AMR 부호화기와 결합된 다전송률 광대역 음성부호화기 설계

김윤주* 이호창* 이인성*
송북대학교 전파공학과

Design of Multi Rate Wideband Speech Coder Using the AMR(Adaptive Multi-Rate) Coder

Eunjoo Kim*, Hocchang Lee*, Insung Lee*
Dept. of Radio Engineering, Chungbuk National Univ.

요 약문
본 논문에서는 AMR(Adaptive Multi-Rate)를 이용하여 광대역 음성부호화기를 설계하였다. 16kHz로 샘플링된 임력 신호를 QMF 필터링에 의해 두 개의 8kHz 샘플링 신호로 변환시킨 후 각 신호의 음성신호는 AMR (Adaptive Multi-Rate)과 ATC(Adaptive Transform Coding)를 사용하여 각각 부호화되어 전송된다. 두 대역으로부터 부호화된 신호는 20.2kbps에서 12.7kbps까지의 전송률을 갖고, 수신단에서는 각 대역을 AMR와 ATC방법으로 역부호화하여 음성신호로 재구성한다. 설계된 광대역 음성부호화기의 성능을 평가하기 위해 ITU-T의 표준인 G.722를 포함하여 MOS 시험을 하였다.

I. 서론
2001년부터 세계적으로 제정한 IMT-2000 이동통신 서비스의 가장 큰 특징은 낮은 대역폭(256kbps)을 사용하여 음성에서 통화까지 다양한 벽이므로 데이터를 전송할 수 있다는 것이다. 기존의 음성 위주의 이동통신 서비스는 데이터, 영상데이터를 포함한 벽이므로 통신이 가능하게 된다. 음성통신을 위해서도 기존의 벽이미 음성 신호의 외관이 동일한 음성 및 통신은 가능하다. 또한 최근에 활발하게 보급되고 있는 고속 인터넷에서는 음성 대역폭의 확대로 음성과 잘 맞는 음직이 음성 및 오디오 전송도 실시간 전송이 가능하게 되었다. 광대역 음성신호는 핵심(200Hz~3400Hz) 충성신호보다 낮은 대역폭(50Hz~7000Hz)을 가지며 16kHz 샘플링 주기를 요구하며 더 beaucoup 음질과 좋은 음질을 나타낸다. 이처럼 보다 자연스러운 음질을 벽이미 통신에서 느낄 수 있어야 한다. 따라서 본 논문에서는 IMT-2000 서비스의 음성부호화기의 표 준으로 제작된 AMR(Adaptive Multi-Rate)를 사용하여 광대역 음성부호화기를 설계하였다. AMR 부호화는 8개의 전송모드로 재날 활용에 적용하여 다목적으로 대처하면서 toll quality를 갖는 부호화 방식이다. 본 논문에서 설계된 광대역 음성부호화 알고리즘은 32kbps마다 전송률이 작고도 음질의 품질이 높은 전송률에서 작동하도록 설계하였으며, 확장된 시스템, 고급전화기와의 중단 전화 등에 사용가능하다. 또한 다전송률 부호화 방식으로 이용

II. 광대역 음성부호화기의 인코더 알고리즘
본 논문에서 설계된 광대역 음성부호화기의 인코더 알고리즘은 그림 1에 나타나 있다. 16kHz로 샘플링된 임력 신호를 QMF 필터에 의해 두 개의 대역으로 나누어지 고, 각각 decimation하여 두 개의 8kHz 샘플링 신호로 변환시킨 후 저색률(0Hz~3400Hz)의 신호와 고색률(3400Hz~7000Hz)의 신호로 나누어 각각 부호화되어 전송된다.4(45)에서 저색률의 신호는 ETSI와 3GPP에서 제안한 IMT-2000 서비스의 음성부호화기 표준으로 제작한 AMR를 이용하여 4.75kbps에서 12.2kbps까지 8개의 전송률로 부호화되어 전송된다. 고색률의 신호는 ATC(Adaptive Transform Coding)방법으로 8kbps의 전송률로 부호화되고, 변환은 DCT (Discrete Cosine Transform) 알고리즘을 이용한다.7 저색률과 고색률의 부호화된 정보는 전자적으로 20.2kbps에서 12.7kbps까지의 전송률을 갖는다.

2.1 저색률 인코더 알고리즘
QMF필터에 의해 분리된 0Hz~3400Hz 대역의 음성 신호는 AMR 인코더로 부호화된다. AMR는 AECIP에 기반을 두고 있다. 또한 전송률이 저속을 가감으로써

그림 1. 광대역 음성부호화기의 인코더
MR-ACELP라고 하기도 한다. 기본 목적은 음성부호화
외 무선 채널의 환경변화에 적응적으로 대처해야 한다.
또한 멀티카이드 손실을 가진다는 장점이 있다. 이에 ETSI
서에서는 전송율을 4.75kbps에서 12.2kbps까지 8개로
나누고 하나의 재생율을 주변경량 인코딩 모드를 갖는
한 가지 동일한 음성부호화기술에 구성된다. 프레임 시즈
는 8kb로 생략된 5ms의 4개의 프레임을 구성
된 20ms크기를 갖는다. 모드에 따른 사용되는 전송율
은 표1에 주어진다.

<table>
<thead>
<tr>
<th>Codec mode</th>
<th>Source codec bit rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>AMR_12.20</td>
<td>12.20 kbps (GSM EFR)</td>
</tr>
<tr>
<td>AMR_16.20</td>
<td>16.20 kbps</td>
</tr>
<tr>
<td>AMR_7.95</td>
<td>7.95 kbps</td>
</tr>
<tr>
<td>AMR_7.40</td>
<td>7.40 kbps (GSM FR1)</td>
</tr>
<tr>
<td>AMR_8.10</td>
<td>8.10 kbps (GSM FR2)</td>
</tr>
<tr>
<td>AMR_8.70</td>
<td>8.70 kbps (GSM FR3)</td>
</tr>
<tr>
<td>AMR_8.90</td>
<td>8.90 kbps</td>
</tr>
<tr>
<td>AMR_8.15</td>
<td>8.15 kbps</td>
</tr>
<tr>
<td>AMR_4.75</td>
<td>4.75 kbps</td>
</tr>
<tr>
<td>AMR_4.70</td>
<td>4.70 kbps</td>
</tr>
</tbody>
</table>

表1 AMR코덱의 모드별 비트율
GSM-EFR : ETSI GSM06.90 TIA/EIA IS-641 TDM A Enhanced Full Rate
AMR_EFR : ARIB 67kbps Enhanced Full Rate
AMR_SID : 프레임이 연속적일 경우

AMR 인코더의 구조는 그림 2에 나타난다. 처음에
모든 모드에 대해서 high-pass filtering을 apply scaling을 하여 전처리 과정을 한다. LP 분석과 양자화
과정은 12.2kbps mode에서는 락 필터 LP 필터를 가
지고 30ms 비대칭 원도우를 이용해서 두 번 실행된다.
이때 lookaheads는 사용되지 않고, Levinson-Durbin 알고리즘을 이용해 계수를 구한 후 압축하니. 나타나
또는 보다 빠른 원도우를 5ms lookahead을 사용한
다. 계산된 LP filter 계수는 양자화와 보장을 위하여
LSP(Linear Spectrum Pair)로 만든다. Algebraic 코드북
정리자는 interleaved single-pulse permutation design
(ISPP)에 기반을 두고 있다. 각 챔버를 통과하고 음성
신호가 각 챔버의 합설립을 통과한 음성신호 사이의
mean square error를 최소화시키기 위한 코드북
인덱스를 검색한다. 모든 코드에 대하여 적절 코드북
이득을 위해 적절 양자화를 실행하고 algebraic 코드북
이득을 위하여 고정계수를 가진 MA-예측장치를 수
정이다. innovation 예측과 예측된 에너지를 구하여
에너지가 적은 값을 찾아 각 factor를 이용하여 양자
화하게 된다.

2.2 고해역 인코더 알고리즘
고해역에서는 GMPE 필터에 의해 분리된 3400kHz-
7000kHz 범위의 음성신호를 각 프레임 단위로 DCT변환
되어 주상수 영역의 값으로 변환된다. 고해역 인코더
에서 사용한 DCT변환은 다음과 같이 주어진다[8][9].

\[
S(k) = \sum_{n=0}^{N} s(n) \cos \left( \frac{(2n+1) \pi k}{N} \right)
\]

여기서 \(s(n)\) : 입력신호
\(S(k)\) : DCT계수

\[
\lambda(k) = \begin{cases} 
1 & k = 0 \\
\sqrt{2} & \text{otherwise} 
\end{cases}
\]

그림 2. AMR 인코더의 구조

그림 3. 고해역 음성부호화기의 인코더

그림 3은 고해역 신호의 인코더 과정으로 보여준다. 핸드
데릴 내에서 음성의 스펙트럼은 모든 대역에 걸쳐 균등하게
분포되어 있지 않으므로, 모든 DCT계수에 같은 크기의
비트를 할당하여 양자화 하는 것은 비효율적이다.
즉, 수 놀리의 스펜트럼이다 다른 부분의 스펜트럼보다
크게 나타나면, 스펜트럼은 높 영역에 더 많은 비
트를 할당하는 것이 효과적이다. DCT계수는 주상수 영
역의 값이므로 스펜트럼의 크기는 DCT계수에 연관하
며 표현된다. 주 주상수 영역의 스펜트럼은 다른 영역
보다 크게 나타난다면, 그 주상수 영역에 해당하는
DCT계수의 에너지는 다른 영역의 에너지보다 큰
값을 갖는다. 이러한 성질의 균등성과 DCT계수의 관계를
이용하여 고해역을 구분한다. 변환된 DCT계수는 각
개의 서브플록으로 나누어지고, 각 서브플록마다 에너
지가 계산된다. 계산된 각 서브플록의 에너지값들은
에너지 벡터 코드를 이용하여 양자화되고, 코드북
인덱스값을 얻는다. 에너지 벡터 코드 인덱스값들은 각
서브플록에 해당하는 비트 정보를 포함하는 비트 벡터
코드북과 연계하여 설계된다. 그러므로 에너지 벡터코
드북의 인덱스는 DCT계수를 바탕 양자화할 때, 각
서브플록에 해당하는 전송 비트 정보를 제공한다 각 서
브플록내의 DCT계수들은 에너지 벡터 코드북의 에너
지값으로 나누어져서 결정하고, 비트 벡터 코드
북의 비트정보를 이용하여 각 서브플록의 양자화
됨 DCT계수는 벡터 양자화된다.

그림 4는 에너지 벡터 코드북과 비트 벡터 코드북 그
리고 벡터 코드북의 관계가 벡터 양자화 구조를 나타낸다.
이하 같은 구조로 고해역의 신호는 양자화되어
8kbps의 전송을 갖도록 설계한다[10][11].
그림 6. 저대역 음성부호화기의 디코더

고대역의 정보는 역변환 컨텍스트와 프레임에 IDCT(Inverse Discrete Cosine Transform)과정을 통해 출력 신호를 만들어낸다. 역변환도 두 템파에 QMF Synthesis Bank에 의해 합치여지며 최종 출력 신호를 만들어낸다

IDCT의 식은 다음과 같이 주어진다[8].

\[ s(n) = \frac{1}{N} \sum_{k=0}^{N-1} S(k) \lambda(k) \cos \left( \frac{(2n+1)\pi k}{2N} \right) \]

여기서, \( s(n) \) : 입력 신호
\( S(k) \) : DCT 계수
\( \lambda(k) = \begin{cases} 1 & k = 0 \\ \sqrt{2} & \text{otherwise} \end{cases} \)

고대역은 한 글록에 해당하는 이중과 비트 활성 정보는 부파조보로부터 얻어진다. 얻어진 이중, 비트 활성 정보는 DCT계수의 배타양자와 인덱스와 결합하여 DCT계수를 복원한다. 복원된 DCT계수는 IDCT의 과정을 통해 원래의 고대역 신호를 복원한다.


그림 4. 벡터 양자화기의 구조

그림 5. 햄태역 음성부호화기의 디코더

Demultiplexer이 의해 분리된 저대역의 신호는 AMR 디코더 알고리즘에 의해 역부호화된다. AMR의 디코더의 구조는 그림6에서 나타내고 있다. LP 저파미터 출력 코드북 벡터, 이득, 고정 코드북 벡터들을 수신하고 디코딩하여 합성음을 얻기 위해 과정을 수행한다. 여기서, 출력은 창조적인 코드북 벡터와 이득에 의해 구해진다. 그리고 음성은 LP 합성 템파를 통과하여 excitation의 패턴을 통해 복원된다.

VI 참고 문헌

[1] F.K. Soong and B.H. Juang, "Line spectrum pair (LSP) and speech data compression", in Proc. ICASSP'84, pp. 1.103-1.106.