 신호의 포먼트 포락선을 의미한다 [3][13].

3.1. 스펙트럼 평탄화 과정

음성신호는 FFT변환을 통해 주파수 영역에서 스펙트

럼 분석이 이루어진다. 그림 1은 본 논문에서 제안한 주파수 평탄화 알고리즘의 블록도이다.

LSP를 구하는 과정을 보면 입력 음성신호를 받고 각 프레임마다 Hamming window를 취한다. 윈도우를 취한 데이터를 이용해 LPC 계수를 찾아내어 LSP 파라미터를 얻어낼 수 있다. 10차의 LSP 파라미터는 각각 주파수 위치를 나타내고 있으며 두 파라미터의 위치간격으로 포먼트가 위치한 주파수 대역을 알아낼 수 있다. 따라서 다음 식 (1)과 같이 나타낼 수 있다.

$$\begin{aligned}
 &LSP_1(i) = LSP(i) - LSP(i+2) - LSP(i) \\
 &LSP_2(i) = LSP(i+2) - LSP(i+1) \\
 &LSP_3(i) = LSP(i+1) - LSP(i) \\
 &D_1(i) = \frac{LSP_1(i)}{LSP(i)} \quad D_2(i) = \frac{LSP_2(i)}{LSP(i)} \quad 1 < i < 8 \quad (1)
 \end{aligned}$$

식 (1)의 D_1 과 D_2 중에 0.35보다 작은 값을 선택하여 각 밴드폭의 위치를 알아내어 저장한다. 이를 순차적으로 반복하여 포먼트의 위치정보를 얻어낼 수 있다. 구해진 밴드폭을 가진 인접한 두 LSP 파라미터를 $LSP_{(det)}$ 라 하면

$$D(n) = \frac{LSP_{(det)}}{2} \quad (2)$$

$D(n)$ 은 식 (2)와 같이 나타낼 수 있다. 주파수 영역에서 좌표로 나타내면 $(D(1), DB(D(1)), (D(2), DB(D(2)), (D$

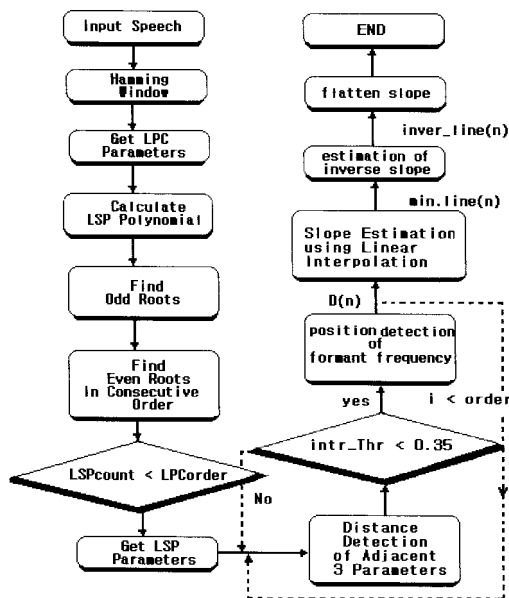


그림 1. 제안한 스펙트럼 평탄화 과정
Fig. 1. Block diagram of proposed Spectrum Flattening algorithm.

(3), $DB(D(3)), \dots, (D(n), DB(D(n))$ 이다.

$$\begin{aligned}
 &Line(m) = DB(D(n)) + (m - D(n))f(D(n)) + \frac{(m - D(n))^2}{2} f''(D(n)) \dots \\
 &\text{단, } f(D(n)) = \frac{DB(D(n+1)) - DB(D(n))}{D(n+1) - D(n)} \quad (3)
 \end{aligned}$$

위의 식 (3)을 이용하여 선형 보간하여 주파수영역에서의 좀더 세밀한 선형적인 기울기를 추정할 수 있다. 얻어낸 기울기를 이용하여 스펙트럼 경사를 평탄화시키기 위해서는 식 (4)를 이용하여 역 기울기를 구한 후에 평탄화시킨다. 식 (5)는 주파수 영역에서 평탄화된 스펙트럼의 결과를 보여준다.

$$\begin{aligned}
 &mindata = \min(DB(D(n))) \\
 &inver_Line(m) = Line(m) - 2(Line(m) - mindata) \quad (4) \\
 &flatten(m) = DB(m) + inver_Line(m), 0 < m < frame \quad (5)
 \end{aligned}$$

3.2. 피치 검출 과정

평탄화된 스펙트럼 신호로부터 기본주파수 (피치)를 구하기 위해 자기상관법을 사용하였다. $P(k)$ 가 로그 스펙트럼신호일 때 자기상관법은 식 (6)과 같이 정의된다.

$$R(m) = \frac{2}{N} \sum_{k=0}^{N/2} P(k)P(k+m) \quad (6)$$

여기서 m 은 주파수 영역에서 지연된 샘플수를 나타낸다. $P(k)$ 는 좌우대칭이므로 자기상관은 FFT 크기인 N 의 1/2만 수행하여도 된다. 시간영역에서 효과적인 자기상관법을 사용하기 위해 전처리과정을 거치듯이 주파수 영

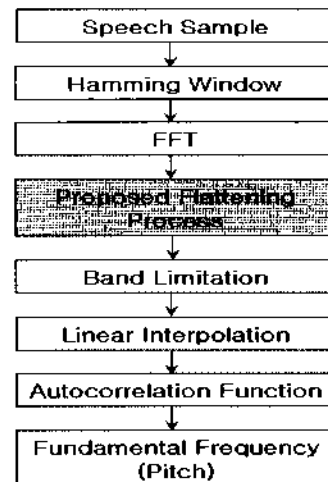


그림 2. 제안한 피치검출 과정
Fig. 2. Block diagram of proposed Pitch Search algorithm.

역에서도 마찬가지로 전처리과정이 필요하다. 먼저 자기 상관법은 안정구간에서 적용되어야 한다. 하지만 고주파 영역에서 안정적이지 못한 고조파가 나타나므로 분석구간을 0 - 1 kHz까지 대역제한시킨다. 또한 주파수 해상도를 고려해야 한다. 주파수 해상도는 FFT 포인트 수에 비례하지만 그 길이는 항상 제한되어 있다. 따라서 주파수 해상도를 보상하기 위해 신호를 선형보간한다. 이는 더욱 정확한 피치검출을 할 수 있게 한다.

IV. 실험 및 결과

4.1. 실험 과정

컴퓨터 시뮬레이션에 이용한 장비는 IBM PC (2.0 GHz) 시스템이며 여기에 음성신호를 입출력하기 위한 상용화된 16비트 A/D 변환기를 인터페이스하여 8 kHz의 표본율로 데이터를 입력하였다. 처리결과와 성능을 측정하기 위해 다음의 대표적인 문장을 연령층이 다양한 남녀 5명의 화자가 각 5번씩 발성하여 시료로 사용하였다. 시료는 두드러진 피크를 가지지 않고 잡음 30 dB를 가진 방에서 녹음하였다.

- 발성1: /인수네 꼬마는 천재소년을 좋아한다./
- 발성2: /예수님께서 천지 창조와 교훈을 말씀하셨다./

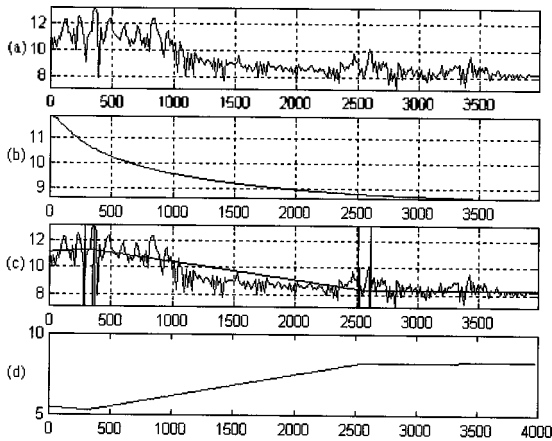


그림 3. 유성음의 기울기 측정
 (a) 유성음의 스펙트럼분석
 (b) 자기상관법을 이용한 기울기 측정
 (c) 제안한 방법을 이용한 기울기 측정
 (d) 제안한 방법의 역기울기 측정
 Fig. 3. Results of gradient estimate of the voice.
 (a) Spectrum analysis of the voiced
 (b) Gradient estimate with ACF
 (c) Gradient estimate with the proposed method
 (d) Inverse gradient estimate with the proposed method

- 발성3: /창공을 날으는 인간의 도전은 끝이 없다./
- 발성4: /숭실대학교 정보통신과 음성통신 연구팀이다./

알고리즘의 처리를 위한 분석 프레임의 길이는 256표본을 사용하였고, 계수가 10자인 특징벡터로써 LSP를 구하였다. 10차 LSP는 2~4개의 포만트를 모델링하게 되고, 제안한 방법에 의해서 대역을 5~9개로 나눈다. 각 구간에서 포락선의 기울기를 구하고, 다시 역기울기를 추정한다. 추정된 역기울기를 통해서 낮은 쪽의 주파수 성분들을 보상하여 준다. 평탄화를 통하여 성도성분의 공진 특성을 잘 나타내어 주게 된다.

그림 3은 유성음의 경우 제안한 방법을 이용하여 기울기를 측정된 결과를 나타낸 것이다. (c)는 제안한 방법으로 LSP를 추출하여 포만트위치를 추정하여 선형보간에 의해 기울기를 측정된 것을 나타낸다. (d)는 스펙트럼 평탄화를 위한 역기울기를 구한 그래프이다. 다음의 그림 4는 보다 분석이 용이한 LPC분석으로 서로 비교한 것이다. (d)그래프에서 보면 거의 decade가 없는 것을 그림에서 확인할 수 있는데 이는 보다 나은 평탄화를 가지고 있음을 나타내고 있다.

그림 5는 유성음과 같이 무성음에 해당하는 프레임들 제안한 방법으로 기울기를 측정된 결과를 나타낸 것이다. (c)는 제안한 방법으로 LSP를 추출하여 포만트위치를 추정하여 선형보간에 의해 기울기를 측정된 것을 나타

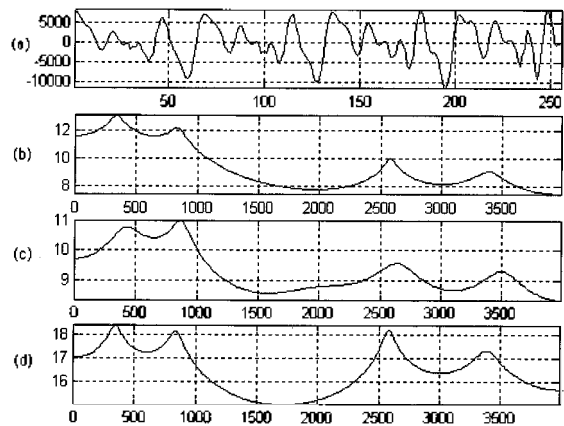


그림 4. 유성음의 LPC분석
 (a) 유성음 파형
 (b) 원음성의 LPC분석
 (c) 자기상관법을 이용한 LPC분석
 (d) 제안한 방법을 이용한 LPC분석
 Fig. 4. Results of LPC analysis of the voice.
 (a) The voice in time domain
 (b) LPC analysis of original speech
 (c) LPC analysis with ACF
 (d) LPC analysis with the proposed method

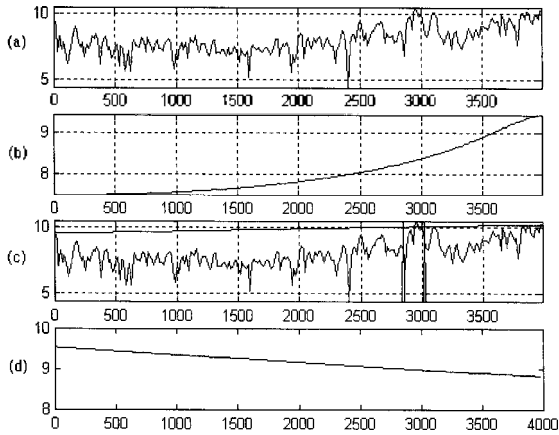


그림 5. 무성음의 기울기 측정
 (a) 무성음의 스펙트럼분석
 (b) 자기상관법을 이용한 기울기 측정
 (c) 제안한 방법을 이용한 기울기 측정
 (d) 제안한 방법의 역기울기 측정
 Fig. 5. Results of gradient estimate of the unvoiced.
 (a) Spectrum analysis of the unvoiced
 (b) Gradient estimate with ACF
 (c) Gradient estimate with the proposed method
 (d) Inverse gradient estimate with the proposed method

낸다. (d)는 스펙트럼 평탄화를 위한 역기울기를 구한 그래프이다. 그림 6은 성도모델을 잘 표현할 수 있는 LPC로 무성음에 대해 비교 분석한 것이다. 그림에서 보는 것과 같이 거의 정확히 보상이 잘되는 것을 알 수 있다.

노이즈 환경에서도 제안된 방법으로 올바르게 피치검출이 가능한지 테스트하기 위해 녹음한 음성신호에 노이즈를 인가한 시료를 가지고 실험하였다. 시간 축 상에서 원래의 음성신호 $x(t)$ 에 잡음신호 $n(t)$ 가 인가된 입력신호 $y(t)$ 는 수식 (9)에 의해서 유도될 수 있다. σ_{nn} 은 SNR만큼 노이즈가 부과된 노이즈의 분산이며, σ_n 은 노이즈신호 $(n(t))$ 의 분산이다. σ_{nn} 은 수식 (8)을 통해서 구할 수 있고, σ_n 은 깨끗한 음성의 분산이며, SNR은 수식 (7)에 의해서 구할 수 있다.

$$SNR = 10 \log_{10} \frac{\text{signal energy}}{\text{reconstruction noise energy}}$$

$$= 10 \log_{10} \left(\frac{\sum_n x(n)^2}{\sum_n (x(n) - \hat{x}(n))^2} \right)$$

$$= 10 \log_{10} \frac{\sigma_s^2}{\sigma_{nn}^2} \quad (7)$$

$$\sigma_{nn}^2 = \sigma_s^2 + 10^{-SNR} \quad (8)$$

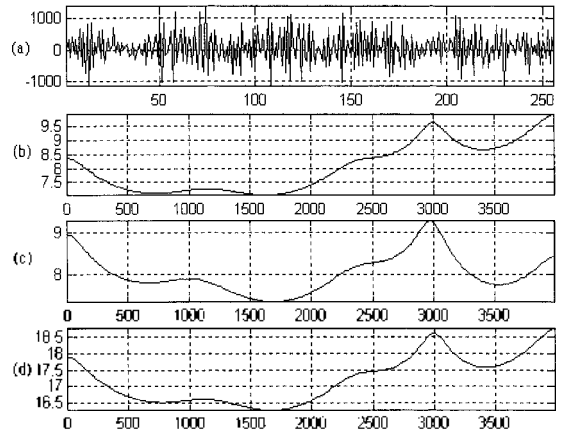


그림 6. 무성음의 기울기 측정
 (a) 무성음 피형
 (b) 원음성의 LPC분석
 (c) 자기상관법을 이용한 LPC분석
 (d) 제안한 방법을 이용한 LPC분석
 Fig. 6. Results of LPC analysis of the unvoiced.
 (a) The unvoiced in time domain
 (b) LPC analysis of original speech
 (c) LPC analysis with ACF
 (d) LPC analysis with the proposed method

$$y(t) = x(t) + n(t) \sqrt{(\sigma_{nn}^2 / \sigma_n^2)} \quad (9)$$

비교 평가 방법은 분산을 이용하였다. 분산을 계산하기 전에 각 결과 신호들은 최대값이 0이 되도록 정규화시키고 평균이 0인 분산을 계산하였다. 본 논문에서 사용한 분산은 식 (10)과 같다. 여기서 N은 FFT포인트 수이고 스펙트럼 신호가 Y축으로 대칭이기 때문에 분산은 N/2까지만 구한다. 또한 k는 주파수 영역에서의 샘플인덱스이고 m은 평균을 의미한다. 이때 m값은 0을 사용하여 0을 기준으로 평탄화의 정도를 평가하였다.

$$\text{Variance} = \frac{2}{N} \sum_{k=1}^{N/2} (x(k) - m)^2 \quad (10)$$

4.2. 결과 평가

제안된 여기 스펙트럼 검출법의 객관적인 비교 평가 및 성능 평가의 비교 대상은 일반적으로 포먼트 정보를 얻기 위해 사용되어온 LPC법, Cepstrum법, lifter function법이다. 스펙트럼을 평탄화하기 위해 각각의 방법과 제안한 방법을 적용하여 분산을 계산하였고 평가하였다.

표 1과 2는 각각 남성화자와 여성화자의 평탄화된 스펙트럼의 분산 값을 나타내었다. 표에서와 같이 기존의 방법과 비교해 제안한 방법의 분산 값이 가장 작은 것을

알 수 있다.

표 3은 LPC법을 이용했을 경우 피치검출에 대한 Gross error rate을 나타낸다. SNR이 각각 30 dB, 20 dB, 10 dB, 6 dB, 3 dB일 때 각 발성에 대한 Gross error rate이다. 평균적인 에러율은 결과값과 같다. 표 4는 제안한 방법으로 스펙트럼을 평탄화하고 피치검출을 하여 얻은 Gross error rate이다. LPC법 보다 평균 1.30 %의 오오율이 감소하였다.

그림 7은 발성 1의 피치 변화도를 나타낸다. 음성시료의 로그 스펙트럼을 평탄화하고 이의 고조파 간격을 측정하여 위와 같은 실험결과를 얻었다. 그림 7- (a)는 발성의 음성신호이다. 이에 대한 프레임별 피치값을 수동으로 찾아 나타낸 것이 그림 7- (b)이다. 이는 기준피치로

사용되었다. 그림 7- (c)와 7- (d)는 각각 LPC법과 제안한 방법을 이용하여 구한 결과 값이다.

V. 결론

음성신호처리분야에서 피치를 정확히 검출하면 음성 인식 시에 화자에 따른 영향을 줄일 수 있기 때문에 인식의 정확도를 높일 수 있고, 음성합성 시에 자연성과 개성을 유지하거나 쉽게 변경할 수 있다. 또한 분석 시에 피치와 동기시켜 분석하면 성문의 영향이 제거된 정확한 성도 파라미터를 얻을 수 있게 된다 [4][6].

본 논문에서는 주파수 영역에서 포먼트의 영향을 효과적으로 제거하여 스펙트럼을 평탄화시키고 FFT의 포인터 수를 늘리지 않고도 주파수 해상도를 높여 피치 검출의 정확성을 높이는 피치 검출법을 제안하였다. 먼저 주파수 대역을 특징벡터 LSP를 기준으로 나누고, 각 대역에서 포락선의 기울기를 구하며, 다시 역기울기를 추정한다. 추정된 역기울기를 통해서 낮은 쪽의 주파수 성분들을 보상하여 준다.

실험 결과 LPC법, Lifter법, Cepstrum법을 이용하여 평탄화시킬 때 제안한 방법보다 평탄화 정도가 떨어짐을 알 수 있었다. 또한 제안한 방법 이외에 가장 양호한 성능을 나타낸 LPC법을 이용하여 피치를 구했을 때 제안한

표 1. 남성화자의 분산값 (dB)

Table 1. Variance of male voice.

	LPC	Cepstrum	lifter function	Proposed method
발성1	280.68	1064.55	672.14	211.68
발성2	244.32	970.21	671.57	205.41
발성3	272.45	995.09	737.41	224.65
발성4	268.71	955.67	582.85	198.33

표 2. 여성화자의 분산값 (dB)

Table 2. Variance of female voice.

	LPC	Cepstrum	lifter function	Proposed method
발성1	436.78	909.22	912.00	324.02
발성2	407.07	896.86	928.65	305.64
발성3	455.54	1068.44	1250.38	395.08
발성4	377.90	745.96	718.58	255.74

표 3. LPC법에 대한 Gross error rate (단위: %)

Table 3. Gross error rate with LPC method.

	30dB	20dB	10dB	6dB	3dB
발성1	2.43	3.77	4.56	4.62	5.40
발성2	1.97	3.18	4.87	6.01	7.36
발성3	3.68	3.63	5.93	6.89	8.40
발성4	3.52	3.91	5.91	7.23	9.10
Average	2.90	3.62	5.32	6.19	7.57

표 4. 제안한 방법에 대한 Gross error rate (단위: %)

Table 4. Gross error rate with the proposed method.

	30dB	20dB	10dB	6dB	3dB
발성1	1.33	1.93	3.13	3.87	4.63
발성2	1.82	2.12	3.52	4.26	5.71
발성3	1.97	2.99	3.84	5.32	6.92
발성4	2.55	2.92	4.25	5.56	7.72
Average	1.92	2.49	3.69	4.75	6.25

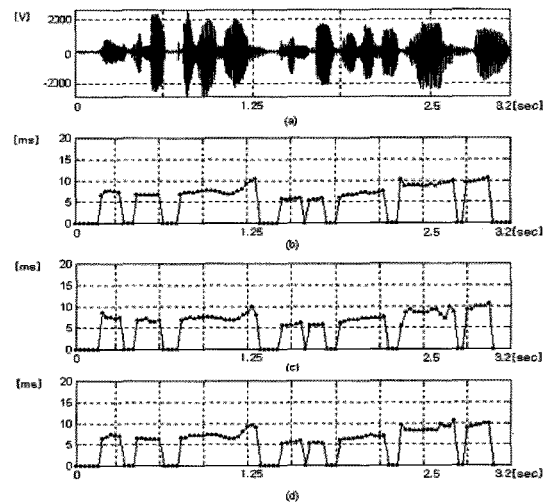


그림 7. 피치 궤적(발성 1)

- (a) 시간영역의 음성신호
- (b) 기준피치
- (c) LPC법
- (d) 제안한 방법

Fig. 7. Pitch contour(utterance 1).

- (a) Speech signal in time domain
- (b) Criterion pitch
- (c) Result of pitch search with LPC method
- (d) Result of Pitch search with the proposed

방법의 조오율이 평균 1.30% 감소하였다. 제안한 방법은 잡음을 부가한 음성의 경우에도 낮은 에러율을 보여 배경 잡음에 강하다는 것을 알 수 있었다.

피치정보는 음성 신호에 있어서 매우 중요하다. 앞으로 본 연구를 바탕으로 배경잡음에도 훨씬 강하고 더 낮은 에러율을 갖는 피치검출기를 구현하기 위해 계속 연구해나가야 할 것이다. 아울러 피치검색 시간과 계산량의 감소, 알고리즘의 단순화를 위한 방법을 강구해야 할 것이다.

참고 문헌

1. C. J. Jung, M. K. Ham and M. J. Bae, "On a pitch alteration technique of speech using the asymmetric weighted window," *IEEE Comm. Society, Proc. of MILCOM'99*, vol. 2, pp. 1439-1443, 1999.
2. P. Bastien, "Pitch shifting and voice transformation techniques," <http://www.tc-helicon.com>, 2001.
3. B. Gold and N. Morgan, *Speech and audio signal processing*, John Wiley & Sons, New York, 2000.
4. A. M. Kondoz, *Digital speech coding for low bit rate communications systems*, John Wiley & Sons, New York, 2004.
5. J. K. Kim and M. J. Bae, "A study of pitch extraction method by using harmonics peak-lifting in speech spectrum," *In ASK, Proceedings of ICSP 2001*, Vol. 1, pp. 617-622, 2001.
6. E. Keller, G. Bailly, A. Monaghan, J. Terken and M. Huckvale, *Improvement in speech synthesis*, John Wiley & Sons, New York, 2002.
7. J. K. Kim, D. S. Na and M. J. Bae, "On a pitch alteration technique in transformation domain of speech signals," *In International Conference on Life System Modelling and Simulation*, Vol. 14, pp. 522-526, 2007.

8. D. G. Kang, M. J. Bae and S. G. Ann, "On Extraction the Center Pitch using the spectrum Leakage Phenomenon for Noise Corrupted Speech Signals," *J. Acoust. Soc. Korea*, vol. 10, no. 1, pp.37-46, 1991.
9. S. Mitra, *Digital signal processing: a computer-based approach*, McGraw-Hill, New York, 2001.
10. 김영규, 서지호, 배명진, "잡음환경에서 서브밴드 스케일링 피치 변경법에 관한 연구," *한국음향학회 2003년도 정기총회 및 추계 학술발표대회 논문집*, 22권, 2호, 35-38쪽, 2003년.
11. W. B. Kleijn and K. K. Paliwal, *Speech Coding and Synthesis*, Elsevier Science, New York, 1998.
12. J. K. Kim, H. S. Hahn, U. J. Yoon and M. J. Bae, "On a Pitch Alteration for Speech Synthesis Systems," *Wireless Personal Communications, Springer Netherlands*, Vol. 50, No. 4, pp. 435-446, Oct, 2009.
13. O. Turk and L. M. Arslan, "Subband Based Voice Conversion," *Proc. of the ICSLP 2002*, Vol. 1, pp. 289-292, 2002.

저자 약력

• 서 지 호 (Ji-Ho Seo)



1998~2003 경성대학교 컴퓨터공학과 (공학사)
 2003~2005 숭실대학교 정보통신공학과 (공학석사)
 2005~2007 온타리오주 선임연구원
 현재 숭실대학교 정보통신공학과 박사과정

• 김 종 국 (Jong-Kuk Kim)

한국음향학회지 제 26권 제8호 참조

• 배 명 진 (Myung-Jin Bae)

한국음향학회지 제 26권 제8호 참조