

에너지정합방법을 이용한 CS-ACELP 음성부호화기의 잡음특성 개선

정희원 이 행 우*, 박 기 영**, 김 중 교***

Improving Noise Performance of CS-ACELP Coder by Energy Matching Method

Haeng-woo Lee*, Ki-young Park**, Chong-kyo Kim*** *Regular Members*

요 약

본 논문은 여러 배경잡음에서 높은 성능을 나타내는 8 kbps CS-ACELP 부호화기의 음질개선방법에 관한 것이다. 파형정합방법을 이용하는 기존의 CELP 부호화기가 순수한 음성신호에 대해서는 우수한 음질을 제공하지만, 배경잡음에 대해서는 큰 음질저하를 가져온다. 따라서 본 논문에서는 음성신호뿐만 아니라 이러한 배경잡음에 대해서도 우수한 음질을 달성하는 에너지정합방법을 소개한다. 이 방법은 부가적인 비트를 요구하지 않으면서 효과적으로 음질을 개선한다. 실험 결과, 객관적 평가에서는 MSEGSR이 0.01~1.1 dB 증가하였고, 주관적 평가에서는 MOS 점수가 평균 0.2, 최대 0.5 점까지 증가하는 것으로 나타났다. 따라서 PPS 알고리즘을 적용한 부호화기의 음질이 최대 12%까지 향상되는 것으로 확인되었다.

핵심용어 : 음성부호화기, CS-ACELP 부호화기, 음질개선

ABSTRACT

This paper describes a method for growing up the speech quality of 8kbps CS-ACELP speech coder under various background noises. While existing CELP coder using waveform matching method provides high quality for clean speech signals, it has significant quality loss for background noises. Therefore, this paper proposes the energy matching method with high quality for background noises as well as speech signals. This method effectively improves the speech quality without requiring any additive bits. As a result of experiments, in the objective estimation method, MSEGSR were increased by 0.01~1.1 dB. Also, in the subjective estimation method, MOS was increased by 0.2 averagely and by 0.5 maximally. Therefore, it was proved that this algorithm achieves the improvement of speech quality by 12% maximally.

Keywords : Speech coder, CS-ACELP coder, Improvement of speech quality

I. 서 론

음성압축의 필요성 때문에 음성부호화기는 오늘날 음성통신이나 음성저장시스템 등 여러 분야에서 광범위하게 활용되고 있다. 특히, 휴대폰과 같은 디지털통신에서 음성부호화기의 사용은 필수적이다.

일반적으로 유선전화나 이동전화는 배경잡음 환경에서 사용되므로 음질개선이 매우 중요한 요소이다. 이동전화의 경우에도 유선전화에서의 성능에 필적하는 음질이 요구되며, 사무실, 승용차, 음식점 등 여러 장소에서 휴대폰을 일상적으로 사용하면서 음질에 대한 요구수준이 크게 높아지고 있다.

* 남서울대학교 전자정보통신공학부(hwlee@nsu.ac.kr),

** 전주공업대학 정보통신과(kypark@jtc.ac.kr),

*** 전북대학교 전자정보공학부(ckkim@moak.chonbuk.ac.kr)

논문번호 : 00442-1128, 접수일자 : 2000년 11월 28일

8 kbps의 전송률에서 현재의 고음질 음성부호화기는 보통 선형예측 Abs(analysis-by-synthesis) 기술^[1]을 사용하고 있다. 이러한 기술을 사용하여 최근에 표준화된 부호화기로는 G.729^{[2][3][4]} CS-ACELP(conjugate-structure algebraic-CELP) 및 G.729A^{[5][6]} CS-ACELP가 있다. CS-ACELP 음성부호화기의 가장 큰 특징은 이득 양자화를 위한 공역 코드북구조와 고정코드북 여기신호의 대수적 표현이다. 이러한 구조는 메모리와 계산량을 감소시키고 전송에러에 강한 특성을 부여함으로써 부호화 및 복호화에 소요되는 지연시간을 25 ms 이하로 줄이고, 8 kbits/s의 전송속도에서도 32 kbits/s ADPCM에 상당하는 음질을 달성한다. 또한, CS-ACELP 부호화기는 비대칭 창을 이용한 LPC 분석으로 look-ahead 구간을 감소시키며, 새로운 방식의 이득 조정과정 등 많은 최신 기술들을 도입하였다. 이 부호화기들은 음성에 대해 거의 유선전화 수준의 고음질을 제공하지만, 배경잡음이 있는 환경에서는 상당한 음질저하를 가져온다. 따라서, 이와 같은 음질저하 문제를 개선하지 않고는 신뢰성 있는 통화품질을 기대하기 어렵다. 본 논문에서는 에너지정합방법을 이용하여 음질을 크게 향상시키는 새로운 음질개선방법을 제안하고자 한다.

논문의 구성은 다음과 같다. II장에서는 G.729A CS-ACELP 음성부호화기에 대해 살펴보고, III장에서는 본 논문에서 제안하는 에너지정합방법에 대하여 설명한다. 그리고 IV장에서는 모의실험 및 결과를 분석하고, 끝으로 V장에서 결론을 맺는다.

II. CS-ACELP 음성부호화기

G.729A는 8 kHz로 샘플링된 16-bit PCM 신호를 입력으로 하고, 그 기본구조는 G.729와 같이 10 ms(80 샘플) 음성 프레임에서 동작하는 8 kbps CS-ACELP 알고리즘에 바탕을 두고 있다. 또한 G.729A로 부호화된 비트열은 G.729로 복호화될 수 있으며, 그 반대의 경우도 가능하다. G.729A는 DSP를 이용한 실시간 구현이 용이하도록 G.729에 몇 가지 고속 알고리즘을 도입함으로써 계산량을 약 50% 정도 감소시켰으며, 음질은 배경잡음을 제외한 대부분의 경우에 G.729에 필적하는 음질을 갖는다. CS-ACELP 부호화기는 그림 1과 같이 기본적으로 합성을 통해서 분석을 하는 Abs 방식의 CELP 부호화기로서, 한 프레임 내의 각 파라미터에 대한 비트 할당은 표 1과 같다.

8 kbps CS-ACELP는 10 ms 크기의 프레임용 가지며, 이것은 다시 5 ms의 부프레임으로 분할된다. 먼저, 선형예측계수를 산출하기 위해 비대칭 창 함수를 사용함으로써 lookahead 구간을 줄여 총 알고리즘 지연시간을 감소시키며, 전처리과정에서 고역통과필터를 사용하여 입력신호의 직류성분을 제거하고, 오버플로우 현상을 방지하기 위해 입력신호의 크기를 1/2로 줄인다.

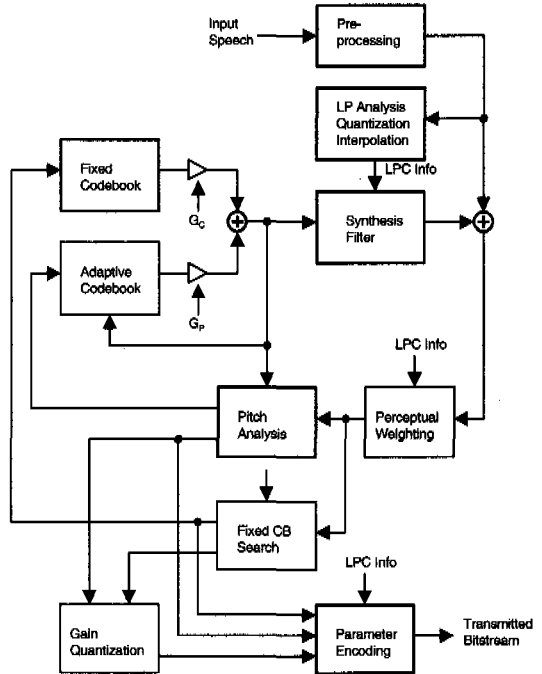


그림 1. CS-ACELP 부호화기의 블록도

표 1. 8 kbps CS-ACELP 부호화기의 비트할당

파라미터		부프레임1	부프레임2	프레임
선스펙트럼 상	이동평균예측			1
	제1단벡터			7
	제2단하위벡터			5
	제2단상위벡터			5
적용코드북	피치지연	8	5	13
	패리티	1		1
고정코드북	색인	13	13	26
	필스부호	4	4	8
이득코드북	제1단	3	3	6
	제2단	4	4	8
합 계		33	29	80

분석과정에서는 Levinson-Durbin 알고리즘을 이용하여 구한 10차의 선형예측계수를 양자화 및 보간 특성이 우수한 LSP(line spectral Pair) 계수로 변환하고, 이 값과 4차의 이동평균 예측기를 통해 예측된 값과의 차이는 2단계 벡터양자화 및 분리 벡터양자화를 거쳐 부호화된다. 이때 서로 다른 계수값을 갖는 2개의 이동평균 예측기가 사용되며, 그 중 각각의 예측기에 의해 구해진 예측값과 LSF 계수값 사이의 가중된 평균자승오차를 최소로 하는 것을 선택한다. 이렇게 양자화된 LSP 계수는 선형 예측 합성필터를 구성하기 위해 부프레임 단위로 보관된 후, 다시 선형예측계수로 복원된다.

피치지연은 프레임 단위의 개회로 검색과 이어지는 부프레임 단위의 폐회로 검색과정을 순차적으로 수행해 적응코드북의 인덱스를 선택한다. 첫 번째 부프레임에서 적응코드북 검색은 개회로 피치지연을 중심으로 몇 개의 지연값에 대해서만 검색을 실시하여 최적의 실수 피치지연을 구한다. 그리고 두 번째 부프레임에서는 검색의 효율성을 높이기 위해 첫 번째 부프레임에서 구한 피치지연값을 기준으로 삼는다. 적응코드북 검색에 의해 입력신호로부터 주기적인 피치성분이 제거되면 여기신호 산출의 마지막 단계인 고정코드북 검색이 이루어진다. 고정코드북은 크기가 1인 다중필스의 부호와 위치로 구성되어 있으며, 각 필스는 트랙(track)이라 불리는 제한된 위치에서만 검색된다. 적응코드북과 고정코드북의 이득은 공액구조를 갖는 2개의 코드북을 사용하여 폐회로 방식으로 양자화된다. 이득코드북 검색은 효율성을 기하기 위해 각 코드북을 단독으로 개회로 검색하여 몇 개의 후보를 정하고, 이 후보들에 대해서만 폐회로 검색을 수행한다. 이러한 방법은 완전검색에 비하여 상당한 계산량의 감소를 가져온다.

복호기에서는 복원된 합성음의 음질향상을 위해 후처리과정을 거친다. 후처리과정에서는 적응 후단 필터링과 교역통과필터링 및 신호크기 조정 등을 수행한다.

G.729A의 주요 부분은 LSP 변환, 인지가중필터, 개회로 피치검색, 적응코드북 검색, 고정코드북 검색 및 후처리과정 등이며, 구체적으로 살펴보면 다음과 같다.

① LSP 변환

G.729에서는 0에서 239까지 균등하게 분할된 60개의 지점에서 Chebyshev 다항식의 값을 계산하고

부호의 변환이 있으면 그 구간을 다시 4개로 나누어 보다 정밀하게 근을 검색한다. G.729A에서는 계산량 감소를 위하여 다항식 근의 검색지점 수를 60에서 50으로 줄이고, 부호 변환시에 정밀 검색 횟수를 4에서 2로 줄인다.

② 인지가중필터

인지가중필터는 다음과 같이 양자화된 선형예측계수를 사용한다.

$$W(z) = \frac{\hat{A}(z)}{\hat{A}(z/\gamma)} \tag{1}$$

여기서, 가중인자 $\gamma = 0.75$ 로 고정한다. 이러한 변화는 가중필터와 선형예측 합성필터의 결합된 형태인 가중합성필터를 $W(z)/\hat{A}(z) = 1/\hat{A}(z/\gamma)$ 로 간략화하고, 따라서 충격파응답과 목표신호 계산 및 필터상태를 갱신하는데 필요한 계산량을 감소시킨다. 가중필터의 간략화는 평탄한 주파수 응답특성을 갖는 입력신호에 대해서는 약간의 음질저하를 가져오지만 대부분의 경우 음질 차이는 거의 없다.

③ 개회로 피치검색

폐회로 적응코드북 검색과정의 계산량을 줄이기 위해서 검색범위는 개회로 검색에서 구한 피치지연 주위로 제한된다. 이러한 개회로 피치검색은 한 프레임당 한번 수행되고, 가중신호를 바로 사용하는 G.729와 달리 저역통과필터를 거친 가중 음성신호 $s_w(n)$ 를 입력으로 사용한다. 이 과정은 3 단계로 구성되는데 첫 단계에서는 [20,40], [40,79], [80,143]의 3 구간에서 식(2)의 주어진 상관값을 최대로 하는 각 구간의 지연값을 찾는다.

$$R(k) = \sum_{n=0}^{30} s_w(2n)s_w(2n-k) \tag{2}$$

다음 단계에서 이렇게 구해진 구간 최대치 $R(t_i)$ 를 다음과 같이 정규화한다.

$$R'(t_i) = \frac{R(t_i)}{\sqrt{\sum_{n=0}^{30} s_w^2(2n-t_i)}} \quad i=1,2,3 \tag{3}$$

마지막으로 정규화된 구간 최대치들은 작은 지연에 선호도를 주고 비교하여 최종 지연값을 선택한다. 결과적으로 식 (2)로 주어진 상관값을 계산하는 과정에서 우수 샘플만을 사용하고, 세 번째 구간 [80,143]에서는 우수 지연에 해당하는 상관값만을

계산한 후 선택된 우수 지연의 ±1에 대한 지연만을 검색함으로써 계산량을 감소시킨다.

④ 적응코드북 검색

적응코드북 검색에서는 첫 번째 부프레임의 [19+1/3,84+2/3] 구간에서 실수 피치지연이 사용되고, [85,143] 구간에서는 정수 피치지연이 사용된다. 두 번째 부프레임에서는 항상 [int(T₁)-5-2/3, int(T₁)+4+2/3] 구간에서 1/3 정밀도를 갖는 실수 피치지연이 사용된다. 적응코드북 검색은 다음 식(4)를 최대로 하는 지연을 선택하는 것으로 이루어진다.

$$R(k) = \frac{\sum_{n=0}^{30} x(n)y_k(n)}{\sqrt{\sum_{n=0}^{30} y_k^2(n)}} \quad (4)$$

여기서, x(n)은 목표신호이고, y_k(n)은 지연 k에서 필터링된 과거의 여기신호이다. G.729A에서는 계산량 감소를 위해 식 (4)의 분자항을 최대로 하는 지연검색방법을 도입한다.

$$R_N(k) = \sum_{n=0}^{30} x(n)y_k(n) = \sum_{n=0}^{30} x_k(n)u_k(n) \quad (5)$$

여기서, x_k(n)은 역방향 필터링된 목표신호, 즉 입력신호와 가중합성필터의 상호상관이고, u_k(n)은 지연 k에서 과거 여기신호이다. 각 부프레임에서 최적 피치지연 T₁과 T₂를 결정하는 과정을 살펴보면, 만약 식 (5)를 사용하여 찾은 최적 정수지연이 85 이상인 경우에는 실수 피치검색과정을 생략하고, 그 이하인 경우에는 최적 정수지연을 중심으로 -1/3, 0, 1/3에서 과거의 여기신호를 보간하여 (5)를 최대로 하는 실수지연을 선택한다. 이와 같은 적응코드북 검색의 간략화로 G.729에 비해 1/3 정도의 오차를 가져오지만, 계산량은 1/2로 감소하고 주관적인 음질에서는 거의 차이가 없다.

⑤ 고정코드북 검색

고정코드북의 펄스 크기는 역방향 필터링된 목표신호의 부호를 이용하여 코드북 검색 전에 미리 정해지는 G.729에서의 SSPA(signal selected pulse amplitude) 방법에 의해 결정된다. 그러나 펄스의 위치는 부프레임당 총 2¹³ 개 중 최대 1440(17.5%) 개의 가능한 위치의 조합에 대해 검색하는 nested-loop 방식의 G.729와 달리, 펄스마다 독립적

인 루프를 사용하는 depth-first 방식을 사용하여 검색 횟수를 최대 320(3.9%)으로 감소시켰다. 이와 같이 고정코드북 검색을 간략화하면 약 0.2 dB의 SNR이 감소하는 음질 저하를 초래하지만, 부호화기 전체 계산량 감소의 50%를 차지한다.

⑥ 후처리과정

후처리과정은 적응 후단필터를 제외하고는 G.729와 동일하다. 적응 후단필터는 장기 후단필터, 단기 후단필터 및 경사보상필터의 세 필터가 직렬로 연결되어 구성된다. G.729와의 주요한 차이는 하모닉 후단 필터링과정이다. 하모닉 후단필터의 장기 피치지연은 G.729에서의 1/8 간격으로 실수 피치지연을 검색하는 과정을 생략하고 전송된 현재 부프레임에서 피치지연의 정수부를 중심으로 ±3 범위의 정수 값만을 검색함으로써 추가적인 계산량을 크게 감소시킨다.

Ⅲ. 에너지정합방법

CS-ACELP 부호화기는 가중화된 자승오차를 최소화시키는 파라미터를 최적으로 선택한다. 가중화된 자승오차식은 다음과 같이 표현된다.

$$D_w = \| W \cdot s - W \cdot H \cdot (g_p \cdot c_p + g_c \cdot c) \|^2 = \| s_w - \tilde{s}_w \|^2 \quad (6)$$

여기서, W와 H는 각각 가중필터 및 합성필터의 행렬이고, s는 부호화되지 않은 음성벡터이며, c_p와 c는 각각 적응 및 고정코드북 벡터, 그리고 g_p와 g_c는 각각 적응 및 고정코드북의 이득이다. 또한, s_w는 가중화된 음성신호이고, \tilde{s}_w 는 가중화된 합성음성신호를 나타낸다.

유성음의 경우, 이 기준으로 우수한 음질을 얻을 수 있다. 그러나 무성음 또는 배경잡음에서 위 식은 효율성이 떨어진다. 즉, 무성음 또는 잡음은 장기 주기성이 부족하기 때문에 적응코드북의 효율성이 낮아진다. 따라서, 잡음에 대해서는 정합도가 떨어지는 파형정합방법보다 에너지정합방법을 사용하는 것이 더 적절하다. 에너지정합방법을 포함한 자승오차식은 다음과 같다.

$$D = (1 - \alpha)D_w + \alpha(\| s_w \| - \| \tilde{s}_w \|)^2 \quad (7)$$

여기서, α는 적응균형인자이다. 위 식은 적응균

형인자를 통하여 파형정합과 에너지정합의 비중을 조절한다. α 는 부프레임마다 입력신호가 유성음인가 또는 무성음이나 잡음인가에 따라서 적응적으로 변화해 간다. 잔차신호에 대한 파형정합도의 크기를 측정하여 파형정합도가 높으면 유성음으로 간주하여 α 값을 작게 조절함으로써 파형정합의 비중을 증가시키고 에너지정합의 비중은 감소시킨다. 그리고 파형정합도가 낮으면 무성음 또는 잡음으로 간주하여 α 값을 크게 조절함으로써 파형정합의 비중은 감소시키고 에너지정합의 비중을 증가시킨다. 현재 부프레임에 대한 파형정합도 η (m)는 다음 식과 같이 평가된다.

$$\eta(m) = 10 \log \frac{\|r\|^2}{\|r - g_p \cdot c_p\|^2} \quad (8)$$

여기서, r 은 선형예측 잔차신호를 나타낸다. η (m) 값은 부프레임마다 크게 변화하기 때문에 여러 부프레임에 대해 평균을 취한 값을 사용한다.

$$\eta_{mid}(m) = \frac{1}{P} \sum_{p=0}^{P-1} \eta(m-p) \quad (9)$$

평균 파형정합도를 나타내는 $\eta_{mid}(m)$ 값에 따라 적응균형인자가 결정된다.

$$\begin{aligned} \alpha &= d & \eta_{mid}(m) < a \\ &= d - \frac{d-c}{b-a} (\eta_{mid}(m) - a) & a \leq \eta_{mid}(m) \leq b \\ &= c & \eta_{mid}(m) > b \end{aligned} \quad (10)$$

여기서, c 는 최소값이고 d 는 최대값이다. 또한, a 의 초기값은 c 이다. a 는 최소값과 최대값 사이를 선형적으로 변화하면서 파형정합과 에너지정합의 비율을 조절하게 된다.

IV. 모의실험 및 분석

본 논문에서 제안한 음질개선 알고리즘의 개선효과를 평가하기 위하여 시뮬레이션 프로그램을 작성하고 여러 잡음에 대하여 모의실험을 수행하였다. 프로그램은 C 언어를 이용하여 PC 상에서 작성하였으며, Microsoft사의 Visual C++ 6.0 컴파일러를 사용하여 컴파일한 후 준비된 입력샘플을 이용하여 DOS 창에서 실행하였다. 시뮬레이션 프로그램은 기능별로 분리된 35개의 부프로그램으로 구성되어 있으며, 전체적으로 약 13,000 lines의 규모이다.

실험대상인 소리자료를 얻기 위해 먼저 음성과

잡음을 현장에서 직접 녹음한 후, 편집과정을 거쳐 잡음속의 음성샘플을 준비하였다. 편집작업은 Syntrium 소프트웨어사의 Cooledit를 이용하여 8 kHz 표본화주파수 및 16-bit 해상도에서 raw PCM 형식으로 음성파일을 처리하여 저장하였다. 그리고 신호의 spectrogram을 관측하기 위하여 SFS 프로그램을 사용하였다. 음질평가는 3개 문장에 대해 남성 및 여성화자별로 5가지 배경잡음에 대하여 실시하였다.

음질의 평가기준은 2가지로 나뉘어 있는데 객관적 평가기준과 주관적 평가기준이 있다. 객관적 평가기준으로는 SEGSNR이 있고, 주관적 평가기준으로는 MOS 점수를 들 수 있다. 본 실험에서는 다소 일관성이 결여된 MOS 대신 객관적 기준인 SEGSNR을 평가하였다. 다만 CS-ACELP 부호화기가 구조적으로 파형부호화방식과 음원부호화방식을 결합한 복합부호화기라는 점에서 SEGSNR을 사용하는 것은 음질평가의 측면에서 다소 부적절한 면이 있으나 간단명료하고 대략적인 음질평가가 가능하기 때문에 많이 이용한다. 일반적으로 SNR은 음성신호중 에너지가 큰 부분에 많은 영향을 받는다. 그러나 음성신호는 에너지가 큰 부분과 작은 부분이 반복되어 나타나는 비정적 신호이므로 SNR을 사용하는 것보다 구간별 SNR을 구하여 그 통계적 특성을 이용하는 SEGSNR을 사용하는 것이 바람직하다.

객관적 평가방법은 SEGSNR을 50 ms 단위로 구한 후, 평가하고자 하는 음성의 전구간에 걸쳐 SEGSNR의 평균을 취한 MSEGSNR(mean SEGSNR)을 사용하였다.

$$SNR(k) = 10 \log \frac{\sum_{i=0}^{39} |sw(i)|^2}{\sum_{i=0}^{39} |x(i) - g_p y(i) - g_c z(i)|^2} \quad (11)$$

$$SEGSNR(n) = \frac{1}{10} \sum_{k=0}^9 SNR(k) \quad (12)$$

여기서, SNR(k)는 k번째 프레임의 신호대잡음비이다. 즉, SEGSNR은 10개의 부프레임에 대한 평균 SNR을 나타내고, MSEGSNR은 이것을 다시 전구간에 걸쳐 평균한 것이다. 이렇게 하면 상대적으로 SNR이 작은 무성음이나 묵음 구간에서의 SNR도 고려가 된다.

실험에 사용된 음성샘플들은 20대 남성 및 여성 화자에 의해 발음된 아래 3개의 문장을 녹음한 것

이고, 배경잡음으로는 인파소음, 교통소음, 바람소리, 엔진소음, 그리고 컷소리 등 5가지 잡음을 사용하였다.

- ① /여기는 음성처리연구실입니다/
- ② /인터넷은 많은 정보를 준다/
- ③ /전주에는 전북대학교가 있습니다/

시험샘플은 일반 녹음기를 이용하여 현장에서 음성과 잡음을 각각 따로 녹음한 후, 음성편집도구를 이용하여 편집과정을 통해 잡음속의 음성을 구현하였다. 다음 그림은 수집한 음성 및 잡음신호에 대한 파형과 spectrogram을 나타낸 것이다.

객관적 평가방식으로서 배경잡음이 존재하는 환경에서 각 문장에 대해 정규 CS-ACELP 방식으로 부호화하였을 경우와 에너지정합방식으로 부호화하였을 경우의 MSEGSR을 측정, 비교하였다. 실험은 각 화자별로, 배경잡음별로, 그리고 문장별로 구분하여 실시하였다. 측정 결과, 잡음이 있는 환경에서는 무잡음 환경에 비해 음질이 10~50% 정도 떨어지는 것으로 나타났다. 다시 말해서, 무잡음 환경

에서는 MSEGSR이 대략 15 dB 정도인데 비해 잡음이 존재하면 약 2~7 dB 정도 감소하였다.

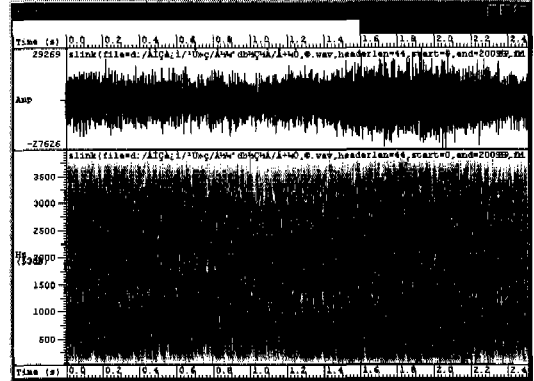


그림 3. 교통소음의 파형 및 스펙트로그램

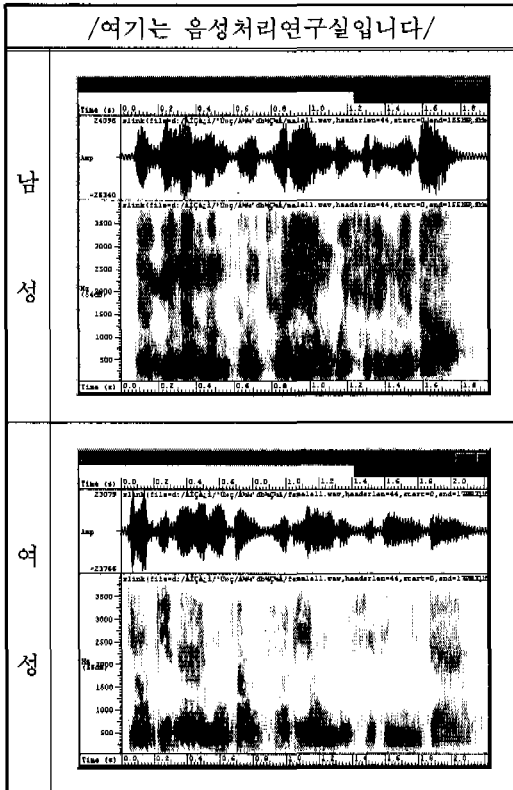


그림 2. 음성신호의 파형 및 스펙트로그램

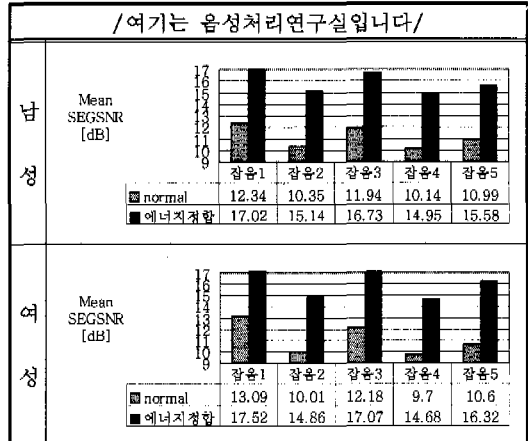


그림 4. 배경잡음이 있는 1번 문장의 MSEGSR

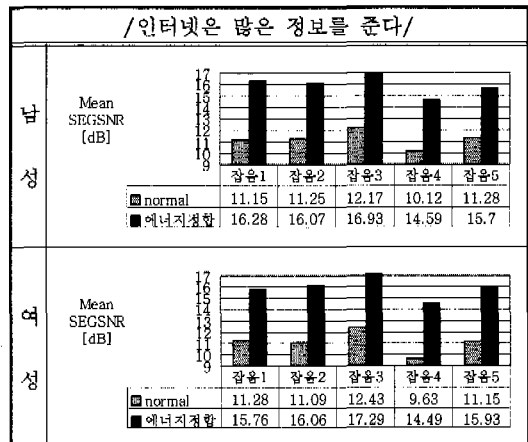


그림 5. 배경잡음이 있는 2번 문장의 MSEGSR

파형정합도 η_{mid} 의 최소값 $a=3$, 최대값 $b=100$, 적응균형인자 α 의 최소값 $c=0.2$, 최대값 $d=0.8$ 로 설정한 상태에서 문장별로 남녀화자에 대해 여러 배경잡음을 적용하였을 때의 MSEGSR을 측정하였다. 실험 결과, 에너지정합방식으로 부호화했을 때가 정규 CS-ACELP 방식으로 부호화했을 때보다 전반적으로 음질이 향상되는 것을 확인하였다. 에너지정합방식으로 부호화하면 정규 CS-ACELP 방식으로 부호화하는 경우보다 최대 59.2%까지 성능이 향상되었다. 즉, 에너지정합방식을 사용하면 4.42~6.28 dB 정도 음질이 개선되는 것을 보여 주었다.

기 위하여 모의실험을 통해 5가지 배경잡음 조건에서 3개 문장에 대한 여성 및 남성화자의 음질평가를 실시하였다. 음질평가기준으로는 SEGSNR을 음성의 전구간에 걸쳐 평균을 취한 MSEGSR의 객관적 성능평가방법을 사용하였다. 실험 결과, 전반적으로 음질이 향상됨을 보였으며, MSEGSR이 4.42~6.28 dB 정도 증가하는 효과를 가져왔다. 따라서 최고 59.2%의 음질개선효과를 나타내는 것으로 확인되었다.

참고 문헌

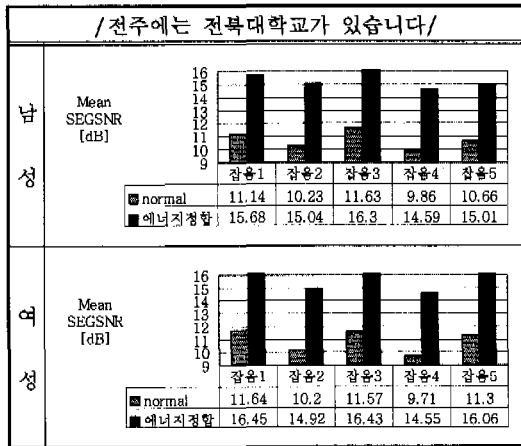


그림 6. 배경잡음이 있는 3번 문장의 MSEGSR

V. 결론

본 논문은 8 kbps CS-ACELP 음성부호화기에 있어서 여러 배경잡음에 대한 음질특성 개선방법에 관한 연구이다. 계산량을 감소시킨 G.729A 표준은 파형정합방법을 사용하여 가중오차를 최소화시키는 최적 솔루션을 선택한다. 그러나 무성음이나 잡음 구간에서는 주기성의 부족으로 파형정합이 잘 이루어지지 않는다. 따라서 이런 경우 파형정합방법보다는 신호의 절대 크기를 증시하는 에너지정합방법이 더 효율적이다. 본 논문에서는 두 가지 방식을 혼합하여 유성음 구간에서는 파형정합방법을 증시하고, 무성음 또는 잡음 구간에서는 음질을 개선하기 위하여 에너지정합방법을 증시하도록 구성하였다. 이렇게 함으로써 유성음 구간에서의 음질은 비슷한 반면, 무성음이나 잡음 구간에서의 음질은 크게 향상되는 효과를 얻었다.

제안한 에너지정합방법의 성능개선정도를 측정하

- [1] W. B. Kleijn and K. K. Paliwal, *Speech Coding and Synthesis*. Amsterdam, Holland, 1995.
- [2] R. Salami, et al, "Description of the Proposed ITU-T 8 kb/s Speech Coding Standard," Proc. IEEE Workshop on Speech Coding for Telecomm., pp. 3-4, 1995.
- [3] ITU-T Recommendation G.729, "Coding of Speech at 8 kbit/s Using Conjugate-Structure Algebraic-Code-Excited-Linear-Prediction (CS-ACELP)," Mar. 1996.
- [4] R. Salami, et al, "Design and Description of CS-ACELP : A Toll Quality 8kb/s Speech Coder," IEEE Trans. Speech and Audio Processing, vol. 6, No. 2, pp. 116-130, 1998.
- [5] ITU-T Recommendation G.729 Annex A, "Coding of Speech at 8 kbit/s Using Conjugate-Structure Algebraic-Code-Excited-Linear-Prediction(CS-ACELP), Annex A Reduced Complexity 8 kbit/s CS-ACELP Codec," Nov. 1996.
- [6] R. Salami, C. Laflamme, B. Bessette, and J-P. Adoul, "Description of ITU-T Recommendation G.729 Annex A : Reduced Complexity 8 kbit/s CS-ACELP Codec," IEEE Prec. Int. Conf. Acoust. Speech and Signal Proc., pp. 775-778, 1997.

이 행 우(Haeng-Woo Lee)

정회원



1985년 2월: 광운대학교

전자공학과 학사

1987년 2월: 서강대학교

전자공학과 석사

2001년 2월: 전북대학교

전자공학과 박사

1987년 2월~1998년 3월: 한국전자통신연구원

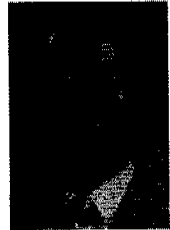
1998년 4월~2001년 1월: 벽성대학 정보통신과

2001년 3월~현재: 남서울대학교 정보통신공학과

<주관심 분야> 음성부호화, 디지털통신, VLSI 설계

박 기 영(Ki-Young Park)

정회원



1978년 2월: 광운대학교

전자계산기공학과 학사

1982년 2월: 광운대학교

전자통신공학과 석사

1996년 2월: 전북대학교

전자공학과 공학박사

1982년 3월~현재: 전주공업대학 정보통신과 교수

<주관심 분야> 음성신호처리 및 VLSI 설계

김 종 교(Chong-Kyo Kim)

정회원



1966년 2월: 전북대학교

전기공학과 학사

1977년 8월: 전북대학교 대학원

전기공학과 석사

1983년 8월: 전북대학교 대학원

전기공학과 공학박사

1979년 4월~현재: 전북대학교 전자정보공학부

<주관심 분야> 음성부호화, 음성인식, 음성합성