

차세대 패킷 전화용 광대역 음성 부호화기의 양자화에 대한 연구

김영보, 정병학, 박호중

광운대학교 전자공학과

Quantization on Wideband Speech Codec for Next Generation Packet Phone

Youngvo Kim, Byounggak Jeong, Hochong Park

Dept. of Electronics Engineering, Kwangwoon University
hcpark@kw.ac.kr

요약

패킷망을 통한 음성 통신이 발달됨에 따라 패킷 스위칭 채널 환경에서 계층적 구조를 가지는 광대역 음성 부호화기의 개발에 대한 요구가 늘어나고 있다. 본 논문에서는 이러한 차세대 패킷 전화용 광대역 음성 부호화기의 상위 대역에 대해서 효율적인 양자화 방법을 제안한다. 먼저 전체 프레임을 다수의 짧은 부프레임으로 구분하고, 각각의 부프레임에 MLT(Modulated Lapped Transform) 변환을 적용하여 주파수 영역으로 변환하여 2차원 구조의 데이터 행렬을 생성한다. 이러한 2차원 구조의 데이터를 크기와 부호로 분리하고, 크기는 2차원 DCT를 사용하여 시간과 주파수 영역에서의 신호 압축을 동시에 얻을 수 있게 하였다. 이와 같은 새로운 구조를 활용하여 기존의 방법보다 Energy Compaction 효과를 높이고 양자화 성능을 향상시킬 수 있었다. 또한 Core Layer의 부호화된 파라미터를 상위 대역의 양자화에 이용함으로써 그 성능을 향상시킬 수 있는 방법을 제안한다.

1. 서론

디지털 음성 통신에서 음성 압축 방법은 통신 시스템의 용량과 통신의 품질을 결정하는 매우 중요한 요인이며, 최근 음성 통신 기술의 발달은 고품질 음성 통신의 구현

을 요구하고 있다. 본 논문은 이러한 요구를 만족시키기 위해 기존의 8kHz로 Sampling 된 4kHz 대역폭의 협대역 음성 신호대신 16kHz로 Sampling된 7kHz 대역폭을 가지는 광대역 음성 신호를 위한 압축기를 제안한다. 또한, 패킷 스위칭 방식의 통신에서 주어진 채널 환경의 변화에 최적인 통신 품질을 제공하기 위해 채널 용량에 따라 전송 계층의 양을 조절가능한 계층적인 구조의 음성 압축기를 제안한다. 이러한 구조는 인터넷을 통한 음성 통신에서 필수적이며 효율적인 통신 서비스를 제공할 것이다. 설계되는 계층적 음성 압축기는 기존의 협대역 통신과의 호환성을 가지기 위해 대역폭 0~4kHz를 하위 대역이라 하고 4~7kHz를 상위 대역이라고 할 때 하위 대역에서는 널리 사용되는 G.729 CS-ACELP 음성 압축기를 사용하고, 이를 기준으로 상위 대역은 제안되는 음성압축기를 사용한다[1].

2. 광대역 신호의 계층 분할

대역폭 계층적인 구조의 음성 압축기를 설계하기 위해서 광대역 입력 신호는 4kHz를 기준으로 하위 대역과 상위대역으로 분리하여 처리한다. 대역 계층을 나누기 위하여 G.722에서도 사용되고 있는 QMF(Quadrature Mirror Filter)를 사용하며, 이것은 보다 나은 주파수 특성을 위하여 48차 FIR 필터로 구성하여 사용한다.[3]

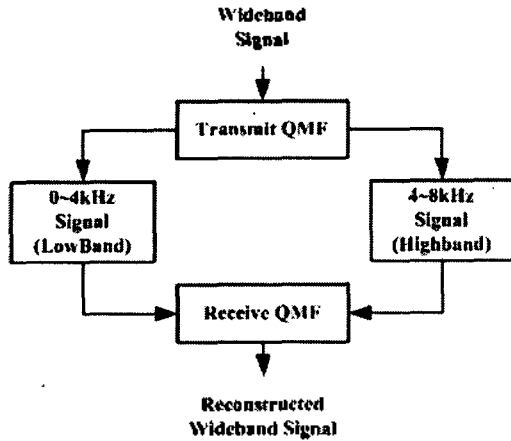


그림 1. QMF를 이용한 대역 계층 분할

3. 제안된 음성 압축기의 구조

3.1 음성 압축기의 전체 구조

본 논문에서 제안한 상위대역 음성 압축기의 전체 규격과 세부적인 Delay는 표 1과 같다. 상위대역 압축기의 전송률은 8kbps이며, 하위대역에 G.729 역시 8kbps를 가지므로 광대역 신호를 위한 압축기의 전송률은 16kbps가 된다. 또한 Sampling Rate가 16kHz인 광대역 신호는 QMF를 통하여 하위대역과 상위대역으로 분리될 때 Downsampling 되므로 각각의 Sampling Rate 역시 8kHz로 변한다. 음성 압축기의 프레임 크기는 30msec이고, 6개의 5msec 부프레임으로 구성되며, Look-Ahead 크기는 5msec이다.

표 1. 상위 대역 압축기의 규격과 Delay

| Sampling Rate | 8kHz | | Delay |
|---------------|--------|--------------|--------------|
| Bit Rate | 8kbps | Transmit QMF | 1.46875 msec |
| Frame Size | 30msec | Look-Ahead | 5.00000 msec |
| Subframe Size | 5msec | Receive QMF | 1.46875 msec |
| Look-Ahead | 5msec | Total Delay | 7.95750 msec |

Encoder와 Decoder의 간단한 전체 구조는 그림 2와 같다. Encoder는 광대역 입력신호를 QMF를 통하여 하위대역과 상위대역으로 분리하여 하위대역의 G.729 Encoder와 상위대역의 설계되는 Encoder에 의해 전송 파라미터를 구한다. 반대로 Decoder에서 하위대역 전송 파라미터는 G.729 Decoder를 통해 하위대역 신호를 복원하고

상위 대역 파라미터는 설계되는 Decoder로 상위대역 신호를 복원하여 QMF를 통해 다시 합성된다.

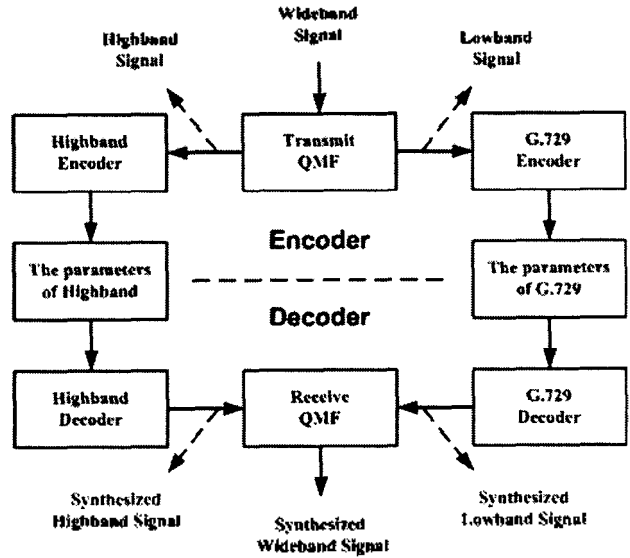


그림 2. 계층적 광대역 음성 압축기의 전체 구조

3.2 상위 대역 압축기의 구조

개발하는 음성 압축기는 Transform Coding 구조에 기반을 둔다. 상위대역 압축기의 전체 프레임 길이는 30msec이고 6개의 부프레임으로 나누어진다. 상위대역의 입력 신호는 5msec 길이의 부프레임 단위로 주파수 영역으로 변환되며 이 때 G.722.1에도 사용되는 MLT(Modulated Lapped Transform)를 사용한다[2]. 효율적인 압축기의 설계를 위하여 MLT 계수들은 직접 양자화에 이용되지 않고 부호(Sign)와 크기(Magnitude) 성분으로 분리된다. 부호가 분리된 크기의 값들은 시간 영역과 주파수 영역에서 보다 높은 상관관계를 가지게 되므로 이차원 DCT를 실행하여 양자화 효율을 높이며, 이차원 DCT 후 얻어진 이차원 data들은 하위대역의 정보와 에너지의 분포도 등을 이용하여 양자화 과정을 거쳐 최종적으로 전송 파라미터들을 생성하게 된다. 본 연구에 모든 실험의 입력 신호로는 NTT DB의 한국어 음성 샘플 8분 분량을 목음을 제거하여 사용하였다.

4. 상위 대역 음성 압축기의 양자화

그림 3은 상위 대역 음성 압축기의 Encoder에 대한 전체 구조를 보여준다.

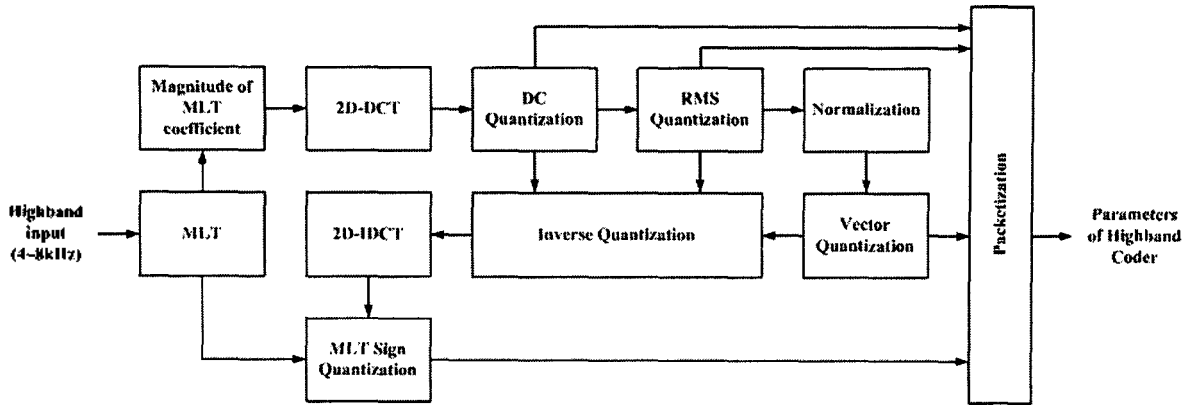


그림 3. 상위 대역 음성 압축기의 Encoder

4.1 MLT 변환

MLT는 그림 4와 같이 50%의 Overlap Window를 사용하기 때문에 현재의 부프레임 5msec와 Look-Ahead 5msec로 구성되어 총 10msec가 되며, 한 프레임은 총 30msec이므로 시간 순서대로 6번의 MLT를 실행하게 된다.

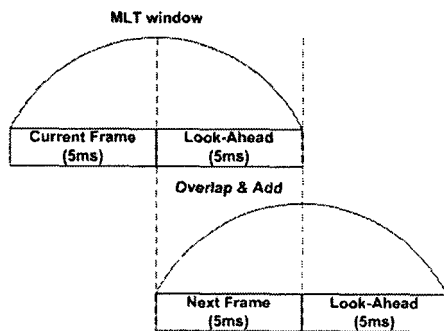


그림 4. 사용되는 MLT의 구조

사용되는 MLT에 입력 sample 수는 80개이고 4~7kHz에 해당하는 MLT 계수는 30개이다. 부프레임은 6개이므로 시간과 주파수의 두 영역의 압축효과를 누리기 위해 그림 5와 같이 이차원 형태로 구성한다.

| | 4-4.6kHz | 4.6-5.2kHz | 5.2-5.9kHz | 5.9-6.4kHz | 6.4-7kHz |
|---------|----------|------------|------------|------------|----------|
| 0-10ms | | | | | |
| 5-15ms | | | | | |
| 10-20ms | | | | | |
| 15-25ms | | | | | |
| 20-30ms | | | | | |
| 25-35ms | | | | | |

그림 5. MLT의 이차원 결합 구조

각 부프레임의 MLT 계수들은 600hz의 간격의 5개 대역으로 분리되어 각 대역 당 계수는 6개가 되며, 한 프레임은 시간의 순서대로 6개의 부프레임을 가지므로 동일 주파수에 대하여 6개의 계수가 시간의 순서대로 존재하게 된다. 이것을 6*6 Matrix 형태의 5개 대역으로 구성

한다. MLT 계수들 사이의 시간과 주파수에 대한 상관관계 계수 값을 측정해보면 평균적으로 0.01 정도로 매우 작다. 그러나 크기(Magnitude)와 부호(Sign)성분으로 분리 후 상관관계를 측정해보면 평균적으로 0.57 정도의 상관관계 계수 값을 가진다. 이러한 점을 이용하여 이차원으로 구성된 MLT 계수들은 오직 그 크기의 정보만이 이차원 DCT로 처리되며, 각 계수의 부호에 관한 정보는 별도로 처리하게 된다.

4.2 이차원 DCT

MLT 계수의 크기 값들은 효율적인 양자화를 위해 시간과 주파수의 두 가지 측면에 대해 DCT를 취하여 Energy Compaction의 효과를 향상시킨다. 그림 6은 이차원 DCT 후 6*6 data의 평균적인 에너지의 dB 값을 보여준다.

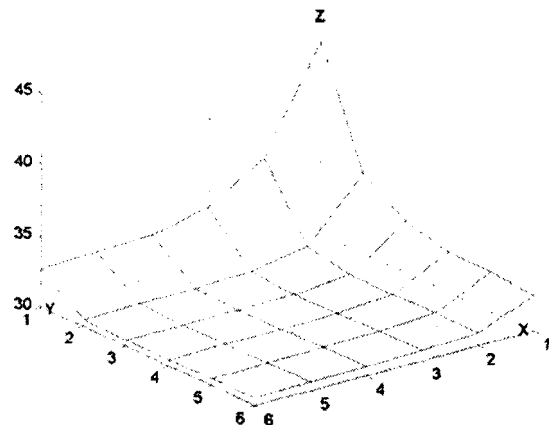


그림 6. 이차원 DCT 후 에너지 분포도

5개 대역의 이차원 data들에 대한 평균적인 값이지만 각 대역 에너지 분포는 모두 유사한 형태를 가진다. DC에 해당하는 X,Y좌표 [1,1]의 값이 매우 크며 DC를 제

외한 다른 값들은 대부분 첫 번째 열과 행에 모여 있다. 이러한 분포는 이차원 DCT로 인하여 시간과 주파수 영역에서 동시에 Energy Compaction의 효과를 가지는 것을 보여주며 양자화 방식 결정에 큰 영향을 주게 된다.

4.3 이차원 DCT data들의 양자화

각 대역 6*6 Matrix 형태의 data들은 에너지가 큰 영역에 더 많은 bit를 할당하여 성능을 높이고 에너지가 작은 영역에는 양자화 성능을 낮추는 대신 더 적은 bit를 할당하기위해 그림 6의 결과에 따라서 평균적 에너지 크기의 순서대로 1차원으로 재배열한다. 이차원 data의 에너지 분포에 따라 DC 위치에 해당하는 값은 개별적으로 양자화를 하며 나머지 35개 data들은 RMS로 정규화하여 Split Vector Quantization(SVQ) 과정을 거치게 된다. RMS 값 역시 개별적인 양자화를 하게 된다. 35개의 data들은 SVQ를 위해 분할이 되는데 처음 11개의 data에 90% 이상의 에너지가 분포하므로 5차와 6차로 Vector를 구성하여 각각 SVQ를 하고 나머지 24개 data들에 대해서는 Codebook 크기를 줄이기 위해 차수가 동일한 3개의 8차 Vector로 구성하여 SVQ를 하였다. 또한 대역별 에너지 집중도를 고려하기 위해 각 대역 RMS의 크기 순서에 따라 일부 data만을 사용하여 복호화하는 실험을 하였다. 그 결과 순위가 낮은 대역의 data들은 일부만을 복원해도 청각적인 면과 에너지 손실 측면에서 모두 우수한 성능을 얻을 수 있었으며 bit 절감의 효율도 얻게 된다.

4.4 MLT 계수 부호 양자화

MLT 계수의 수와 같이 부호는 부프레임 마다 30개씩 존재하므로 한 프레임에서 부호는 총 180개가 존재한다. 제안된 음성 압축기의 bit 제한으로 인하여 부호를 위해서 92bit가 할당되었기 때문에 복호화된 MLT 계수 크기의 순서대로 92개의 MLT 계수에 대해서만 부호를 할당하게 된다. 부호를 전달 받지 못하는 88개의 계수들에 대해서는 여러 방법을 적용해보았지만 Random 성질로 인하여 이득을 얻지 못하였기 때문에 Random 함수를 이용하여 생성된 부호를 사용하였다.

4.5 DC와 RMS의 양자화

Log를 취한 상태에서 각 대역의 DC 사이에는 매우 높은 상관 관계가 존재하며 첫 번째 대역의 DC는 G.729에서의 Codebook Gain 값과 높은 상관 관계를 가진다. 이러한 상관 관계를 이용하기위해 예측오차를 Scalar 양자화한다. RMS 값 역시 Log를 취한 상태에서 각 대역에 DC와의 상관 관계가 매우 높은 점을 이용하여 예측 오차에 대하여 Scalar 양자화를 한다.

5. 결론

본 논문은 계층적 구조를 위한 음성 압축기와 양자화 방법을 제안하였다. 제안된 압축기는 시간과 주파수 영역에서의 상관관계를 적극적으로 이용한 구조를 가지며 양자화 방법 또한 이차원 변환 구조의 특성을 중점으로 연구되었다. G.722.1 광대역 음성압축기와 제안된 음성 압축기를 WPESQ(Wideband Perceptual Evaluation of Speech Quality)로 비교할 때 제안된 음성 압축기에 G.729는 성능이 떨어지고 중요한 점은 상위대역에 관한 결과이기 때문에 왜곡이 덜한 하위대역의 정보를 위해 G.729 Annex E 11.8kbps를 사용하여 비교해보았다. 그 결과 제안된 계층적 음성압축기의 상위대역은 8kbps의 적은 bit rate를 가지면서도 효율적인 결과를 보여주었으며, 아직 미세하게 남아있는 잡음의 문제를 해결하기 위하여 지속적인 연구가 진행 중에 있다.

참고문헌

- [1] ITU-T Rec. G.729, "Coding of speech at 8kbit/s using conjugate-structure algebraic-code-excited linear-prediction(CS-ACELP)," Mar. 1996.
- [2] ITU-T Rec. G.722.1, "Coding at 24 and 32 kbit/s for hands-free operation in systems with low frame loss," 1999
- [3] ITU-T Rec. G.722, "7 kHz audio-coding within 64 kbit/s." 1988.