

AMR과 EVRC 음성 부호화기간의 비탠덤 방식을 이용한 상호 부호화

Tandemless Transcoding for AMR and EVRC Speech Coders

이 선 일*, 유 창 동*
(Sunil Lee*, Chang D. Yoo*)

* 한국과학기술원 전자전산학과

(접수일자: 2002년 2월 27일; 채택일자: 2002년 7월 29일)

본 논문에서는 AMR과 EVRC 음성 부호화기간의 비탠덤 (Tandemless) 방식을 이용한 상호 부호화 방법이 제안되었다. 제안된 방법은 기존의 탠덤 (Tandem) 방식의 상호 부호화 방법과 달리 음성 신호를 다시 복호화했다가 부호화하지 않고, CELP 계열의 음성 부호화기들이 공통적으로 사용하는 파라미터들을 직접 변환한다. 상호 부호화는 LSP 변환, 적응 코드북을 위한 피치 지연 값 및 적응 코드북 이득 변환, 고정 코드북 벡터 및 고정 코드북 이득 변환으로 구성되어 있다. 제안된 방법을 객관적, 주관적 방법으로 평가한 결과 기존의 탠덤 방식에 비하여 적은 계산량과 지연 시간으로 탠덤 방식과 최소 동등, 혹은 우월한 음질을 얻을 수 있다는 것을 확인했다.

핵심용어: 음성 부호화, AMR, EVRC, 상호 부호화, 탠덤 방식, 비탠덤 방식

투고분야: 음성처리 분야 (2.2)

Novel tandemless transcoding method for AMR and EVRC speech coders is proposed in this paper. In contrast to conventional tandem method, the parameters which is used commonly in speech coder where CELP algorithm is adapted are directly transcribed. The proposed algorithm is composed of LSP transcoding, pitch delay transcoding, gains transcoding and fixed codebook vector transcoding. Evaluation results show that the novel algorithm achieves better speech quality than tandem method and reduce computational complexity and delay.

Keywords: Speech coding, AMR, EVRC, Transcoding, Tandem, Tandemless

ASK subject classification: Speech signal processing (2.2)

I. 서론

최근 몇 년간에 걸친 제 3세대 (3rd Generation) 무선 이동통신 시스템, 일명 IMT-2000 (International Mobile Telecommunication 2000)의 표준화 작업 결과 서로 다른 두 방식 (동기식과 비동기식)이 함께 표준안으로 선정된 바 있다. 또한, 우리나라에서의 IMT-2000 서비스 사업자 선정 결과 두 개의 비동기식 방식의 사업자와 한 개의 동기식 사업자가 선정되었다. 결국 제 3세대 무선

이동 통신망은 비동기식과 동기식이 함께 운영되게 될 것이다. 이에 따라 서로 다른 두 통신망에서 각기 사용되는 음성 부호화 방식간의 호환성에 대한 연구가 요구된다. 본 논문에서는 비동기식 IMT-2000 시스템의 표준 음성 부호화기인 Adaptive Multi-Rate (AMR)[1] 음성부호화기와 2세대 동기식 이동 통신망에서 표준 음성 부호화기로 널리 쓰이고 있는 Enhanced Variable Rate Codec (EVRC)[2] 음성부호화기간의 상호 부호화 방법을 제안한다. Algebraic Code-Excited Linear Prediction (ACELP) 알고리즘에 기반을 둔 AMR 음성부호화기는 최소 4.75 Kbit/s에서 최대 12.2 Kbit/s에 이르는 8개의 전송율을 가지는 가변 음성부호화기이다. 이에 반해 EVRC

책임저자: 이선일 (cowboysun@kaist.ac.kr)
305-701 대전광역시 유성구 구성동 373-1
한국과학기술원 LG-Semicon Hall 유창동 교수님 연구실
(전화: 042-869-5470; 팩스: 042-862-0559)

음성부호화기는 Relaxation CELP (RCELP)[3] 알고리즘에 기반을 두고 있으며 0.8 Kbit/s에서 8.55 Kbit/s에 이르는 3개의 동작 모드를 가지고 있다.

일반적으로 서로 다른 두 음성부호화 표준간의 호환성 문제를 해결할 수 있는 가장 단순한 방법은 한쪽 방식으로 부호화(Encoding)된 것을 음성 신호로 복호화(Decoding)한 다음 이를 다시 변환하고자 하는 방식으로 부호화하는 것이다. 이를 텐덤 방식이라 한다. 그림 1의 (a)에서 보듯이 텐덤 방식에 의한 음성부호화기(Coder) A에서 음성부호화기 B로의 상호 부호화는 음성부호화기 A에 의해 부호화된 음성 신호를 음성부호화기 A의 복호화기를 써서 펄스 부호 변조(PCM: Pulse Coded Modulation) 형식의 음성 신호로 복호화한 후 이를 다시 음성부호화기 B에 의해 부호화한다. 음성부호화기 B에서 음성부호화기 A로의 상호 부호화는 그 역의 과정이다. 그러나 이러한 텐덤 방식에 의한 상호 부호화는 두 번의 부/복호화 과정으로 인해 음질의 저하(Quality loss)가 일어날 뿐 아니라 지연 시간(Delay)과 계산량(Computational Complexity) 또한 증가하는 문제점을 안고 있다.

이러한 문제점들을 개선하기 위한 방법으로 한쪽 음성 부호화 표준으로 부호화하여 구한 비트열(Bit-stream)

을 다른 쪽 방식의 음성부호화기를 위한 비트열로 직접 변환하는 비텐덤(Tandemless) 방법이 있다. 그림 1의 (b)에서 볼 수 있듯이 비텐덤 방식에 의한 음성부호화기 A에서 음성부호화기 B로의 상호 부호화는 음성부호화기 A에 의해 부호화된 비트열에서 음성 신호를 생성하지 않고 직접 음성부호화기 B에 의해 부호화된 비트열을 생성한다. 음성부호화기 B에서 음성부호화기 A로의 상호 부호화는 그 역의 과정이다. 제안된 비텐덤 방식에 의한 상호 부호화에선 Code-Excited Linear Prediction (CELP) 계열의 음성 부호화기에서 공통적으로 부호화되는 음성 특징 파라미터들을 이러한 비텐덤 방법을 써서 직접 변환하게 된다. 비텐덤 상호 부호화 방법은 두 번의 부/복호화 과정을 모두 거치는 텐덤 방법보다 짧은 지연 시간과 적은 계산량으로 동등한 품질의 음성 신호를 복원할 수 있다.

본 논문에서는 이러한 비텐덤 방법을 사용하는 AMR과 EVRC 음성부호화기간의 상호 부호화 알고리즘을 제안한다. 제안된 상호 부호화 알고리즘은 크게 LSP 변환과정, 적응 코드북을 위한 피치 변환과정, 적응 코드북 및 고정 코드북 이득 변환 과정, 그리고 고정 코드북 변환 과정으로 구성된다. 여러 가지 음성 신호로 모의 실험한 결과,

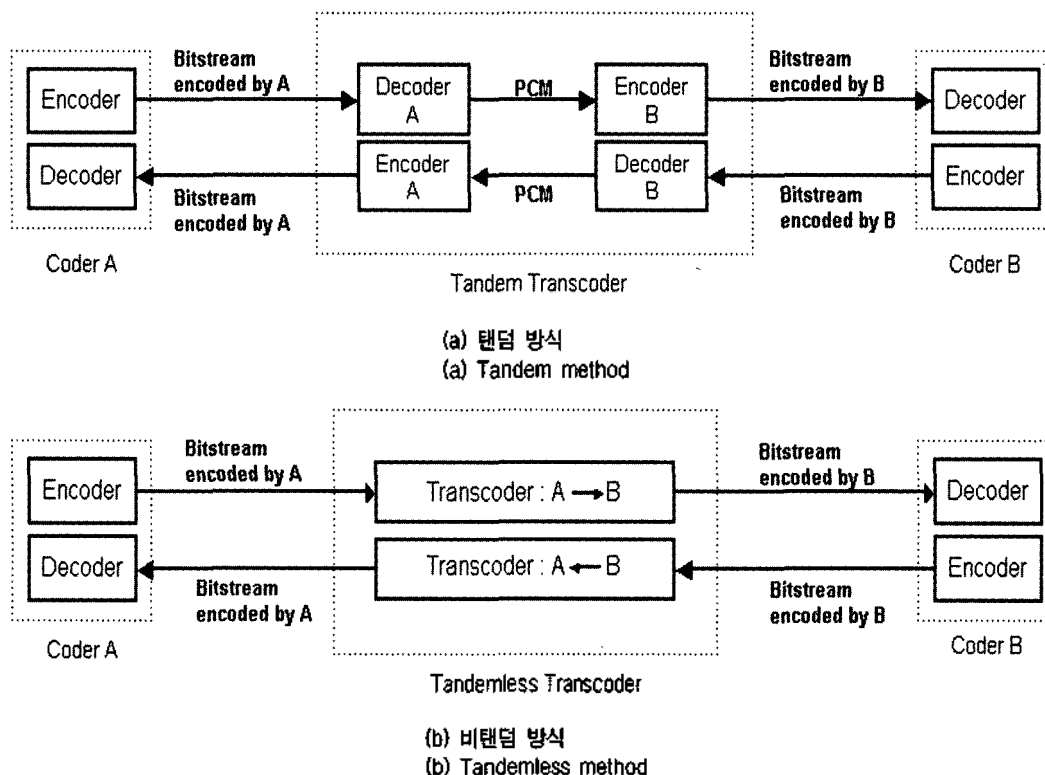


그림 1. 텐덤방식과 비텐덤방식의 비교
Fig. 1. Comparison between tandem and tandemless method.

본 논문에서 제안한 상호 부호화 알고리즘이 두 번의 부/복호화 과정을 거친 탠덤 방법보다 짧은 전달 지연 시간과 적은 계산량으로 동등한 음질의 음성 신호를 복호화하는 것을 확인하였다.

본 논문의 구성은 다음과 같다. 제 2 장에서는 상호 부호화 대상인 AMR 음성부호화기와 EVRC 음성부호화기에 대해서 살펴보고, 제 3 장에서는 AMR과 EVRC 음성부호화기 간의 상호 부호화 과정을 설명한다. 제 4 장에서는 주/객관적 평가를 통하여 제안된 알고리즘과 기존의 탠덤 방식의 의한 상호 부호화의 성능을 비교한다. 끝으로 제 5 장에서 결론을 맺는다.

II. AMR과 EVRC 음성 부호화기

2.1. AMR 음성 부호화기

3rd Generation Partnership Project (3GPP)의 Adaptive Multi-Rate (AMR) 음성부호화기는 제 3 세대 이동통신 시스템인 비동기식 IMT-2000의 표준 음성부호화기이다. AMR 음성부호화기는 합성에 의한 분석 (Abs: Analysis-by-Synthesis) 방식에 기반한 ACELP를 음성부호화 알고리즘으로 사용하고 있으며, 주어진 채널 환경에서 최적의 음질을 유지할 수 있도록 모두 8개의 서로 다른 음성 부호화 모드—12.2, 10.2, 7.95, 7.4, 6.7, 5.9, 5.15, 4.75 Kbit/s—로 동작한다.

AMR 음성부호화기는 음성신호를 프레임 단위로 합성에 의한 분석 방법을 이용한 선형예측 방법을 통해 부호화하며, 여기신호는 ACELP 방법에 의해 적응 코드북 (ACB: Adaptive Codebook)과 고정 코드북 (FCB: Fixed Codebook), 양 코드북으로부터의 코드 벡터들의 합으로 모델링한다. 즉, AMR은 현재 입력 음성 프레임으로부터 선형예측계수 (LPC: Linear Prediction Coefficients), 적응 코드북 (ACB: Adaptive Codebook) 및 적응 코드북 이득, 고정 코드북 (FCB: Fixed Codebook) 및 고정 코드북 이득과 같은 음성 특징 파라미터들을 추출하여 부호화한 후 전송한다. AMR의 프레임 길이는 20 ms이며 12.2 kbit/s 모드를 제외한 나머지 전송 모드에서는 5 ms의 예견구간 (look-ahead)을 가지고 있다. 따라서 최대 25 ms의 알고리즘 지연이 생기게 된다.

AMR 음성 부호화기는 다음과 같은 순서로 음성을 부호화한다. 샘플링 주파수가 8000 samples/s이고, 13-bit 선형 펄스 부호 변조 형식인 입력 음성 신호는 먼저 전처리 과정을 거치면서 80 Hz 이하의 저주파 성분이 제거되

고 오버플로우 (Overflow) 방지를 위해 크기가 1/2 로 줄어든다. 전처리 과정을 거친 음성은 프레임당 두 번 (12.2 kbit/s), 혹은 한 번 (나머지 모드)씩 선형 예측 분석이 이루어진다. 여기서 구해진 선형예측계수들은 양자화 (Quantization) 과정과 전송 중의 오류에 강인하게 만들기 위해 Line Spectrum Pair (LSP)로 변환된 후 부프레임 (Subframe) 별로 보간되고, 한편으로는 벡터 양자화되어 그 인덱스가 전송된다. 양자화된 LSP 역시 부프레임 별로 보간된 후 선형예측계수로 변환되어 각 부프레임을 위한 합성 필터를 구성하게 된다. 적응 코드북 분석은 매 부프레임당 한번씩, 두단계에 걸쳐 수행된다. 먼저 개회로 피치 지연 검색을 통해 정수 피치 지연을 구한 후, 폐회로 피치 지연 검색을 통해 구해진 정수 피치 지연 부근에서 1/3 또는 1/6의 정확도를 찾는 최적의 피치 지연 값을 부샘플 (Sub-sample) 단위로 찾아낸다. 한편 적응 코드북 이득은 구해진 적응 코드북 벡터를 이용하여 계산된다. 고정 코드북 분석은 매 부프레임당 한 번씩 수행된다. 고정 코드북 분석을 통해 고정 코드북을 구성하는 각 펄스의 부호와 위치가 결정된다. 이렇게 찾아진 고정 코드북 벡터를 이용하여 고정 코드북 이득 (FCB gain)이 계산된다. 적응 코드북 이득과 고정 코드북 이득은 12.2 Kbit/s 모드와 7.95 Kbit/s 모드에선 스칼라 양자화되고, 나머지 모드에선 함께 벡터 양자화되어 그 인덱스가 전송된다.

한편 음성이 복호화될 때에는 먼저 수신된 인덱스를 이용하여 적응 코드북과 고정 코드북을 구하고, 이들에 각각의 이득을 곱한 후 더하여 여기 신호를 복원해낸다. 이 여기 신호를 복호화한 LSP에서 얻은 합성 필터에 통과 시킨 후 후처리 과정을 거치면 음성을 복원하게 되는 것이다.

2.2. EVRC 음성 부호화기

제 2세대 코드 분할 다중 접속 (CDMA: Code Division Multiple Access) 방식 이동통신 시스템의 표준 음성 부호화기로 사용되고 있는 IS-127 Enhanced Variable Rate Codec (EVRC) 음성부호화기는 AMR 음성부호화기와 달리 Relaxation CELP (RCELP) 알고리즘에 기반을 두고 있다.

RCELP 알고리즘은 기존 CELP 계열의 합성에 의한 분석 방법과 비슷하나 부호화하려는 원음에 변형을 가하는 일반화된 합성에 의한 분석 (generalized Abs)[9] 방법을 사용한다. 이때 원음은 프레임 단위로 구한 정수 피치를 선형 보간하여 얻은 피치 킨투어에 적합하도록 변형된다.

기존의 합성에 의한 분석 방법에서는 피치를 정확히 찾을수록 부호화 성능이 개선되므로 폐회로 검색을 통해 매 부프레임마다 부샘플 단위로 피치 지연을 검색한다. 이에 반해 RCELP 방법에서는 정수 피치 지연을 개회로 피치 지연 검색만으로 결정하여 프레임 단위로 전송하므로 부프레임마다 부샘플단위로 피치 지연값을 부호화하는 기존의 방식에 비해 피치 지연에 할당된 비트를 줄여 고정 코드북과 같은 여기신호나 채널 오류 방지를 위한 비트보다 많은 비트를 할당함으로써 높은 음질을 유지할 수 있다.

EVRC 음성부호화기는 일반적인 CELP 계열의 음성 부호화기에서처럼 현재 입력 음성 프레임으로부터 선형 예측 계수 (LPC), 적응 코드북 (피치 지연) 및 이득, 고정 코드북 및 이득과 같은 음성 모델 특징 파라미터들을 부호화한다. EVRC 음성부호화기는 20 ms의 프레임 길이와 10 ms의 예견구간 (look-ahead)을 가지고, 채널 대역폭의 효율적인 사용을 위하여 세가지 전송률—8.55 Kbit/s (Rate 1), 4.0 Kbit/s (Rate 1/2), 0.8 Kbit/s (Rate 1/8)—로 동작한다.

EVRC 음성부호화기가 음성을 부호화하는 과정은 다음과 같다. 먼저 직류 (DC) 성분을 제거하기 위한 고대역 통과 필터링과 잡음 억제 과정으로 구성된 전처리 과정을 거친 음성 신호는 20 ms 프레임 단위로 선형 예측 분석을 통해 선형예측계수를 구하고 (LPC analysis) 이를 보간과 양자화 에러에 강인한 LSP로 변환한다. 전처리된 음성 신호를 현재와 이전 프레임에서 얻은 LSP 파라미터를 보간하여 얻은 역필터에 통과시켜 잔여 신호 (Residual Signal)를 얻은 후 개회로 검색으로 피치 지연을 추정한다. 현재 프레임과 이전 프레임의 피치 지연을 선형 보간하여 선형 피치 컨투어를 계산한 후 이를 이용하여 잔여 신호를 수정한다. 수정된 잔여 신호를 지각가중 합성 필터에 통과시켜 수정된 가중 음성 신호를 구한다. 수정된 가중 음성 신호에서 적응 코드북의 기여를 제거하여 고정 코드북 검색을 위한 목적 신호를 생성한다. 고정 코드북은 AMR에서와 마찬가지로 ACELP 방식으로 검색된다.

2.3. AMR과 EVRC 음성 부호화기의 비교

표 1은 AMR과 EVRC 음성 부호화기의 주요 유사점 및 차이점을 정리한 것이다.

표 1. AMR과 EVRC 음성부호화기의 비교
Table 1. Comparison of AMR and EVRC Speech Coders.

	AMR	EVRC
Audio Interface	8000 samples/s 13-bit uniform PCM	8000 samples/s 13-bit uniform PCM
Coding Scheme	Analysis by Synthesis	Generalized Analysis by Synthesis (RCELP)
Analysis size	20ms frame size 4 subframes (5ms)	20ms frame size 3 subframes
LPC analysis	5ms or no look-ahead	10ms look-ahead
LSP quantization	Predictive Split VQ (2,2,2,2,2) / (3,3,4)	Split VQ (2,2,3,3) / (3,3,4)
Pitch delay	Update every subframe (Fractional pitch)	Update every frame (Integer pitch)
Fixed Codebook	ACELP	ACELP
Gains	Scalar quantization Vector quantization	Scalar quantization

III. AMR과 EVRC 음성부호화기간의 비텐덤 방식을 이용한 상호 부호화 알고리즘

3.1. AMR에서 EVRC로의 상호 부호화

AMR에서 EVRC 음성부호화기로의 상호 부호화는 LSP 변환, 적응 코드북을 위한 피치 지연 변환, 이득 변환, 고정 코드북 변환으로 구성된다. AMR 음성부호화기의 8개 모드 중 7.95 kbps 모드 이상은 EVRC 음성부호화기의 Rate 1로 대응되고, 7.40 kbps 모드 이하는 Rate 1/2로 대응된다. AMR 음성부호화기의 DTX 모드와 EVRC 음성부호화기의 Rate 1/8 은 상호 부호화의 효율을 높이기 위해 내부적으로 사용된다.

3.1.1. LSP 변환

제안된 LSP 변환과정에선 AMR 프레임의 LSP 정보를 그림 2와 같은 선형 보간법을 통해 EVRC 프레임의 LSP 정보로 변환한다. AMR과 EVRC 음성부호화기는 한 프레임의 길이가 20 ms (160 샘플)로 동일하지만 선형 예측 분석시 적용되는 윈도우의 모양과 길이, 그리고 예견구간 (look-ahead)의 길이가 서로 다르다. 상호 부호화 과정에서는 이러한 점을 감안하여 AMR 음성부호화기의 비트열로부터 현재 프레임을 위한 LSP, $\Omega_A^{(m)}$ 를 복호화한 후 기억시켜 놓은 이전 프레임의 LSP, $\Omega_A^{(m-1)}$ 와 합계식 (1)의 같은 선형 보간을 적용하여 EVRC 음성부호화기를 위한 LSP, $\Omega_B^{(m)}$ 로 직접 전환한다.

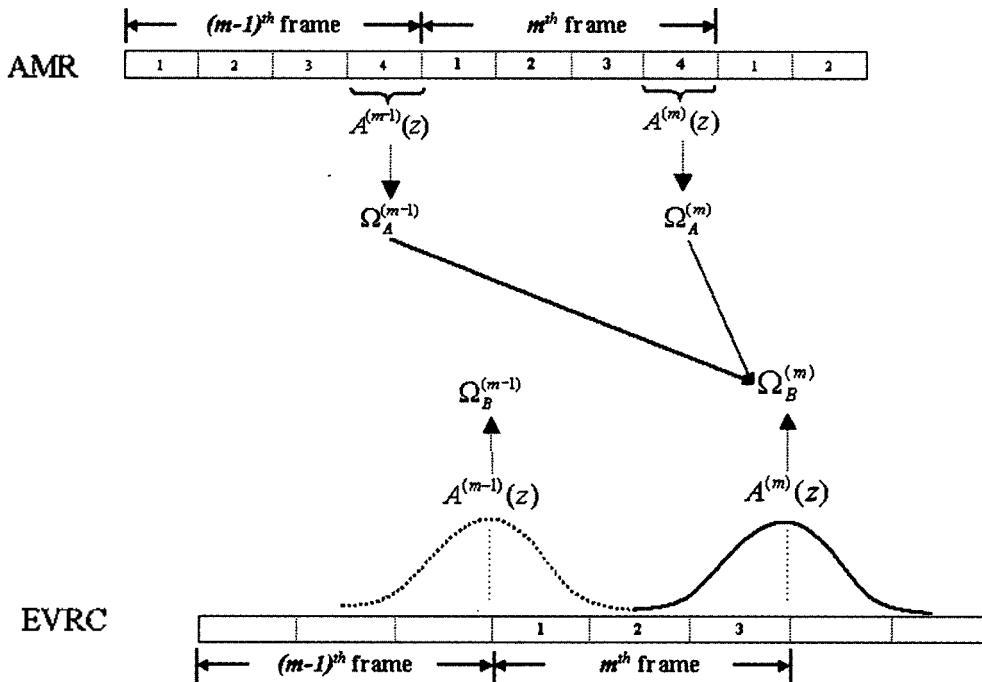


그림 2. AMR에서 EVRC로의 선형 보간을 이용한 LSP 변환
 Fig. 2. Transcoding of LSP from AMR to EVRC using the linear interpolation.

$$\Omega_B^{(m)} = \mu \Omega_A^{(m)} + (1 - \mu) \Omega_A^{(m-1)} \quad (1)$$

여기서 $\Omega_A^{(m)}$ 와 $\Omega_B^{(m)}$ 는 각각 AMR과 EVRC 음성부호화기의 m번째 프레임의 LSP이다. 그리고 μ 는 AMR과 EVRC의 현재 프레임의 위치를 고려한 선형 보간 가중치로서 0.8 ~ 0.9의 값을 갖는다.

제안된 방법을 사용하면 AMR 음성부호화기의 비트열을 음성 신호로 복호화했다가 다시 EVRC 음성부호화기로 부호화하는 과정에서 선형 예측 계수를 구하기 위해 자기상관계수를 계산하고, Levinson Durbin 알고리즘으로 선형 예측 계수를 구하는 과정, 그리고 선형 예측 계수를 구한 후 이를 다시 LSP로 변환하는 과정이 필요없게 되므로 계산량과 계산시간이 크게 줄어든다. 게다가 이렇게 LSP를 직접 변환함으로써 두번의 LSP 양자화 과정을 거치면서 발생하는 LSP 양자화 오차를 겪지 않아도 되므로 음질 저하가 일어나지 않는 이점이 있다.

3.1.2. 적응 코드북을 위한 피치 변환

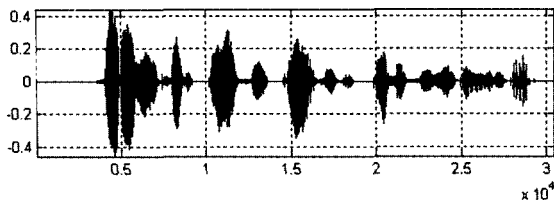
LSP의 상호 부호화가 끝나면 식 (2)를 이용하여 이득의 비, g_r 을 계산한다.

$$g_r = \frac{g_p}{g_c} \quad (2)$$

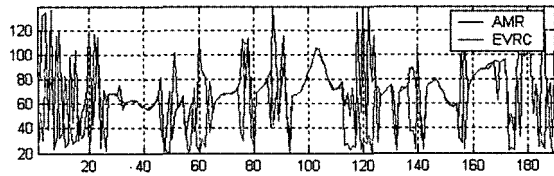
여기서 g_p 는 적응 코드북 이득이고, g_c 는 고정 코드북 이득이다.

그림 3은 원음성과 피치 지연 값, 그리고 이득 비의 관계를 보여주고 있다. 이 그림을 보면 이득의 비가 큰 구간은 음성이 없는 구간으로 AMR과 EVRC 음성부호화기에서 구한 피치 지연 값이 크게 차이가 나고, 반면 이득의 비가 작은 구간은 음성이 있는 부분으로 AMR과 EVRC 음성부호화기에서 구한 피치값의 오차도 작다는 것을 확인할 수 있다. 여기서 우리는 이득의 비가 큰 프레임의 피치 지연값은 음성이 없는 구간이므로 굳이 복잡한 과정을 거쳐 변환할 필요가 없으며, 이득의 비가 작은 프레임은 AMR에서 구한 피치 지연 값을 이용하여 변환의 효율을 높일 수 있다는 것을 알 수 있다. 이 중 이득의 비가 큰 경우의 상호 부호화는 고정 코드북 변환과 결합되어 3.1.5 절에서 다루어진다.

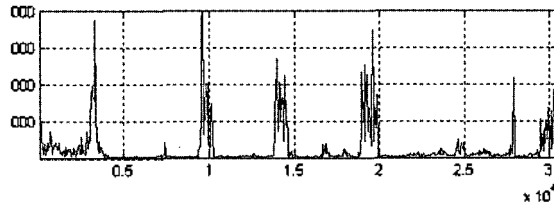
이득의 비가 기준값 이하인 경우, 즉 음성이 있는 구간은 새로운 피치 지연값을 구해야 한다. 그림 3에서 확인했듯이 음성이 있는 구간에서는 AMR과 EVRC 음성부호화기에서 구한 피치 지연 값이 비슷하다. 그러나 이 값을 그대로 사용하면 복원된 음성의 품질이 크게 떨어진다. 또한 LSP를 변환할 때처럼 프레임의 시간차를 고려하여 피치를 선형 보간해도 음질이 매우 나쁘다. 그러므로 새로운 피치 지연값을 검색할 필요가 있다.



(a) 원음성 파형
(a) Original speech waveform



(b) AMR과 EVRC에서 구한 피치 지연 값 비교
(b) Comparison of pitch delay values obtained by AMR and EVRC coder



(c) 이득의 비
(c) Gain ratio

그림 3. 원음성 파형, 피치 지연과 이득의 비의 관계
Fig. 3. Relation among original speech waveform, pitch delay, and gain rate.

그러나 그림 3을 보면 새로운 피치 지연값이 AMR 음성 부호화기 형식의 비트열에서 복호화하여 얻은 피치 지연값 부근에 있을 것으로 예상되므로 검색 구간을 제한

함으로써 계산량을 줄일 수 있다. 피치 지연값 검색은 식 (3)을 따른다.

$$R(D) = \sum_{n=0}^{160-D} \epsilon[n+n_0]\epsilon[n+n_0+D] \quad (3)$$

$$\max\{20, T_o - 5\} \leq D \leq \min\{120, T_o + 5\}$$

여기서 T_o 는 복호화된 AMR 음성부호화기의 피치 지연값에서 정수 부분만을 취한 값이다. 우리는 식 (3)의 $R(D)$ 를 최대로 하는 D_{max} 를 두 가지의 n_0 , 80과 160 값에 대하여 구한 후 EVRC 음성부호화기에서 최적의 피치 지연을 구하는 알고리즘을 그대로 사용하여 EVRC 음성부호화기를 위한 정수 피치 지연 r 를 구한다. 식 (3)에서 $\epsilon[n]$ 은 잔여 신호이다. 이는 EVRC 음성부호화기에서 피치 지연 검색이 잔여 신호 영역에서 이루어지기 때문에 이를 따른 것이다. 그러나 본 알고리즘에서는 잔여 신호 대신 AMR 형식의 비트열을 복호화하여 얻을 수 있는 여기 신호, $u[n]$ 을 이용하여 피치 지연값을 검색한다. 여기 신호는 식 (4)를 이용하여 구할 수 있다.

$$u[n] = g_p v[n] + g_c d[n] \quad (4)$$

여기서 $v[n]$ 은 적응 코드북 벡터, $d[n]$ 은 고정 코드북 벡터이고 g_p 와 g_c 는 각각 적응 코드북, 고정 코드북 이득으로 모두 AMR로 부호화된 비트열에서 복호화한 값들이다.

3.1.3. 이득 변환

적응 코드북 이득은 선형 보간과 같은 방법으로 직접

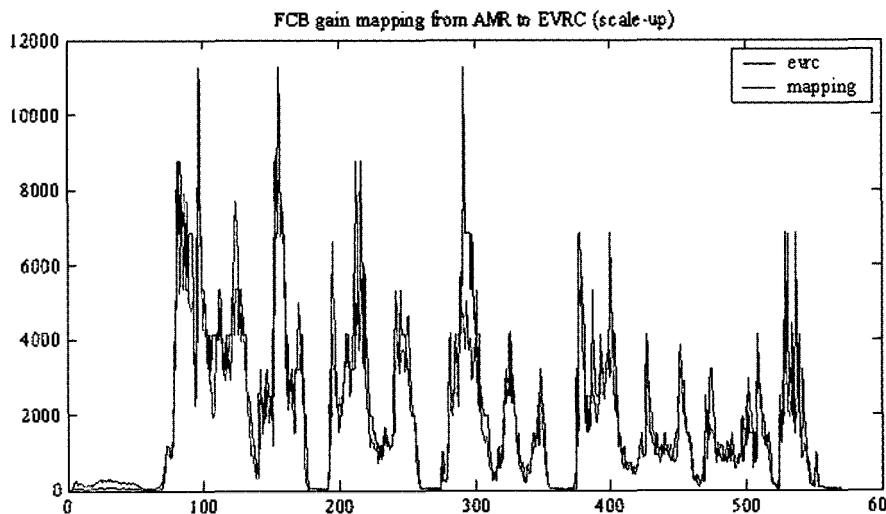


그림 4. 고정 코드북 이득의 변환
Fig. 4. Transcoding of fixed codebook gain.

변환할 경우 음질이 크게 떨어진다[4]. 그러므로 적응 코드북 이득은 변환된 피치 지연값을 바탕으로 직접 계산해야 한다. 그런데 우리는 3.1.3에서 설명된 알고리즘을 이용하여 비교적 정확한 피치 지연값을 얻었으므로 적응 코드북 이득은 큰 부가적 계산량 없이 쉽게 계산할 수 있다.

적용 코드북 이득과 달리 고정 코드북 이득은 선형 보간과 스케일 조절을 통해 변환이 가능하다. 그림 4는 같은 음성에 대하여 EVRC에서 구한 고정 코드북 이득과 AMR에서 구한 고정 코드북 이득을 변환한 것을 비교하고 있다. 그래프 상으로도 두 값이 큰 차이가 없다는 것을 볼 수 있으며 음질 또한 저하시키지 않음을 확인했다.

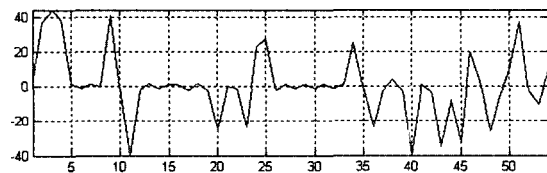
3.1.4. 고정 코드북 벡터 변환

그림 5와 6은 각각 어떤 부 프레임의 적응 코드북, 여기 신호, 고정 코드북 벡터를 보여주고 있다.

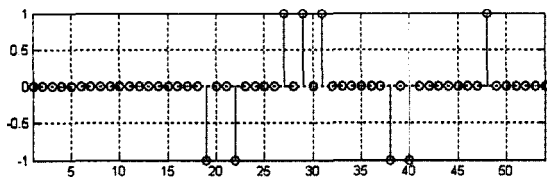
그림 5의 부 프레임의 경우 적응 코드북 벡터가 거의 여기 신호 벡터와 흡사한 것을 알 수 있는데, 이는 고정 코드북 이득이 작아 여기 신호에 대한 고정 코드북 벡터의 기여가 작기 때문이다. 이에 반해 그림 6의 부 프레임에선 고정 코드북 벡터에 비하여 적응 코드북 벡터의 여기 신호 벡터에 대한 기여가 작는데, 이는 적응 코드북 이득

이 작기 때문이다. 이들이 다소 극단적인 예이기는 하지만 여기서 우리는 그림 6의 부 프레임같은 경우엔 복잡하게 고정 코드북 벡터를 변환할 필요가 없고, 그 판정 기준으로 앞에서 정의된 이득의 비를 사용할 수 있다는 것을 알 수 있다. 즉 이득의 비가 크면 적응 코드북이 우세한 지역, 작으면 고정 코드북이 우세한 지역으로 판정을 내려 고정 코드북의 영향이 미미한 프레임에선 변환 과정을 생략함으로써 상호 부호화의 효율을 높이는 것이다.

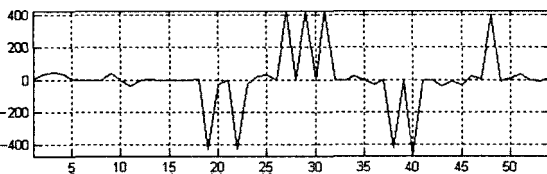
이득의 비가 큰 경우의 상호 부호화는 피치 지연값의 변환과 결합되어 3.1.5에서 다루어진다. 이득의 비가 작은 경우엔 새로운 고정 코드북이 검색되어야 한다. 그런데 고정 코드북 벡터는 피치 지연값과 달리 이득의 비가 작은 프레임에서도 서로 큰 상관관계를 지니고 있지 않다. 그림 7은 같은 음성의 동일한 프레임에 대해 AMR과 EVRC가 검색한 고정 코드북 벡터를 비교한 것이다. 0 아닌 펄스의 수의 차이를 고려하더라도 서로 상관관계가 거의 없음을 알 수 있다. 이는 주파수 영역에서 살펴봐도 마찬가지로, 상대방 음성 부호화기에서 구한 고정 코드북을 그대로 사용하면 음질이 크게 떨어졌다. 그러므로 고정 코드북의 변환은 불가피하게 탠덤 방식과 유사한 전체 검색 (Full Search)을 통해 이루어져야 한다. 그러나 이득



(a) 적응 코드북 벡터
(a) Adaptive codebook



(b) 고정 코드북 벡터
(b) Fixed codebook

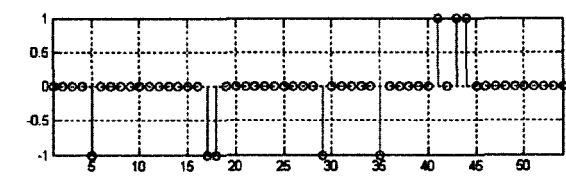


(c) 여기 신호 벡터
(c) Excitation signal vector

그림 5. 적응 코드북 우세 프레임
Fig. 5. Adaptive codebook dominant fram.



(a) 적응 코드북 벡터
(a) Adaptive codebook vector



(b) 고정 코드북 벡터
(b) Fixed codebook vector



(c) 여기 신호 벡터
(c) Excitation signal vector

그림 6. 고정 코드북 우세 프레임
Fig. 6. Fixed codebook dominant frame.

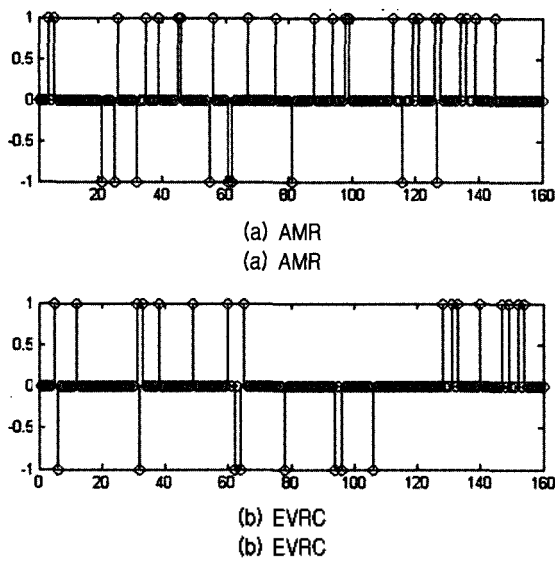


그림 7. 고정 코드북 비교
Fig. 7. Comparison of fixed codebook.

의 비를 판정하여 고정 코드북 검색이 불필요한 프레임들을 찾아냄으로써 평균 계산량은 크게 줄일 수 있었다.

3.1.5. Rate 1/8로의 변환

앞의 3.1.2 절과 3.1.4 절에서 이득의 비가 기준값보다 크다고 판정된 프레임은 음성이 없는 구간이므로 복잡한 변환 과정을 생략함으로써 상호 부호화의 효율을 높인다. 그런데 EVRC 음성 부호화기에는 음성이 없는 구간을 부호화하는 Rate 1/8이 존재한다. 우리는 EVRC가 Rate를

결정하는 알고리즘을 적용할 필요없이 이득의 비만을 계산하여 그 값이 어떤 기준값을 넘는 경우 이를 Rate 1/8로 부호화한다. 그리고 AMR에서 DTX 모드로 부호화된 프레임도 Rate 1/8로 대응시킨다. EVRC의 Rate 1/8에서는 LSP와 프레임 에너지만이 부호화되는데 이중 LSP는 3.1.1 절에서 상호 부호화한 것을 사용하고, 프레임 에너지는 AMR에서 구한 고정 코드북 이득을 3.1.3 절에서 설명한 고정 코드북 이득 상호 부호화 방법으로 변환한 값을 사용한다. 음성이 없는 구간에서 복잡한 상호 부호화 과정없이 EVRC의 Rate 1/8을 이용한 변환을 함으로써 계산량과 지연 시간을 크게 줄일 수 있었으며, 이로 인한 음질의 저하 또한 없음을 확인하였다.

3.2. EVRC에서 AMR로의 상호 부호화

EVRC에서 AMR로의 상호 부호화도 AMR에서 EVRC로의 경우와 마찬가지로 LSP 변환, 적응 코드북을 위한 피치 지연 변환, 고정 코드북 벡터 변환, 그리고 이득 변환으로 구성된다. EVRC는 자체적으로 비트율 결정 알고리즘을 가지고 있어서 음성에 따라 비트율을 결정하지만 AMR의 모드는 자체적으로 정해지지 않고 외부에서 입력 받는다. 그러므로 실제적 상황을 고려하여 EVRC의 비트율을 특정 AMR 모드로 대응시키지 않고 변환할 모드를 입력받아 사용한다. 단, EVRC의 Rate 1/8로 부호화된 프레임은 임의로 AMR의 모드중 가장 비트율이 낮은 4.75 kbps 모드에 대응시켜 변환한다.

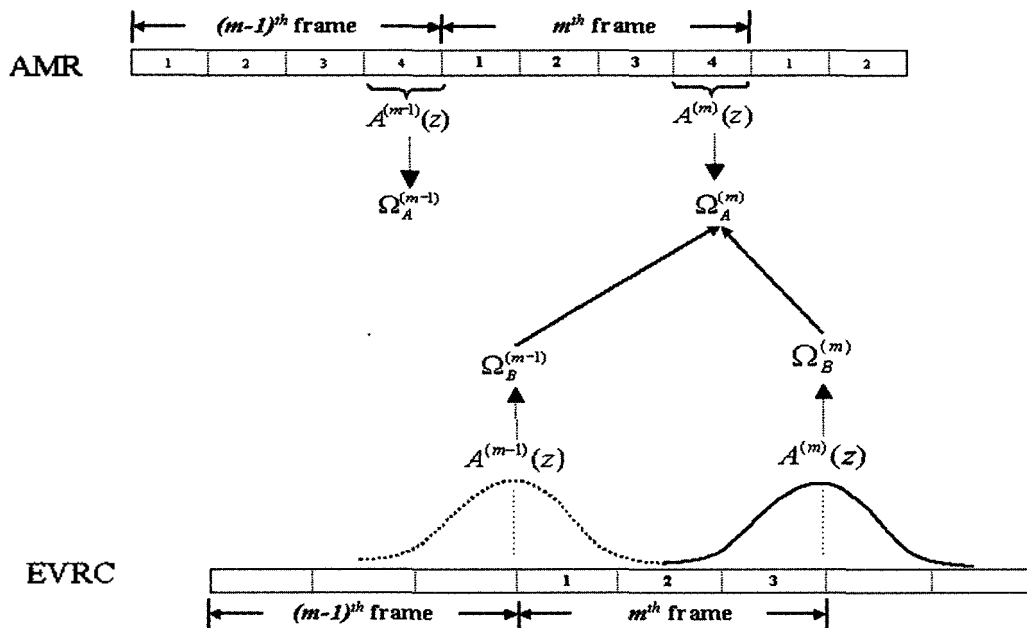


그림 8. EVRC에서 AMR로의 선형 보간을 이용한 LSP 변환
Fig. 8. Transcoding of LSP using linear interpolation from EVRC to AMR.

3.2.1. LSP 변환

제한된 LSP 변환과정에선 EVRC 프레임의 LSP 정보를 그림 8과 같은 선형 보간법을 통해 AMR 프레임의 LSP 정보로 변환한다.

여기서 $\Omega_A^{(m)}$ 과 $\Omega_B^{(m)}$ 는 각각 AMR과 EVRC의 m번째 프레임의 LSP를 의미한다. 선형 보간은 식 (5)와 같다.

$$\Omega_A^{(m)} = \mu \Omega_B^{(m)} + (1 - \mu) \Omega_B^{(m-1)} \quad (5)$$

AMR 음성부호화기에선 12.2 kbps 모드를 제외하고는 프레임마다 한 번씩 선형 예측 계수를 구하여 이를 LSP로 변환한다. 이때 구해진 LSP는 비대칭적인 윈도우의 특성상 네번째 부프레임의 LSP라 간주할 수 있다. 우리는 AMR에서 EVRC 음성부호화기로의 변환과 마찬가지로 양 음성부호화기들의 예견 구간 등을 고려하여 0.9 ~ 0.95의 값을 갖는 μ 를 선택한다.

제한된 방법에 의하면 EVRC에서 AMR 음성부호화기로의 변환도 반대 방향으로의 변환과 마찬가지로 이중 부/복호화 과정이 필요없게 되어 계산량과 계산시간을 크게 감소시킬 수 있으며 부/복호화 과정에서 일어나는 음질 저하도 겪지 않게 된다. 또한 AMR이 선형 예측 계수 계산에서 필요로 하는 5 ms의 예견구간이 필요없게 되므로 지연 시간이 감소하게 된다.

3.2.2. Rate 1/8로 부호화된 프레임의 변환

AMR에서 EVRC 음성부호화기로 상호 부호화시, 우리는 이득의 비를 계산하여 해당 프레임이 음성을 포함하고 있는지 판정했다. 하지만 EVRC에서 AMR 음성부호화기로의 상호 부호화 시에는 이 과정이 불필요하다. EVRC 음성부호화기는 음성을 부호화할 때에 음성이 없는 구간을 Rate 1/8로 부호화하기 때문이다. 우리는 Rate 1/8로 부호화된 프레임을 AMR 음성부호화기의 부호화 모드들 중 비트율이 가장 낮은 4.75 kbps 모드에 대응시킨다.

EVRC 음성부호화기의 Rate 1/8에서는 LSP와 프레임 에너지만이 부호화된다. 이 중 LSP는 3.2.1 절의 방법으로 변환한다. 다시 AMR 음성부호화기 형식으로 부호화하기 위해서는 다른 파라미터들인 피치 지연값, 이득, 그리고 고정 코드북을 적절히 만들어주어야 한다. 그런데 Rate 1/8로 부호화되는 부분은 음성이 없는 부분이므로 이들 파라미터들을 복잡하게 검색할 필요는 없다. 피치는 전 프레임의 피치값을 사용한다. 이득은 프레임 에너지를 고려하여 결정된 후 양자화되며, 고정 코드북 벡터는 4.75 kbps 모드의 고정 코드북 벡터 형식에 맞는 한도 내에서 무작위로 선택된다. 4.75 kbps 모드에서는 이득이 매 부

프레임마다 양자화되는 것이 아니라 첫번째 및 세번째 부프레임에서만 양자화되므로 이 사실도 고려한다.

3.2.3. 적응 코드북을 위한 피치 변환

피치 변환은 AMR에서 EVRC 음성부호화기로의 경우와 마찬가지로 상대방 음성 부호화기에서 구한 피치 지연값을 이용하여 이루어진다. 먼저 EVRC 음성부호화기의 적응 코드북과 고정 코드북 성분을 복호화한 후 이들을 더한 여기 신호를 생성한다. 한편, 변환된 LSP로부터 부프레임을 위한 지각가중 합성필터를 만든 후 여기신호를 통과시켜 가중된 음성 신호를 생성하고 AMR 음성부호화기의 피치 지연 검색과정을 이용해서 피치를 찾게 된다. 단 Rate 1 또는 Rate 1/2로 부호화된 프레임은 음성이 있는 구간인데 이 구간에서는 AMR과 EVRC 음성부호화기에서 구한 피치 정보가 큰 상관관계를 가지므로 이를 이용할 수 있다. EVRC 음성부호화기에서 복호화한 정수 피치값을 AMR 음성부호화기의 피치 지연 검색 중 1단계 검색에서 구해지는 정수 피치 지연값으로 놓고 피치 지연을 검색하는 것이다. 이렇게 피치 지연 검색 구간을 제한함으로써 음질의 저하없이 계산량을 줄일 수 있었다.

3.2.4. 이득 변환

이득 변환과정은 AMR에서 EVRC 음성부호화기로의 경우와 같다. 적응 코드북 이득은 변환된 피치 지연값을 이용하여 직접 계산한다. 그리고 고정 코드북 이득은 EVRC에서 얻은 고정 코드북 이득을 선형 보간한 후 스케일을 조정하여 사용하였다. 이로써 음질의 저하는 없다는 것을 확인했다.

3.2.5. 고정 코드북 변환

고정 코드북은 3.1.4절에 언급된 것과 같은 이유로 텐덤 방식과 유사하게 전체 검색되어 변환된다. 그러나 Rate 1/8로 부호화된 프레임을 별도의 고정 코드북 검색 없이 변환함으로써 음질은 유지하면서도 평균 계산량은 감소시킬 수 있었다.

IV. 성능 평가

4.1. 계산량

표 2와 3은 8초 길이의 음성 파일을 각각 텐덤 방식과 비텐덤 방식으로 변환하였을 때 걸리는 처리 시간을 비교한 것이다. 남, 녀 화자당 16개씩 총 32개의 음성 파일에

표 2. 계산량 비교 (AMR→EVRC)

Table 2. Comparison of computational complexity (AMR→EVRC).

Method	Time (sec)	Complexity (%)
Tandem	9.18	100.0
Tandemless	6.59	71.9

표 3. 계산량 비교 (EVRC→AMR)

Table 3. Comparison of computational complexity (EVRC→AMR).

Method	Time (sec)	Complexity (%)
Tandem	7.03	100.0
Tandemless	3.90	55.5

대해 실험하였으며 처리 시간은 그 평균을 구한 것이다. 표 2와 3의 결과를 보면 비텐덤 방식이 텐덤 방식에 비하여 AMR에서 EVRC로의 상호 부호화의 경우엔 28.1%, EVRC에서 AMR로의 상호 부호화의 경우엔 44.5% 정도 계산량이 적다는 것을 확인할 수 있다.

4.2. 지연 시간

식 (6)과 (7)은 텐덤 방식 (D_{AB}^d, D_{BA}^d)과 비텐덤 방식 (D_{AB}^u, D_{BA}^u)의 총 지연 시간을 나타낸다.

$$D_{AB}^d = 35 + \alpha_A + \beta_A + \alpha_B + \beta_B$$

$$D_{AB}^u = 25 + \alpha_A + P_{AB} + \beta_B \tag{6}$$

$$D_{BA}^d = 35 + \alpha_B + \beta_B + \alpha_A + \beta_A$$

$$D_{BA}^u = 25 + \alpha_B + P_{BA} + \beta_A \tag{7}$$

여기서 A는 AMR 음성부호화기를, B는 EVRC 음성부호화기를 의미하며 AB와 BA는 각각 A에서 B, B에서 A로의 변환을 나타낸다. α_i 와 β_i ($i=A$ or B)는 각각 부호화기와 복호화기의 처리 시간이며, P_{AB} 와 P_{BA} 는 제안된 비텐덤 방식의 상호 부호화 알고리즘의 처리 시간이다.

선형예측 분석시 AMR과 EVRC 음성부호화기는 각각 5 ms와 10 ms의 예견구간을 필요로 한다. 따라서 텐덤 방식에 의한 상호 부호화인 D_{AB}^d 와 D_{BA}^d 에는 부/복호화기의 처리 시간을 제외하고도 35 ms (=20 ms 프레임 길이 + 5 ms AMR lookahead + 10 ms EVRC lookahead)의 지연 시간이 존재하게 된다. 이에 반해 비텐덤 방식의 경우 D_{AB}^u 은 EVRC 음성부호화기에서의 선형예측 분석 과정이 필요없게 되므로 25 ms의 지연 시간이 존재하고, D_{BA}^u 은 반대로 AMR 음성부호화기에서의 선형예측 분석 과정이 필요없게 되므로 30 ms의 지연 시간이 존재한다. 또

한, $P_{AB} \ll \beta_B + \alpha_A$ 이고 $P_{BA} \ll \beta_B + \alpha_A$ 이므로 총 지연 시간은 크게 감소되는 효과가 있다.

4.3. 객관적 음질 평가 - ASDM

본 연구에서는 객관적 음질의 평가 방법을 위해 선형 예측 계수에 기반을 둔 두 음성의 오차를 측정하는 평균 스펙트럼 왜곡 측정 (ASDM: Average Spectral Distortion Measure)[5]을 사용하였다. ASDM은 식 (8)과 같이 주어진다.

$$ASDM = \frac{1}{M_r} \sum_{m=1}^{M_r} \sqrt{\frac{1}{\pi} \int_0^\pi [10 \log_{10} S^m(\omega) - 10 \log_{10} \widehat{S}^m(\omega)]^2 d\omega}$$

(dB) (8)

여기서 $S^m(\omega)$ 와 $\widehat{S}^m(\omega)$ 는 각각 m차 선형 예측 계수 분석 프레임의 기준 선형 예측 계수 스펙트럼과 비교 대상 선형 예측 계수 스펙트럼을 의미한다. M_r 는 계산된 프레임의 총 개수이다. 실험을 위해 남/녀 화자가 말한 3초 분량의 문장을 각각 텐덤 방식과 비텐덤 방식으로 변환하여 그 ASDM 값을 구했다.

표 4, 5를 보면 화자에 상관없이 비텐덤 방식의 평균 스펙트럼 왜곡 측정이 텐덤 방식의 그것에 비하여 더 적음을 알 수 있다. 이는 비텐덤 방식으로 상호 부호화된 선형 예측 계수의 스펙트럼이 원 선형 예측 계수 스펙트럼과 더 유사하다는 것을 의미한다.

표 4. AMR (7.95kbps)에서 EVRC (Rate 1)로의 ASDM 값
Table 4. ASDM value (AMR 7.95 kbps → EVRC Rate 1).

ASDM (dB)	Male	Female
Tandem	1.0707	1.8018
Tandemless	0.6883	1.1647

표 5. EVRC (Rate 1)에서 AMR (7.95 kbps)로의 ASDM 값
Table 5. ASDM value (EVRC Rate → AMR 7.95 kbps 1).

ASDM (dB)	Male	Female
Tandem	1.1082	1.8120
Tandemless	0.6593	1.0107

4.4. 주관적 음질 평가 - 선호도 평가

주관적 음질 평가를 위하여 10명의 청자에게 8 kHz로 표본화된 남/여 화자 음성 파일을 텐덤 방식과 제안된 비텐덤 방식으로 각각 변환한 후 복원하여 2번씩 들려 주고 음질의 선호도를 조사하였다. 음질 평가에 이용된 음성

표 6. AMR에서 EVRC로의 상호 부호화 음질 평가 결과
Table 6. Results of preference test for tandemless trans-coding from AMR to EVRC.

Preference	Male	Female
Tandem	30 %	30 %
Tandemless	40 %	30 %
No Preference	30 %	40 %

표 7. EVRC에서 AMR로의 상호 부호화 음질 평가 결과
Table 7. Results of preference test for tandemless trans-coding from EVRC to AMR.

Preference	Male	Female
Tandem	40 %	25 %
Tandemless	20 %	30 %
No Preference	40 %	45 %

은 남자 4명, 여자 4명, 총 8명의 화자가 발음한 것으로 화자당 4개씩, 총 32개의 파일들로 구성되어 있으며 한 파일은 두 문장으로 이루어져 있다.

표 6과 표 7의 결과를 보면 AMR에서 EVRC로의 상호 부호화와 EVRC에서 AMR로의 상호 부호화 모두 비탠덤 방식이 탠덤 방식과 최소한 동등하거나 더 나은 음질을 나타내고 있다는 것을 확인할 수 있다. 표 7을 보면 EVRC에서 AMR로의 상호 부호화의 경우 비탠덤 방식에 대한 선호도가 반대 방향으로의 상호 부호화에 비해 다소 낮은 것을 볼 수 있는데 이는 AMR이 피치 지연값 오차에 더 민감한 합성에 의한 분석법에 기반하고 있기 때문이다.

V. 결론

본 논문에서는 서로 다른 음성부호화 표준을 사용하는 통신망에서의 호환성 문제 해결을 위해 음성신호를 탠덤 방식으로 변환함으로써 야기되는 음질의 저하, 계산량 및 지연 시간 증가 등의 문제를 최소화할 수 있는 비탠덤 방식의 상호 부호화 알고리즘을 제안하였다. 제안된 알고리즘은 AMR로 부호화된 음성 패킷과 EVRC로 부호화된 음성 패킷을 파라미터 도메인에서 상호 부호화함으로써 음성 도메인에서 상호 부호화하는 탠덤 방식보다 적은 계산량과 짧은 지연 시간을 가지며, 음질 또한 탠덤 방식과 동등하거나 혹은 그 이상이다.

같은 프레임 길이 (20 ms)를 갖는 AMR과 EVRC 음성 부호화기는 서로 다른 예견구간을 고려하여 상호 부호화된다. 상호 부호화 과정은 이득의 비 (AMR에서 EVRC로

의 상호 부호화) 혹은 EVRC의 부호화 비트율 (EVRC에서 AMR로의 상호 부호화)에 의한 판정으로 정확한 파라미터 변환이 필요한 경우와 그렇지 않은 경우로 크게 나뉘어질 수 있다. 판정에 의하여 해당 프레임이 음성이 없는 구간으로 확인된 경우는 Rate 1/8로 부호화하거나 (AMR에서 EVRC로의 상호 부호화) 이득 및 고정 코드북 벡터를 고정된 채로 4.75 kbps 모드로 (EVRC에서 AMR로의 상호 부호화) 부호화함으로써 음질 저하를 막으면서 상호 부호화의 효율을 높인다. 정확한 파라미터 변환이 필요한 경우는 먼저 선형 보간을 이용하여 LSP를 변환하고, 상대방의 비트열에서 복호화한 피치값 부근에서 정확한 피치를 검색하는 방식으로 피치를 변환한다. 고정 코드북은 탠덤 방식과 동일하게 검색하여 변환하고 이득은 위의 과정에서 얻어진 신호들을 이용하여 계산하되, 고정 코드북 이득은 선형 보간 및 스케일 조정으로 변환한다.

본 연구에서 제안한 상호 부호화 알고리즘을 주/객관적 방법으로 평가해 본 결과, 양방향 모두 탠덤 방식과 비교해 계산량과 지연 시간을 감소시키면서도 음질이 탠덤 방식과 동등하거나 그 이상임을 확인할 수 있었다.

참고 문헌

- 3GPP TS 26.090 V5.0.0, Adaptive Multi-Rate (AMR) speech codec: Transcoding functions, June, 2002.
- TIA/EIA/IS-127, Enhanced variable rate codec, speech service option 3 for wideband spread spectrum digital systems, 1997.
- W. B. Kleijn, and P. Kroon, "The RCELP Speech-Coding Algorithm," *European Trans. on Telecom.*, 5 (5), 573-582, 1994.
- H.-G. Kang, H.-K. Kim and R. V. Cox, "Improving transcoding capability of speech coders in clean and frame erased channel environments," *IEEE Proc. 2000 IEEE Workshop on Speech Coding*, 78-80, 2000.
- K. K. Paliwal, and B. S. Atal, "Efficient vector quantization of LPC parameters at 24 bits/frame," *IEEE Trans. Speech and Audio Processing*, 1 (1), 3-14, 1993.
- 윤성완, 정성교, 박영철, 윤대희, "8kbps G.729A에서 5.3kbps G.723.1로의 상호 부호화 알고리즘," *신호처리학회 학술대회 논문집* 13 (1), 823-826, 2000.
- A. M. Kondoz, *Digital Speech coding for Low Bit Rate Communications Systems*, John Wiley & Sons Inc., 59-75, 84-96, 335-341, 1994.
- M. R. Schroeder, B. S. Atal, "Code-excited linear prediction (CELP) high-quality speech at very low bit rates," *Int. Conf. on Acoustics, Speech and Signal Processing*, 937-940, 1985.
- W. B. Kleijn, *Analysis-by-Synthesis Speech Coding Based on Relaxed Waveform-Matching Constraints*, Ph. D. dissertation, Delft University of Technology, 1991.

저자 약력

● 이 선 일 (Sunil Lee)



2001년 2월: 연세대학교 기계전자공학부 전기및전자
전공 (공학사)

2001년 2월 ~ 현재: 한국과학기술원 전자전산학과 석
사과정

※ 주관심분야: 음성 코딩

● 유 창 동 (Chang D. Yoo)

한국음향학회지 제20권 제3호 참조