

2.4 kbps 하모닉-CELP 코더를 위한 웨이블릿 피치 검출기

Wavelet-based Pitch Detector for 2.4 kbps Harmonic-CELP Coder

방 상 운*, 이 인 성*, 권 오 주**
(Sang-Un Bang*, In-Sung Lee*, O-Joo Kwon**)

*충북대학교 전파공학과, **국방과학연구소
(접수일자: 2003년 8월 18일; 채택일자: 2003년 10월 9일)

본 논문은 2.4 kbps 하모닉-CELP 부호화기를 위한 피치 검출기의 설계 방법과 전이 시점을 검출하고 그 값을 기준으로 유/무성음 변환 구간에 대한 합성 윈도우를 달리하여 효과적인 파형 보간이 이루어지도록 하기 위한 방법을 제안하였다. 하모닉-CELP 부호화기에서 유성음 구간은 과거와 현재 프레임의 표준 파형을 보간하여 이루어지므로 전이 구간에서 피치 주기가 반으로 줄거나 두 배로 예측되어질 경우, 피치주기의 심한 변화량에 의해 파형 왜곡 및 프레임 경계에서의 불연속을 발생시킨다. 또한 하모닉 합성을 할 때 삼각 윈도우에 의한 중첩-합산 (overlap-add) 방법을 사용하기 때문에 전이 구간에서 유성음 구간의 신호가 순간적인 증가 (감소)를 할 경우 삼각 윈도우의 영향으로 합성 여기 신호가 선형 증가 (감소)하는 단점이 있다. 우선 피치 검출기의 설계는 정확한 피치의 검출을 하되 피치 더블링에 의한 프레임 불연속성을 막기 위해 1차 혼성 검색법을 사용하였으며, ACF에 의한 2차 검색으로 피치의 정확도를 높였다. 그리고 삼각 윈도우에 의해 합성 파형이 선형 증가하던 문제는 웨이블릿에 의해 검출된 GCI를 이용하여 전이 시점을 검출한 후, 그 값을 기준으로 사다리꼴 윈도우 설정을 하여 해결하였다. 실험 결과 파형 보간 코더에서 가장 문제가 되었던 피치 더블링이 사라졌으며, 피치 검색 오차율은 ACF 검출법에 비해 5.4% 개선되었고 웨이블릿에 의한 검출법에 비해 2.66% 개선되었다. 전이 구간에서의 MOS값은 0.13 향상되었다.

핵심용어: 표준 파형 보간 코더, 웨이블릿, 피치 검출기

투고분야: 음성처리 분야 (2.2)

This paper presents the methods that design the Wavelet-based pitch detector for 2.4 kbps Harmonic-CELP Coder, and that achieve the effective waveform interpolation by decision window shape of the transition region. Waveform interpolation coder operates by encoding one pitch-period-sized segment, a prototype segment, of speech for each frame, generate the smooth waveform interpolation between the prototype segments for voiced frame. But, harmonic synthesis of the prototype waveforms between previous frame and current frame occur not only waveform errors but also discontinuity at frame boundary on that case of pitch halving or doubling. In addition, in transition region since waveform interpolation coder synthesizes the excitation waveform by using overlap-add with triangularity window, therefore, Harmonic-CELP fail to model the instantaneous increasing speech and synthesis waveform linearly increases. First of all, in order to detect the precise pitch period, we use the hybrid 1st pitch detector, and increase the precision by using 2nd ACF-pitch detector. Next, in order to modify excitation window, we detect the onset, offset of frame by GCI. As the result, pitch doubling is removed and pitch error rate is decreased 5.4% in comparison with ACF, and is decreased 2.66% in comparison with wavelet detector. MOS test improve 0.13 at transition region.

Keywords: Prototype waveform interpolation coder, Wavelet, Pitch detection

ASK subject classification: Speech signal processing (2,2)

I. 서론

1997년 Texas Instruments에 의해 개발된 MELP[1] 부호화기가 2.4 kbps DoD코더로 표준화된 이후 저전송률 부호화 알고리즘에 대한 개발은 국내외에서 군통신과 개인 단말기 통신 분야에 걸쳐 활발히 이루어지고 있다. MELP는 혼합 부호화 방식을 사용한 대표적인 보코더로서 기존 MBE, 정현파 모델링 및 PWI처럼 유/무성음 분류에 의해 음성을 부호화하던 방식에서 벗어나 혼합 영역에 대한 별도의 처리 기법을 추가함으로써 다양한 모드 결정이 가능하도록 한 것이 특징이다[2-4].

이러한 MELP의 혼합 영역 부호화 기법은 기존 보코더의 장점을 결합하여 모드 결정에 따른 최적화 파라미터를 추출하도록 하는 계기가 되었다. 그 결과 STC-MBE, PWI-STC, PWI-ZFE과 같은 기법의 알고리즘이 개발되었으며[8,9], 하모닉-CELP구조를 이용하여 시간과 주파수 영역에서 파라미터를 추출하기 위한 연구 또한 진행되고 있다[10]. 더욱 흥미로운 것은 MELP코더를 제안한 McCree 역시 MELP와 CELP를 결합시킨 보코더를 제안하였다는 것이다[11]. 즉, 음성의 파형 부호화 방법과 하모닉 파라메트릭 기법의 결합이 최근 저전송률 부호화 알고리즘의 화두임을 알 수 있다.

하모닉-CELP 부호화기는 유성음 구간에 대해 표준 파형을 생성한 후 파형 간의 보간을 하는 PWI구조를 가지며, 무성음 구간은 CELP와 마찬가지로 코드북에서의 펄스와 이득값을 이용하여 잡음과 하모닉 분석에서 표현하기 힘든 비주기 펄스를 표현하는 방법을 사용한다. 하모닉 부호화를 위해 이처럼 표준 파형에 의한 보간 기법을 사용하는 이유는 피치주기에 해당하는 구간에 대해서만 파라미터화하여 전송하기 때문에 보코더의 전송률을 굉장히 낮출 수 있기 때문이다. 또한 인접한 두 표준 파형 간의 보간에 의해 합성 파형을 만들거 때문에 합성 원도우의 효과적인 설정과 음성의 시작점과 끝점을 지정함으로써 원본 파형의 특성을 잘 살린 연속 종합 신호를 생성할 수 있는 장점이 있다.

그러나 파형의 합성이 피치를 중심으로 이루어지기 때문에 잘못된 피치를 사용할 경우 파형 왜곡이 심하게 나타날 수 있으며, 특히 피치가 더블링되어 추정될 경우 프레임간의 불연속성이 발생할 수 있다. 또한 표준 파형의 생성을 위한 하모닉 부호화 방법은 파라미터를 추출할 때 피치에 의한 기본주파수를 설정하여 이것의 배수되는 성분에 대한 에러 최소화 기법을 사용한다. 그리고 양자화 과정에서 하모닉 크기값 역시 피치 주기에 따른 가변

차원을 가지기 때문에, PWI-STC와 같은 파형 보간 보코더의 구성은 정확한 피치를 검출한 이후 가능하다는 것을 알 수 있다.

추출된 피치로 표준 파형을 생성한 후에는 삼각 원도우에 의한 중첩-합산으로 여기신호의 보간이 이루어진다. 그러나 하모닉-CELP부호화기는 전이 구간의 여기신호를 합성할 때, 무성음에서 유성음으로 변환할 경우 전반부 80샘플은 CELP에 의한 방법을 사용하며, 유성음 구간에 대한 하모닉 합성은 전이 프레임 전체를 통해 이루어진다. 즉 하모닉 합성신호의 전반 80샘플은 삼각 원도우 특성에 의해 무성음에 대한 영향을 적게 한 후, 무성음 밑으로 숨겨지도록 하는 방법을 쓴다[19]. 하지만 하모닉 여기 신호의 합성이 삼각 원도우에 의해 이루어지므로, 유/무성음 변환 구간에서는 유성음 신호의 순간적인 변화를 따라가지 못한다. 또한 유성음과 무성음에 대한 경계가 이미 정해져 있기 때문에 하모닉에 의해 합성되어야 할 부분이 무성음으로 합성되거나, 그 반대로 되는 경우가 발생하므로 전이 시점에 대한 검색이 필요하다.

이러한 강인한 피치 검색기의 설계를 위해 본 논문의 2장 3장에서는 피치 검색기의 설계 과정에 대해 다룰 것이다. 그리고 4장에서는 음성의 시작점과 끝점을 검출함으로써 전이구간에 대한 효과적인 파형 보간이 이루어지도록 하기 위해, 웨이블릿 피치 검색기에서 얻은 GCI를 통해 음성 시작점과 끝점의 추출하여 원도우를 설정하는 것에 대해 다루겠다. 마지막으로 5장 및 6장에서 하모닉-CELP 구조의 보코더[19]에 위의 방법을 적용한 결과를 나타내도록 하겠다.

II. 웨이블릿 기반의 피치 추정

저전송률 파라메트릭 코더에서 가장 문제가 되는 것은 음성의 시작점, 끝점, 그리고 피치 변화가 심한 전이구간이다. 이러한 구간에서 비 사건 검출(non event detection)은 과거 프레임과 현 프레임 사이의 상관도가 거의 없어 피치 추정을 하기에는 부적합한 모델이다. 단, 미래 프레임을 미리 취함으로써 후단이 유성음인지 무성음인지를 판별할 경우 이에 대해 적절한 분석이 가능하지만, 이는 한 프레임에 대한 지연을 일으키게 된다. 따라서 현 프레임만으로 전이구간에서의 피치를 추정하고, 전이가 나타나는 시점을 찾아주어 유/무성음에 대한 별도의 처리를 하여 주어야 한다. 이를 위해 Mallat에 의해 제안된 2차 회기(quadratic spline) 웨이블릿 함수를 이용하여

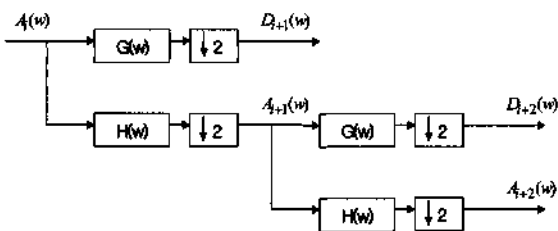
GCI를 검출함으로써 피치를 얻었다[12,13]. 또한, 음성 음의 시작점과 끝은 피치 검출을 위해 사용되었던 GCI를 이용하여 추정할 수 있었다.

2.1. 웨이블릿 변환

웨이블릿 변환을 이용하여 신호를 분석하는 과정은 그림 1과 같이 부밴드 (sub-band) 필터링 기법에 기반을 둔 이산 직교 웨이블릿 (DyWT)의 피라미드 구조로 이루어진 필터 뱅크의 형태로 나타낼 수 있다.[12] 이것은 입력을 A_i 라고 놓고, 고역 필터 G 와 저역 필터 H 를 통과시켜 출력 A_{i+1} , D_{i+1} 를 얻고, 다시 A_{i+1} 를 필터에 통과시켜 A_{i+2} , D_{i+2} 를 얻는 과정의 반복에 의해 웨이블릿 함수 A_i 와 스케일 함수 D_i 출력을 얻는 것을 말한다. i 는 신호의 스케일이다. 이러한 저역 통과 필터와 고역 통과 필터를 통한 출력은 신호의 대략적인 성분의 표현과 상세 성분의 표현이라고 할 수 있으며, 상세 성분 즉, 출력 D_i 를 이용하여 피치를 추정하게 되는 것이다.

저역, 고역 필터의 각 계수는 표 1과 같이 나타내며 각 필터를 웨이블릿 기저 (basis)로서 표현할 경우 식 (1)과 같이 나타낼 수 있다. 이것은 각 스케일마다 $2^k - 1$ 개의 0을 삽입하고, $2^{k/2}$ 로서 스케일링하는 것으로 웨이블릿 함수와 스케일 함수

$$\begin{aligned} \varphi(t) &= 2^{k/2} \varphi(2^k t) \\ \psi(t) &= 2^{k/2} \psi(2^k t) \end{aligned} \quad (1)$$



G : highpass wavelet filter H : complementary lowpass wavelet filter
 D : wavelet function A : scale function

그림 1. 필터 뱅크에 의한 웨이블릿 변환 과정
 Fig. 1. Wavelet transform by using filter bank.

표 1. 2차 회기 웨이블릿에 대한 필터 계수
 Table 1. Filter coefficient of the quadratic spline wavelet.

k	$h(k)$	$g(k)$
-1	0.125	0
0	0.375	-2.0
1	0.375	2.0
2	0.125	0

$\varphi(t)$: Lowpass function, Wavelet function
 $\psi(t)$: Highpass function, scale function

를 얻어내는 것을 말한다[10]. 이러한 과정은 고정된 필터 계수를 이용하여 컨벌루션만으로 밴드별 특성을 얻어낼 수 있어 계산량을 감소시켜 주며, 잡음에 대해 강인함을 가지는 특성을 부여한다. k 는 스케일 값이다.

단, 실험에서는 각 필터의 입력을 전단의 출력으로 하는 반복과정을 거치지 않고, 변환을 할 때 각 필터의 스케일 함수와 웨이블릿 함수는 필터간의 컨벌루션을 미리 한 후 음성을 통과시킴으로써 얻었다. 이러한 과정은 20 ms의 음성에 대해 원하는 밴드의 출력을 얻을 때까지 모든 필터계수를 곱할 필요없이 필터 계수간의 컨벌루션을 미리 취해주고 음성은 한번만 통과시켜 계산량을 감소시킬 수 있기 때문이다. 또한 웨이블릿 필터의 출력은 FIR 필터의 특성을 가지고 있기 때문에 웨이블릿 차수의 1/2만큼의 지연을 없애주어 분석하였다[14].

2.2. 피치의 추정

웨이블릿 변환에 의해 얻은 스케일 3, 4, 5의 웨이블릿 함수 D_i 은 500-1000 hz (스케일 3), 250-500 hz (스케일 4), 125-250 hz (스케일 5)의 대역을 통과시킨 출력이다. 이러한 필터의 특성을 그림 2(b)에 각 스케일별로 나타내었다. 이렇게 필터의 대역을 정해준 것은, 일반적으로 피치 주기가 20에서 128샘플이라는 사실에 기인한 것이다.

또한, GCI를 검출하기 위한 임계값은 스케일이 증가할수록 진폭은 감소하게 되는 점을 고려하여 실험적으로 스케일 3에서의 지역 최대값 추출시 τ_0 는 프레임 최대값의 0.4로, 스케일 4의 임계값 τ_1 는 0.45로, 스케일 5의 임계값 τ_2 는 0.38로 하였다.

특히 스케일 함수를 구하는 과정은 웨이블릿 계수 사이에 0을 삽입하기 때문에 추출되어진 지역 최대값은 스케일이 증가할수록 벌어지게 되므로 각 스케일 함수간의 비교를 할 때 스케일 4를 기준으로 각 스케일의 지역 최대값이 10샘플 이내에 있을 경우만을 GCI로 설정을 하여 주었다. 이렇게 구한 GCI간의 시간 간격 및 발생 시점을 조사함으로써 전이의 발생 시점과 그 구간에서의 정확한 피치를 추정할 수 있었다[14]. 추출된 피치의 오차율과 스케일별 검출율을 아래의 표 2에 나타내었다.

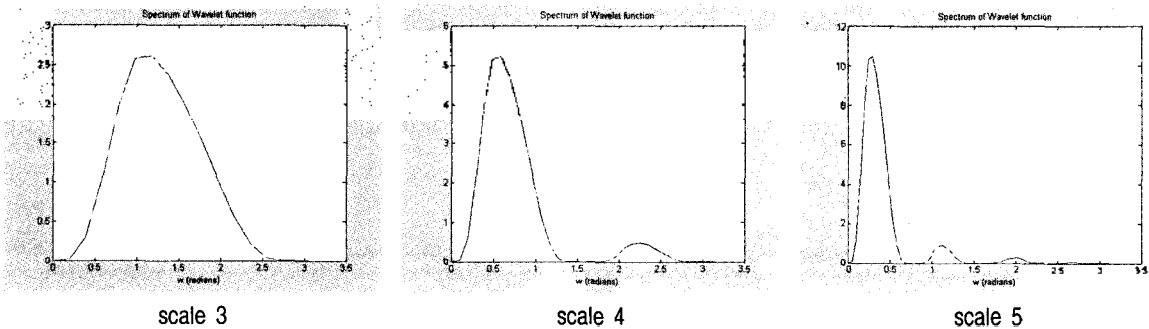
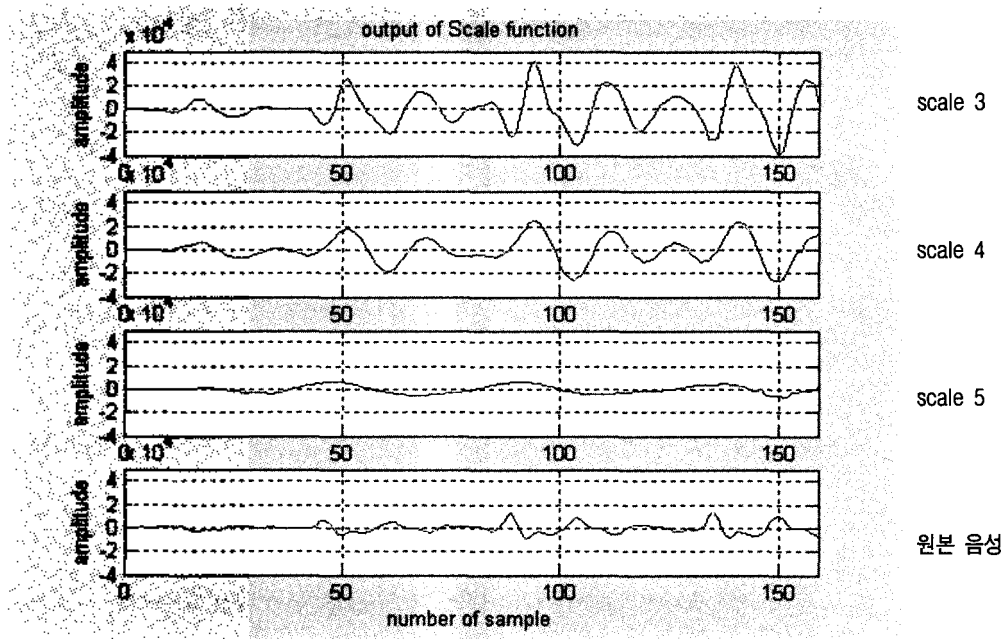


그림 2. (a) scale 3, 4, 5에서의 wavelet 함수 출력, (b) 각 스케일 별 밴드 특성
 Fig 2. (a) output of scale 3, 4, 5 wavelet function, (b) band characteristic of each scale.

표 2. 피치 에러율 및 스케일별 검출율
 Table 2. Pitch error rates and detection rate of each.

Pitch detector	Error rates (%)
· ACF (Inmarsat)	7.2
· Wavelet	4.46
Scale	Pitch detection rate (%)
Scale 3	98.4%
Scale 4	97.2%
Scale 5	91%

III. 웨이블릿 피치 검출기와 ACF의 결합

앞에서 말했듯 PWI는 피치에 의한 표준 파형의 생성 후 파형 간의 보간에 의해 음성을 합성하기 때문에 원래 피치의 절반이 되거나, 두 배가 되었을 경우 합성된 음성에서 왜곡이 발생하게 된다. 실제 MOS실험에서 피치가 절반이 될 경우 왜곡은 거의 느껴지지 않으나, 두 배가

되었을 경우 프레임 경계에서 불연속성이 심하게 나타나는 것을 확인할 수 있었다. 따라서 정확한 피치를 검출하되, 가능한 한 피치 더블링이 발생하지 않는 것에 관점을 맞추어 피치 검출기를 설계하였다.

3.1. 웨이블릿 및 주파수 영역 변환 1차 검색과 ACF에 의한 2차 검색

기존 코더는 주파수영역에서 피치를 검색한 후 ACF와의 결합에 의한 1, 2차 검색 방법을 사용하고 있다[19]. 이러한 방법의 장점은 유성음 구간에서 피치에 의한 펄스는 주파수 영역에서 하모닉으로 나타나기 때문에, 1차 검색에서 음성의 주기성을 찾아내기가 쉽다는 것이다. 이후 2차 검색에서 ACF에 의해 1차 검색에서 찾아진 피치 주위를 한 번 더 검색하여 정확한 피치를 추정한다. 그러나 그림 3에서 확인할 수 있듯이, 음성의 시작지점에서 피치 더블링이 자주 발생하는 것을 알 수 있다.

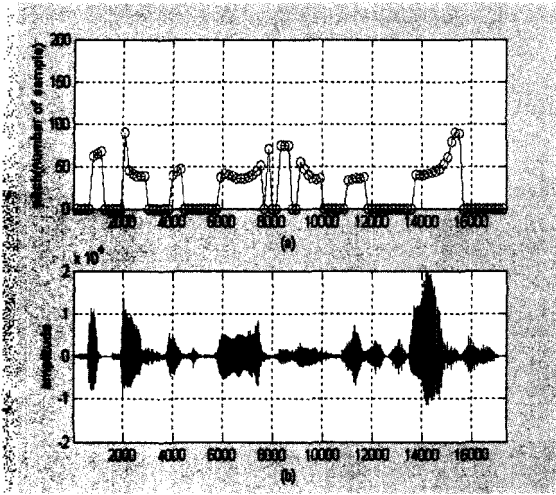
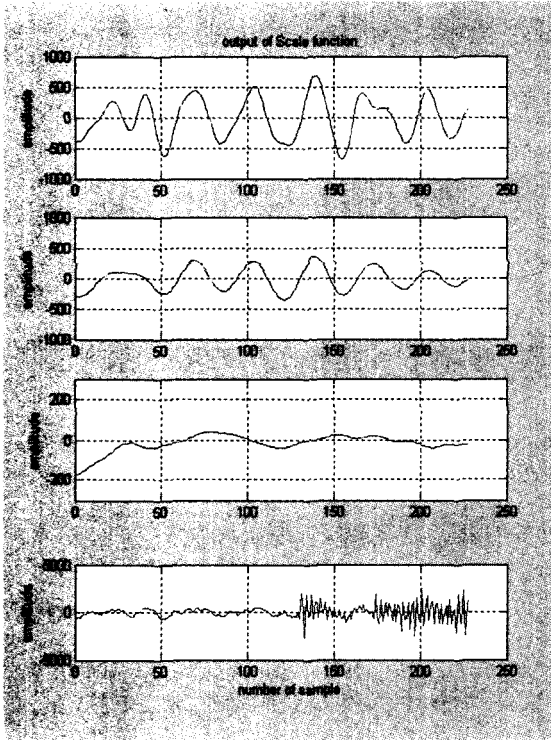


그림 3. 기존 피치 검색기에 의한 피치
Fig. 3. Pitch of existing pitch detector.

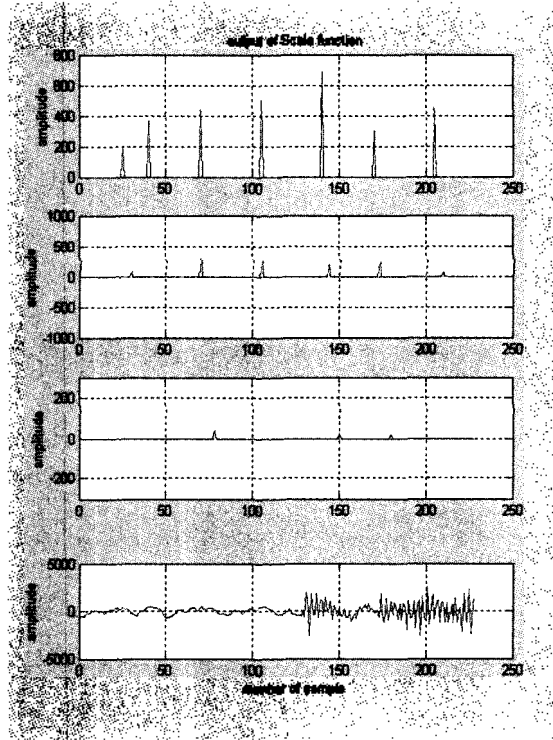
또한, 웨이블릿만을 사용할 경우 그림 4에서처럼 유/무성음혼합 영역에서 GCI는 주기성을 찾기가 힘들다. 따라서 GCI간에 평균을 취할 경우 부정확한 피치 검출로 인한 왜곡을 발생하게 된다.

이러한 서로의 장단점을 보완하기 위해 그림 5와 같이 유/무성음 변환 구간은 웨이블릿에 의한 1차 검색으로 처리를 하였다. 그리고 일반 유성음과 혼합영역은 LPC 잔여 신호의 스펙트럴 파워를 이용한 자기 상관 함수에 의해 검색하는 혼성 1차 검색 방법을 이용하였으며, ACF에 의한 2차 검색으로 피치값의 정확도를 높였다.

그림 6은 피치의 추정과정과 전이 시점을 검출하는 순서도이다. 피치의 추정 과정은 그림 5를 통해 이미 설명이 되었으므로 음성 시작점, 음성 끝점의 추정과 웨이블릿 검



(a) 스케일 함수의 출력
(a) Output of wavelet function



(b) 각 스케일의 지역 최대값
(b) Local maxima of each scale

그림 4. 유무성음 혼재 구간에서의 웨이블릿에 의한 피치 검색
Fig. 4. Wavelet pitch detector of voice/unvoice mixed region.

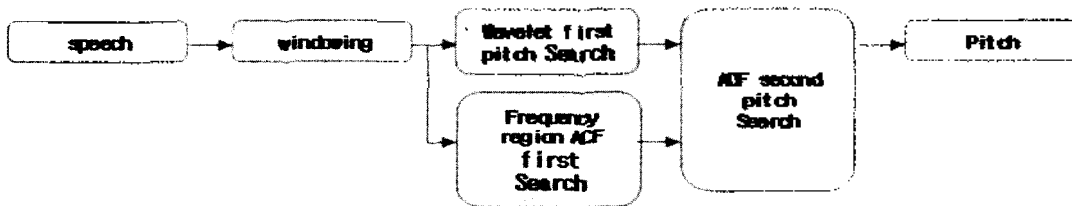


그림 5. 제안된 피치검색기의 블록도
Fig. 5. Block diagram for adjusting pitch detector.

출시 GCI가 3개 미만일 경우 피치 추정방법에 대해서만 알아보겠다. 전이 시점은 피치를 추정하기 위해 웨이블렛 함수로부터 얻었던 GCI를 사용한다. 즉 무성음에서 유성음으로 변환될 경우 가장 처음에 얻었던 GCI를 사용하고, 무성음에서 유성음으로 변환하는 경우는 마지막을 취하는 것이다. 만약 검출된 음성 시작점, 음성 끝점 지점의 값이 80을 넘어갈 경우 0으로 설정을 한다. 왜냐하면 무성음에서 유성음으로 (유성음에서 무성음으로) 변환하는 구간에서는 프레임 전반부 80샘플 (후반부 80샘플)을 이미 무성음 부분에서 여기 신호로 추출하였기 때문에 그 이상을 넘어가지 않도록 한 것이다. 이것에 자세한 설명은 4장에서 하도록 하겠다. 또한 추출된 GCI가 3개 미만일 경우 주기성을 찾기 힘들기 때문에 자기 상관에 의한 검색법을 사용하도록 한다.

IV. 수정된 피치 검출기에 의한 음성 시작점, 음성 끝점 포인트의 추정과 이의 적용

4.1. 전이 구간 윈도우 설정과 음성 합성

기존 하모닉-CELP코더(19)는 그림 7(a)에서의 윈도우를 가지고 합성이 이루어진다. 우선 LPC계수는 (a)의 상단처럼 프레임을 8단으로 나누어 2.5 ms마다 과거와 현재의

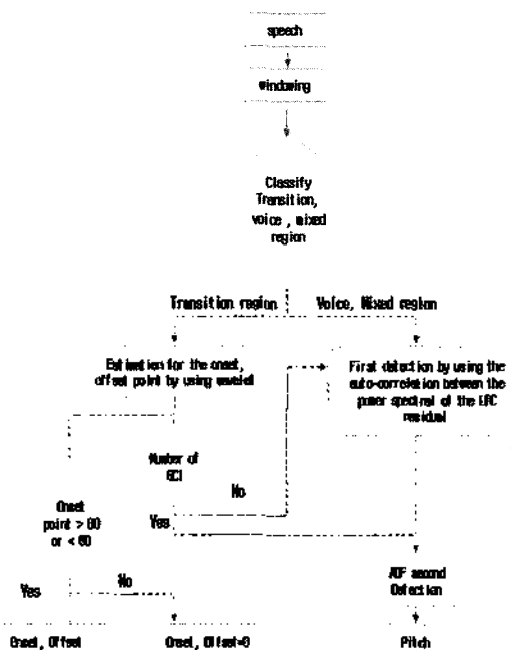
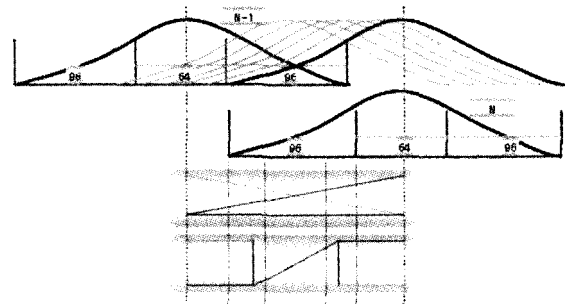


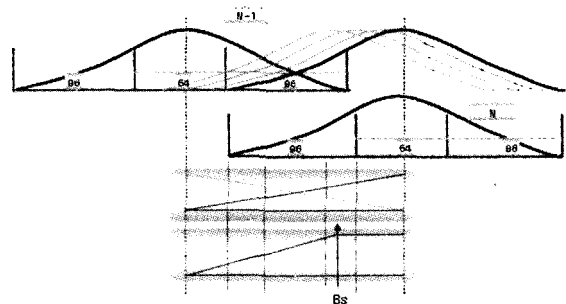
그림 6. 피치와 전이 시점 검출을 위한 순서도
Fig. 6. Flow chart for detection both pitch and transition point.

LSP를 보간함으로써 얻는다. 그리고 합성 잔여 신호를 LPC 역필터링한 후 피치 변화가 작을 경우 삼각 윈도우를 이용한 선형 보간 중첩-합산 합성방법을 사용하며, 클 경우 하단의 사다리꼴 윈도우에 의해 합성이 이루어진다.

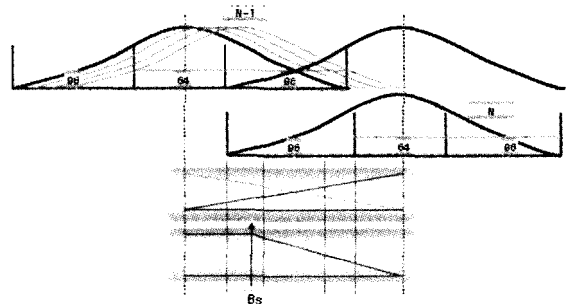
반면 무성음 영역은 전체 256샘플의 윈도우를 씌운 현재 프레임에서 과거 48샘플과 미래 48샘플을 제외한 현 입력만으로 분석/합성이 이루어진다. 특히 무성음에서 유성음 변환 구간은 위의 무성음 합성 프레임의 구조적 특성 때문에 과거 무성음 80샘플의 여기 출력을 합성 프레임 160샘플 중 전반부에 위치하도록 하며, 유성음 구간에서의 하모닉 여기 신호는 삼각 윈도우를 씌워 무성음과 합쳐지게 된다. 이러한 방법은 유/무성음 변환 구간에 대



(a) 기존 혼합코더의 하모닉 합성 구조
(a) Existing coder



(b) 전이 구간을 위한 수정된 합성 구조 (음성 시작점)
(b) Adjusting coder (onset)



(c) 전이 구간을 위한 수정된 합성 구조 (음성 끝점)
(c) Adjusting coder (offset)

그림 7. 전이 구간내 하모닉 합성 구조
Fig. 7. Harmonic synthesis structure.

한 처리를 위한 것으로 앞단 80샘플의 하모닉 성분은 윈도우 특성에 의해 영향분을 작게 만들어준 후 잡음 아래로 묻혀지게 하는 방법이다. 그러나 하모닉 합성을 할 때

삼각 윈도우에 의한 증첩-합산 방법을 사용하기 때문에 원본 음성이 순간적인 증가를 할 경우 삼각 윈도우의 영향으로 합성 여기 신호가 선형 증가하는 단점이 있다.

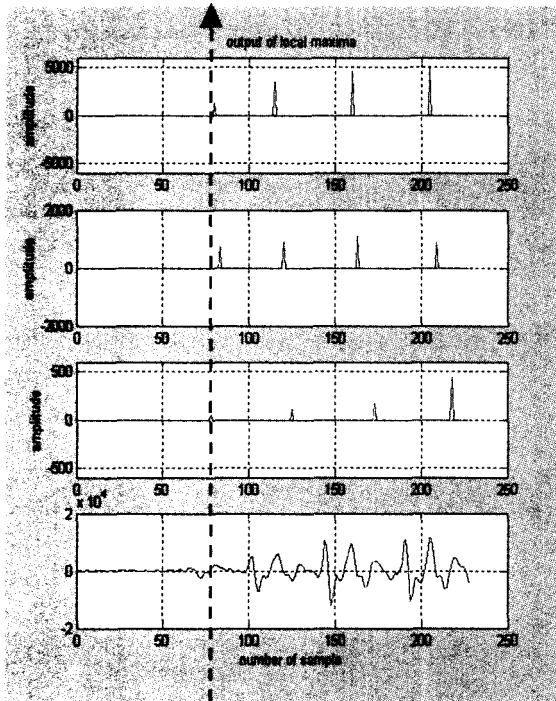


그림 8. 음성 시작점 포인트의 검출 방법
Fig. 8. Detect of onset point.

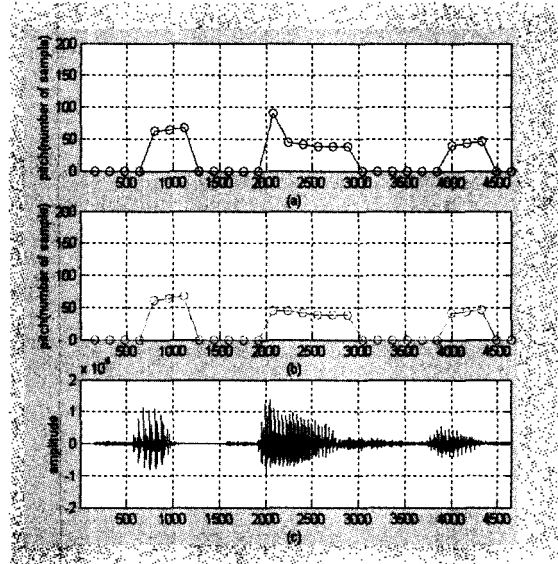


그림 9. 피치 검출 결과 비교, (a) 제안된 피치 검출기의 결과, (b) 기존 피치 검출기의 결과, (c) 입력 음성
Fig. 9. Comparison of pitch detector, (a) Adjusting pitch detector output, (b) Existing pitch detector output, (c) Input speech.

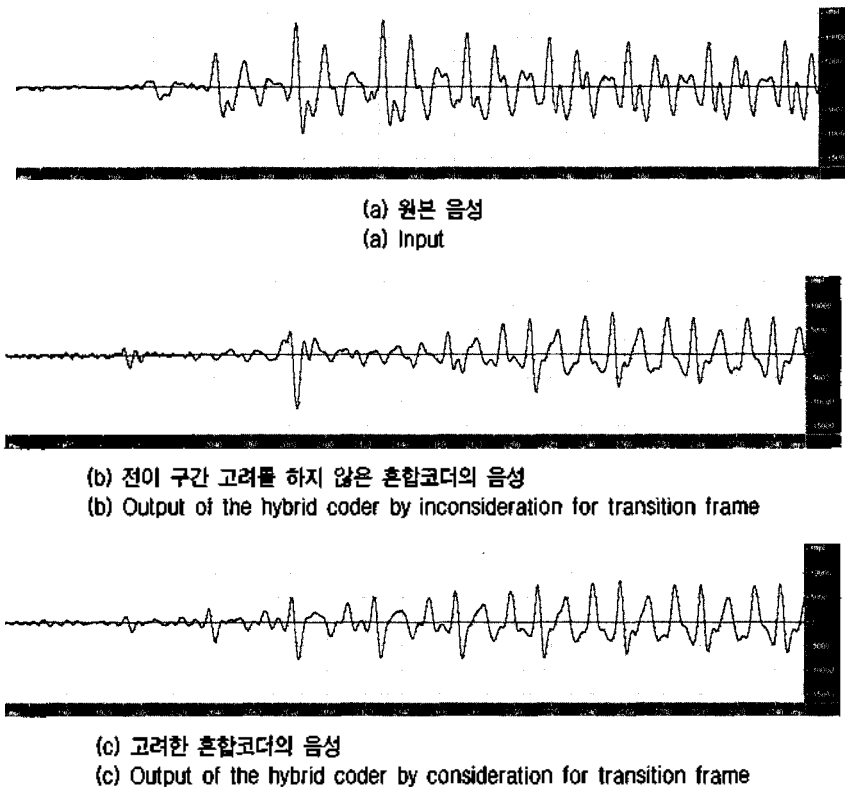


그림 10. 합성 파형 비교
Fig. 10. Comparison of the synthesis result.

이러한 문제점을 해결하기 위해 그림 8과 같이 웨이블릿에 의한 B_s점을 찾아내어 그림 7(b)와 (c) 하단의 구조를 갖는 윈도우로 합성이 이루어지도록 하였다. 먼저 (b)는 음성의 시작 포인트(B_s)의 지정을 통해 유성음 특성이 나타나는 구간 이후의 값부터 가중치를 두도록 하는 방식을 사용하였다. (c) 또한 마찬가지로 음성의 끝점 B_e를 지정하고, 위의 과정처럼 보간과 윈도우를 설정하여 유성음을 잘 살릴 수 있도록 설정을 하여 주었다.

V. 실험 결과

5.1. 전이 구간 피치 검색기 적용 결과

그림 9(a)는 주파수 영역에서의 1차 검색 후 2차로 교차 상관(cross correlation)에 의해 검색을 한 후 피치이다. 음성의 2번째 구간(2000샘플 주위)에서 (b)와 비교해서 피치 주기가 두 배로 되어 있는 것이 확인되었으며, 이는 유/무성음 변환 구간 이후 유성음에서의 급격한 피치 변화를 일으키게 된다. 따라서 그림 7(a)에서의 사다리꼴 윈도우를 사용하여 현 단의 전반부를 합성하지만, 그림 10(b)처럼 프레임간의 연속성이 이루어지지 않는 것을 확인할 수 있다. 이러한 문제점은 웨이블릿-ACF 피치 검출기를 사용함으로써 해결이 된다.

그림 9(b)에 비해 (a)에서 피치 검출은 2000샘플에서 나타나는 두 배의 피치를 없애준 것을 확인할 수 있었으며, 그림 10(c)처럼 합성한 파형의 연속성이 이루어지는 것을 확인하였다. 전이 구간에 대해 제안된 방법을 적용

표 3. 피치 에러율
Table 3. Pitch error rates.

Pitch detector	Error rates (%)
· ACF (Inmarsat)	7.2
· Wavelet	4.46
· Wavelet-ACF	1.8

한 결과, 표 3에서와 같이 피치 검색 오차율은 ACF 검출법에 비해 5.4%개선되었으며, 웨이블릿에 의한 검출법에 비해 2.66%개선되었다.

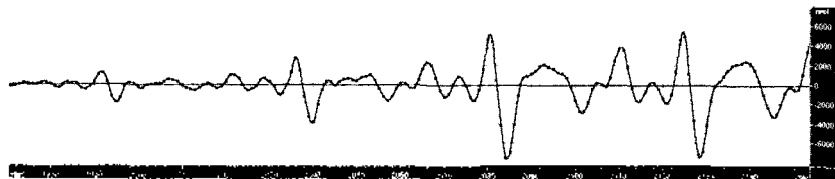
5.2. 전이 구간의 LPC 합성 수정 결과

그림 7(b)와 (c)의 하단 윈도우를 이용하고, 음성의 시작과 끝점의 검출하여 무성음의 영향을 줄이도록 가중치를 조절한 결과를 그림 11과 12에 나타내었다.

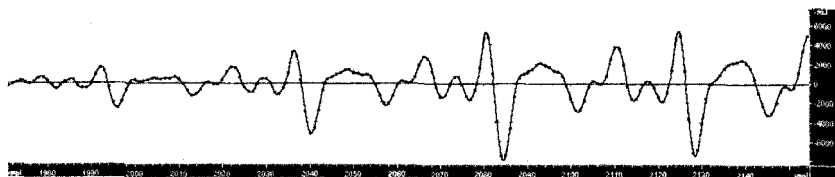
그림 11 (a)와 그림 12(a)에 비해 그림 11(b)와 그림 12(b)는 음성의 시작점과 끝점에서 점차적인 증가와 감소 부분이 많이 개선된 것을 확인할 수 있었다.

5.3. MOS 테스트 및 피치 추정 결과

마지막으로 표 3에는 MOS 테스트 결과를 나타내었다. MOS 테스트는 'PESQ-ITU-T Recommendation P.862 Version 1.2-2 August 2002' 프로그램을 이용하여 결과를 얻었다. 우선 기존의 보코더에 비해 음성의 연속성이 잘 유지되는 것을 확인하기 위해 음성의 시작점과 끝점에 대한 부분을 추출하여 실험을 하였다. 실험 결과 음성의 시작 부분은 하모닉-CELP구조의 특성상 랜덤 위상으로



(a) 무성음을 영향을 배제하지 않고, LSP의 보간이 이루어진 경우
(a) Output of the LSP interpolation by inconsideration for unvoiced frame



(b) 무성음 영향을 배제하고, LSP 보간을 한 경우
(b) Output of LSP interpolation by consideration for unvoiced frame

그림 11. 합성 파형 비교 (음성 시작점 구간)

Fig. 11. Comparison of the synthesis result for onset transition.

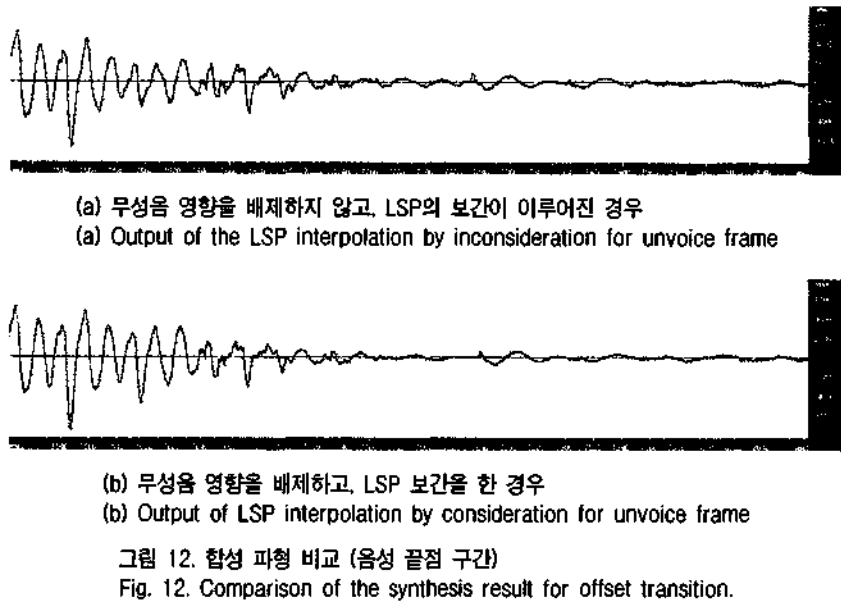


표 4. 기존 코더와의 비교

Table 4. Comparison between the adjusting coder and the existing coder.

vocoder	MOS value
· Existing GH-CELP	2.374
· Adjusting method	2.501

하모닉 정현파를 합성하기 때문에 파형의 왜곡이 발생한다. 따라서 MOS값은 전체 문장에 대한 MOS값보다 낮게 추출이 되므로, 정확한 피치의 추출에 의해 기본 피크치를 잘 표현하였는가에 대해서만 확인하였다. 이를 표 4에 나타내었다.

그리고 그림 13(a)처럼 기존 코더에서 나타났던 더블링을 없애 줌으로써 (b)와 같이 프레임의 연속성이 유지되는 것을 확인하였다. 실제 음성을 들을 경우 불연속성이 나타나는 구간은 강한 잡음처럼 탁탁튀는 소리가 나기 때문에 음질을 굉장히 떨어뜨린다.

VI. 결론

본 논문은 저전송률 파형 보간 코더에서 나타나는 잘못된 피치의 추정으로 인한 프레임간의 불연속성을 해결하며, 낮은 지연 (delay)을 갖는 보코더를 설계하기 위해 웨이블릿에 기반을 둔 피치 검출기를 연구하였다. 또한 전이 프레임의 유/무성음 혼재 구간에서 무성음 합성 신호가 유성음의 합성에 영향을 미치지 않도록 하기 위한 연구를 하였다. 제안된 방법은 기존 코더와의 비교했을 경우 음성의 시작과 끝부분에 대해 MOS 0.13 정도의 값이 향상되는 것을 확인하였다. 하지만 아직 유성음 합성단에서 음성 시작점은 랜덤 위상을 발생시킨 후 선형 위상으로서 합성하기 때문에 원본 파형과는 다른 모양을 가지고 있다. 이러한 점을 보완하기 위해 시간 영역 모델링에 근간을 둔 파형 생성 방법이 필요함을 알 수 있었다.

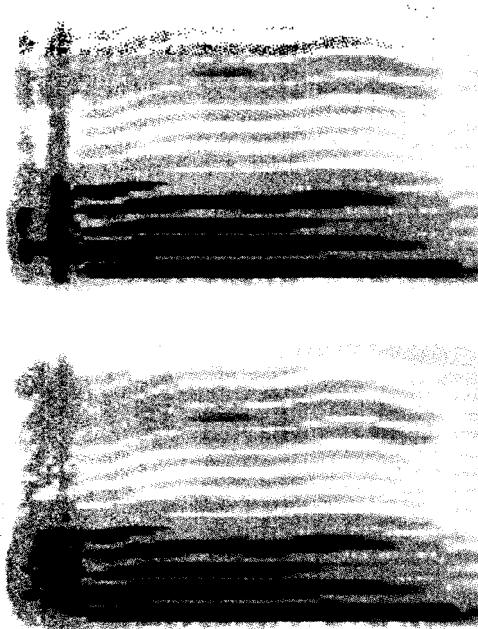


그림 13. (a) 기존 피치 검색기를 사용한 음성시작점 스펙트로그램
(b) 제안된 피치 검색기를 사용한 음성시작점 스펙트로그램

Fig 13. (a) speech spectrogram of onset region by using existing pitch detector (b) speech spectrogram of onset region by using adjusting pitch detector.

참고 문헌

1. A. McCree, K. Truong, E. George, T. Barnwell, and V. Viswanathan, "A 2.4 kbit/s coder candidate for the new U.S. federal standard," *Proc. IEEE International Conference on Acoustics, Speech and Signal Processing*, 200-203, Atlanta, 1996.
2. D. W. Griffin and J. S. Lim, "Multiband excitation vocoder," *IEEE Trans. Acoust., Speech, Signal Processing*, 36 (8), 1223-1235, 1988.
3. R. J. McAulay and T. F. Quatieri, "The application of subband coding to improve quality and robustness of the sinusoidal transform coder," *Proc. ICASSP 93*, 2, 439-442, 1993.
4. W. B. Kleijn and J. Haagen, "A speech coder based on decomposition of characteristic waveforms," *Proc. ICASSP 95*, 508-511, 1995.
5. K. A. Teague, B. Leach, and W. Andrews, "Development of a high-quality MBE based vocoder for implementation at 2400 bps," *Proc. IEEE Wichita Conf. Communications, Networking and Signal Processing*, 129-133, April 1994.
6. H. Hassanein, A. Brind Amour, S. Dry, and K. Bryden, "Frequency selective harmonic coding at 2400 bps," *Proc. 37th Midwest Symp. Circuits and Systems*, 2, 1436-1439, 1995.
7. W. B. Kleijn, "Encoding speech using prototype waveforms," *IEEE Trans. Speech Audio Processing*, 1, 386-399, Oct. 1993.
8. K. Yaghmaie and A. M. Kondoz, "Multiband prototype waveform analysis synthesis for very low bitrate speech coding," *Proc. ICASSP 97*, 1571-1574, 1997.
9. D. J. Hiotakakos and C. S. Xydeas, "Low bit rate coding using an inter-polated zinc excitation model", *Proc. ICCS 94*, 865-869, 1994.
10. E. Shlomot, V. Cuperman, and A. Gersho, "Combined harmonic and waveform coding of speech at low bit rate," *Proc. ICASSP 98*, 585-588, 1998.
11. J. Stachurski, A. McCree, V. Viswanathan, A. Heikkinen, A. Ramo, S. Himanen, and P. Blocher, "HYBRID MELP/CELP coding at bit rates from 6.4 TO 2.4 kb/s," DSP Solutions R&D Center, Texas Instruments, Dallas, Texas, USA, 2003.
12. S. Mallat, W. L. Hwang, "Singularity detection and processing with wavelets," *IEEE trans. on IT*, 38 (2), 617- 643, 1992.
13. S. Mallat and S. Zhong, Characterization of signals from multiscale edges, *IEEE Trans. Pattern Anal. Machine Intell.*, 14, 710-732, July 1992.
14. 손영호, 배건성, "웨이블릿 변환을 이용한 유성음/무성음/목음 분류," *음성통신 및 신호처리 워크샵 논문집*, 449-453, 1998.
15. S. Kadambe and G. F. Boudreau-Bartels, "Application of the wavelet transform for Pitch detection of Speech Signal," *IEEE Trans Information Theory*, 38 (2), Mar. 1992.
16. A. M. Kondoz, "Code-excited linear predictive coding," *Digital Speech*, Chap. 6, 174-212, 1994.
17. ITU-T Recommendation G.729, "Coding of speech at 8kbps using conjugate-structure algebraic code excited linear prediction (CS-ACELP)," June 1995.
18. F. C. A. Brooks, and Lajos Hanzo, "A multiband excited waveform-interpolated 2,35-kbps speech codec for bandlimited channels," *IEEE Trans on VT*, 49 (3), May 2000.
19. 김종학, 이인성, "하모닉 코딩과 CELP방법을 이용한 저 전송률 음성 부호화 방법 Low Rate Speech Coding Using the Harmonic Coding Combined with CELP Coding," *한국음향학회지 THE JOURNAL OF THE ACOUSTICAL SOCIETY OF KOREA*, 19 (3), 26-34, 1225-4428, 2000.

저자 약력

• 방 상 운 (Sang-Un Bang)



2001년 2월: 충북대학교 전자공학과 학사
2001년 3월~현재: 충북대학교 전파공학과 석사

• 이 인 성 (In-Sung Lee)



1983년 2월: 연세대학교 전자공학과 학사
1985년 2월: 연세대학교 전자공학과 학사
1992년 12월: Texas A&M University 전기공학과 박사
1986년 5월~1987년 7월: 한국 통신 연구 개발단 전 입 연구원
1997년 3월~1999년 8월: 한국 전자통신 연구원 이 동통신 기술 연구단 선임 연구원
1995년~현재: 충북대 전기전자컴퓨터 공학부 부교수

• 권 오 주 (O-Joo Kwon)



1989년 2월: 경북대학교 전자공학과 공학사
1993년 2월: 경북대학교 대학원 전자공학과 공학석사
2002년 2월: 경북대학교 대학원 전자공학과 공학박사
1997년 9월~1999년 2월: Matra Marconi Space UK
1993년 3월~현재: 국방과학연구소 선임연구원
* 주관심분야: 위성통신, 이동통신, 디지털신호처리