하모닉 코딩과 CELP 방법을 이용한 저 전송률 음성 부호화 방법

Low Rate Speech Coding Using the Harmonic Coding Combined with CELP Coding

김 종 학* 이 인 성*
(Jonghark Kim*, Insung Lee*)

※ 이 논문은 1998년 한국학술진흥재단의 공모과제 연구비에 의하여 연구되었음.

요 약

본 논문에서는 선형예측 잔여신호에 대한 하모닉 벡터 여기 코딩에, 시간 대역 분리 혼합 코딩을 결합한 4kbps 음성코더를 제안한다. 하모닉 벡터 여기 코딩은 유성음 구간에서 하모닉 여기 코딩을 사용하며, 무성음 구간에 대해서는 분석-합성 구조의 벡터 여기 코딩을 사용한다. 그러나, 이러한 양단 모드 코딩 방법은 유성음과 무성음이 혼재하는 전이 구간에서는 비 효과적이므로, 유/무성음 모드 코딩 이외의 새로운 방법이 요구된다. 이에, 전이 구간을 위한 시간 분리 전이 코딩을 설계하였으며, 여기서, 유/무성음 결정 알고리즘은 단위 구간 내의 유성음과 무성음의 존속기간을 결정하고, 이전 구간의 유/무성음 결정에 따라 하모닉-하모닉 코딩과 벡터-하모닉 코딩을 선택적으로 사용한다. 복호화기에서는 하모닉 크기값들의 IFFT 과정을 통해 유성음 여기신호가 효과적으로 합성되며, 무성음 여기신호는 역 벡터 양자화를 통해 만들어진다. 재 복원된 음성 신호는 중첩합산 방법에 의해 합성된다.
핵심용어: 하모닉, 전이, 혼합, 파형, 보간, 여기

ABSTRACT

In this paper, we propose a 4kbps speech coder that combines the harmonic vector excitation coding with time-separated transition coding. The harmonic vector excitation coding uses the harmonic excitation coding in the voiced frame and uses the vector excitation coding with the structure of analysis-by-synthesis in the unvoiced frame, respectively. But two mode coding method is not effective for transition frame mixed in voiced and unvoiced signal and a new method beyond using unvoiced/voiced mode coding is needed. Thus, we designed a time-separated transition coding method for transition frame in which a voiced/unvoiced decision algorithm separates unvoiced and voiced duration in a frame, and harmonic-harmonic excitation coding and vector-harmonic excitation coding method is selectively used depending on the previous frame U/V decision. In the decoder, the voiced excitation signals are generated efficiently through the inverse FFT of harmonic magnitudes and the unvoiced excitation signals are made by the inverse vector quantization. The reconstructed speech signal are synthesized by the Overlap/Add method.
Key words: Hamonic, Transition, Hybrid, Waveform, Interpolation, Excitation
투고분야: 음성처리(2.2, 2.4)

I. 서 론

통신 시스템의 용량 증대 및 질 좋은 서비스를 위한 연구가 활발히 이루어지는 가운데, 그 발전 추세가 비교적 빠른 보코더 부분 역시 많은 연구가 진행되어 왔다[1][2][3]. 특히 보코더 기술에 있어서 가장 중요한 문제점인 전송률과 음질개선에 대한 연구는 과거부터 계속되어온 과제이며, 앞으로도 개선되어야 할 부분이다. 이러한 논점에 맞춰 근래까지 CELP(Code Excitation Linear Predictive)코더는 음성코더의 주류를 이루어왔으며, 최근에는 여기필터에 분석-합성 방법을 사용하여 대략 6kbps의 전송률에서도 고음질의 음성을 복원할 수 있게 되었다. 그러나, 저 전송률에서의 CLEP 시스템은 모델 파라미터의 양자화 과정에서 심한 왜곡을 초래해 음질을 크게 저하시키는 단점을 가지고 있다. 또한 배경잡음이 포함되었을때 음질이 급속

* 충북대학교 전파공학과
접수일자 : 1999년 2월 12일

히 저하되는 현상도 나타난다. 이러한 점을 고려하여 보코더 성능을 향상시키기 위한 연구가 활발히 이루어지고 있으며, 특히 음성 모델링과 여기신호의 양자화가 그 주요 논점을 이루고 있다. 결국, 이에 대한 효과적인 대안으로 선형예측 잔여신호의 주파수 영역 하모닉 코딩방법이 제안되었으며, 현재 이러한 접근방법은 이미 저전송률에서 비교적 좋은 음질을 보여주고 있다. 2.4kbps에서 동작하는 새로운 미국 연방 표준(U.S. Federal Standard) 음성코더인 MELP(Mixed Excitation Linear Predictive Coding)[4]와 HVXC(Harmonic Vector Excitation Coding)[5]를 그 대표적인 예로 들 수 있겠다. 이러한 하모닉 코딩은 유성음의 주기적인 신호 복원에 대해서 적합하고, 또한 기본 주파수가 대략 100Hz보다 작은 무성음에 대해서도 효과적인 것으로 알려져 있다. 그러나 하모닉 모델 또한 유성음의 개시, 파열음, 비주기적인 펄스 부분인 전이구간에서는 비효과적이라는 단점이 있다. 또한 주파수 영역 관찰에 있어 이런 유/무성음이 혼재하는 구간을 유/무성음 모드로 양단적 부호화 하는 것은 비효율적인 모델이 된다.

최근 이러한 전이 구간을 고려한 혼합방식의 부호화가 고려되고 있으나 유성음 구간내에서의 전이 표현 및 복잡도가 낮은 합성방법에 적합한 예는 되지 못하고 있다[9]. 따라서, 본 연구에서는 이를 고려하여 전이구간에서 시간기점을 사이로 하모닉 코딩과 벡터 여기 코딩을 함께 쓰는 방법을 제시한다. 기존의 선형예측 기반 하모닉 코딩의 유/무성음 혼재 구간을 유성음 단독 구간으로 하여 얻어지는 기계적인 소리와 하모닉 크기의 벡터 양자화 과정에서의 비효율적인 비트 할당 등의 단점을 보완하게 된다. 제안된 코더는 10차 LSP 계수, 전이시점, 유성음 하모닉 크기 벡터 양자화값, 무성음 벡터여기 양자화 값, 전이구역 하모닉 및 벡터여기 양자화값을 구하는 부호화 과정과 유/무성음 및 전이구간의 여기신호를 합성하고 LPC(Linear Prediction Coefficient) 합성필터와 후단여파기를 거치는 복호화과정으로 분류된다. 이러한 일련의 과정을 다음 2, 3장에 걸쳐 설명한다.

II. 혼합 코딩의 기본구조

혼합코더의 부호화기와 복호화기의 기본구조를 그림 1에 도시하였다. 우선 부호화기는 유/무성음 그리고 전이구간을 코딩하기 위한 3가지 코더로 분류되며, 각 분류코더의 입력신호는 LPC 잔여신호가 된다. 이 잔여신호를 얻기 위해 선형예측(Linear Prediction) 분석방법이 사용되며, 여기서 얻은 LPC 는 LSP(Line Spectrum Pairs)로 변환된 다음, 전형적 2단계 벡터양자화와 상호예측 벡터양자화 중 선택적으로 양자화 된다. 최종적으로 양자화 된 LSP계수를 보간하여 입력음성신호를 역 LPC 필터링 시킴으로써 잔여신호를 추출하게 된다.

추출된 잔여신호는 유/무성음과 전이구간을 판별하는 분류기에 의해 각 대응 코더에 의해 부호화된다. 분류 값과 이에 따른 각 코더의 파라미터 값 그리고 LSP 인덱스가 복호기에 전송되며, 복호기는 이 분류 값에 따라 유성음 구간 일 경우 파형 합성기와 파형 보간 과정을 거치며, 무성음 구간에서는 시간영역 합성과정을 거치게된다. 또한, 전이 구간에서는 시간대역 분리 합성과정으로 여기신호를 만들어내며, 최종적으로는 LPC 합성필터와 후단여파기를 거쳐 복원 신호를 만들어 낸다.

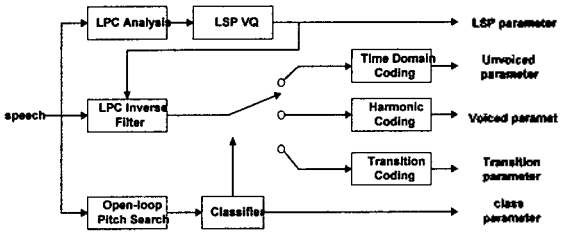


그림 1. 혼합 코더의 부호화 과정
Fig. 1. Block diagram of the hybrid encoder.

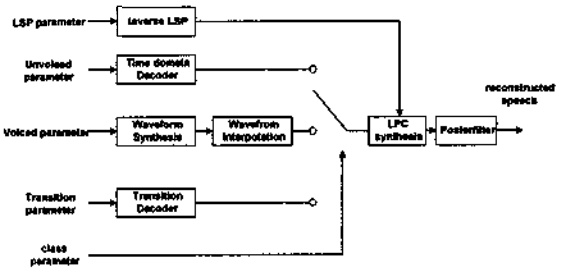


그림 2. 혼합 코더의 복호화 과정
Fig. 2. Block diagram of the hybrid decoder.

2.1. 하모닉 여기 코더

여기신호 표현은 정현(sine)파형 모델을 기초로 하여 다음과 같이 표현될 수 있다[6][7].

$$s(n)=\sum_{l=1}^{L} A_l \cos(w_l n + \phi_l) \qquad (1)$$

A_l과 ϕ_l 은 주파수가 w_l인 정현파 성분에 대한 크기와 위상을 나타내며, L 은 정현파형 들의 갯수를 나타낸다. 유성음 구간의 여기신호는 하모닉 부분이 대부분의 음성신호 정보를 포함하고 있어, 적절한 스펙트럴 기본 모델을 이용하여 근사화 시킬 수 있다. 다음은 선형위상 합성을 가지는 근사 모델을 나타낸다[7][8].

$$s^k(n)=\sum_{l=1}^{L_k} A_l^k \cos(l w_0^k n + \psi^k(l, w_0^k, n) + \Phi_l^k) \qquad (2)$$

k와 L_k는 프레임 번호와 각 프레임 당 하모닉 갯수를 나타낸다. w_0는 피치 각주파수를 나타내며 Φ_l^k는 k번째 프레임, l번째 하모닉의 이산위상을 나타낸다. k번째 프

레임 하모닉 크기를 나타내는 A_l^k와 w_0는 복호기에 전송되는 정보이며, 해밍윈도우의 256-point FFT한 값을 기준 모델로 하여 다음 값이 최소화되는 스펙트럴과 피치 파라미터 값으로 폐 루프 검색 방법으로 결정된다.

$$e_l = \sum_{i=a_l}^{b_l} (|X(i)| - |A_l||B(i)|)^2 \quad (3)$$

$X(i)$는 원 신호의 스펙트럼, $B(i)$는 스펙트럴 기준 모델을 뜻한다. 여기서 구한 하모닉 크기들의 갯수는 약 8에서 60개로 가변적이기 때문에 벡터양자화하기 위해서는 오버 샘플링을 구현하는 차원변환기를 필요로 하게된다.

위와는 달리 선형위상인 $\phi^k(l, w_0^k, n)$는 복호기에서 합성되는 정보이며, 다음과 같은 식으로 선형위상을 합성할 수 있다[7][8][9].

$$\phi^k(l, w_0, n) = \phi^{k-1}(l, w_0^{k-1}, n) + \frac{l(w_0^{k-1} + w_0^k)}{2} n \quad (4)$$

윗 식에서 볼 수 있듯이, 선형위상은 이전 프레임과 현 프레임의 시간에 따른 피치 각주파수를 선형으로 보간하여 얻어진다. 인간의 청각 시스템은 위상 연속성이 보존되는 동안 선형 위상에 둔감하며, 부정확한 또는 완전히 판이한 이산 위상을 허용할 수 있다. 이러한 지각적 특성은 저 전송률 코딩에 있어 하모닉 모델의 연속성에 대한 중요한 조건이 된다. 따라서, 합성 위상은 측정된 위상을 대체할 수 있게 된다[9].

복호기에서의 파형 합성은 부호화기로부터 전송된 스펙트럴 파라미터와 피치 파라미터 값을 사용하여 수행하게 된다. 우선, 기준 파형을 합성하기 위해, 스펙트럴 파라미터를 역 양자화 과정을 통해 하모닉 크기들을 추출한다. 그런 다음 식 (2),(4)에서 제시한 선형 위상합성방법을 사용하여 각 하모닉 크기들에 해당하는 위상정보를 만들어 낸 후, 128-point IFFT를 통해 기준 파형을 만들어 낸다. 이렇게 만들어진 기준 파형은 피치정보를 포함하지 않은 상태이기 때문에 순환형태로 재구성한 다음, 피치 주기로부터 얻은 오버 샘플링 비율로 피치변화를 고려하여 보간, 샘플링하여 최종 여기신호를 얻어낸다[10].

$$ov = \frac{256}{2T_p} = \frac{256/4}{T_p/2} = \frac{64}{l} \quad (5)$$

$$p_{ov}[n] = \sum_{i=0}^{n} \left(\frac{N-i}{N} ov^{k-1} + \frac{i}{N} ov^k \right) \quad (6)$$

윗 식은, 각각 오버 샘플율(ov)과 샘플링위치($p_{ov}[n]$)를 나타낸다. 여기서, N은 프레임 길이, T_p는 피치주기, l은 하모닉 갯수, k는 프레임 번호를 뜻한다.

순환 파형을 만들어내는 과정 중 프레임간의 부드러운 연속성을 보장하기 위해 현 순환 파형의 시작 지점을 바꾸는 방법이 필요하게 된다. 이를 위해 offset 값을 정의하며, 이를 이용하여 각 프레임 순환 파형의 끝 지점을 한 개의 기준 파형 구간 길이인 128로 맞춤으로써 다음 프레임의 첫 지점과 자연스럽게 이어질 수 있도록 하게 된다[10].

$$w'^{k-1}(n) = w^{k-1}(\mathrm{mod}(n, 128)) \qquad n = 0, \cdots, L-1 \quad (7)$$

$$w'^{k}(n) = w^{k}(\mathrm{mod}(offset + n, 128)) \qquad n = 0, \cdots, L-1 \quad (8)$$

$$offset = 128 - \mathrm{mod}(L, 128) \quad (9)$$

여기서, L은 N개의 샘플을 복원시키기 위해 오버샘플에 사용되는 데이터 갯수이며, $w'^k(n)$ mod(x, y)는 x를 y로 나눈 나머지 값을 나타낸다. 또한 은 k번째 순환 파형을, $w^k(n)$은 k번째 기준 파형을 나타낸다.

2.2. 벡터 여기 코더

무성음 구간의 잡음과 같은 여기신호를 부호화하기 위해 stochastic 코드북을 사용한 분석/합성 방법이 사용되며, 다음과 같은 왜곡측정치가 최소가 되는 이득과 모양벡터를 찾아낸다[11].

$$e = \sum_{n=0}^{N-1} (ref(n) - G \times syn(n))^2 \quad (10)$$

여기서, $ref(n)$은 지각 가중된 LPC 합성필터를 사용한 입력신호에서 ZIR(Zero Input Response)를 제거한 ZSR(Zero State Response)이며, $syn(n)$ 은 코드북값에 의한 ZSR이다.

III. 전이 구간 코더

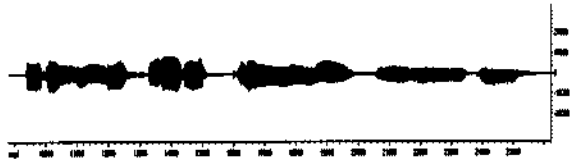
전이 구간은 유/무성음이 혼재하는 구간으로 대개 유성음에서 무성음으로 또는 그 반대로 넘어가는 시점에 있다. 이런 구간에서 하모닉 코딩의 선형 보간 overlap/add 합성방법은 부드럽게 이어지는 부분보다는 에너지가 급변하는 부분에서 파형 이득과 피치왜곡 등의 단점을 나타낸다. 따라서, 전이 구간에서는 이런 에너지 급변시점의 검출하며, 구간을 분리 코딩하는 방법이 요구된다.

3.1. 전이 구간 및 시점 검출

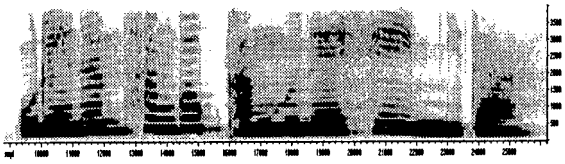
일반 유/무성음 검출은 하모닉 코딩을 적용시켰을 경우, 스펙트럴 크기 값들의 추정된 정확도, 피치, 주파수 균등(Frequency Balance)값의 인자들로 결정 될 수 있다[6]. 이런 유/무성음 판별 후 전이 구간검출이 시도되며, 전이 모드는 유/무성음 모드에 대해 우선권을 가진다.

전이 검출은 160 샘플의 임의의 시점을 기준으로 과거와 미래의 에너지비가 급변하는 정도를 측정하기 위해

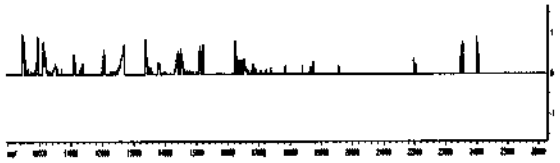
다음과 같이 n시점에 대한 과거-미래 에너지 비율 값을 구한다.



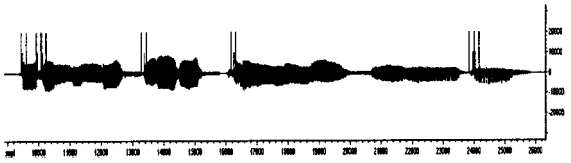
(a) 여자음성 A 원본신호
(a) Original speech signal of a female A,



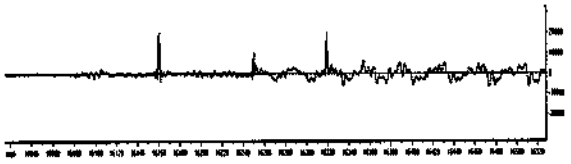
(b) 여자음성 A 대한 Spectrogram
(b) Spectrogram of the original speech signal of a female A,



(c) 여자음성 A에 대한 조건 (14)식 $Er[n]$ 값
(c) Graph of the fuction $Er[n]$ in a female speech A,



(d) 여자음성 A에 대한 전이구간 및 전이시점 검출
(d) Transition-mode detection and transition-point detection in a female speech A,



(e) 여자음성 A에 대한 4번째 전이구간의 확대
(e) Magnify of the fourth transition frame in a female speech A.

그림 3. 전이구간 및 전이시점에 대한 검출 예
Fig. 3. Example of the transition-mode detection and the transition-point detection;

$$MI[n] = MIN\left(\sum_{i=0}^{T_p} x^2[n+i],\ \sum_{i=0}^{T_p} x^2[n-i]\right) \quad (11)$$

$$MA[n] = MAX\left(\sum_{i=0}^{T_p} x^2[n+i],\ \sum_{i=0}^{T_p} x^2[n-i]\right) \quad (12)$$

$$Er[n] = 1 - \frac{MI[n]}{MA[n]} \quad (13)$$

T_p는 개루프 피치주기이며, $x[n]$은 DC제거 필터를 통과한 후의 음성신호를 나타낸다. $MIN(x, y)$는 x, y 중 작은 수를, $MAX(x, y)$는 x, y중 큰 수를 택하는 함수이다. 또한, 실제 과거-미래 에너지 비율은 높지만, 에너지 차이가 지각적으로 분별되지 못하는 경우를 고려하여, 다음 두 가지 조건을 만족하는 경우, 전이구간이라 판별한다.

$$MA[n] - MI[n] > TH1 \quad (14)$$

$$Er[n] > TH2 \quad (15)$$

여기서, $TH1$, $TH2$는 실험적인 상수값이 된다. 위 조건을 만족하는 경우, 전이 시점을 구하는 과정이 포함되며, 프레임내의 $Er[n]$이 가장 큰 곳을 전이 시점으로 파라미터화 한다. 이에 대한 예를 그림 3에 나타내었다.

3.2. 하모닉-하모닉 여기 코딩부분

전이 구간에서의 코딩은 이전구간이 무성음 구간인지 아닌지에 대한 조건에 따라 결정된다. 우선, 이전 구간이 무성음 구간인 경우 전이시점을 기준으로 과거 전이구간 샘플들은 벡터여기코딩이 사용되며, 미래 전이구간 샘플들에 대해서는 하모닉 여기코딩이 사용된다. 반대로, 무성음이 아닌 경우에 대해서는 양 샘플들 모두 하모닉 코딩이 적용된다. 본 음성 부호화기에서는 전자를 전이1로 후자에 대해서는 전이2로 분류하며, 각각을 벡터-하모닉 코딩, 하모닉-하모닉 코딩으로 나눈다.

하모닉 코딩을 위한 목적신호로는 전이시점의 과거에 대해서는 미리 기억된 과거샘플들이 사용되며, 미래에 대해서는 현 구간에 대한 샘플수가 적을 수 있기 때문에 미래 구간의 96샘플이 사용된다. 미래구간을 위해 추출된 목적신호는 96에서 256사이의 가변크기를 가지므로 추출된 목적신호를 중앙에 배치하고, 나머지 부분은 0을 첨가함으로써 256크기의 최종 목적신호를 만들어낸다. 이러한 각각의 최종 목적신호는 해밍윈도우를 씌운 뒤 256-point DFT과정을 거친다. 그런 다음 스펙트럴 정보를 폐루프 피치 검출과정으로 추출하며, 각각의 최종 피치 정보도 계산되어진다.

분리된 각각의 과거, 현재 전이구간의 스펙트럴 및 피치 정보는 과형 복원을 위한 파라미터로 쓰이게 된다. 우선, 피치주파수의 배수마다 추출된 스펙트럴 크기 값들인 스

펙트럴 파라미터를 사용해 그 크기 값들을 1-point 마다 일렬로 배열하고 나머지 64-point 까지는 0으로 채움으로써 대칭인 128-point 목적 샘플을 만든다. 그런 다음, 128-point IDFT를 적용함으로써 기준 파형을 구하게 된다. 이렇게 만들어진 기준 파형은 과거 및 미래 전이구간의 파형을 합성하는데 쓰이게 된다. 과거 전이구간 파형은 식 (16),(17)에 의해 과거 전이구간 기준 파형 $w_p(n)$으로부터 유도된 $w_{sp}(n)$와 $w_{nsp}(n)$을 식 (19)에 적용하여 순환 파형 $w'_p(n)$으로 만든 후 오버샘플링 과정을 통해 합성된다. $w_{sp}(n)$부분은 유성음 구간에서와 동일한 방법이며, $w_{nsp}(n)$부분은 전이구간의 급격한 변화를 고려하기 위해 제안된 방법으로, 접합 부분의 1 피치주기 과거 전이구간 기준 파형 $w_{sp}(n)$과 새로운 미래 전이구간 기준 파형 $w_{nsf}(n)$을 선형 보간한 것이다. 이러한 선형 보간 방법은 접합부에서의 크기에 대한 연속성을 보장하기 위해서이다.

이에 대한 예를 그림 4에 도시하였으며, 'New past waveform'이 새로운 미래 기준 파형 $w_{nsf}(n)$과 과거 1 피치주기의 과거 기준 파형 $w_{nsp}(n)$을 선형 보간한 파형의 예를 나타낸다. 여기서, 새로운 미래 기준 파형 $w_{nsf}(n)$은 접합부분에 대한 위상일치 문제를 해결하기 위해 미래전이 구간 파형 $w_f(n)$을 쉬프트시킨 파형을 나타낸다. 전이시점 접합부분에서의 위상연속성을 보장하기 위해, 전이시점에서의 과거 1 피치 주기의 과거 기준 파형 $w_{sp}(n)$과 미래 전이구간 기준 파형 $w_f(n)$의 상관도가 최대화되는 $offset_f$ 값을 선택하게 된다. $offset_f$로 쉬프트된 이러한 새로운 미래 기준 파형 $w_{nsf}(n)$은 전이시점 접합부분의 과거 기준 파형 $w_{nsp}(n)$을 위한 선형 보간으로 사용될 뿐만 아니라, 미래전이 구간 파형에 대한 합성을 위해 식 (17)과 같은 순환 파형 $w'_f(n)$ 합성과정과 식 (5),(6)과 같은 오버 샘플링 과정에 사용된다.

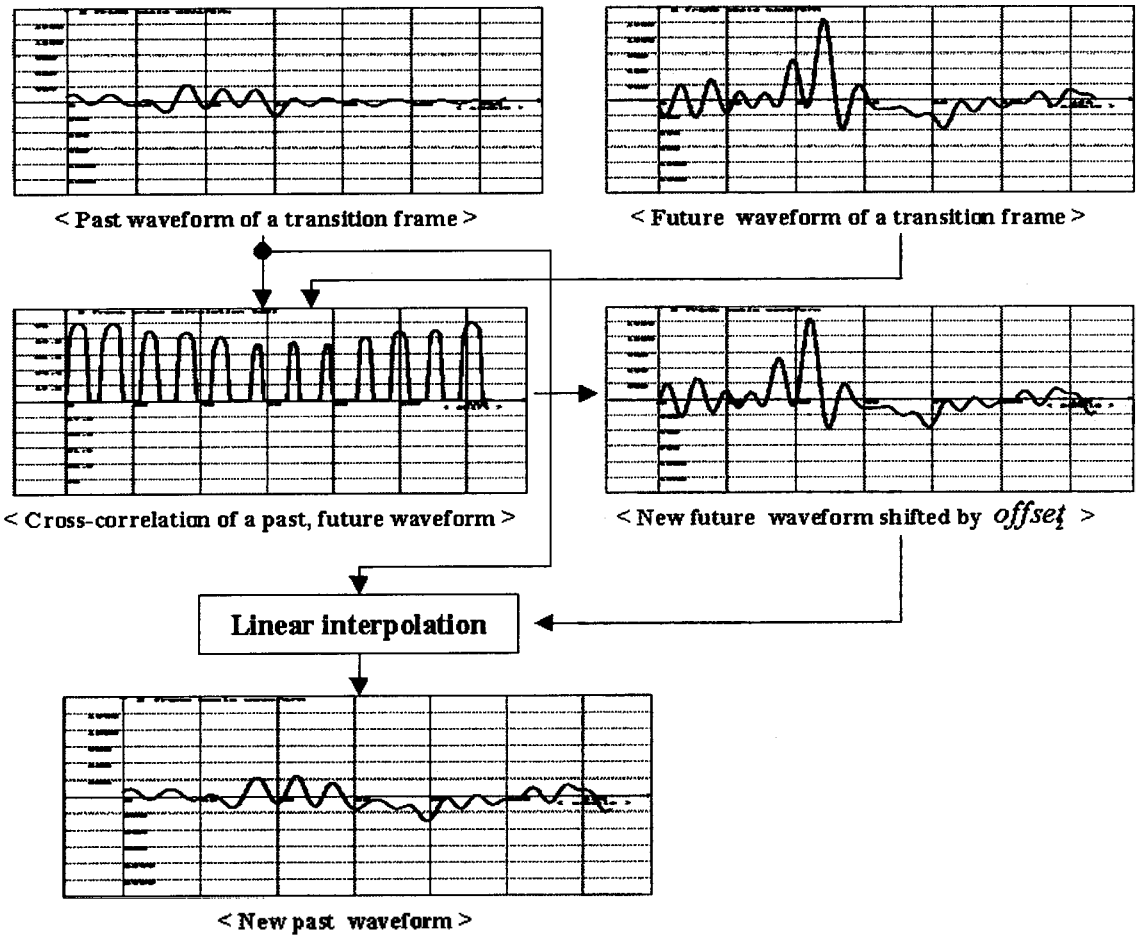


그림 4. 전이구간에서 위상일치를 위한 과거 기준 파형의 설정
Fig. 4. Past waveform evaluation for the phase matching in a transition frame.

$$w'_p(n) = \begin{cases} w_{sp}(\bmod(n,128)) & n \le L_p - 128 \\ w_{nsp}(\bmod(n,128)) & n > L_p - 128 \end{cases} \quad n=0,\cdots,L_p-1 \tag{16}$$

$$w'_f(n) = w_{nsf}(\bmod(n,128)) \quad n=0,\cdots,L_f-1 \tag{17}$$

$$w_{nsp}(n) = LI(w_{sp}(n), w_{nsf}(\bmod(n,128))) \quad n=0,\cdots,127 \tag{18}$$

$$w_{nsf}(n) = w_f(\bmod(offset_f+n,128)) \quad n=0,\cdots,127 \tag{19}$$

$$offset_f = \max_t \left[\sum_{n=0}^{127} w_{sp}(n) w_f(n+t)\right] \tag{20}$$

$$w_{sp}(n) = w_p(\bmod(offset_p+n,128)) \quad n=0,\cdots,127 \tag{21}$$

$$offset_p = 128 - \bmod(L_p,128) \tag{22}$$

LI : 선형보간

여기서, $w'_p(n)$은 전이시점 이전의 과거 순환 파형을, $w'_f(n)$은 전이시점 이후의 미래 순환 파형을 나타낸다. $w_{nsp}(n)$은 전이시점 접합부분의 새로운 과거 기준 파형을, $w_{nsf}(l)$은 전이시점 이후의 새로운 기준 파형을 나타낸다. 또한, $w_p(n)$은 전이시점 이전의 과거 기준 파형을, $w_f(l)$은 전이시점 이후의 기준 파형을 나타낸다. L_p와 L_f 는 전이시점까지 오버샘플에 사용되는 과거 기준 파형의 샘플 수와 미래 기준 과형 샘플 수를 나타내고. 각 기준 파형은 0에서 127사이에 값을 가지며, 그 외의 구간에 대해서는 0을 갖는다.

전이시점 이후의 미래 전이구간과 유성음 프레임과의 접합부분에서는 기존의 유성음 구간 사이의 접합방식에 대해 약간의 변형을 필요로 한다. 전이 구간에서의 위상을 일치시키기 위해 쉬프트 과정을 거쳤으므로, 식 (23)에서와 같이 미래 전이구간의 끝점과 유성음 시작부분에서 선형 보간으로 사용될 미래전이 구간 기준 파형 시작점을 일치시키는 쉬프트 과정이 첨가된다. $offset_{ff}$는 미래전이 구간 기준 파형 끝점을 나타낸다. 그 이후의 합성과정은 식 (8),(9)에서와 동일하다.

$$w'^{k-1}(n) = w_{nsf}(\bmod(offset_{ff}+n,128)) \quad n=0,\cdots,L-1 \tag{23}$$

$$w'^{k}(n) = w'^{k}(\bmod(offset+n,128)) \quad n=0,\cdots,L-1 \tag{24}$$

$$offset_{ff} = 128 - \bmod(L_f,128) \tag{25}$$

$$offset = 128 - \bmod(L,128) \tag{26}$$

전이시점 이후의 미래전이 구간에 무성음구간이 올 경우 또한, 유성음구간과 같은 쉬프트과정을 거친다. 무성음 구간이므로 현재 순환 파형 w'^{k} 는 모두 0 이며, 벡터여기 코딩에의해 합성된 무성음 신호가 선형 보간 된다.

3.3. 벡터-하모닉 여기 코딩부분

이 부분은 무성음에서 유성음으로 전이할 경우에 해당하며, 주로 음성의 시작부분에 해당된다. 이런 부분은 정체 무성음이 유지되다가 전이시점에서 갑작스런 유성음 구간으로 바뀌는 부분을 나타낸다. 따라서, 하모닉-하모닉 코딩보다는 벡터-하모닉 코딩을 필요로 하며, 전이시점의 과거부분을 복원대상으로 하여 복호화기가 구성된다. 이전 구간이 무성음 구간이기 때문에 연속성을 고려하여 전이시점의 과거부분을 무성음 구간으로 간주한다. 벡터여기 부분 역시 무성음 구간에서의 stochastic 코드북 구조를 사용하며, 전이구간 전 샘플을 목적 대상으로 한다.

전이1로 분류되는 무성음-유성음 전이구간은 전이시점 이후의 미래 전이구간에 대해서는 유성음-유성음 전이구간인 하모닉-하모닉 코딩에서와 같은 합성방법을 적용한다. 이 경우, 하모닉-하모닉 코딩에서의 과거 전이 구간 기준파형은 모두 0가 되며, 벡터 여기 코딩으로부터 합성된 무성음 여기신호가 선형 보간의 대상이 된다. 미래 전이구간과 이후의 접합에 대한 방법은 하모닉-하모닉 코딩에서와 동일하다.

IV. 양자화 및 비트 할당

양자화 과정은 크게 LSP 및 유/무성음 양자화 과정과 여기신호 양자화 과정으로 구성된다. 우선 LSP 양자화 과정은 3 단계 분할 벡터 양자화(3-Stage Split Vector Quantization)를 적용하여, 24 비트를 할당하였고, 유성음/무성음/전이1/전이2를 표현하기위한 모드 값에는 2bit가 할당된다. 따라서, 4kbps에 할당된 총 80 비트 중 54 비트가 여기신호 표현을 위해 할당되며, 3가지 코딩 파라미터에 따라 비트 할당이 바뀌게 된다. 이중 하모닉 여기 코딩은 피치 값 및 하모닉 양자화 값을 그 파라미터로 가지며, 하모닉 양자화를 위해 2 단계 분할 벡터 양자화(2-Stage Split Vector Quantization)를 사용한다. 비트 할당은 피치 값에 7 비트, 1 단계에 Gain 5비트 Shape 10 비트를, 2 단계에는 1 단계의 에러값을 Full search 코드북 4개를 사용하여 모두 32 비트를 할당하였다. 벡터 여기 코딩은 10ms 단위의 2개의 부 프레임으로 구성되며, 각 프레임은 같은 2 단계 벡터 양자화를 적용한다. 1 단계는 Gain 4 비트 Shape 7 비트를, 2 단계는 다시 5ms 단위로 하여 Gain 3비트, Shape 5 비트를 적용하였다. 전이 구간의 전이2로 분류되는 하모닉-하모닉 여기 코딩은 각 구간에서의 추출된 피치를 7비트,7비트 총 14비트를 할당하였고, 1 단계 양자화를 각각 적용하고, 남은 3 비트를 전이 시점을 전송하는데 이용하였다. 또한 전이1으로 분류되는

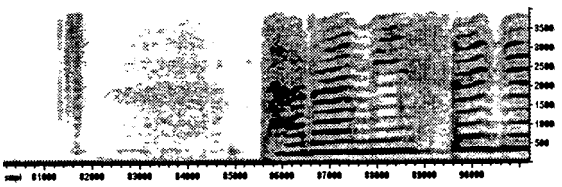
벡터-하모닉 여기 코딩은 벡터코딩의 1 단계 양자화에 쓰인 22 비트와 하모닉 코딩의 피치 7 비트 할당 및 1단계 양자화에 쓰인 22 비트를 적용하였고, 역시 남은 10 비트를 전이시점을 표현하는데 할당하였다. 이 부분에서 전이시점을 위한 10 비트는 실제로는 너무 큰 할당 값이므로, 일부를 다른 용도로 쓸 수 있을 것이다. 지금까지 설명한 비트 할당을 표 1에 나타내었다.

표 1. 제안된 4kbps 코더의 비트 할당
Table 1. Bit allocation of the proposed 4kbps coder.

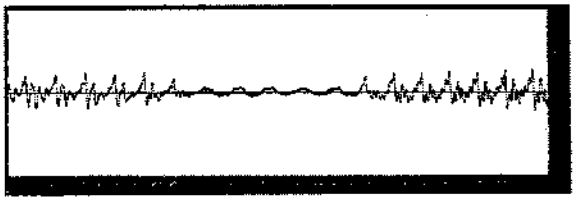
<table>
<tr><td></td><td colspan="8">비트수</td></tr>
<tr><td>10 차 LSP 양자화값</td><td colspan="8">24</td></tr>
<tr><td>V/UV 상태값</td><td colspan="8">2</td></tr>
<tr><td></td><td>분류</td><td colspan="2">무성음</td><td>유성음</td><td>분류</td><td colspan="2">전이 1</td><td>전이 2</td></tr>
<tr><td rowspan="7">Excitaion 관련 인덱스값</td><td rowspan="2">피 치</td><td colspan="2" rowspan="2">0</td><td rowspan="2">7</td><td>피치</td><td colspan="2">7</td><td>7+7=14</td></tr>
<tr><td>전이시점</td><td colspan="2">10</td><td>3</td></tr>
<tr><td rowspan="2">1 단계</td><td>10ms</td><td>11</td><td rowspan="2">15</td><td rowspan="2">과거 구간 1 단계</td><td>10ms</td><td>11</td><td rowspan="2">15</td></tr>
<tr><td>10ms</td><td>11</td><td>10ms</td><td>11</td></tr>
<tr><td rowspan="4">2 단계</td><td>5ms</td><td>8</td><td rowspan="4">32</td><td rowspan="4">미래 구간 1 단계</td><td rowspan="4">15</td><td rowspan="4">15</td></tr>
<tr><td>5ms</td><td>8</td></tr>
<tr><td>5ms</td><td>8</td></tr>
<tr><td>5ms</td><td>8</td></tr>
<tr><td>총 비트수</td><td colspan="8">80</td></tr>
</table>

V. 실험 결과

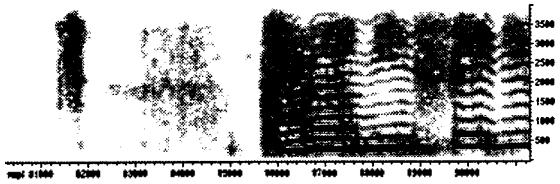
다음 그림 5는 "바람이 세게 불고 있습니다" 중 "바람이 세게" 에 해당하는 부분이다. 원음 스펙트로그램에서 볼 수 있듯이 에너지가 급변하는 "바람" 부분은 확대된 전이구간 파형인 그림 5(b),(d),(f)에서 전이구간이 고려되지 않은 코더에 비해 개선되었음을 볼 수 있다. 이는 연속되는 전이 구간 사이의 이득 불일치를 개선하였음을 의미한다. "강당의 학생들이 꽉 차 있었다" 중 "강당의 학생들"에 해당하는 부분이며, 발음이 시작되는 부분이 그 주목 대상이 된다. 여기서, 전이 구간이 고려되지 않은 코더의 경우 피치의 불연속성이 관찰되며, 이를 개선한 전이 코더의 스펙트로그램과 그 확대부분 파형인 그림 6(b),(d),(f)를 볼 수 있다.



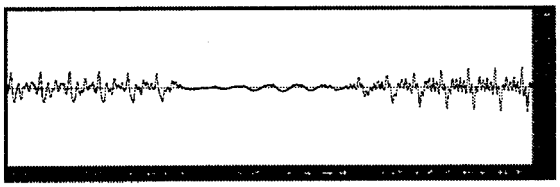
(a) 원음 스펙트로그램
(a) Spectrogram of the original speech,



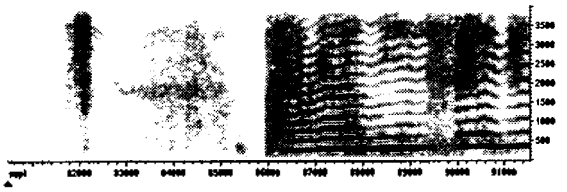
(b) 원음 파형
(b) Waveform of the original speech,



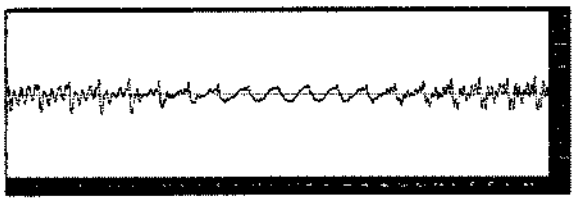
(c) 전이코딩을 사용한 복원신호 스펙트로그램
(c) Spectrogram of the reconstructed speech using the transition coding,



(d) 전이코딩을 사용한 복원신호 파형
(d) Waveform of the reconstructed speech using the transition coding,

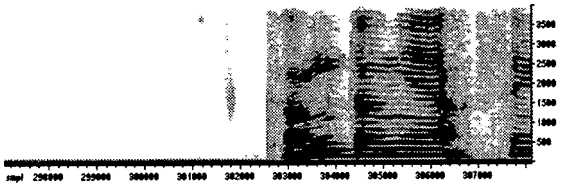


(e) 전이코딩을 사용하지 않은 복원신호 스펙트로그램
(e) Spectrogram of the reconstructed speech non-using the transition coding,

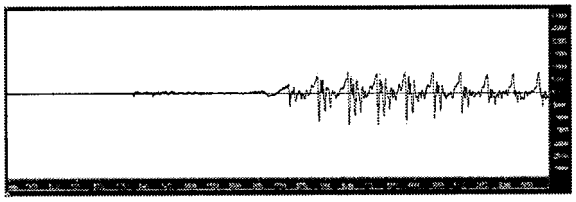


(f) 전이코딩을 사용하지 않은 복원신호 파형
(f) Waveform of the reconstructed speech non-using the transition coding,

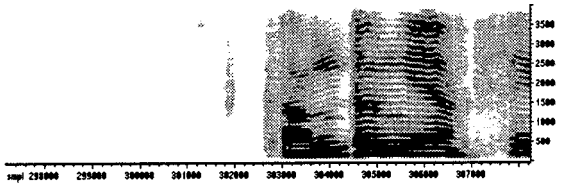
그림 5. 전이 구간의 이득 일치 비교
Fig. 5. Comparison for the gain matching in a transition frame;



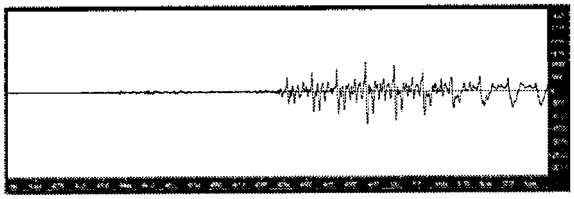
(a) 원음 스펙트로그램
(a) Spectrogram of the original speech,



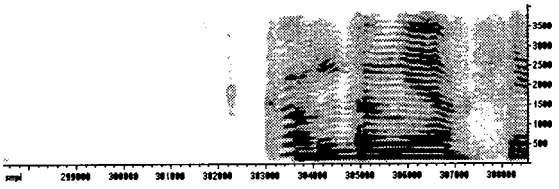
(b) 원음 파형
(b) Waveform of the original speech,



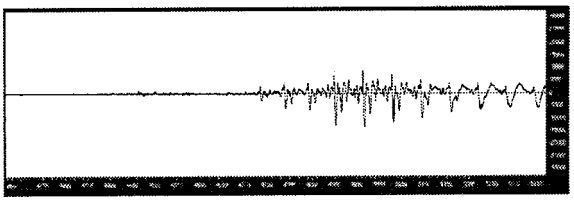
(c) 전이코딩을 사용한 복원신호 스펙트로그램
(c) Spectrogram of the reconstructed speech using the transition coding,



(d) 전이코딩을 사용한 복원신호 파형
(d) Waveform of the reconstructed speech using the transition coding,



(e) 전이코딩을 사용하지 않은 복원신호 스펙트로그램
(e) Spectrogram of the reconstructed speech non-using the transition coding,



(f) 전이코딩을 사용하지 않은 복원신호 파형
(f) Waveform of the reconstructed speech non-using the transition coding.

그림 6. 전이 구간의 피치 왜곡 비교
Fig. 6. Comparison for the pitch distortion in a transition frame;

본 음성부호화기는 20㎳ 단위로 처리하며 96샘플의 중첩으로 기인한 32㎳의 인코딩 지연을 가지며. 표 2에 본 논문에서 제안된 음성 부호화기의 MOS(Mean Opinion Score) 시험 결과를 나타내었다. MOS 시험은 ITU 표준안인 CS-ACELP[12]와 8kbps QCELP[13] 그리고, 4.8kbps DOD-CELP[14]를 비교 대상으로 하였으며, 음질 시험에는 발화자가 다른 남자, 여자 각각 10개 문장을 사용하여, 8명의 청취자를 대상으로 수행하였다. 실험 결과, 제안된 4kbps 음성코더는 8kbps QCELP 알고리즘보다 나은 음질을 보였으며, CS-ACELP 에 비해서는 약 MOS 0.41 정도 낮게 평가되었다.

표 2. MOS 시험 결과
Table 2. Result of MOS test.

분류	남자	여자	전체
원음	4.76	4.70	4.72
CS-ACELP	4.19	3.70	3.94
8Kbps QCELP	3.61	3.38	3.48
4.8Kbps DOD-CELP	3.03	2.55	2.79
전이 구간을 고려한 혼합코더	3.73	3.33	3.53
전이 구간을 고려치 않은 혼합코더	3.55	3.24	3.39

VI. 결 론

본 논문에서 제시된 코더는 선형 보간으로부터 기인하는 하모닉 코더의 단점인 연속된 전이 구간에서의 이득 불일치와 피치왜곡의 개선을 목적으로 설계되었다. 또한, 그 실험 결과, 복원신호의 스펙트로그램과 음질 테스트로부터 개선여부를 확인할 수 있었다. 4kbps 음성 부호화기 음질은 8kbps QCELP 보다 나은 음질을 보였다. 이러한, 전이 구간 코더는 좀더 낮은 전송률에서 좋은 음질을 얻기 위한 새로운 모델로 제시될 수 있으며, 좀더 효과적인

검출 및 접목을 이룰 경우 명료한 음질을 얻을 수 있을 것으로 기대된다.

참 고 문 헌

1. R. V. Cox, "Speech Coding Standards," *Speech Coding and Synthesis, Chapter 2*, pp. 49-78, W. B. Kleijn, and K. K. Paliwell Eds., Elsevier, 1995.
2. D. Childer, R. V. Cox, R. DeMori etc, "The Past, Present, and Future of Speech Processing," *IEEE Signal Processing Magazine*, pp. 24-48, B. H. Juang Eds., May 1998.
3 A. M. Kondoz, "Coding Strategies and Standards," *Digital Speech, Chapter 5*, pp. 117-140, 1994.
4. A. V. McCree, K. Trung, E, B, George, T. P. Banwell and V. Viswanathan, "A 2.4kbit/s MELP Coder Candidate for the New U.S. Federal Standard," in *Proc. ICASSP-96*, vol. 1, pp. 200-203, May 1996.
5. L. Nishiguchi, K. Iijima and J. Matsumoto, "Harmonic vector excitation coding of speech at 2.0 kbps," in *Proc. IEEE Workshop on Speech Coding For Telecommunications*, pp. 39-40, 1997.
6. R. J. McAulay, and T. F. Quartieri, "Sinusoidal coding," *Speech Coding and Synthesis*, Chapter 4, pp. 121-173, W. B. Kleijn, and K. K. Paliwell Eds., Elsevier, 1995.
7. R. J. McAulay and T. F. Quatieri, "Speech analysis/synthesis based on a sinusoidal representation," *IEEE Transaction on Acoustic, Speech, and Signal Processing*, vol. 1, ASSP-34, pp. 744-754, August 1986.
8. P. Hedelin, "Phase compensation in all-pole speech analysis," in *Proc. IEEE ICASSP-88*, pp. 339-342, 1988.
9. E. Shlomot, Vladimir Cuperman and A. Gersho, "Combined Harmonic and Waveform Coding of Speech at Low Bit Rate," in *Proc, ICASSP-98*, pp. 585-588, 1998.
10. Masayuki, Nishiguchi and J. Matsumotor, "Harmonic and Noise Coding of LPC Residuals with Classified Vector Quantization," in *Proc, ICASSP-95*, pp. 484-487, 1995.
11. A. M. Kondoz, "Code-Excited Linear Predictive Coding," *Digital Speech, Chapter 6*, pp. 174-212, 1994.
12. R. Salami, C. Laflamme, J. Adoul etc, "Design and Descrition of CS-ACELP: A Toll Quality 8 kb/s Speech Coder," *IEEE Transaction on Acoustic, Speech, and Signal Processing*, vol. 6, No 2, pp. 744-754, March 1998.
13. P. J. A. DeJaco, W. Gardner and C. Lee, "QCELP: North American CDMA digital cellular variable rate speech coding standard," in *Proc. IEEE Workshop on speech Coding for Telecommunications*, (Sainte-Adele. Quebec), pp. 5-6, 1993.
14. J. P. Campbell, V. C. Welch, and T. E. Tremain, "The new 4800 bps voice coding standard," in *Proc. Military Speech Tech.*, pp. 64-70, 1989.

▲김 종 학(Jonghark Kim)
1998년 2월 : 충북대학교 전자공학과 학사
1998년 3월~현재 : 충북대학교 전파공학과 석사과정
※주관심분야 : 음성 오디오 부호화, 영상압축, 적응필터

▲이 인 성(Insung Lee)
1983년 2월 : 연세대학교 전자공학과 학사
1985년 2월 : 연세대학교 전자공학과 석사
1992년 12월 : Texas A&M University 전기공학과 박사
1986년 5월~1987년 7월 : 한국통신 연구개발단 전임연구원
1993년 2월~1995년 9월 : 한국전자통신연구원 이동통신 기술연구단 선임연구원
1995년~현재 : 충북대 전기전자공학부 부교수
※주관심분야 : 음성 및 영상신호압축, 이동통신, 적응필터