

고성능 CELP 음성 압축기를 위한 새로운 코드북 구조

New Codebook Structure For A High-Quality CELP Speech Coder

박 호 중*, 권 순 영*

(Ho Chong Park*, Soon Young Kwon*)

요 약

본 논문에서는 고성능 CELP 음성 압축기를 위한 "Baseline 코드벡터"와 "Implied 코드벡터"로 구성되는 새로운 구조의 코드북을 제안한다. Implied 코드벡터는 피치 주기 이전의 합성음으로부터 구하여지며 여기(勵起)신호의 피치 구조를 강화하여 합성음의 음질을 향상시킨다. Implied 코드벡터는 전달되지 않고 인코더 및 디코더에서 각각 합성음을 이용하여 독립적으로 구하여진다. 또한 펄스와 랜덤 성분을 모두 가지는 복합 여기방식을 이용하여 음질을 더욱 향상시킨다. 제안된 코드북 구조를 이용하여 10msec 프레임을 가지는 8kbps CELP 음성 압축기를 설계하여 하나의 DSP 칩에 실시간 구현하였고, 이것의 성능을 SNRseg와 MOS로 측정하였다. 평균 SNRseg는 12.14dB로 G.729 CS-ACELP의 SNRseg보다 6dB 높고, 조용한 환경에서의 MOS는 3.80으로 G.729 CS-ACELP의 MOS보다 0.02 높다.

ABSTRACT

This paper proposes new codebook structure for a high-quality CELP speech coder which consists of "Baseline Codevector" and "Implied Codevector." The implied codevector is determined from the synthesized speech delayed by the pitch period, which enhances the pitch structure of excitation signal resulting in high-quality synthesized speech. The implied codevector is not transmitted, but extracted from the synthesized speech encoder and decoder independently. In addition, the mixed excitation with both pulse and random codebook is used to further enhance the speech quality. A 8 kbps CELP speech coder with 10msec frame based on the proposed codebook structure is designed and implemented on a single DSP chip in real-time. The performance of the coder is measured in terms of SNRseg and MOS. The average SNRseg is 14dB which is 0.6 higher than that of G.729 CS-ACELP, and the MOS for clean speech is 3.80 which is 0.02 higher than that of G.729 CS-ACELP.

I. 개 요

LPC 방식의 음성 압축기[1, 2]가 개발된 이후 CELP 음성 압축기[3]가 16kbps 이하의 전송율에서 가장 효율적인 압축 방식으로 이용되어 왔다. 그 예로 ITU의 표준 음성 압축기인 G.728 LD-CELP[4]과 G.729 CS-ACELP[5], TIA/EIA의 표준 음성 압축기인 IS-54 VSELP[6], IS-96A QCELP[7], IS-127 EVRC[8] 등의 CELP 방식 음성 압축기가 개발되었고, 이들은 현재 디지털 통신 시스템 등에 적용되어 서비스 중이다. 그러나 전송률 8kbps와 그 이하의 음성압축기들은 아직 실제 인간의 음성과 같은 자연스러운 음성을 합성하지 못하고 있고, 이로 인하여 이들 음성압축기를 사용하는 디지털 통신 시스템은 처음 예상 하였던 음질을 제공하지 못하고 있는 실정이다.

CELP 음성 압축기는 Analysis-by-Synthesis 방법을 이용하여 입력 음성의 예측 잔여 신호를 코드북으로 양자화하는 방식에 기초를 둔다. 즉, 코드북으로부터 선택된 코드벡터(codevector)에 이득(Gain)을 곱하여 원하는 여기 신호(Excitation Signal)를 만들고 이를 피치 필터와 포먼트 필터(Formant Filter)를 통과시켜 음성신호를 만든다. CELP 음성 압축기는 코드북의 구조에 따라 크게 구별이 되고, 또한 이에 따라 합성된 음성의 음질이 큰 차이를 나타낸다.

그동안 여러 형태의 코드북이 제안되어 왔는데, 랜덤 코드북을 사용하는 LD-CELP, VSELP와 QCELP, Algebraic 코드북을 사용하는 CS-ACELP, EVER 등이 개발되었고, 또한 합성된 음성의 피치 구조를 향상시키기 위한 PSI-CELP[9]도 개발되었다. 그러나 이러한 구조의 코드북들이 다양한 특성을 가지는 여기 신호를 표현하는데 한계가 있어 아직까지 8kbps 이하에서 "Toll-Quality"의 음성 압축기가 개발되지 못하였다.

*광운대학교 전자공학과

접수일자: 1997년 12월 22일

본 논문에서는 두개의 코드벡터 조합으로 구성된 새로운 구조의 코드북을 제안한다. 하나의 코드벡터는 일반적인 CELP 구조와 마찬가지로 현재 프레임의 입력신호를 기준으로 구하여지고 이를 "Baseline 코드벡터"라 하며, 나머지 코드벡터는 합성 신호를 현재 프레임의 피치 주기 만큼 지연 시킨 신호를 기준으로 구하여지고 이를 "Implied 코드벡터"라 한다. 이 두 코드벡터에 각각의 이득을 곱하고 최종적으로 두 신호를 합하여 여기 신호를 얻는다. Implied 코드벡터는 피치 주기 이전의 신호를 기준으로 구하여지므로 여기 신호의 피치 성분을 보장하여 주는 역할을 하여 음질 향상을 가져온다. 그동안 피치 구조 향상을 위한 코드북 개선 방법들이 제안되어 왔으나 이들은 항상 일정한 정도의 피치 성분을 강조하는 반면 [5, 8, 9], 본 논문에서는 제안된 방법은 Implied 코드벡터의 이득을 조정하여 가변적으로 피치 구조의 강도를 조정할 수 있다. 또한 Implied 코드벡터는 현재의 입력신호가 아닌 과거의 합성 신호를 기준으로 구하여지므로 직접 전달할 필요없고 디코더에서도 동일한 Implied 코드벡터를 구할 수 있다.

제안된 코드북은 음질을 더욱 향상시키기 위하여 펄스 코드북과 랜덤 코드북의 조합으로 구성된다. 즉 Baseline 코드벡터와 Implied 코드벡터의 합으로 구성되는 여기 신호에서, 각 코드벡터가 서로 다른 성질(펄스 또는 랜덤)의 코드북으로부터 선택되도록 한다. 따라서 여기 신호는 항상 랜덤과 펄스 성분을 모두 가지는 복합 여기 신호가 되어 보다 다양한 성질의 여기 신호를 표현할 수 있어 음질의 향상을 가져올 수 있다. 코드벡터 검색은 코드북 부프레임 경계에서의 불연속성을 극복하고 검색 성능을 향상시키기 위하여 인접한 두 코드북 부프레임에 포함되는 4개의 코드벡터를 Joint Optimization 기술을 사용하여 검색한다.

본 논문은 다음과 같이 구성된다. 2장에서는 제안된 코드북 구조에 대하여 자세히 설명하고 3장에서는 제안된 이용한 음성압축기의 실제 설계와 구현, 그리고 성능을 보여준다. 4장에서는 결론을 내린다.

II. 새로운 코드북 구조

2.1. 배경

CELP 음성 압축기는 주어진 코드북에서 최적의 코드벡터를 선택하여 필요한 여기 신호를 만들어 낸다. 따라서 한정된 크기의 코드북을 이용하여 필요항 여기 신호를 적절히 표현하는 것이 CELP 음성 압축기의 성능을 좌우하는 요인이 된다. 그러나, 음성 압축기의 전송율을 줄이고 코드북으로부터 최적의 코드벡터를 검색하기 위한 계산량을 줄이기 위하여 코드북의 크기와 구조가 극히 제한된다.

IS-54 VSELP는 코드벡터를 여러 Base 신호의 벡터 합으로 표시한다. 이와같은 구조를 이용하면 Base 신호에

의한 합성 결과만으로 Base 신호의 벡터 합으로 구성된 실제 코드벡터에 의한 합성 결과를 쉽게 알 수 있으므로 최적의 코드벡터를 쉽게 찾을 수 있다. 각 Base 신호는 기본적으로 랜덤 신호로 이루어져 있다. IS-96A QCELP는 순환 코드북을 이용한다. 각 코드벡터는 서로 한 Sample 씩 Shift된 관계로 구성되며, 한 코드벡터에 의한 합성 신호를 알면 옆의 코드벡터에 의한 합성 신호를 Shifting 관계를 이용하여 쉽게 구할 수 있다. 그러나 이 방법은 검색 속도는 빠르지만 제한된 순환 구조 때문에 여러 특성의 여기 신호를 만들 수 없는 단점이 있다.

G.729 CS-ACELP, GSM EFR Coder[12], IS-127 EVRC 등에 사용되는 ACELP 방식의 코드북은 펄스 구조를 가지며, 정해진 수의 펄스만으로 여기 신호를 만든다. 이 구조는 이론적으로는 상당히 큰 크기의 코드북을 가지나 전체 가능한 코드북을 전부 검색하는 것은 계산량이 너무 많아 불가능하므로 실제로는 가능성이 있는 일부 코드북 영역만을 검색하고 펄스의 위치를 제한하여 계산량을 줄이도록 하며, 따라서 항상 최적의 코드벡터를 찾는 것은 아니다. 또한 각 펄스의 크기를 서로 틀리게 하려면 전송율이 증가하고 계산량이 많아지므로 코드북 부프레임 내에서는 모든 펄스가 같은 전대 크기(부호는 다를 수 있음)을 가지게 하는 단점이 있다.

인코더에서 입력 신호를 분석한 결과 생성되는 잔여 신호, 즉 코드북으로 표현하여야 하는 신호는 이론적으로 백색 잡음이어야 하나, 실제로는 분석이 완벽하지 못하고 양자화 오차 등에 의하여 어떤 특성을 가지게 된다. 특히 피치 성분이 잔여 신호에 남아 있게 되는데, 보통 코드북으로 이러한 피치 성분을 정확히 표현하는데 한계가 있다(특히 피치 주기가 짧을때). 이와 같은 문제를 해결하기 위하여 PSI-CELP, G.729, IS-127 등에서는 코드벡터 자체를 피치 주기만큼 반복시키는 방법을 제안하였고, 이를 통하여 음질을 향상시킬 수 있었다. 그러나 이는 잔여 신호 중에 피치 성분이 많이 남아 있는 경우에는 효과가 크지만, 그렇지 않은 경우(여성의 단음 모음의 경우 종종 발견 됨)에는 오히려 음질을 저하시키는 단점이 있다.

2.2. 새로운 코드북 구조

본 논문에서는 제안하는 코드북 구조는 위에서 언급한 코드북의 제약점들을 해결하는데 중점을 두고 설계하였다. 그림 1은 제안된 코드북 구조를 가지는 음성 압축기의 디코더 동작을 보여준다. Implied 코드벡터는 포트먼트 필터의 출력 합성음을 현재의 피치 주기만큼 지연시킨 신호를 목표 신호로 하여 Analysis-by-Synthesis 방법에 의하여 구한다. 코드북은 펄스 코드북과 랜덤 코드북으로 구성되어 있고 Implied 코드벡터는 인코더로부터 전달된 Baseline 코드벡터와 다른 성질(펄스 또는 랜덤)의 코드북에서만 검색한다. 다음, 전달된 Baseline 코드벡터와 이득을 곱하고, 직접 구하여진 Implied 코드벡터와 전달된 이득을 곱하여 두 신호의 합으로 여기 신호를 만들

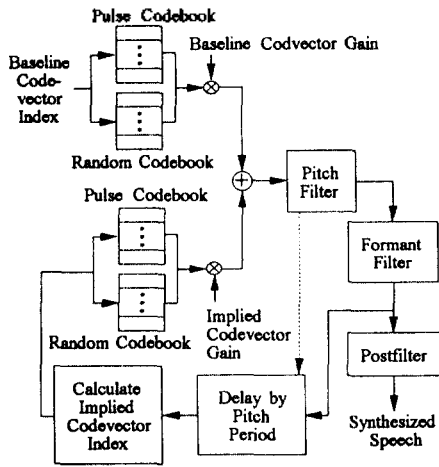


그림 1. 제안된 코드북 구조를 가지는 디코더의 동작

고 피치 필터와 포먼트 필터를 통과시켜 합성음을 만든다. Implied 코드벡터 이득은 인코더로부터 직접 전달되는데, 이는 전달 오류에 의하여 인코더와 디코더 사이의 동기화가 어긋났을 때 양 쪽의 Implied 코드벡터 이득이 틀려지는 것을 방지하기 위함이다. 합성음의 음질은 코드벡터보다 이득에 더 민감하므로 Implied 코드벡터의 이득만 전달하고 코드벡터는 전달하지 않고 디코더에서 내부적으로 구한다.

이와 같은 구조의 코드북을 사용하면 Implied 코드벡터에 의한 여기 신호와 피치 성분 강화로 합성음의 피치 구조가 향상되고, 복합 여기 방식으로 인한 적절한 여기 신호의 표현이 가능하여 일반적인 코드북에 비하여 상당한 음질 향상을 가져온다. 특히 Implied 코드벡터의 이득을 조정하여 피치 성분을 입력신호에 적용하여 가변적으로 조절 가능하고, 코드벡터들의 이득을 조정하여 여기 신호의 펄스 성분과 랜덤 성분 비율을 가변시킬 수 있는 장점이 있다.

입력 신호의 피치 성분이 강할 경우 현재 프레임의 입력 신호와 피치 주기 이전의 입력 신호 사이에는 상당히 큰 상관관계 있고, 따라서 합성신호 사이에서도 이러한 상관관계가 있게 된다. Implied 코드벡터는 피치 주기 이전의 합성된 신호를 근간으로 구하여지므로 여기 신호에 피치 성분을 포함시키는 역할을 하여 출력 합성음의 피치 구조가 개선되어 음질의 향상을 가져 온다. 또한 무성음 구간, 천어구간 등과 같이 현재 신호와 피치 주기 이전 신호와의 상관관계가 적을 때는 Implied 코드벡터의 역할이 적어지며 이는 Implied 코드벡터의 이득이 작아지는 현상으로 나타난다. 따라서 Implied 코드벡터의 이득을 이용하여 피치 구조를 가변적으로 향상시킬 수 있다. 이 방법은 지금까지의 고정된 피치 구조 향상 방법에 비하여 입력 신호의 특성에 따라 적용하여 동작하므로 보다 효과적인 피치 구조 개선을 할 수 있다.

복합 여기 방식의 장점은 많이 알려져왔으나[10, 11] 두 종류의 신호를 전달하는데 많은 Bit가 필요한 문제점이 있어 널리 사용되지 못하였다. 제안된 코드북은 이 문제를 Implied 코드벡터 개념을 도입하여 하나의 성분은 직접 전달하지 않음으로써 적은 Bit로 복합 여기방식을 구현하고 있다.

이 구조의 코드북은 과거의 신호를 이용하여 Implied 코드벡터를 찾으므로 인코더와 디코더 사이의 동기가 맞지 않으면 서로 다른 Implied 코드벡터를 찾게 되어(Implied 코드벡터 불일치) 음질 저하가 발생한다. 이를 방지하기 위하여 피치 성분이 작은 무성음 또는 묵음 구간에서는 Implied 코드벡터의 목표 신호를 과거의 합성음 대신 Pseudo Random Sequence로 선택한다. 이 경우 인코더와 디코더에서 동일한 목표 신호를 선택하여야 하며, 예로 양 쪽이 동일하게 얻을 수 있는 포먼트 필터의 출력 응답신호를 이용하여 목표 신호를 만들 수 있다. 이 경우 현 프레임의 전달 오류가 없는 한 양 쪽의 동기가 맞춰지게 된다. 또한 동기가 맞지 않을 때의 음질 저하를 줄이기 위하여 Implied 코드벡터의 이득은 직접 인코더로부터 전달된다.

Implied 코드벡터는 전달 되지 않고 내부적으로 만들어 사용하므로 이론적으로는 코드북을 이용한 양자화 없이 피치 주기 이전 잔여 신호를 그대로 Implied 코드벡터로 사용할 수 있다. 그러나 이와같이 양자화 되지 않은 원래의 신호를 코드벡터를 사용하면 인코더와 디코더 사이의 미세한 비동기(非同期)에 의하여서도 Implied 코드벡터 불일치가 발생하여 비동기가 계속 다음으로 전달되는 단점이 있다. 즉, Implied 코드벡터를 코드북을 이용하여 양자화 함으로써 양자화 오차를 이용하여 동기가 어긋나는 것을 방지할 수 있게 된다. Implied 코드북의 크기가 커지면 양자화 오차가 줄어들고 극단적으로 크기가 무한대가 되면 양자화 없이 신호를 그대로 코드벡터로 사용하는 결과가 된다. Implied 코드북의 크기, 즉 양자화 오차에 의한 음질저하와 Implied 코드벡터 불일치에 의한 음질 저하의 상관관계, 그리고 최적의 코드북 크기에 대해서는 현재 계속 연구 중에 있다. 또한 잔여 신호를 전체 특성만을 포함하는 간단한 신호를 만들어 Implied 코드벡터로 사용하는 방법도 있으나, 이 경우 단순화 오차에 의한 음질 저하를 예측할 수 없는 문제점이 있다. 따라서 위에서 설명한 방식과 같이 코드북을 이용한 Analysis-by-Synthesis 방법에 의하여 최적의 Implied 코드벡터를 구하는 방법이 가장 좋은 음질의 합성음을 만든다. 물론 Analysis-Synthesis 방법을 사용하면 추가 검색이 필요하여 계산량이 증가하는 단점은 있다.

2.3. 코드벡터 및 이득 검색 방법

이 장에서는 인코더에서 Baseline 코드벡터와 Implied 코드벡터, 그리고 이들의 양자화된 이득을 검색하는 방

법을 설명한다. 검색은 Implied 코드벡터, Baseline 코드벡터, 두 코드벡터의 양자화된 이득의 순서로 진행된다. 디코더의 Implied 코드벡터 검색은 인코더에서와 동일한 방식으로 실행된다.

현재의 피치 부프레임에서 피치 주기를 구하고 이것을 출력 합성신호 비퍼에 적용하여 피치 주기 만큼 지연된 합성신호를 얻는다. 이 신호를 마치 입력신호인 것처럼 가정하고 목표 신호를 구하고 일반적인 Analysis-by-Synthesis 방법으로 랜덤 코드북에서 한개, 펄스 코드북에서 한개, 모두 두개의 최적 Implied 코드벡터를 찾는다. 이때, 코드벡터의 이득은 양자화하지 않은 최적값을 이용한다.

Baseline 코드벡터는 현재의 입력 신호를 이용하여 구하는데, 위에서 구한 Implied 코드벡터가 최종 여기 신호의 일부분으로 작용하므로 이 부분의 역할을 고려하여 구한다. 즉, 현재의 입력신호를 이용하여 목표 신호를 먼저 구하고 이 신호에서 Implied 코드벡터에 의한 역할을 제외한 나머지 목표 신호를 구하여 이를 기준으로 최적의 Baseline 코드벡터를 구한다. 이때도 이득은 양자화하지 않은 값을 이용한다. 위에서 두 종류의 Implied 코드벡터(펄스와 랜덤)가 구하여졌으므로 각각에 대하여 최적 Baseline 코드벡터를 구하여야 하며, 이때 Implied-Baseline 코드벡터 조합이 서로 다른 성질의 신호가 되는 경우만 검색한다. 이 결과 최종 여기 신호는 항상 펄스 + 랜덤 구조를 가지게 되고 검색시간도 줄어든다.

보다 청각적으로 정확한 여기 신호를 만들기 위하여 Baseline 코드벡터 검색은 두개의 코드북 부프레임에 대하여 Delayed Joint Optimization을 한다. 이때, 두개의 코드북 부프레임이 하나의 피치 부프레임에 해당한다고 가정한다. 즉, 매 코드북 부프레임에서 구한 두개의 Implied 코드벡터 각각에 대하여 Baseline 코드벡터를 복수로 구하고(보통 2~3개), 두 코드북 부프레임에서 가능한 모든 코드벡터들의 조합에 모든 가능한 양자화된 이득을 적용하여 최적의 코드벡터 조합 하나의 이득의 양자화된 이득을 구한다. 즉, 아래의 식을 최소화 시키는 코드벡터 조합과 양자화된 이득을 구한다. 이와같은 검색 방법을 이용하면 부프레임 경계에서의 여기신호 불연속성이 줄어들고 보다 넓은 구간에서 최적화를 하므로 특히 천이구간에서 성능이 향상된다.

$$\sum_{n \in \text{pitch subframe}} \left\{ x(n) - (G_{b_{p1}} C_{b_{1,jk}}(n) + G_{i_{p2}} C_{i_{1,j}}(n) + G_{b_{p3}} C_{b_{2,uv}}(n) + G_{i_{p4}} C_{i_{2,u}}(n)) \right\}^2$$

여기서 $x(n)$ 은 목표 신호, G_{b_p} 와 G_{i_p} 는 각각 Baseline 코드벡터와 Implied 코드벡터의 양자화된 이득이며, 아래첨자 p 는 양자화 Index를 가리킨다. 예로, Baseline 코드벡터 이득에 4 Bit가 할당되면 $p1$ 과 $p3$ 은 0부터 15까지의 값을 가진다. $C_{i_{1,j}}(n)$ 와 $C_{i_{2,u}}(n)$ 은 각각 첫번째와 두번째

코드북 부프레임의 Implied 코드벡터이며 j 와 u 는 펄스와 랜덤을 구별하는 첨자이다.(즉, $j, u \in \{1, 2\}$). $C_{b_{1,jk}}(n)$ 은 첫번째 부프레임의 Implied 코드벡터 $C_{i_{1,j}}(n)$ 에 대한 Baseline 코드벡터를 나타내며 첨자 k 가 복수로 선택된 여러개의 후보를 표시하고, $C_{b_{2,uv}}(n)$ 은 두번째 부프레임에 대한 동일한 의미를 나타낸다. 따라서 위 식을 최소화 시키는 $j, k, u, v, p1, p2, p3, p4$ 값들을 찾으면 최적의 여기 신호가 구하여진다. 단일 코드벡터에서 시간 축 n 이 유효한 영역을 벗어난 경우는 그 신호 값을 0으로 가정한다.

위의 식을 최소화하는 값들의 검색 시간을 줄이기 위하여 다음의 방법을 사용한다. 각 Implied 코드벡터와 Baseline 코드벡터를 구할 때 양자화 되지 않은 최적의 이득이 이미 구하여져 있다. 따라서 모든 양자화된 이득을 검색하는 대신, 즉 모든 p index를 검색하는 대신 양자화 되지 않은 최적의 이득값과 동일한 부호의 양자화된 이득만을 검색하면 검색 시간이 상당히 줄어들고 검색 결과도 최적에 매우 근접한 결과를 가져온다.

III. 음성압축기의 설계 및 구현

펄스, 랜덤 코드북을 이용하고 Implied 코드벡터를 이용한 피치 구조 개선은 Bit 수의 큰 증가없이 CELP 음성 압축기의 성능을 크게 향상시킨다. 이러한 구조의 코드북을 이용하여 실제 CELP 음성압축기를 설계하고 실시간 구현을 하여 성능을 분석하였다. 표 1은 설계된 음성 압축기의 규격을 보여주고 표 2는 음성 압축기의 Bit 할당을 보여준다. 특히, 음성 압축기가 디지털 통신에 사용될 때 음성 지연으로 인한 통신 품질 저하를 방지하기 위하여 프레임의 크기를 10msec로 하였다.

3.1. 코드북 설계

음성 압축기의 코드북 부프레임 크기가 20 Sample (2.5msec)이므로 하나의 펄스만으로도 피치 성분을 충분히 표현할 수 있다. 따라서 펄스 코드북은 20개의 코드벡터로 구성되며 i -th 코드벡터 $P_i(n)$ 은 다음과 같이 표현되며, 이 구조의 펄스 코드북은 Baseline과 Implied에서 동일하게 사용된다.

$$P_i(n) = \begin{cases} 1 & \text{if } n=i \\ 0 & \text{else} \end{cases}$$

이와같은 단일 펄스 코드벡터를 이용하면 검색이 빠르고 각 펄스의 크기를 독립적으로 정할 수 있는 장점이 있다. 랜덤 코드북은 단위 에너지를 가지는 단순한 백색잡음 신호로 구성되며 크기는 44로 한다. Baseline과 Implied에 동일한 랜덤 코드북이 사용된다.

코드북 부프레임의 Baseline 코드벡터에 6 Bit가 할당되어 Index는 0부터 63이며, 펄스 코드벡터의 Index는 0부터 19, 랜덤 코드벡터의 Index는 20부터 63으로 구성되

어 있어 Baseline 코드벡터 Index로부터 코드벡터의 상질을 알 수 있다. 코드북 이득은 매 피치 부프레임 단위로 벡터 양자화한다. 즉, 두 코드북 부프레임의 Baseling 코드벡터 이득, Implied 코드벡터 이득으로 4차 벡터를 구

표 1. 8kbps 음성 압축기의 규격

항목	값
Data Rate	8000 bit/sec
Sampling Rate	8kHz
Frame Size	80 samples(10msec)
Pitch Subframe Size	40 samples(5msec)
Codebook Subframe Size	20 samples(2.5msec)
LPC Filter Order	10
Pitch Predictor Order	1

표 2. 8kbps 음성 압축기의 Bit 할당

Parameter	Pitch Subframe 1		Pitch Subframe 2		Frame
	CB1	CB2	CB3	CB4	
Low LSP (lsp1 ~ lsp4)	-		9		9
High LSP (lsp5 ~ lsp10)	-		9		9
Pitch Period	8		5		13
Pitch Gain	3		2		5
Baseline Codevector Index	6	6	6	6	24
Codevector Gain	10		10		20
합계	33		37		80

3.2. 포먼트 필터와 피치 필터 설계

포먼트 필터는 자기상관계 방법으로 10개의 LPC를 구하고 LSP로 변환 한 후 Moving Average 예측하여 오차값을 구하고 이를 4차와 6차의 벡터로 나누고 각 벡터를 9 bit로 Split-VQ한다. 이는 피치 부프레임 2에서만 이루어지고 피치 부프레임 1에서는 부프레임 사이의 LSP 보간으로 LPC를 구한다.

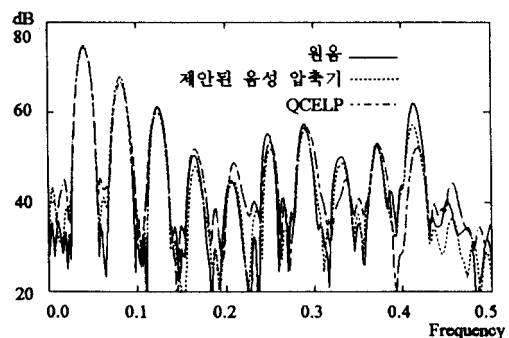
피치 필터는 One-Pole 필터이고 매 프레임마다 Open-Loop 피치를 구하고 각 피치 부프레임마다 Open-Loop 피치 주위에서 Closed-Loop 피치 주기를 구하며, 피치 주기를 해상도는 1/2과 1/3 Sample로 한다. 피치 이득은 비선형 Scalar 양자화하며, 피치 주기와 이득 모두 피치 부프레임 1에서는 절대값을 구하고 피치 부프레임 2에서는 부프레임 1에 대한 상대값을 구한다.

3.3. 실시간 구현

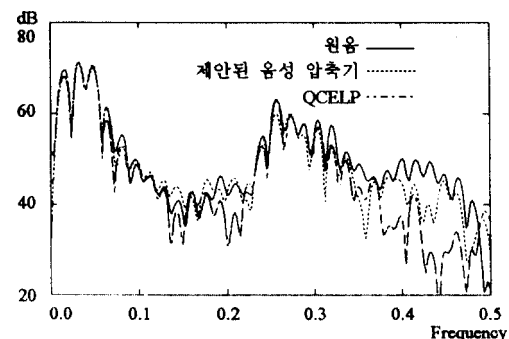
위에서 설계된 CELP 음성 압축기를 Texas Instrument의 TMS320C31 60MHz DSP 칩 하나를 이용하여 Full-Duplex로 실시간 구현하였다. 22K Word의 External Memory와 25 MIPS가 사용되며, 두개의 코드북을 사용하고 특히 랜덤 코드북을 사용하므로 코드 벡터 검색에 상당히 많은 계산이 필요하였다. 아직 검색 부분의 계산량 최적화를 하지 못한 상태이며, 최적화를 마치면 CS-ACELP, EVRC 등과 비슷한 계산량(약 20 MIPS)으로 줄일 수 있을 것으로 기대된다.

3.4. 음성 압축기의 성능 분석

음성 압축기의 성능을 합성음의 스펙트럼 분석, Segmental SNR(SNRseg)을 이용한 Objective Test, MOS를 이용한 Subjective Test로 나누어 측정하였다. 그림 1은 여성과 남성의 유성음 구간에 대하여 제안된 음성 압축기의 성능을 IS-96A QCELP와 비교하여 보여준다. (a)는 여성 유성음에 대한 스펙트럼이고 (b)는 남성 유성음에 대한 스펙트럼이다. 스펙트럼의 Peak 위치와 크기의 관점에서 보면 (a)의 경우 주파수 0.21 근처, 0.33 근처, 0.4 이상에서 제안된 음성 압축기의 성능이 QCELP보다 특히 우수하고, (b)의 경우 주파수 0.2 근처, 0.38 이상에서 특히 제안된 음성 압축기의 성능이 우수하다. 표 3은 제안된 음성 압축기와 G.729 CS-ACELP의 성능을 동일한 조건에서



(a) 여성 유성음의 스펙트럼



(b) 남성 유성음의 스펙트럼

그림 2. 제안된 음성 압축기와 IS-96A QCELP의 합성음 스펙트럼 비교. (a)여성 유성음의 스펙트럼. (b)남성 유성음의 스펙트럼.

표 3. SNRseg 비교표

File	Length (sec)	CS-ACELP (dB)	제안된 음성 압축기 (dB)
1	22	11.25	11.65
2	336	10.52	11.45
3	370	12.50	12.80
평균	-	11.55	12.14

표 4. 음성 압축기의 MOS 비교표

신호	원음	제안된 음성 압축기	CS-ACELP	VSELP	QCELP
MOS	4.08	3.80	3.78	3.44	3.32

SNRseg로 비교한 것으로 모든 경우에서 SNRseg가 향상된 것을 알 수 있으며, 평균적으로 CS-ACELP에 비해 0.6dB 향상되었다.

조용한 환경에서 측정된 MOS가 표 4에 정리되어 있다. 다른 음성 압축기와의 성능 비교를 위하여 VSELP, QCELP, CS-ACELP의 MOS도 동시에 측정하였고, 객관적이고 신뢰성 있는 결과를 얻기 위하여 미국의 MOS 측정 전문업체인 Texas 소재 Dynastat에 의뢰하여 측정하였다. 모든 음성 압축기에 대하여 동일한 음성 데이터를 사용하고 동일한 환경과 조건에서 압축/복원하여 MOS 측정하였고 사용 언어는 모두 영어이다. 제안된 음성 압축기의 성능이 VSELP와 QCELP에 비해 매우 우수한 것을 볼 수 있고, CS-ACELP에 비해 MOS가 0.02 향상된 것을 볼 수 있다.

위의 결과에서 알 수 있듯이 주관적 평가, 객관적 평가에서 모두 제안된 음성 압축기가 기존의 다른 8kbps 표준 음성 압축기보다 성능이 우수하며, 특히 10msec 프레임 크기를 가지는 G.729 CS-ACELP 보다도 성능이 약간 우수하다. 이로 바꾸어 제안된 코드북 구조를 이용하여 음성 압축기의 성능을 크게 향상시킬 수 있는 것을 알 수 있다.

IV. 결 론

본 논문에서는 펄스와 랜덤 성분을 모두 가지고 피치 구조가 향상된 새로운 코드북을 제안하였다. 이 코드북은 Baseline 코드벡터와 Implied 코드벡터로 구성되며, Implied 코드벡터는 합성음을 피치 주기 만큼 지연시킨 신호를 기본으로 구하여진다. 따라서 Implied 코드벡터는 디코더로 전달될 필요가 없고 디코더에서 직접 구하여지며, 이 방법으로 전송율을 줄일 수 있다. 또한 피치 주기 이전의 신호를 분석하여 Implied 코드벡터를 구하므로 여기 신호의 피치 구조가 향상되어 우수한 음질을 합성

할 수 있다. 또한, 복합 여기 방식을 이용하여 다양한 특성의 여기 신호를 적절히 표현할 수 있어 음질 향상을 가져온다. 이러한 코드북을 가지는 8kbps의 CELP 음성 압축기를 설계하여 실시간 구현하고 성능을 분석한 결과 VSELP, QCELP에 비하여 음질이 상당히 향상되고, G.729 CS-ACELP보다 약간 음질이 향상된 것을 Objective Test와 Subjective Test를 통하여 확인하였다. 전송 오류에 의한 Implied 코드벡터 불일치문제, 비음성 신호에 대한 성능 등에 대한 연구가 현재 진행 중이다.

참 고 문 헌

1. B. S. Atal and S. L. Hanauer, "Speech analysis and synthesis by linear prediction of the speech waver," *J. Acoustic Society of America*, vol.50, pp.637-655, Aug. 1971.
2. T. E. Tremain, "The government standard linear predictive coding algorithm: LPC-10," *Speech Technology*, pp.40-49, April 1982.
3. M. R. Schroeder and B. S. Atal, "Code-excited linear prediction: High quality speech at very low rates," *Proc. IEEE ICASSP-85*, pp.937-940, 1985.
4. "ITU Recommendation G.728: Coding of speech at 16kbit/s using Low-Delay Code Excited Linear Prediction," 1992.
5. "ITU Recommendation G.729: Coding of speech at 8kbit/s using Conjugate-Structure Algebraic-Code-Excited Linear Predictive(CS-ACELP)," 1996.
6. I. Gerson and M. Jansuk, "Vector sum excited linear prediction(VSELP) speech coding at 8kb/s," *Proc. IEEE ICASSP-90*, pp.461-464, 1990.
7. TIA/EIA/IS-96A, *Speech Service Option Standard for Wideband Spectrum Digital Cellular System*, 1994.
8. TIA/EIA/IS-127, *Enhanced Variable Rate Codec, Speech Service Option 3 for Wideband Spread Spectrum Digital Systems*, 1997.
9. K. Mano et al, "Design of a pitch synchronous innovation CELP coder for mobile communications," *IEEE Journal on Selected Areas in Communications*, vol.13, pp.31-41, Jan. 1995.
10. J. Makhoul, R. Viswanathan, R. Schwartz and A. W. F. Huggins, "A mixed-source model for speech compression and synthesis," *J. Acoust. Soc. Amer.*, vol.64, pp.1577-1581, Dec. 1978.
11. S. Y. Kwon and A. J. Goldberg, "An enhanced LPC vocoder with no voiced/unvoiced switch," *IEEE Trans. Acoustics, Speech and Signal Processing*, vol.32, pp.851-858, Aug. 1984.
12. GSM 06.60, *Enhanced Full Rate(EFR) speech transcoding*, 1996.

▲박 호 증(Hochong Park)

1986년 2월: 서울대학교 전자공학과(공학사)

1987년 12월: University of Wisconsin-Madison, Dept. of
Electrical and Computer Engineering(M.S.)

1993년 5월: University of Wisconsin-Madison, Dept. of
Electrical and Computer Engineering(Ph.D.)

1993년 9월~1997년 8월: 삼성전자 정보통신본부 선임연
구원

1997년 9월~현재: 광운대학교 전자공학과 조교수

※주관심분야: 음성 및 영상 신호처리, 비선형 신호처리

▲권 순 영(Soon Young Kwon)

1964년 2월: 서울대학교 전자공학과(공학사)

1971년: University of Houston, Dept of Electrical Engi-
neering(M.S.)

1975년: University of Houston, Dept of Electrical Engi-
neering(Ph.D.)

1994년~현재: Telogy Networks Inc., U.S.A., Director

※주관심분야: 음성 신호처리, 통신 시스템