

ACELP 음성 부호화기를 위한 개선된 대수 코드북구조 와 고정 코드북 이득의 차분 양자화 기법

정희원 김 상 조*, 서 방 원*, 김 형 명*

An improved algebraic-codebook structure and a differential quantization of the fixed-codebook gain for ACELP coders

Sang-Jo Kim*, Bang-Won Seo*, Hyung-Myung Kim* *Regular Members*

요 약

본 논문에서는 ACELP 음성 부호화기를 위한 효율적인 대수 코드북 구조와 고정 코드북 이득에 차분 양자화를 적용하는 방법을 제안하였다. 대수 코드북은 잔류 신호를 표시하는 코드북의 펄스 개수가 증가함에 따라 음질은 향상되나 펄스 위치를 탐색하는데 필요한 계산량이 급격히 증가한다는 단점이 있다. 이러한 단점을 보완하기 위해 펄스 위치가 서로 다른 두개의 대수 코드북을 이용하여 펄스 위치를 탐색하는데 필요한 계산량을 감소시키는 방법을 제안하였다. 또한, 현재 부프레임과 이전 부프레임의 고정 코드북 이득간의 상관관계를 이용한 차분 양자화 방법을 적용하여 코드북 이득에 할당되는 비트수를 줄이도록 하였다. 모의실험을 통하여 제안된 방법이 기존의 ACELP 음성 부호화기보다 낮은 전송률로 같은 성능을 낼 수 있으며 채널오류가 전파되지 않음을 확인하였다.

ABSTRACT

We propose an efficient codebook structure for algebraic CELP (ACELP) speech coder and a differential quantization method for fixed codebook gain. As the number of pulses of an algebraic codebook increases, the speech quality can be improved but the computational load in searching the codebook increases exponentially. We propose two-algebraic-codebook structure to reduce the computational load in searching the codevector. An efficient differential quantization scheme has been proposed to reduce the bit rate for the fixed codebook gain based on the correlation between the fixed codebook gains of current subframe and previous one. Simulation results show that the proposed scheme can achieve the same performance as the conventional ACELP with the reduced bit rate and that it is robust to the channel error propagation.

I. 서 론

무선 통신 (wireless communication)에 사용되는 주파수 자원은 사용 분야에 따라 제한되며, 사용할 수 있는 대역폭 (bandwidth)도 한정적이다. 이러한

이유에서 이동통신이나 위성통신과 같은 분야에서는 대역폭을 효율적으로 사용하기 위한 연구가 많이 이루어지고 있다. 음성 부호화기에서도 대역폭을 효율적으로 사용하기 위해 낮은 전송률에서 좋은 음질을 갖는 음성 부호화기에 대한 연구가 절

*한국과학기술원 전기 및 전자공학과
논문번호 : 98079-0220, 접수일자 : 1998년 2월 20일

실히 필요하다. 지금까지 개발된 음성부호화 방식은 크게 파형 부호화 (waveform coding), 보코딩 (voice coding), 그리고 혼합 부호화 (hybrid coding)로 분류할 수 있다 [1][2][3].

이러한 음성 부호화 방식의 주된 기능은 음성 신호를 분석하여 잉여 성분 (redundancy)을 제거한 후, 제거되지 않고 남아있는 잔류 성분들을 심리적으로 타당한 방법을 이용하여 부호화하는 것이다. 파형 부호화 방식은 부호화 과정에 필요한 계산량이 적고 음질이 우수하지만 전송률이 높다는 단점이 있고, 보코딩 방식은 낮은 전송률을 갖지만 음질이 좋지 않다는 단점이 있다. 혼합 부호화 방식은 파형 부호화 방식과 보코딩 방식의 장점을 결합시킨 형태로 비교적 낮은 전송률에서 좋은 음질을 제공한다. 그러나, 혼합 부호화 방식은 부호화 과정에 소요되는 계산량이 많다는 단점이 있다 [4][5].

부호 여기 선형 예측 (Code Excited Linear Prediction : CELP) 부호화기는 혼합 부호화 방식으로 분류되는 부호화기로서, 가장 널리 사용되고 있으며 연구도 활발히 이루어지고 있다. CELP 부호화기는 LPC 분석, 피치 (pitch) 탐색, 고정 코드북 (fixed codebook) 탐색의 세 과정으로 구성되는데, 고정 코드북 구조에 따라 여러 방식의 부호화기가 제안되었다 [6][7][8]. 부호화기의 성능과 계산량은 고정 코드북의 구조에 따라 여러 형태로 나타난다. 고정 코드북에 대수 코드북을 사용하는 경우를 대수 CELP (Algebraic CELP : ACELP) 방식 부호화기라고 부른다.

본 논문에서는 최근 많은 연구가 이루어지고 있는 이 ACELP 부호화기의 성능을 향상시키기 위하여 다음과 같이 두 가지 방법을 제안한다. 첫째는 고정 코드북에 사용되는 대수 코드북의 개선된 구조를 제안하고 그 성능을 분석한다. 대수 코드북은 영이 아닌 몇 개의 펄스 (pulse)로 구성되며 펄스의 개수가 많을수록 음질은 향상되나 펄스 위치를 탐색하는데 필요한 계산량이 급격히 증가한다는 단점이 있다. 본 논문에서는 펄스 위치가 서로 다른 두개의 대수 코드북을 이용하여 펄스 위치를 탐색

하는데 필요한 계산량을 감소시키는 방법과, 두개의 대수 코드북 이득의 상관관계를 이용하여 하나의 이득만을 전송함으로써 전송 비트 (bit)수를 줄이는 방법을 제안한다. 둘째는 고정 코드북 이득을 차분 양자화하는 방법이다. 현재 부프레임 (subframe)의 고정 코드북 이득과 이전 부프레임의 고정 코드북 이득간의 상관관계를 분석한 후, 이를 이용하여 고정 코드북 이득에 차분 양자화 (differential quantization)방법을 적용하여 코드북 이득에 할당되는 비트수를 줄인다. 또한 차분 양자화 방식을 사용하는 경우 채널 오류에 의한 영향을 알아보기 위하여 프레임 손실 (frame erasure)이 없는 경우와 프레임 손실이 있는 경우로 나누어 성능을 분석하고, 채널 오류가 있는 경우에도 성능 저하를 줄이는 방법을 제시한다.

본 논문의 구성은 다음과 같다. II장에서 ACELP 음성 부호화기의 구조에 대하여 설명하고, III장에서 개선된 대수 코드북과 고정 코드북 이득에 차분 양자화를 적용하는 방법을 제안한다. IV장에서 기존의 방법과 제안된 방법을 모의실험 (simulation)을 통하여 분석하고, 마지막 V장에서 결론을 맺는다.

II. ACELP 음성 부호화기

ACELP 음성 부호화기는 LPC 분석, 피치 탐색, 고정 코드북 탐색의 세 과정으로 구성된다. 먼저 LPC 분석을 통하여 시변 선형 필터의 계수를 추출하고 입력 음성에서 단구간 상관을 제거하여 잔류 신호를 얻는다. 피치 탐색 과정에서는 잔류 신호에 존재하는 주기 성분을 추출하여 잔류 신호에서 주기 성분을 제거한다. 마지막으로 고정 코드북 탐색 과정에서 입력 신호와 합성 신호간의 평균 제곱근 오차를 최소화하는 잔류 신호에 대한 고정 코드북의 값을 선택한다 [9].

ITU-T 권고안 G.723.1의 표준으로 선택된 ACELP 음성 부호화기는 대수 코드북을 사용하여 여기 신호를 생성한다. 입력 신호는 8000 Hz로 표본화 (sampling)한 16 비트 선형 PCM 신호이며 복호

기의 출력 신호도 동일하다. 그림 1은 ACELP 부호화기의 구조를 나타낸 것이다.

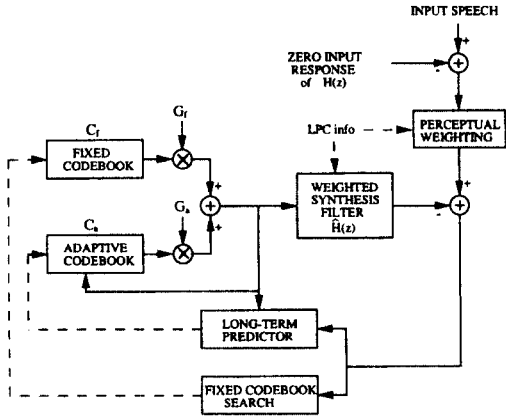


그림 1. ACELP 부호화기 구조

2.1 LPC 분석

LPC 분석에서는 10차 선형 예측기를 사용하며 현재 부프레임의 가운데를 중심으로 180표본 크기의 해밍 윈도우를 씌운다. 윈도우를 통과한 신호로부터 공분산을 구한 후 레빈슨-더빈 (Levinson-Durbin) 순환 알고리즘을 사용하여 각 부프레임마다 10차 LPC 계수들을 계산한다. LPC 합성 필터 (synthesis filter)는 다음과 같다.

$$A_i(z) = \frac{1}{1 - \sum_{j=1}^{10} a_{ij}z^{-j}}, \quad 1 \leq i \leq 4. \quad (1)$$

여기서 i 는 부프레임 색인이다. 부프레임마다 구한 LPC 계수들 모두 전송하는 대신 네번째 부프레임의 LPC 계수만을 전송하고 나머지 부프레임의 LPC 계수들은 이전 네번째 부프레임의 LPC 계수와 현재 네번째 부프레임의 LPC 계수들을 보간 (interpolation) 하여 구한다. LPC 계수들을 전송하기 전에 LSP (Line Spectrum Pairs) 주파수들로 변환하여 벡터 양자화한다. 복호기에서는 전송된 벡터 양자화 색인에 해당하는 LSP 주파수들을 구한 후 보간에 의해 각 부프레임의 LSP 주파수들을 구한다. 여기서 구한 LSP 주파수들은 합성 필터의 LPC

계수들로 다시 변환된다.

2.2 피치 탐색

각 프레임마다 두개의 개루프 피치 주기를 구한다. 하나는 처음 두 부프레임을 위한 것이고 다른 하나는 나머지 두 부프레임을 위한 것이다. 개루프 피치 주기, L_{ol} ,는 인지 가장 입력 음성으로부터 계산되며, 상관 척도를 최대로 하는 18에서 142사이의 값을 갖는 색인이 선택된다. 상관 척도, $C_{ol}(j)$ 는 다음과 같다.

$$C_{ol}(j) = \frac{[\sum_{n=0}^{119} f(n) \cdot f(n-j)]^2}{\sum_{n=0}^{119} f(n-j) \cdot f(n-j)}, \quad 18 \leq j \leq 142. \quad (2)$$

여기서 $f(n)$ 은 인지 가장 입력 음성이다. 개루프 피치 주기와 5차 피치 예측기를 이용하여 각 부프레임마다 페루프 피치 주기를 계산한다. 첫번째와 세번째 부프레임의 페루프 피치 주기는 개루프 피치 주기 근처 (-1, 0, 1)에서 선택되고 7 비트로 부호화된다. 그리고, 두번째와 네번째 부프레임의 페루프 피치 주기는 2 비트로 부호화 되는데 이전 부프레임의 페루프 피치 주기와 차이 값을 전송하며 그 값은 -1, 0, 1, 2 중에서 선택된다. 피치 이득은 피치 이득 양자화를 위한 170개의 코드워드로 구성된 코드북 중에서 최적의 코드워드가 선택되며 선택된 색인을 전송한다.

2.3 고정 코드북 탐색

ACELP 부호화기에서는 고정 코드북으로 대수 코드북을 사용하는데, 대수 코드북을 사용하면 CELP 음성 부호화기의 고정 코드북 탐색에 필요한 계산량과 코드북을 위한 메모리를 줄일 수 있다.

코드북에 의해서 생성된 여기 신호는 합성 필터를 통과한 후 이득을 곱하면 합성 신호가 생성된다. 입력 음성 신호는 LPC 예측 필터를 거치면서 단구간 상관이 제거되고, 피치 탐색 과정을 통하여 장구간 상관도 제거된다. 단구간 상관과 장구간 상관

제거된 잔류 신호는 고정 코드북에 의해서 양자화 된 후 전송된다. 코드북 탐색에서는 합성 신호와 가중 음성 신호간의 평균 제곱근 오차를 최소화 하는 코드북 색인과 이득을 구한다. 평균 제곱근 오차는 다음과 같다.

$$E_k = \|r - G H v_k\|^2. \quad (3)$$

여기서 r 은 현재 부프레임의 가중 입력 음성 신호에서 가중 합성 필터의 영상태 응답과 피치 주기 성분을 제거한 목적 벡터이고, G 는 코드북 이득, 그리고 v_k 는 색인 k 에 해당하는 대수 코드워드 (algebraic codeword)이다. H 는 대각 성분이 $h(0)$ 이고 하위 대각 성분들이 $h(1), \dots, h(59)$ 인 하위 삼각 토플리츠 길썸 행렬 (lower triangular Toeplitz convolution matrix)이다. $h(n)$ 은 가중 합성 필터 (weighted synthesis filter)의 영상태 응답 (impulse response)이다.

III. 제안된 대수 코드북과 대수 코드북 이득의 차분 양자화

3.1 제안된 대수 코드북 구조와 탐색 과정

대수 코드북은 영이 아닌 몇 개의 펄스들로 구성된다. 이때 펄스의 위치는 코드북의 구조에 의해 제한되며 성능도 코드북 구조에 따라 달라진다. 펄스 위치에 제한이 없는 경우에는 펄스 위치를 나타내는데 많은 비트가 사용되어서 비효율적이기 때문에 펄스의 위치를 적절히 제한할 필요가 있다. 일반적으로 펄스의 개수가 많을수록 성능은 향상되나 펄스의 위치를 정하기 위한 계산량이 크게 증가한다. 이런 이유에서 적은 계산량에서 우수한 성능을 나타내는 코드북 구조의 제안이 필요하다.

보편적으로 대수 코드북은 코드북 탐색에 필요한 계산량을 고려하여 영이 아닌 펄스를 4개 이하를 사용하고 있는데 본 논문에서는 합성음의 음질을 향상시키기 위하여 코드북을 구성하는 영이 아닌 펄스를 6개를 사용한다. 펄스의 개수가 증가함에 따라 코드북 탐색에 필요한 계산량도 증가한다.

표 1. 기존의 대수 코드북 (코드북A)

부호	위치
±1	0, 6, 12, 18, 24, 30, 36, 42, 48, 54
±1	1, 7, 13, 19, 25, 31, 37, 43, 49, 55
±1	2, 8, 14, 20, 26, 32, 38, 44, 50, 56
±1	3, 9, 15, 21, 27, 33, 39, 45, 51, 57
±1	4, 10, 16, 22, 28, 34, 40, 46, 52, 58
±1	5, 11, 17, 23, 29, 35, 41, 47, 53, 59

표 2. 제안된 대수 코드북 (코드북B)

부호	위치
±1	0, 6, 12, 18, 24, 30, 36, 42, 48, 54
±1	2, 8, 14, 20, 26, 32, 38, 44, 50, 56
±1	4, 10, 16, 22, 28, 34, 40, 46, 52, 58

(a) 펄스의 위치가 짝수인 경우

부호	위치
±1	1, 7, 13, 19, 25, 31, 37, 43, 49, 55
±1	3, 9, 15, 21, 27, 33, 39, 45, 51, 57
±1	5, 11, 17, 23, 29, 35, 41, 47, 53, 59

(b) 펄스의 위치가 홀수인 경우

표 3. 제안된 대수 코드북 (코드북C)

첫번째 코드북		두번째 코드북	
부호	위치	부호	위치
±1	0, 6, 12, 18, 24, 30, 36, 42, 48, 54	±1	1, 13, 25, 37, 49
±1	2, 8, 14, 20, 26, 32, 38, 44, 50, 56	±1	5, 17, 29, 41, 53
±1	4, 10, 16, 22, 28, 34, 40, 46, 52, 58	±1	9, 21, 33, 45, 57

(a) Grid1 = 0, Grid2 = 0 인 경우

첫번째 코드북		두번째 코드북	
부호	위치	부호	위치
±1	0, 6, 12, 18, 24, 30, 36, 42, 48, 54	±1	3, 15, 27, 39, 51
±1	2, 8, 14, 20, 26, 32, 38, 44, 50, 56	±1	7, 19, 31, 43, 55
±1	4, 10, 16, 22, 28, 34, 40, 46, 52, 58	±1	11, 23, 35, 47, 59

(b) Grid1 = 0, Grid2 = 1 인 경우

첫번째 코드북		두번째 코드북	
부호	위치	부호	위치
±1	1, 7, 13, 19, 25, 31, 37, 43, 49, 55	±1	0, 12, 24, 36, 48
±1	3, 9, 15, 21, 27, 33, 39, 45, 51, 57	±1	4, 16, 28, 40, 52
±1	5, 11, 17, 23, 29, 35, 41, 47, 53, 59	±1	8, 20, 32, 44, 56

(c) Grid1 = 1, Grid2 = 0 인 경우

첫번째 코드북		두번째 코드북	
부호	위치	부호	위치
±1	1, 7, 13, 19, 25, 31, 37, 43, 49, 55	±1	2, 14, 26, 38, 50
±1	3, 9, 15, 21, 27, 33, 39, 45, 51, 57	±1	6, 18, 30, 42, 54
±1	5, 11, 17, 23, 29, 35, 41, 47, 53, 59	±1	10, 22, 34, 46, 58

(d) Grid1 = 1, Grid2 = 1 인 경우

이런 문제점을 해결하기 위하여 두개로 구성된 대수 코드북 구조를 제시한다. 대수 코드북의 값들은 N 개의 영이 아닌 펄스들로 구성되고 각각의 펄스들은 M 개의 위치에서 선택된다. 본 논문에서 사용되는 코드북은 N 이 6이고 M 이 10이다.

표 1은 하나의 대수 코드북으로 구성된 기존의 코드북 구조이며, 표 2는 두개의 대수 코드북으로 구성된 제안된 코드북 구조로서 각각의 코드북은 모두 홀수 또는 짝수의 펄스 위치로 구성된다. 기존의 대수 코드북은 6개의 펄스 위치를 결정하기 위해서 10^6 의 펄스 위치 탐색이 필요하다. 그러나, 제안된 대수 코드북은 3개의 펄스 위치를 먼저 결정한 후 나머지 3개의 펄스 위치를 결정하기 때문에 펄스 위치 탐색 회수가 3×10^3 으로 줄어든다.

먼저 3개의 펄스 위치를 표 2의 (a)와 (b)를 모두 탐색하여 결정하는데 여기서 선택된 펄스의 위치는 모두 홀수이거나 짝수이다. 앞에서 선택된 3개의 펄스 위치가 홀수(또는 짝수)이면 나머지 3개의 펄스 위치는 짝수(또는 홀수)중에서 선택된다.

표 1의 코드북을 사용하는 경우에는 6개의 펄스를 전송하기 위하여 필요한 비트수는 20 비트이며 각각의 펄스들의 부호를 위해서 1 비트씩 필요하기 때문에 전체 펄스 정보에 필요한 비트수는 26 비트이다. 표 2의 코드북을 사용하는 경우에는 첫번째 3개의 펄스 위치가 홀수인지 짝수인지를 결정하기 위하여 1 비트가 더 필요하기 때문에 전체 27 비트가 필요하다. 코드북에 사용되는 비트수를 줄이기 위하여 표 3과 같은 코드북을 제안한다. 표 3에서는 첫번째 3개의 펄스를 결정하는 코드북은 M 이 10이고, 나머지 3개의 펄스를 결정하는 코드북

은 M 이 5이다. 이 경우 전체 펄스 위치를 위한 비트수는 17 비트이고 부호를 나타내기 위해 6 비트가 필요하다. 그리고, 4개의 코드북 중에서 하나가 선택되기 때문에 Grid 정보가 2 비트가 더 필요하여 전체 코드북을 위한 비트수는 25이다

그림 2는 제안된 대수 코드북 탐색 과정을 블록도로 나타낸 것으로 코드북 탐색 과정은 다음과 같다. 먼저 첫번째 코드북에서 3개의 펄스 위치를 다음과 같이 결정한다. 첫번째 코드북에 의해서 생성된 코드워드, v_{k_1} ,는 가중 합성 필터를 통과하여 합성 여기 신호를 생성한다. 합성 여기 신호에 코드북 이득을 곱하여 합성 신호를 생성한다. 가중 입력 음성 신호와 합성 신호간의 차를 구한 후 이 값의 평균 제곱근 오차 (mean square error)를 최소화하는 색인, k_1 ,과 이득, G_1 ,을 구한다. 평균 제곱근 오차, E_{k_1} ,는 다음과 같다.

$$E_{k_1} = \|r - G_1 H v_{k_1}\|^2 \tag{4}$$

여기서 G_1 은 첫번째 코드북 이득이며, v_{k_1} 은 색인 k_1 에 해당하는 대수 코드워드이다. 식 (4)를 최소화 하는 색인 k_1 '는 식 (5)의 τ_{k_1} 을 최대화하는 값으로 선택된다.

$$\tau_{k_1} = \frac{C_{k_1}^2}{\epsilon_{k_1}} = \frac{(d^T v_{k_1})^2}{v_{k_1}^T \Phi v_{k_1}} \tag{5}$$

여기서 $d = H^T r$ 는 목적 벡터 입력과 영상태 응답간의 상관도를 나타내는 벡터이고, $\Phi = H^T H$ 는 영상태 응답의 공분산 행렬이다. 코드북을 탐색하기 전에 벡터 d 와 행렬 Φ 를 먼저 계산한다. 벡터 d 의 값들은 다음과 같이 계산되고,

$$d(j) = \sum_{n=j}^{59} r(n) \cdot h(n-j), \quad 0 \leq j \leq 59, \tag{6}$$

행렬 Φ 의 값들은 다음과 같이 계산된다.

$$\Phi(i, j) = \sum_{n=j}^{59} h(n-i) \cdot h(n-j), \quad j \geq i, \quad 0 \leq i \leq 59. \tag{7}$$

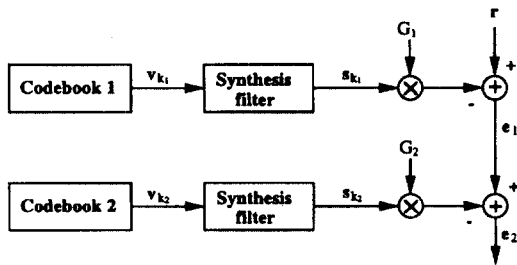


그림 2. 제안된 대수 코드북 탐색도

제안된 대수 코드북 구조는 코드북 탐색 과정을 매우 빨리 수행할 수 있는데, 이는 여기 벡터 (excitation vector), v_{k_i} ,가 단지 3개의 영이 아닌 펄스로 구성되어 있기 때문이다. 식 (4)의 상관도는 다음과 같이 나타난다.

$$C_{k_i} = \alpha_0 d(m_0) + \alpha_1 d(m_1) + \alpha_2 d(m_2) \quad (8)$$

여기서, m_i 와 a_i 는 각각 i 번째 펄스의 위치와 부호 (\pm)이다. 식 (4)의 짝수 위치에 있는 펄스들로 구성된 코드워드의 에너지 (energy)는 다음과 같이 나타난다.

$$\begin{aligned} \epsilon_{k_i} = & \Phi(m_0, m_0) \\ & + \Phi(m_1, m_1) + 2\alpha_0 \alpha_1 \Phi(m_0, m_1) \\ & + \Phi(m_2, m_2) + 2[\alpha_0 \alpha_2 \Phi(m_0, m_2) + \alpha_1 \alpha_2 \Phi(m_1, m_2)]. \end{aligned} \quad (9)$$

홀수 위치에 있는 펄스들로 구성된 코드워드의 에너지는 각각의 펄스를 시간 축으로 한 위치씩 앞으로 이동한 코드워드의 에너지로 대체한다.

첫번째 코드북에 의한 최적의 코드워드, v_{k_i} ,가 가중 합성 필터를 통과한 신호를 s_{k_i} 라고 하면, 첫번째 코드북 이득을 다음과 같이 구할 수 있다.

$$G_1 = \frac{C_{\max}}{E_{\max}}. \quad (10)$$

여기서 C_{\max} 와 E_{\max} 는 다음과 같다.

$$\begin{aligned} C_{\max} &= \sum_{n=0}^{59} r(n) \cdot s_{k_i}^*(n), \\ E_{\max} &= \sum_{n=0}^{59} s_{k_i}^*(n) \cdot s_{k_i}^*(n). \end{aligned} \quad (11)$$

코드북 이득은 로그 양자화기 (logarithmic quantizer)에 의해 양자화되는데, 양자화된 코드북 이득을 \hat{G}_1 으로 나타낸다. 코드북 이득에 사용되는 양자화기는 3.2 dB 간격으로 24단계로 구성

되어 있다.

첫번째 코드북이 결정된 후 두번째 코드북의 탐색이 이루어지는데, 첫번째 코드북이 홀수 펄스 위치에서 선택되면 두번째 코드북은 짝수 펄스 위치에서 선택하고, 첫번째 코드북이 짝수 펄스 위치에서 선택되면 두번째 코드북은 홀수 펄스 위치에서 선택된다. 위에서 구한 색인, k_i ,에 해당하는 오차 신호, e_i ,를 다음과 같이 나타낼 수 있다.

$$e_i = r - \hat{G}_1 \cdot s_{k_i} \quad (12)$$

이 오차 신호를 목적 신호로 하여 두번째 코드북 탐색이 이루어진다. 그림 3은 첫번째 코드북 이득에 대한 두번째 코드북 이득의 관계를 나타낸 것으로 상관성이 많음을 알 수 있다. 이러한 상관관계를 이용하여 두번째 코드북의 이득, G_2 ,을 양자화된 첫번째 코드북의 이득, \hat{G}_1 ,과 코드북 이득 비례상수를 사용하여 식 (13)과 같이 정의하였다.

$$G_2 = g_{12} \hat{G}_1. \quad (13)$$

여기서 g_{12} 는 코드북 이득 비례상수로서 0.6이다. 코드북 이득 비례상수는 첫번째 코드북 이득과 두번째 코드북 이득간의 비 (G_2 / G_1)를 구하여 많은 실험값들의 평균값으로 선택하였다. 두번째 코드

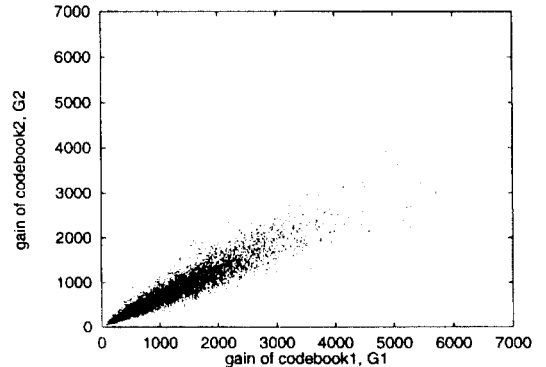


그림 3. 첫번째 코드북 이득에 대한 두번째 코드북 이득의 관계

북의 이득은 식 (13)과 같이 결정되기 때문에 두번째 코드북의 이득을 전송할 필요가 없다.

두번째 코드북 탐색에서는 다음 식을 최소화하는 색인, k_2' 을 결정한다.

$$E_{k_2} = \|e_1 - G_2 H v_{k_2}\|^2. \quad (14)$$

여기서 v_{k_2} 는 두번째 코드북에 의해서 생성된 코드워드이고, 식 (14)를 최소화하는 색인 k_2' 는 식 (15)의 τ_{k_2} 를 최대화하는 값으로 선택된다.

$$\begin{aligned} \tau_{k_2} &= 2G_2 C_{k_2} - G_2^2 \epsilon_{k_2} \\ &= 2G_2 (d'^T v_{k_2}) - G_2^2 (v_{k_2}^T \Phi v_{k_2}). \end{aligned} \quad (15)$$

여기서 벡터 d' 의 값들은 다음과 같이 계산되고,

$$d'(j) = \sum_{n=j}^{59} e_1(n) \cdot h(n-j), \quad 0 \leq j \leq 59, \quad (16)$$

식 (15)의 C_{k_2} 는 다음과 같이 계산된다.

$$C_{k_2} = \alpha_0 d'(m_0) + \alpha_1 d'(m_1) + \alpha_2 d'(m_2). \quad (17)$$

ϵ_{k_2} 는 첫번째 코드북 탐색 과정에서 사용한 식 (9)와 같이 계산된다.

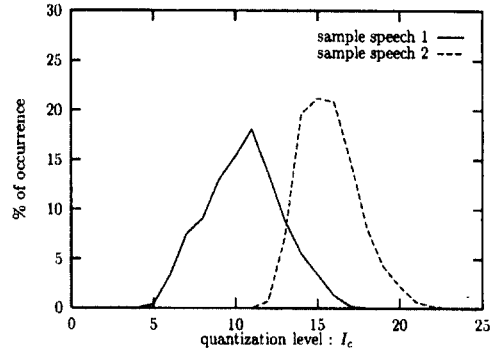
복호기에서는 전송된 펄스 정보로부터 다음과 같이 여기 신호, ex ,를 생성한다.

$$ex = G_1 (v_{k_1}' + g_{12} v_{k_2}'). \quad (18)$$

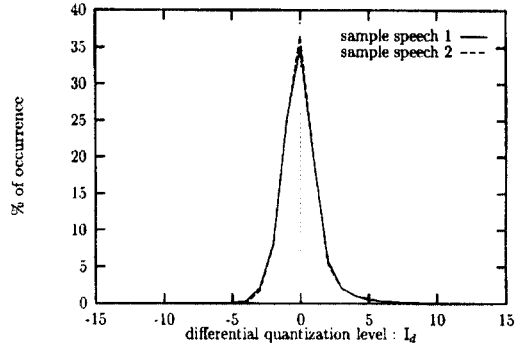
3.2 고정 코드북 이득의 차분 양자화 (differential quantization)

고정 코드북 이득은 입력 신호와 잔류 신호의 크기에 따라 여러 형태로 나타난다. 그림 4에서 (a)는 고정 코드북 이득을 24단계로 스칼라 양자화한 것이고, (b)는 현재 부프레임의 이득과 이전 부프레임의 이득간의 차분 정보를 나타낸 것이다.

표본 음성1 (sample speech1)은 입력 신호의 크기



(a)



(b)

그림 4. 고정 코드북 이득의 양자화

(a) 양자화 단계에 대한 이득의 분포

(b) 인접 부프레임간의 차분 이득의 분포

가 작은 경우이고 표본 음성2 (sample speech2)는 입력 신호의 크기가 큰 경우이다. 이런 서로 다른 입력 신호에 대해서 현재 부프레임의 이득과 이전 부프레임의 이득간의 차분 정보를 구해보면 그림 4의 (b)와 같이 '0' 주위에 가장 많이 분포함을 알 수 있다. 이러한 사실을 이용하여 고정 코드북 이득을 그대로 양자화해서 전송하는 대신 고정 코드북 이득의 차분 정보를 전송하면 성능 저하 없이 양자화에 필요한 비트수를 줄일 수 있다.

현재 부프레임 고정 코드북 이득에 해당하는 색인을 I_c 라고 하고, 이전 부프레임 고정 코드북 이득에 해당하는 색인을 I_b 라고 하면, 이득의 차분 정보,

I_d 는 다음과 같이 정의된다.

$$I_d = \begin{cases} I_c - I_p, & -I_{\min} \leq I_c - I_p \leq I_{\max} \\ -I_{\min}, & I_c - I_p < -I_{\min} \\ +I_{\max}, & I_c - I_p > I_{\max} \end{cases} \quad (19)$$

여기서 I_{\max} 은 양수의 값을 갖는 차분 정보 중 최대값을 나타내고, I_{\min} 은 음수의 값을 갖는 차분 정보 중에서 절대값이 최대인 값을 나타낸다. 즉, 차분 정보는 $-I_{\min}$ 과 I_{\max} 사이에 존재하게 되며, 부호기에서는 고정 코드북 이득을 전송하는 대신 차분 정보를 전송하고, 복호기에서는 이전 부프레임의 이득 정보와 전송된 차분 정보를 이용해서 현재 부프레임의 이득 정보를 다음과 같이 복원한다.

$$I_c = I_p + I_d \quad (20)$$

IV. 성능 분석

4.1 환경

본 논문에서는 성능을 분석하기 위하여 8 kHz로 표본화 된 16 비트 음성 신호를 사용하였다. 성능 분석의 신뢰도를 높이기 위해 길이가 1분 이상인 남녀 음성 10개를 사용하였으며, 성능 분석 결과로 나타난 값들은 남녀 음성 10개에 대한 평균값이다. 음성 부호화기의 성능 평가를 위한 객관적인 성능 척도 (performance measure)로는 $segSNR$ 을 사용하였다. $segSNR$ 은 각 부프레임의 SNR 을 평균한 값으로 다음과 같다 [10].

$$segSNR = \frac{1}{N_s} \sum_{i=1}^{N_s} SNR_i, \quad (21)$$

$$SNR_i = 10 \log \left(\frac{\|s\|^2}{\|s - \hat{s}\|^2} \right), \quad 1 \leq i \leq N_s.$$

여기서 s 는 입력 음성 벡터이고 \hat{s} 는 합성된 음성 벡터이며, N_s 는 부프레임 개수이다.

채널 오류에 의한 성능 저하를 알아보기 위하여 프레임 손실이 있는 경우에 대하여 성능을 분석하였다. 프레임 손실의 빈도는 랜덤 (random)하게 발생하며, 복호화 과정에서 어느 한 프레임이 손실되었다고 판별되면 그 프레임의 정보는 이전 프레임의 정보로 대체된다 [11].

4.2 제안된 대수 코드북의 성능 분석

코드북의 구조에 따라 계산량과 음질이 달라지는데, 여기서는 4가지 경우의 코드북에 대해서 비교하였다. 각각의 코드북은 다음과 같다.

코드북A: 표 1 코드북

코드북B: 표 2 코드북

코드북C: 표 3 코드북

코드북D: 표 2에서 첫번째 코드북으로만 구성

표 4. 코드북 구조에 따른 비트 할당과 성능

	코드북A	코드북B	코드북C	코드북D
코드북에 필요한 비트수	26	27	25	14
펄스위치 탐색회수	10^6	$3 \cdot 10^3$	$2(10^3 + 5^3)$	$2 \cdot 10^3$
segSNR [dB]	12.379	12.443	12.126	10.062

코드북A는 하나의 코드북에서 6개의 펄스를 결정하는 구조이고, 코드북B와 코드북C는 첫번째 코드북에서 3개의 펄스를 결정하고 두번째 코드북에서 나머지 3개의 펄스를 결정하는 구조이다. 코드북D는 하나의 코드북에서 3개의 펄스를 결정하는 구조이다.

코드북A와 코드북B는 펄스의 개수와 위치는 같으나 코드북 구조에 따라 탐색 방법이 다르다. 코드북A의 경우는 6개의 펄스를 동시에 탐색하여 6개의 펄스 모두를 최적화하는 이득을 구하기 때문에 펄스 탐색 회수가 매우 많다. 하지만 코드북B는 최적의 펄스 3개를 탐색한 후, 나머지 3개의 펄스를 탐색하기 때문에 코드북A에 비하여 펄스 탐색 회수가 상당히 줄어든다. 음질에서도 코드북B가 코드북A보다 조금 좋게 나타나는데 그 이유는 다음과 같다. 6개의 펄스를 모두 최적화하여 코

드북 이득을 정하는 것보다 3개의 펄스에 대한 이득을 먼저 구한 후 식 (13)을 이용하여 나머지 3개의 펄스에 대한 이득을 구하는 것이 평균 제공근 오차를 더 작게 하기 때문이다. 코드북B가 코드북A보다 한 비트가 더 많다는 단점은 있으나 음질면에서 조금 우수하고 특히 계산량이 매우 작기 때문에 실시간 구현에 용이하다. 코드북C의 경우는 음질면에서 다소 떨어지지만 계산량과 할당되는 비트수가 작다는 장점이 있다. 코드북D는 첫번째 코드북만을 사용한 것으로, 두번째 코드북에 의한 효과를 알아보기 위해서 비교하였다. 두번째 코드북을 추가로 사용하면 첫번째 코드북만을 사용하는 경우에 비하여 음질을 높일 수 있음을 알 수 있다.

4.3 고정 코드북 이득의 차분 양자화 방법에 대한 성능 분석

그림 4에서 알 수 있듯이 현재 부프레임에 대한 이득과 이전 부프레임에 대한 이득간의 변화는 크지 않다. 이를 바탕으로 차분 정보, I_n 의 범위에 따라 성능의 변화를 살펴보았다. 차분 정보는 $-I_{min}$ 과 I_{max} 사이 값으로 제한된다. 그림 5는 I_{max} 와 I_{min} 의 값을 변화시키면서 성능을 분석한 것으로, I_{max} 와 I_{min} 이 증가함에 따라 $segSNR$ 도 증가하지만 어느 한계 이상의 값에서는 거의 변화가 없음을 알 수 있다. 본 논문에서는 I_{max} 와 I_{min} 을 각각 3과 2로 선택하여 차분 정보를 6단계로 양자화하여 전송함으로써 기존의 24단계로 양자화하여 전송하는 경우보다 전송 비트를 부프레임마다 2 비트씩 줄일 수 있다.

채널 오류가 있는 경우에 대하여 차분 양자화 방법의 성능을 분석하기 위하여 차분 양자화를 적용하는 부프레임의 범위를 다음과 같이 3가지 경우로 나누었다.

- 첫번째 경우 : 차분 양자화 방식을 적용하지 않고 기존의 양자화 방식을 적용한 경우
- 두번째 경우 : 첫번째 부프레임은 기존의 방식을 적용하고, 두번째, 세번째, 그리고 네번째 부프레임은 차분 양자화 방식을 적용한 경우
- 세번째 경우 : 모든 부프레임에 차분 양자화 방법을 적용한 경우

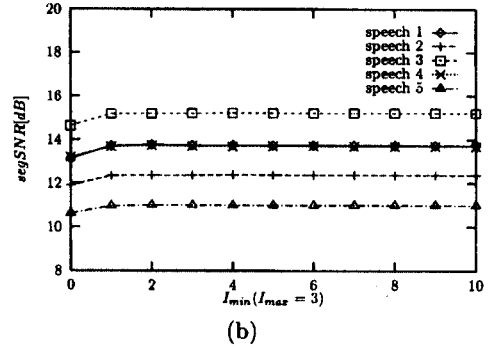
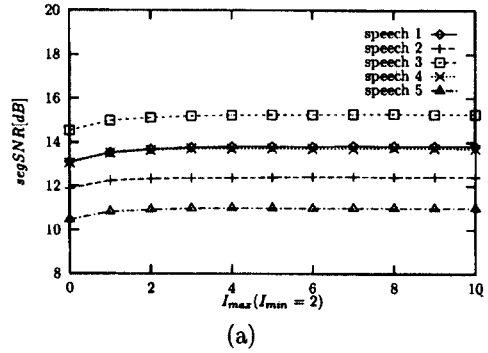


그림 5. 고정 코드북 이득의 차분 양자화 범위에 따른 성능 분석
 (a) $I_{min}=2$ 인 경우 I_{max} 의 변화에 대한 성능 분석
 (b) $I_{max}=3$ 인 경우 I_{min} 의 변화에 대한 성능 분석

수신된 신호는 프레임 단위로 오류를 판정하게 되는데, 오류가 있는 프레임을 오류 프레임이라고 한다. 오류 프레임이 발생할 경우 복호기에서 합성된 음성의 품질을 향상시키기 위하여 오류 프레임에 대한 보간 처리를 수행한다. 본 논문에서는 오류 프레임이 발생하면 오류 프레임의 정보를 삭제하고 오류 프레임의 정보를 이전의 프레임 정보로 대체하는 보간 방법을 사용하였다. 현재 프레임의 정보는 이전 프레임의 정보와 매우 비슷하기 때문에 이러한 보간 방법이 매우 유용하다.

그림 6은 프레임 손실 비율에 따른 성능을 분석한 것으로 프레임 손실이 없을 때에는 $segSNR$ 값이 모든 경우에 대하여 비슷한 값을 가지므로 차분 양

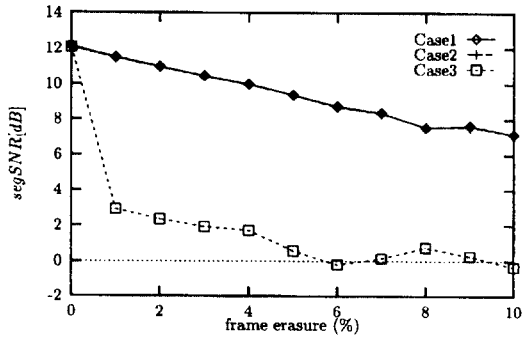


그림 6. 프레임 손실에 따른 성능 분석

자화를 사용할 경우에도 성능 저하가 없음을 알 수 있다. 그러나, 프레임 손실이 있을 때에는 첫번째 경우와 두번째 경우는 비슷한 성능을 나타내지만 세번째 경우는 성능 저하가 많이 일어남을 알 수 있다. 모든 부프레임에 차분 양자화를 적용하는 경우에는 오류가 발생하면 그 오류에 대한 효과가 누적되어 오류가 없는 프레임에도 영향을 주어 성능이 많이 저하된다. 본 논문에서는 성능 저하가 거의 없으면서 전송 비트를 많이 줄일 수 있는 두번째 경우를 선택하였다. 음성 부호화기에서는 채널 오류를 프레임 단위로 판별하기 때문에 첫번째 부 프레임의 이득에는 차분 양자화 방법을 적용하지 않고 나머지 세개의 부프레임 이득에 차분 양자화 방법을 적용함으로써 채널 오류의 파급 효과를 해결하였다.

4.4 기존의 부호화기와 제안된 부호화기의 성능 비교

차분 양자화 기법과 제안된 대수 코드북을 모두 사용한 경우를 기존의 부호화기인 ITU-T 권고안 G.723.1 (MP-MLQ : 6.3kbps)과 비교하였다. 차분 양자화 방법은 두번째 경우를 선택하였으며, I_{max} 과 I_{min} 은 각각 3과 2로 정하였다. 코드북 구조는 코드북B와 코드북C 경우를 선택하여 비교하였다. 코드북 이득 비례상수는 0.6으로 정하였다. 표 5는 여러 부호화기에 대한 $segSNR$ 을 나타낸 것으로 제

안된 방법 모두 기존의 G.723.1 부호화기보다 성능이 향상되었음을 알 수 있다.

표 5. 기존의 부호화기와 성능 비교

	G.723.1	코드북B	코드북C
전송률 [kbps]	6.3	6.4	6.13
segSNR [dB]	12.046	12.413	12.073

제안된 코드북 탐색에 필요한 시간을 ITU-T 권고안 G.723.1 (MP-MLQ : 6.3kbps)와 비교하였다. C언어로 짠 프로그램을 Sun SPARCstation 20에서 실행하여 얻은 결과로서, 코드북 탐색에 걸린 시간과 전체 부호화에 걸린 실제 시간을 프레임 단위로 나타내었다. 여기서 프레임은 30msec로 구성되어 있다.

표 6. 한 프레임 부호화에 필요한 시간

	G.723.1 (6.3 kbps)	코드북B (6.4 kbps)	코드북C (6.13 kbps)
코드북 탐색 시간 [msec/frame]	19.8	7.9	6.1
전체 부호화 시간 [msec/frame]	39.8	27.9	26.1

표 6에서 보는 바와 같이 제안된 코드북 탐색에 걸리는 시간이 기존의 ITU-T 권고안G.723.1 (MP-MLQ : 6.3kbps)에 비해서 상당히 줄어들었음을 알 수 있다. 기존의 부호화기는 Sun SPARCstation 20 환경에서 30msec/frame을 초과함으로 실시간 처리를 할 수 없지만, 제안된 코드북을 이용하면 실시간 처리가 가능하다.

V. 결 론

본 논문에서는 코드북 탐색에 필요한 계산량을 감소시킬 수 있는 대수 코드북 구조를 제안하였으며, 고정 코드북 이득을 차분 양자화하여 고정 코드북에 할당되는 비트수를 줄였다.

ACELP 음성 부호화기에서는 영이 아닌 펄스의 수가 많을수록 음질이 좋지만, 펄스 위치 정보를 나타내기 위한 비트수가 증가하고 펄스 위치 탐색에 필요한 계산량이 급격히 증가한다는 단점이 있

다. 본 논문에서 제안한 대수 코드북 구조에서는 펄스의 위치를 두개의 코드북에 의해서 결정하였다. 각각의 코드북에서 전체 펄스의 절반만을 결정하기 때문에 탐색에 필요한 계산량을 많이 감소시킬 수 있었다. 그리고, 첫번째 코드북 이득과 두번째 코드북 이득간의 상관관계를 이용하여 두번째 코드북 이득을 양자화된 첫번째 코드북 이득에 코드북 이득 비례상수를 곱하여 추정함으로써, 두번째 코드북 이득을 전송하는데 소요되는 비트를 절약하였다. 두번째 코드북 이득을 추정함에 따른 음질 저하를 분석한 결과 두번째 코드북 이득을 전송하는 경우에 비하여 음질 저하가 거의 없음을 알 수 있었다.

현재 부프레임의 고정 코드북 이득과 이전 부프레임의 고정 코드북 이득간의 변화가 크지 않다는 점을 이용하여 고정 코드북 이득을 차분 양자화하여 코드북 이득에 할당되는 비트수를 줄였다. 채널 오류에 의한 영향을 알아보기 위해 프레임 손실이 없는 경우와 프레임 손실이 있는 경우로 나누어 성능을 분석하였다. 모든 부프레임의 이득을 차분 양자화하면 채널 오류가 있는 경우에 성능 저하가 많음을 알 수 있었으며, 이러한 채널 오류에 의한 성능 저하를 막기 위한 방법을 제시하였다. ACELP 음성 부호화기에서는 프레임 단위로 채널 오류를 판별하여 채널 오류가 발생하면 그 프레임의 정보를 이전 프레임의 정보로 대체했다. 이 경우 첫번째 부프레임을 제외한 나머지 부프레임에만 차분 양자화를 적용하면, 채널 오류의 전파에 의한 성능 저하가 없음을 알 수 있었다.

제안된 방법을 적용하는 경우는 기존의 부호화기에 비해 음질 저하 없이 계산량을 감소시키면서 낮은 전송률로 부호화할 수 있었다. 본 논문에서 제안된 방법은 음성 부호화기의 실시간 구현에 용이하며, 낮은 전송률의 부호화기에 적용할 수 있을 것으로 기대된다.

참 고 문 헌

1. L. R. Rabiner and R. W. Schafer, Digital

Processing of Speech Signals. Englewood Cliffs, New Jersey: Prentice-Hall, 1978.

2. N. S. Jayant and P. Noll, Digital Coding of Waveforms. Englewood Cliffs, New Jersey: Prentice-Hall, 1984.

3. T. E. Tremain, "The government standard linear predictive coding algorithm : LPC-10," Speech Technol., pp. 40--49, Apr. 1982.

4. J. D. Deller, Jr, J. G. Proakis, J. H. L. Hansen, Discrete-Time Processing of Speech Signals, New York: Macmillan Publishing Company, 1993.

5. J. Makhoul, "Linear prediction : A tutorial review," Proc. IEEE, vol. 63, pp. 561--580, Apr. 1975.

6. M. R. Schroeder, "Code-excited linear prediction (CELP): high quality speech at low rates," Proc. ICASSP-85, pp. 937--940, 1985.

7. P. Kroon and E. F. Deprettere, "A class of analysis-by-synthesis predictive coder for high quality speech coding at rates between 4.8 and 16 kbits/s," IEEE J. Select. Areas Commun., vol. 6, no. 2, pp. 353--363, Feb. 1988.

8. EIA/TIA-PN2398 (IS-54), "The 8kbits/s VSELP algorithm," 1989.

9. Draft Recommendation G.723, "Dual rate speech coder for multimedia communications transmitting at 5.3 and 6.3 kbit/s," Telecommunication Standardization Sector, ITU-T.

10. A. S. Spanias, "Speech coding : A tutorial review," Proc. IEEE, vol.82, pp. 1541--1582, Oct. 1994.

11. TIA/EIA Interim Standard IS-96, "Speech service option standard for wideband spread spectrum digital cellular system", Telecommunication Industry Association, Apr. 1994.



김 상 조 (Sang Jo Kim) 정회원
1973년 5월 15일생
1995년 2월 : 경북대학교 전자공
학과 학사
1998년 2월 : 한국과학기술원 전
기 및 전자공학과
졸업 석사

1998년 2월 - 현재 : 한국통신 가입자망연구소 전임
연구원

<관심 분야> 음성 부호화기, 음성 인식, 초고속 가
입자망



서 방 원 (Bang Won Seo) 정회원
1975년 2월 21일생
1997년 2월 : 한국과학기술원 전
기 및 전자공학과 학사
1997년 3월 - 현재 : 한국과학기
술원 전기 및 전자
공학과 석사과정

<관심 분야> 음성 부호화기, CDMA Multi-user
Detector



김 형 명 (Hyung Myung Kim) 정회원
1952년 10월 24일생
1974년 2월 : 서울대학교 공학사
1982년 4월 : 미국 Pittsburgh 대학
전기공학과 석사
1985년 12월 : 미국 Pittsburgh 대
학 전기공학과 공
학박사

1986년 4월 - 1992년 8월 : 한국과학기술원 전기 및
전자공학과 조교수

1992년 9월 - 현재 : 한국과학기술원 전기 및 전자
공학과 부교수

<관심 분야> 디지털 신호와 영상처리 다차원시스
템 이론 비디오신호 전송통신 이론이
동 통신 기술 분야