

論文2002-39SP-6-14

AMR과 EVRC 음성부호화기를 위한 파라미터 직접 변환 방식의 상호부호화 알고리즘

(Transcoding Algorithm for AMR and EVRC Vocoders Via Direct Parameter Transformation)

李善一*, 柳昌東**

(Sunil Lee and Chang D. Yoo)

요약

본 논문에서는 AMR과 EVRC 음성부호화기를 위한 새로운 파라미터 직접 변환 방식의 상호부호화 알고리즘을 제안한다. 상호부호화를 위하여 부가적인 복호화, 부호화 과정을 거쳐야하는 기존의 Tandem 방식과 달리 제안된 파라미터 직접 변환 방식에서는 양 음성부호화기가 음성을 부호화하기 위하여 공통적으로 사용하는 파라미터들이 직접 변환된다. 제안된 알고리즘은 파라미터 복호화, 프레임 분류, 모드 결정, 그리고 두가지 프레임형을 위한 상호부호화기로 구성된다. 상호부호화기는 LSP, 프레임 에너지, 적응 코드북을 위한 피치 지연, 고정 코드북 벡터, 그리고 양 코드북의 이득을 변환한다. 제안된 알고리즘을 다양한 방법으로 평가해본 결과 기존의 Tandem 방식과 비교하여 계산량과 지연 시간을 줄이면서도 동등한 음질을 구현함을 확인할 수 있었다.

Abstract

In this paper, a novel transcoding algorithm for the Adaptive Multi Rate(AMR) and the Enhanced Variable Rate Codec(EVRC) vocoders via direct parameter transformation is proposed. In contrast to the conventional tandem transcoding algorithm, the proposed algorithm converts the parameters of one coder to the other without going through the decoding and encoding processes. The proposed algorithm consists of the parameter decoding, frame classification, mode decision, and transcoders for two frame types. The transcoders convert the parameters such as LSP, frame energy, pitch delay for the adaptive codebook, fixed codebook vector, and codebook gains. Evaluation results show that while exhibiting better computational and delay characteristics, the proposed algorithm produces equivalent speech quality to that produced by the tandem transcoding algorithm.

Keywords : AMR, EVRC, transcoding, direct parameter transformation, speech coding

I. 서론

오늘날엔 다양한 유/무선 통신망이 존재한다. 그리고

* 學生會員, ** 正會員, 韓國科學技術院 電子電算學科
(Dept. of Electrical Engineering and Computer Science, Korea Advanced Institute of Science and Technology)

接受日字:2002年10月1日, 수정완료일:2002年10月22日

각 통신망은 고유의 특성에 적합한 음성부호화 표준을 채택하여 사용하고 있다. 예를 들어 일반 유선 전화망은 펄스 부호 변조(Pulse Code Modulation, PCM) 방식의 ITU G.711^[9] 음성부호화기를 사용하는 데에 반해 최근 활발히 보급되고 있는 Voice over Internet Protocol (VoIP)망에서는 Code Excited Linear Prediction (CELP)^[15]방식의 ITU G.723.1^[10]과 ITU G.729A^[11] 음성부호화기가 널리 쓰이고 있다. 또한 이동 통신 환경에서는 역시 CELP 방식인 Adaptive

Multi Rate (AMR)^[1]와 Enhanced Variable Rate Codec (EVRC)^[8] 음성부호화기 등이 표준으로 사용되고 있다.

최근 유/무선 음성 통신망과 VoIP 방식을 이용하는 데이터 통신망간의 통합과 효율적인 연동이 중요한 과제로 제기되고 있다. 그러나 서로 다른 음성부호화 표준은 서로 호환이 되지 않기 때문에 각 통신망을 연동시키기 위해서는 상호부호화(Transcoding) 과정이 반드시 필요하다. 이와 같은 상황은 음질을 저하시키지 않으면서도 계산량이 적고, 지연 시간 또한 짧은 효율적인 상호부호화 알고리즘의 개발을 요구하고 있다. 일반적으로 서로 다른 음성부호화 표준간의 상호부호화를 가장 간단히 해결할 수 있는 방법은 Tandem 방식이다. <그림 1(a)>는 Tandem 방식을 이용하여 음성부호화기 A와 음성부호화기 B간의 상호부호화를 수행하는 과정을 나타낸 것이다. 그림을 보면 알 수 있듯이 Tandem 방식에서는 먼저 원천 음성부호화기(Source Coder)로부터 전송된 비트열(Bit-stream)을 복호화(Decoding)하여 PCM 형식의 음성으로 합성한 후, 이를 다시 목적 음성부호화기(Target Coder) 형식으로 부호화(Encoding)하는 과정으로 상호부호화를 수행한다. 하지만 Tandem 방식으로 상호부호화를 수행할 경우 불필요한 복호화와 부호화 과정 때문에 계산량과 지연 시간이 늘어나고, 음질 또한 저하되는 문제점이 발생한다. 이를 해결하고자 하는 것이 바로 파라미터 직접 변환 방식의 상호부호화 알고리즘이다. 파라미터 직접 변환 방식은 상호부호화의 대상이 되는 음성부호화기들이 CELP 등의 같은 음성부호화 알고리즘에 기반하고 있어 음성을 부호화할 때에 같은 파라미터

(Parameter) 집합을 사용할 때에 적용이 가능하다. <그림 1(b)>는 파라미터 직접 변환 방식으로 상호부호화를 수행하는 과정을 나타낸 것이다. 그림을 보면 파라미터 직접 변환 방식의 경우 Tandem 방식과 달리 중간에 음성을 합성하지 않고, 부호화된 파라미터들을 직접 변환하는 것을 알 수 있다. 이러한 직접적인 변환을 통해 Tandem 방식의 문제점인 계산량과 지연 시간 증가를 해결하고, 나아가 음질 저하 또한 막는 것이 파라미터 직접 변환 방식의 목표라 할 수 있다. 기존의 파라미터 직접 변환 방식에 기반한 상호부호화를 다룬 논문들로는^[17,23~25] 등이 있으며, Qualcomm사가 부분적으로 파라미터 직접 변환 방식을 이용하여 출원한 미국 특허도^[26] 있다.

본 논문은 AMR 음성부호화기와 EVRC 음성부호화기를 위한 새로운 파라미터 직접 변환 방식의 상호부호화 알고리즘을 제안한다. AMR 음성부호화기는 3세대 비동기식 이동 통신망의 음성부호화 표준이며 합성에 의한 분석(Analysis-by-Synthesis, AbS)에 기반한 Algebraic CELP(ACELP) 알고리즘으로 음성을 부호화한다. 이에 반해 EVRC 음성부호화기는 2세대 및 2.5세대 CDMA 이동 통신망의 음성 부호화 표준으로써 일반화된 합성에 의한 분석(Generalized AbS)^[13]에 기반한 Relaxation CELP(RCELP)^[14] 알고리즘으로 음성을 부호화한다. 기존에 쓰여진 AMR과 EVRC 음성부호화기간의 파라미터 직접 변환 방식의 상호부호화에 관한 논문으로는 [23]이 있다. 본 논문은 [23]에서 제안된 알고리즘을 여러 면에서 수정, 개선하여 성능을 향상시켰다. 프레임 분류를 부호화기 내부의 기능을 이용하게 하였고, AMR과 EVRC 음성부호화기에서 배경 잡음 부호화를 위해 쓰이는 모드까지 모두 고려하여 제안된 상호부호화 알고리즘이 실제 환경에서도 그 성능을 유지할 수 있도록 하였다. 그리고 다소간의 음질 저하를 가져왔던 2분법적 고정 코드북 벡터 변환 알고리즘을 개선하여 음질을 향상 시켰고 계산량 또한 감소시켰다.

제안된 상호부호화 알고리즘은 파라미터 복호화, 프레임 분류, 모드 결정, 그리고 두가지 프레임형을 위한 상호부호화기로 구성된다. 상호부호화기는 AMR과 EVRC 음성부호화기에서 공통적으로 쓰이는 Line Spectral Pair(LSP),^[16] 적응 코드북(Adaptive Codebook)을 위한 피치 지연(Pitch Delay), 고정 코드북 벡터(Fixed Codebook Vector), 코드북 이득(Codebook

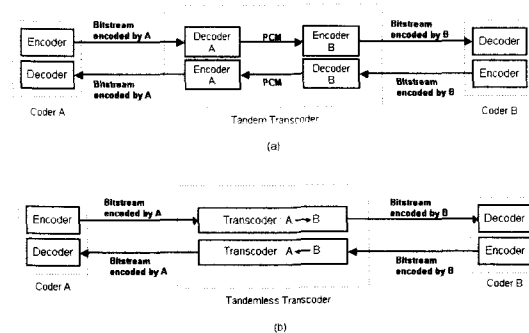


그림 1. 상호부호화 방식

(a) Tandem (b) 파라미터 직접 변환

Fig. 1. Transcoding Method

(a) Tandem

(b) Direct Parameter Transformation

Gains), 그리고 프레임 에너지(Frame Energy) 등을 변환한다.

본 논문은 다음과 같이 구성되어 있다. 먼저 II 장에서는 상호부호화의 대상이 되는 AMR과 EVRC 음성부호화에 대하여 간략히 알아본다. 그리고 III장과 IV장에서는 AMR에서 EVRC로의, 그리고 EVRC에서 AMR로의 상호부호화 알고리즘을 자세히 설명한다. V장에서는 제안된 파라미터 직접 변환 방식의 상호부호화 알고리즘을 다양한 방법으로 평가한 결과를 제시하며, 마지막으로 VI장에서 결론으로써 논문을 끝맺는다.

II. AMR과 EVRC 음성부호화기

1. AMR 음성부호화기

AMR 음성부호화기는 부호화기 및 복호화기,^[2] Voice Activity Detector(VAD)와 Comfort Noise Generator(CNG)를 포함하는 Source Controlled Rate(SCR) 모듈,^[3~5] 그리고 전송 과정의 오류(Error)와 패킷 손실(Packet Loss)에 강한 통신을 구현하기 위한 오류 은닉(Error Concealment) 알고리즘^[6]으로 구성되어 있다. 본 논문에서는 전송 과정에 오류가 없다고 가정하므로 오류 은닉 알고리즘은 고려하지 않는다. AMR 음성부호화기는 4.75 kbit/s에서 12.2 kbit/s에 이르는 비트율로 동작하는 8개의 모드(Mode)와 배경 잡음 구간을 효율적으로 전송하기 위한 저비트율 부호화 모드로 이루어져있다. 부호화 모드는 외부 네트워크로부터의 입력에 의해 매 프레임(Frame)마다 결정된다. AMR 음성부호화기는 8000 samples/s의 주파수로 표본화된 13 비트 선형(Uniform) PCM 형식의 음성을 입력 신호로 받는다. 음성 프레임의 길이는 160 샘플에 해당하는 20ms 이며 한 프레임은 길이가 5ms인 총 4개의 부프레임(Subframe)으로 나뉘어진다.

AMR 음성부호화기는 입력 음성을 다음의 과정으로 부호화한다. 먼저 입력 음성은 전처리(Pre-processing) 과정을 거친다. 전처리 과정에서 입력 음성은 고대역 필터를 통과한 후 Overflow 방지를 위해 그 크기가 반으로 줄어들게 된다. 전처리 과정을 거친 후엔 선형 예측 분석(Linear Prediction Analysis)이 수행된다. 선형 예측 분석은 12.2 kbit/s 모드에서는 한 프레임당 두번, 그 외의 모드에서는 한 프레임에 한번만 수행된다. 12.2 kbit/s 모드를 제외한 모드들에선 선형 예측 분석을 위

해서 5ms의 예견(Look-ahead) 구간이 필요하다. 구해진 선형 예측 계수(Linear Prediction Coefficients, LPC)는 선형 보간(Linear Interpolation), 양자화(Quantization) 및 전송 과정에서의 오류에 강인하게 만들기 위하여 LSP로 변환된 후 양자화되어 전송된다. AMR 음성부호화기에서 선형 예측 합성 필터(Linear Prediction Synthesis Filter)의 입력인 여기 신호(Excitation Signal)는 적응 코드북 벡터와 고정 코드북 벡터에 각 코드북의 이득을 곱하여 더한 형태로 모델링된다. 양 코드북을 위한 파라미터들은 매 부프레임마다 계산되어 전송된다. 선형 예측 분석 후, 적응 코드북이 검색된다. 적응 코드북 검색은 두 단계에 걸쳐 수행된다. 첫 단계인 개회로 피치 분석(Open-loop Pitch Analysis)은 5.15와 4.75 kbit/s 모드를 제외하면 매 프레임마다 두번씩 수행된다. 피치 지연 추정은 시각가중된(Perceptually Weighted) 음성 신호 영역에서 이루어진다. 개회 피치 분석으로 개회로 피치 지연 값이 구해지면 그 값 부근에서 폐회로(Closed-loop) 피치 분석을 통하여 피치 지연 값과 그 이득을 계산한다. 이때 피치 지연값은 모드에 따라 1/6 혹은 1/3의 정확도까지 계산된다. 마지막 단계는 고정 코드북 검색이다. 고정 코드북은 Interleaved Single-Pulse Permutation(ISPP) 구조에 기반하고 있다. 고정 코드북 검색 과정을 통해 최적의 여기 신호를 구성할 수 있는 펄스 위치와 부호가 결정된다.

AMR 음성부호화기는 음성이 없는 배경 잡음 구간을 낮은 비트율로 전송함으로써 통신 채널의 효율을 높일 수 있도록 SCR을 채택하고 있다. SCR은 수신단에서 입력 신호에 음성이 존재하는지를 판단할 수 있도록 VAD를 채택하며, 수신단에서 배경잡음 구간을 합성하기 위한 CNG 또한 포함하고 있다. AMR 음성부호화기는 두 가지의 VAD를 선택적으로 사용할 수 있도록 제공하고 있으며, 수신단의 CNG 모듈이 이용할 수 있도록 배경 잡음 구간에서 잡음의 특성 정보를 주기적으로 전송한다.

AMR 음성부호화기의 ANSI-C 소스 코드는[7]에서 제공된다. 제안된 파라미터 직접 변환 방식의 상호부호화 알고리즘을 구현하기 위해 이를 부분적으로 참고하였다.

2. EVRC 음성부호화기

EVRC 음성부호화기는 세가지 비트율-8.55(Rate 1),

4.0(Rate 1/2), 그리고 0.8(Rate 1/8) kbit/s-로 동작하는 다중모드 음성부호화기이다. 부호화 모드는 매 프레임마다 결정되지만 AMR 음성부호화기와 달리 자체적인 비트율 결정 알고리즘(Rate Decision Algorithm, RDA)에 의해 모드가 결정된다. Rate 1/8 모드는 음성이 아닌 배경 잡음을 부호화하기 위하여 사용된다. EVRC 음성부호화기는 AMR 음성부호화기와 마찬가지로 8000 samples/s의 주파수로 표본화된 13 비트 선형 PCM 형식의 음성을 입력으로 취한다. 음성 프레임의 길이는 160 샘플에 해당되는 20ms 이며 한 프레임은 각각 6.625(53 샘플), 6.625, 그리고 6.75(54 샘플) ms 길이의 세 부프레임으로 나뉘어진다.

EVRC 음성부호화기는 다음과 같이 음성을 부호화한다. 입력 음성은 먼저 고대역 필터 통과, 스케일 조정, 그리고 잡음 억제(Noise Suppression)로 이루어진 전처리 과정을 통과한다. 그리고 선형 예측 분석을 수행한 후 구해진 LPC를 LSP로 변환하여 양자화한다. 선형 예측 분석에서는 10ms의 예견 구간이 요구된다. 선형 예측 분석 후에는 적응 코드북이 검색된다. 앞에서 설명된 바와 같이 EVRC는 RCELP 알고리즘에 기반하고 있기 때문에 매 프레임당 하나의 정수 피치 지연만을 추정해 전송한다. EVRC 음성부호화기의 고정 코드북 구조 및 검색법은 AMR 음성부호화기와 유사하다. 그러나 Rate 1 모드에서는 계산량 감소를 위해 depth-first tree 검색 구조가 사용된다는 점에 차이가 있다.

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

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

EVRC 음성부호화기는 SCR 모듈을 포함하고 있지 않지만 대신 Rate 1/8 모드가 배경 잡음을 부호화하는 역할을 수행한다.

3. AMR과 EVRC 음성부호화기의 비교

<표 1>은 AMR과 EVRC 음성부호화기의 주요 유사점 및 차이점을 정리한 것이다. 표를 보면 두 음성부호화기의 가장 큰 차이점은 프레임 당 부프레임의 개수 및 부프레임의 길이, 적응 코드북을 위한 피치 지연의 검색 회수, 정밀도 및 검색법 등에 있음을 확인할 수 있다. 그리고 양 음성부호화기의 서로 다른 적응 코드북 검색 과정으로 인하여 두 음성 부호화기가 검색하는 고정 코드북 벡터의 모양 또한 크게 다를 것임을 예상할 수 있다. 기타 표에서 언급되지 않은 양 음성부호화기의 각종 양자화방식상의 차이는 제안된 상호부호화 알고리즘의 구현 시에 많은 어려움을 가져다 주기도 했다.

III. AMR에서 EVRC 음성부호화기로의 상호부호화

1. 전체적인 구조

<그림 2>는 제안된 파라미터 직접 변환 방식의 상호부호화 알고리즘의 구조도를 나타낸 것이다. 이 구조는 AMR에서 EVRC 음성부호화기로의 상호부호화, 그리고 그 반대의 경우 모두 공통된 것이다. 단, 외부 모드 입력(External Mode Input)은 EVRC에서 AMR 음성부호화기로의 상호부호화에만 존재한다.

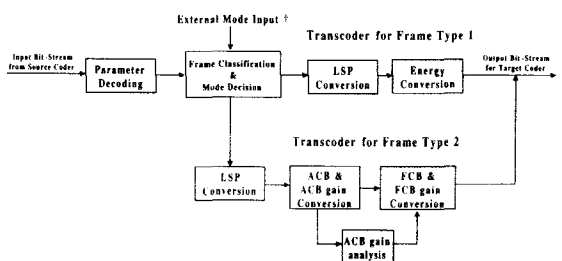


그림 2. 제안된 파라미터 직접 변환 방식의 상호부호화 알고리즘의 블록도(†는 EVRC에서 AMR로의 상호부호화에만 존재)

Fig. 2. Block Diagram of the proposed transcoding algorithm († exists only in the transcoding from EVRC to AMR).

제안된 알고리즘의 입력은 AMR 음성부호화기에 의해 부호화된 비트열(Input Bit-stream)이다. 입력 비트열은 먼저 파라미터 복호화(Parameter Decoding) 과정을 거친다. 이 부분은 AMR 음성부호화기의 부호화기를 부분적으로 이용하여 구현하였다. 파라미터 복호화 과정에서는 변환해야 할 파라미터들과 필요한 참조 신호들을 부호화하고, 만들어낸다. 복호화되는 파라미터들은 LSP, 피치 지연, 고정 코드북 벡터, 그리고 양 코드북의 이득이다. 이 밖에 부호화된 피치 지연 값과 고정 코드북 벡터 및 코드북 이득을 이용하여 여기 신호도 만들어 낸다. 이 신호는 피치 지연 변환 시에 사용된다. 그 밖에 입력 음성 프레임이 AMR의 어떤 모드로 부호화되었는지에 대한 정보도 파라미터 복호화 과정에서 얻어진다. 파라미터 복호화가 끝나면 프레임 분류(Frame Classification)가 수행되어 입력 음성 프레임을 제 1 프레임형(Frame Type 1)과 제 2 프레임형(Frame Type 2)으로 분류한다. 여기서 제 1 프레임형은 음성이 아닌 배경 잡음이 부호화 되어 있는 프레임이고, 제 2 프레임형은 그 외에 음성이 부호화 되어 있는 프레임이다. 판정을 위해서는 AMR 음성부호화기의 SCR 모듈에서 판정되어 전송된 정보를 이용한다. 프레임 분류가 끝나면 모드 결정(Mode Decision)이 이루어진다. 제 1 프레임형은 음성이 아닌 내용을 담고 있기 때문에 일괄적으로 EVRC 음성부호화기의 Rate 1/8 모드로 모드가 결정된다. 제 2 프레임형은 AMR 모드에 따라 7.95 kbit/s 모드 이상은 Rate 1 모드로, 그 이하의 모드는 Rate 1/2 모드로 대응시킨다. 모드 결정까지 끝나면 각 프레임은 분류 결과에 따라 각각 제 1 프레임형을 위한 상호부호화기(Transcoder for Frame Type 1)와 제 2 프레임형을 위한 상호부호화기(Transcoder for Frame Type 2)에 입력되어 상호부호화 과정을 거치게 된다. 다음 절 등에서 그 자세한 상호부호화 과정 및 알고리즘이 설명된다.

2. LSP 변환

LSP는 제 1 및 제 2 프레임형 모두 전송되는 파라미터이므로 두 프레임형을 위한 상호부호화기 모두에서 변환이 이루어진다. 제안된 상호부호화 알고리즘에서는 LSP를 간단한 선형 보간(Linear Interpolation)으로 변환한다. <그림 3>은 AMR에서 EVRC 음성부호화기로의 상호부호화에서 LSP가 변환되는 과정을 나타낸 것이다. 선형 보간의 식은 다음과 같다.

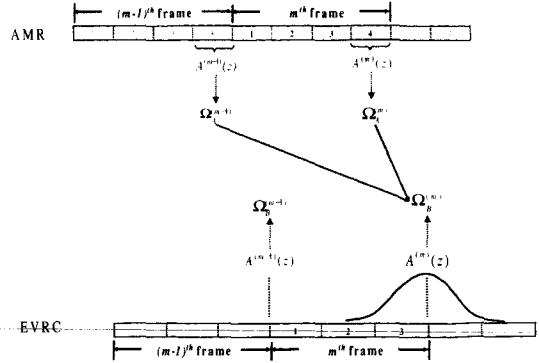


그림 3. AMR에서 EVRC로의 LSP 변환
Fig. 3. LSP transcoding from AMR to EVRC.

$$\Omega_B^{(m)} = \mu_1 \Omega_A^{(m)} + (1 - \mu_1) \Omega_A^{(m-1)} \quad (1)$$

여기서 $\Omega_A^{(m)}$ 과 $\Omega_A^{(m-1)}$ 는 각각 AMR 음성부호화기의 (m)번째와 (m-1)번째 프레임의 LSP이고, $\Omega_B^{(m)}$ 는 EVRC 음성부호화기의 (m)번째 프레임의 LSP이다. 식을 보면 $\Omega_A^{(m)}$ 과 $\Omega_A^{(m-1)}$ 을 계수 μ_1 을 이용하여 선형 보간, 우리가 목적하는 $\Omega_B^{(m)}$ 를 얻어내고 있음을 알 수 있다. AMR과 EVRC 음성부호화기의 음성 프레임의 길이는 20ms로 동일하지만 예전 구간의 길이와 선형 예측 분석 시 사용되는 윈도우(Window)의 모양 및 길이 차이 등 때문에 같은 (m)번째 프레임이라도 약간의 시간차가 존재하게 된다. 선형 보간 계수 μ_1 은 이를 고려하여 실험적으로 결정된다. 본 논문에서는 LPC 스펙트럼의 오차를 측정하는 척도 중의 하나인 평균 스펙트럼 왜곡 측정(Average Spectral Distortion Measure, ASDM)^[10]을 측정하여 원래 EVRC 음성부호화기에서 계산된 LPC 스펙트럼과 가장 유사한 스펙트럼을 구현할 수 있도록 상호부호화가 이루어지게 하는 계수 μ_1 을 선택하였으며 그 값은 0.84였다.

제안된 알고리즘을 사용하면 부가적인 선형 예측 분석 과정이 필요하지 않게 되므로 예전 구간에 의한 알고리즘 지연이 없어서 전체 지연 시간이 짧아지고, 계산량 또한 줄어드는 장점이 있다.

3. 피치 지연 및 적응 코드북 이득 변환

피치 지연 및 적응 코드북 이득 변환은 제 2 프레임형을 위한 상호부호화기에서만 이루어진다. AMR과 EVRC 음성부호화기는 피치 지연을 구하는 방법이 크게 다르다. 합성에 의한 분석 알고리즘에 기반한 AMR

음성부호화기는 음질을 유지하기 위해 피치 지연값을 정확히 검색해서 빠르게 갱신해 줄 필요가 있다. 그렇기 때문에 피치 지연 검색 및 전송이 5ms의 부프레임 단위로 이루어지며 1/3 혹은 1/6 부샘플(Sub-sample) 단위의 정밀도로 피치 지연 값이 검색된다. 이에 반해 EVRC 음성부호화기는 일반화된 합성에 의한 분석 알고리즘에 기반하고 있기 때문에 개회로 추정으로 구한 정수 피치 지연을 프레임 단위로만 전송한다. 단순히 AMR 음성부호화기에서 구한 피치 지연값의 정수 부분을 취하여 변환된 값으로 사용하거나, 단순한 선형 보간법으로 피치 지연값을 변환하면 결과 음성의 품질이 크게 떨어진다. 즉, LSP 파라미터와 달리 피치 지연은 순수한 파라미터간의 변환이 불가능하다는 것이다. 그러므로 부수적인 피치 지연 검색 과정이 불가피하게 요구된다. 이때 상호부호화의 목적 음성부호화기가 EVRC 음성부호화기이므로 피치 지연의 전체적인 구조는 EVRC 음성부호화기의 것을 따른다. EVRC 음성부호화기에서는 피치 지연 검색을 잔여 신호(Residual Signal) 영역에서 수행한다. 본 논문에서는 파라미터 복호화 과정에서 얻은 여기 신호를 잔여 신호로 간주하고 이 영역에서 피치 지연 검색을 수행하였다. 그러나 만약 피치 지연을 전체 검색(Full Search)한다면 이는 Tandem 방식과 동일한 것이 된다. 본 논문에서는 AMR 음성부호화기에서 구한 피치 지연값을 이용하여

피치 지연 검색의 구간을 줄임으로써 계산량을 줄이는 방법을 도입하였다. <그림 4>는 남성 화자가 발음한 8 초 분량의 음성 파형과 해당 음성에 대하여 AMR과 EVRC 음성부호화기가 구한 피치 지연값을 비교하여 나타낸 것이다. 음성 파형에 함께 그린 것은 AMR 음성부호화기의 두가지 VAD 선택중 하나인 VAD1에 의한 음성 유무 판정 결과이다. AMR 음성부호화기의 피치 지연값은 DTX 모드를 동작시키지 않은 10.2 kbit/s 모드의 결과에서 얻은 것이고, EVRC 음성부호화기의 피치 지연값은 부호화 모드를 Rate 1 으로 한정시키고 얻은 것이다. 그림을 보면 음성이 있는 구간에서는 AMR과 EVRC 음성부호화기에서 구한 피치 지연 값이 서로 유사함을 알 수 있다. 실제로 여러 음성들에 대하여 통계적으로 구해보아도 음성이 있는 구간에서 AMR과 EVRC 음성부호화기에서 구한 피치 지연값의 오차는 평균 10 샘플 미만임을 확인할 수 있다. 이 사실을 이용하여 EVRC 음성부호화기를 위한 피치 지연값을 검색할 때 AMR 음성부호화기로부터 복호화하여 얻은 피치 지연값을 이용한다. 먼저 파라미터 복호화에서 얻은 각 부프레임의 피치 지연 4개에서 부샘플(Sub-sample) 단위를 무시한 다음 산술 평균을 취하여 평균 피치 지연값인 T 를 얻는다. 그리고 이 값 주위에서 EVRC 음성부호화기를 위한 최적의 피치 지연값인 τ 를 검색한다. 이를 위해 먼저 식 (2)의 $R(D)$ 를 최대화하는 D 값을 AMR 음성부호화기로부터 얻은 평균 피치 지연 T 부근에서 검색한다.

$$R(D) = \sum_{n=0}^{159-D} \epsilon[n+n_0] \epsilon[n+n_0+D] \quad (2)$$

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

여기서 $\epsilon[n]$ 은 파라미터 복호화 과정에서 얻은 여기 신호이며 D 는 두 가지 n_0 값 80과 160에서 검색된다. 두 가지 n_0 값에서 구한 D 값 중 최적의 피치 지연 τ 를 선택하는 과정은 EVRC 음성부호화기에서 피치 지연을 검색하는 과정을 그대로 따랐다. 이렇게 피치 지연 검색의 구간을 줄이는 방법으로 피치 지연을 변환함으로써 계산량을 줄이면서도 결과 음성의 품질을 유지할 수 있었다.

본 논문에서는 적응 코드북 이득을 별도의 알고리즘 없이 새로 계산하는 방법으로 변환하였다. EVRC 음성부호화기에서 적응 코드북 이득은 피치 지연 검색 과

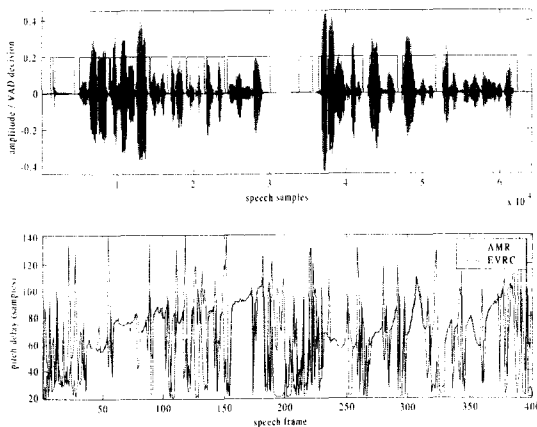


그림 4. 원음성 파형과 VAD 결과(위), AMR과 EVRC에서 구한 피치 지연의 비교(아래)

Fig. 4. Original speech waveform and result of VAD(top), comparison of pitch delays Obtained by AMR and EVRC speech Coders (bottom).

정에서 필요한 신호와 피치 지연 검색의 결과값 및 관련 신호들로 쉽게 계산이 가능하기 때문이다. 그리고 피치 지연값은 고정 코드북 벡터의 변환에서 중요하게 쓰이기 때문에 정확성을 기하는 차원에서도 새로이 계산하는 방법으로 변환하였다.

4. 고정 코드북 벡터 및 고정 코드북 이득 변환

<그림 5>는 같은 음성에 대하여 AMR과 EVRC 음성부호화기가 검색한 고정 코드북 벡터를 비교한 것이다. <그림 5>를 보면 같은 음성에 대해 검색된 것임에도 불구하고 양 고정 코드북 벡터가 전혀 상관관계를 지니고 있지 않음을 알 수 있다. 양 음성부호화기에서 검색된 고정 코드북 벡터들이 통계적으로 어떤 관계를 가지는지 수학적으로 검증하기 위하여 많은 음성 샘플에 대하여 검색된 고정 코드북 벡터들을 모은 뒤 두 고정 코드북 벡터 간의 교차 상관(Cross Correlation)을 구하여 그 분포를 살펴보았다. 그 결과 AMR과 EVRC 음성부호화기에서 검색된 고정 코드북 벡터는 통계적으로 서로 독립임을 확인하였다.^[20-22] 즉, 양 음성부호화기에서 구한 고정 코드북 벡터는 상대방 음성부호화기에서 검색될 고정 코드북 벡터에 대한 정보를 전혀 지니고 있지 않다는 것이다. 그러므로 추가적인 고정 코드북 벡터 검색은 불가피하다는 것을 알 수 있다.

고정 코드북 검색은 AMR과 EVRC 음성부호화기 모두 부호화 과정에서 가장 큰 계산량이 요구되는 부분이다. 그러므로 상호부호화 과정에서 고정 코드북 검색이 불가피하게 수행되어야 한다고 해도 전체 검색을

한다면 Tandem 방식에 비하여 계산량 면의 장점이 없게 된다. 그러므로 본 논문에서는 고정 코드북 검색 시 검색 공간(search space)을 줄여 계산량을 절감하는 알고리즘을 도입하였다. 이 알고리즘은 고정 코드북 벡터를 구성하는 펄스 위치 조합의 가지 수를 제한하여 계산량을 줄이는 것이다. 검색 공간의 제한에는 고정 코드북 검색 시 내부적으로 생성되어 펄스의 부호 등을 미리 결정 하는 데에 사용되는 참조 신호(reference signal)를 이용하였다. 하지만 EVRC 음성부호화기의 Rate 1 모드는 depth-first tree 구조로 고정 코드북 검색을 수행하기 때문에 계산량 감소폭이 적었고, 상대적으로 계산량을 많이 절감할 수 있는 Rate 1/2 모드는 전체 프레임에서 사용되는 빈도가 Rate 1 혹은 Rate 1/8 모드에 비하여 현저히 낮기 때문에 AMR에서 EVRC 음성부호화기로의 상호부호화가 반대 방향에 비하여 Tandem 방식 대비 계산량 절감폭이 작은 결과를 얻었다. AMR 음성부호화기의 고정 코드북 검색 알고리즘은 검색 공간 제한으로 인한 계산량 감소폭이 상대적으로 컸기 때문이다. 이는 성능 평가에서 검증된다. 검색 공간 제한 과정에서 제한폭은 변환된 적응 코드북 이득으로 결정하였다. 적응 코드북 이득은 정의상 적응 코드북 벡터가 여기 신호를 얼마나 잘 모델링하였는지를 나타내는 값이므로 적응 코드북 이득 값이 크면 고정 코드북 검색의 공간을 줄여 계산량을 절감하고, 적응 코드북 이득 값이 작으면 고정 코드북 검색 공간을 늘려 적응 코드북 벡터가 모델링 하지 못한 부

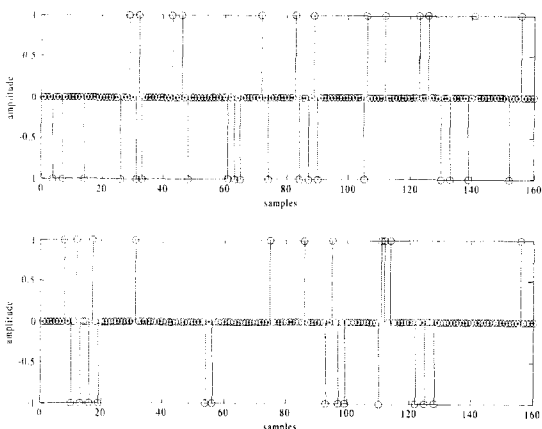


그림 5. 고정 코드북 벡터의 비교
AMR (위), EVRC (아래)
Fig. 5. Comparison of fixed codebook vectors.
AMR (top), EVRC (bottom)

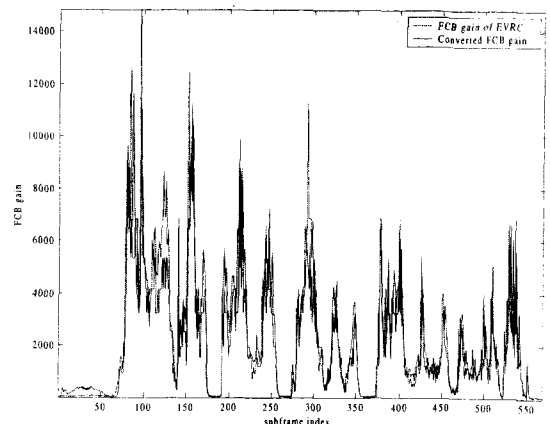


그림 6. EVRC의 고정 코드북 이득(실선)과 변환된 고정 코드북 이득(점선)의 비교
Fig. 6. Comparison of the FCB gain calculated by EVRC coder (solid line) and converted FCB gain(dotted line).

분을 고정 코드북 벡터에서 보완할 수 있도록 하였다. 적응 코드북 이득 값을 이용한 검색 공간 제한폭 변화는 실험적으로 결정하였다.

AMR에서 EVRC 음성부호화기로의 상호부호화에서 고정 코드북 이득은 간단한 선형 보간과 스케일 조정으로 변환이 가능하였다. <그림 6>은 EVRC 음성부호화기에서 계산한 고정 코드북 이득과, AMR에서 구한 고정 코드북 이득을 제안된 알고리즘으로 변환한 것을 비교한 것이다. <그림 6>을 보면 두 값이 거의 유사하다는 것을 확인할 수 있다. 실제로 변환된 값을 이용하였을 경우 음질 저하 또한 없다는 것을 확인하였다.

5. 프레임 에너지 변환

음성이 아닌 배경 잡음 구간을 효율적으로 전송하기 위해 AMR 음성부호화기는 SCR 모듈에 의해 동작하는 저비트율 전송 모드를, EVRC 음성부호화기는 Rate 1/8 모드를 사용한다. 이 모드들에선 수신단에서 comfort noise를 생성하는 데에 필요한 파라미터인 LSP와 프레임 에너지만이 계산되어 전송된다. 프레임 에너지 변환은 제 1 프레임 형을 위한 상호부호화에서만 수행된다.

AMR 음성부호화기에서 프레임 에너지는 8개의 프레임이 연속으로 VAD에 의하여 배경 잡음 구간으로 판정된 경우 각 프레임들에서 계산된 로그 프레임 에너지(Logarithmic Frame Energy)의 평균을 취하여 계산된다. (i)번째 프레임의 로그 프레임 에너지는

$$en_{\log}(i) = \frac{1}{2} \log \left(\frac{1}{L} \sum_{n=0}^{L-1} s^2[n] \right) \quad (3)$$

위의 식으로 계산된다. 여기서 $s[n]$ 은 고대역 통과 필터를 거친 입력 음성 신호이고, L 은 한 프레임의 길이이다. 전송되는 평균 프레임 에너지 $en_{\log}^{mean}(i)$ 은 최근 8 프레임의 로그 프레임 에너지의 산술 평균을 취하여 계산된다. 전송된 프레임 에너지는 수신단에서 comfort noise를 발생시킬 때 로그 영역에서 선형 영역으로 변환된 후 선형 예측 합성 필터의 이득으로 정규화되어 고정 코드북 벡터의 이득으로 사용된다.

EVRC 음성부호화기에서는 AMR과 달리 프레임 에너지가 부프레임 별로 계산된다. 각 부프레임을 위한 프레임 에너지 이득(Frame Energy Gain)은 충격 응답의 에너지를 잔차 신호의 평균의 비로 나누어 계산한다. 프레임 에너지는 양자화 전에 로그 영역으로 변환

되었다가 수신단에서는 다시 선형 영역으로 복원되어 comfort noise 발생 시 이득으로 사용된다.

두 음성부호화기에서 프레임 에너지는 서로 다른 식으로 계산되지만 수신단에서 사용되는 방식은 동일하다. 본 논문에서는 적절한 스케일 조정만을 하는 방식으로 프레임 에너지를 변환하였다. 단, AMR에서 EVRC 음성부호화기로의 상호부호화의 경우, EVRC의 세 부프레임을 위한 프레임 에너지를 같은 값으로 설정하였다. 이 방식으로 프레임 에너지를 변환한 결과 음질에 전혀 저하가 없었다. 그러나 본 논문은 모든 모의 실험에서 입력 음성에 잡음이 없고(clean speech), 전송 과정의 오류도 없다고 가정했으므로 그런 상황에서 음질의 저하를 가져오지 않는 정상적인 comfort noise의 생성이 이루어지는 지는 후의 실험으로 확인해보아야 할 것이다.

IV. EVRC에서 AMR 음성부호화기로의 상호부호화

1. 전체적인 구조

EVRC에서 AMR 음성부호화기로의 상호부호화의 구조도 <그림 2>에 나타난 바와 같다. 입력은 EVRC 음성부호화기에서 부호화된 비트열이다. 입력된 비트열은 먼저 파라미터 복호화 과정을 거친다. 복호화되는 파라미터들은 LSP, 피치 지연, 고정 코드북 벡터, 그리고 양 코드북의 이득이다. EVRC에서 AMR 음성부호화기로의 상호부호화의 경우 파라미터 복호화 과정에서 음성을 합성해 낸다. 음성의 합성은 복호화된 파라미터들로 여기 신호를 만들고, 이 여기 신호를 복호화한 LPC로 구성된 합성 필터에 통과시켜 이루어진다. 이렇게 합성된 음성을 가중 필터에 통과시켜 각각 가중된 음성을 얻는다. 이 음성은 피치 지연 변환에 사용된다. 입력 음성 프레임이 EVRC 음성부호화기의 어떤 모드로 부호화 된 것인지에 대한 정보도 파라미터 복호화 과정에서 얻어진다. 프레임 분류 과정에서 입력 프레임이 제 1 프레임형과 제 2 프레임형으로 분류되는 것은 AMR에서 EVRC로의 상호부호화와 동일하다. EVRC 음성부호화기는 별도의 SCR 모듈이 없지만 음성이 없는 구간을 Rate 1/8 모드로 부호화하므로 이 정보를 이용하여 프레임을 분류한다. 프레임 분류 후에는 모드가 결정된다. 제 1 프레임형은 음성이 없는 구간이므로

AMR 음성부호화기의 SCR 모듈이 동작하는 바에 따라 저 비트율 전송 모드로 모드가 결정된다. 반면에 제 2 프레임형은 외부 모드 입력에 따라 각 프레임의 부호화 모드가 임의로 결정된다. 이는 모드가 내부적인 모드 대응으로 결정되는 AMR에서 EVRC 음성부호화기의 상호부호화와 다른 부분인데 AMR 음성부호화기는 부호화 모드가 외부 네트워크의 입력에 의해 결정된다는 점을 반영하여 이렇게 외부에 의해 결정되도록 하였다. 모드 결정이 끝나면 각 프레임 형을 위한 상호부호화기에 의하여 상호부호화가 이루어지게 된다. 각 과정은 다음 절 들에서 자세히 설명된다.

2. LSP 변환

EVRC에서 AMR 음성부호화기로의 LSP 변환도 AMR에서 EVRC로의 경우와 마찬가지로 간단한 선형 보간을 통해 이루어진다. <그림 7>은 EVRC에서 AMR 음성부호화기로의 LSP 변환 과정을 나타낸다. 구체적인 선형 보간의 식은 다음과 같다.

$$\Omega_A^{(m)} = \mu_2 \Omega_B^{(m)} + (1 - \mu_2) \Omega_B^{(m-1)} \quad (4)$$

식을 보면 EVRC 음성부호화기의 (m)번째와 (m-1)번째 프레임의 LSP인 $\Omega_B^{(m)}$ 과 $\Omega_B^{(m-1)}$ 를 계수 μ_2 를 이용하여 선형 보간, 우리가 목적하는 AMR 음성부호화기의 (m)번째 프레임의 LSP인 $\Omega_A^{(m)}$ 를 구하는 것을 확인할 수 있다. 선형 보간 계수 μ_2 는 AMR에서 EVRC로의 경우와 마찬가지로 ASDM을 계산하여 실험적으로 결정하였으며 그 값은 0.95였다.

제안된 알고리즘을 사용함으로써 AMR에서 EVRC로의 상호부호화와 같은 이유로 지연 시간과 계산량을

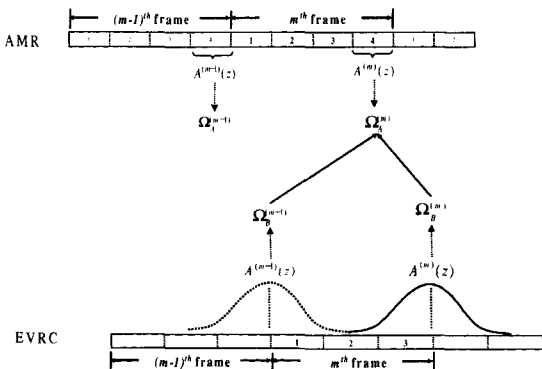


그림 7. EVRC에서 AMR로의 LSP 변환
Fig. 7. LSP transcoding from EVRC to AMR.

절감할 수 있었다.

3. 피치 지연 및 적응 코드북 이득 변환

EVRC에서 AMR 음성부호화기로의 피치 지연 변환은 EVRC 음성부호화기로부터 얻은 정수 피치 지연 부근에서 AMR 음성부호화기를 위한 피치 지연을 검색함으로써 이루어진다. 이때 목적 음성부호화기가 AMR이므로 전체적인 피치 지연 변환 과정은 AMR의 피치 지연 검색 과정의 구조를 따른다.

AMR 음성부호화기는 두 단계에 걸쳐 피치 지연값을 추정한다. 1단계는 개회로 피치 지연 추정으로 대략의 정수 피치 지연을 계산하는 과정이고, 2단계는 앞 단계에서 구한 정수 피치 지연 부근에서 부샘플 단위의 정확한 피치 지연값을 폐회로(Closed-loop) 방식으로 검색하는 과정이다. 제안된 알고리즘에서는 1단계의 개회로 피치 지연 검색을 생략하고 EVRC 음성부호화기에서 구한 정수 피치 지연값을 2단계 검색에서 검색 공간을 제한하는 값으로 바로 사용케하였다. AMR 음성부호화기에서 피치 지연 검색은 각각 가중된 음성 영역에서 이루어지는데 이는 앞의 파라미터 복호화 과정에서 이미 구했으므로 그대로 사용하면 된다. 이와 같은 방식으로 피치 지연을 변환함으로써 계산량을 줄이면서도 결과 음성의 품질을 유지할 수 있었다.

적응 코드북 이득은 AMR에서 EVRC 음성부호화기로의 상호부호화와 마찬가지로 직접 계산함으로써 변환된다.

4. 고정 코드북 벡터 및 고정 코드북 이득 변환

EVRC에서 AMR 음성부호화기로의 상호부호화도 AMR에서 EVRC로의 경우와 마찬가지로 음질의 저하를 막기 위해 부가적인 고정 코드북 벡터 검색 과정이 필요하다. 본 논문에서는 계산량을 줄이기 위해 고정 코드북 검색 시 검색 공간을 줄이는 방식을 도입하였다. 검색 공간을 제한하는 기준으로는 고정 코드북 검색 과정에서 내부적으로 생성되어 펄스의 부호 결정 등에 사용되는 참조 신호를 사용하였다.

AMR 음성부호화기의 10.2 kbit/s 모드에 고속화 알고리즘을 적용한 결과를 <표 2>에 정리하였다. 이 모드의 실험에서는 10.2 kbit/s 모드에서 펄스 위치를 검색하는 576가지의 조합을 그 1/3인 196가지로 줄였다. 표의 결과를 보면 남/녀 화자가 발음한 입력 음성 모두에 대하여 큰 음질 저하 없이 계산량을 전체적으로 11%가량 절감했다는 것을 확인할 수 있다. 10.2 kbit/s 에서

고정 코드북 검색이 부호화 과정에 필요한 계산량의 절반 가량을 차지한다는 점을 고려하면 고정 코드북 검색에 필요한 계산량의 20% 가량을 절감했다는 것을 알 수 있다. 다른 모드들에 대해서도 같은 방식으로 검색 공간을 제한하여 계산량을 절감하였다.

표 2. 고속화 알고리즘의 적용 결과
(10.2 kbit/s 모드의 경우)

Table 2. Performance of the fast algorithm.
(case of 10.2 kbit/s mode)

Original	11.84	11.81	4.059	3.881
Fast	10.50	10.47	4.071	3.856

검색 공간 제한의 정도를 결정하기 위해서는 AMR에서 EVRC로의 상호부호화와 마찬가지로 적응 코드북 이득을 이용하였다. 이는 AMR 음성부호화기에서도 적응 코드북 이득이 적응 코드북 벡터가 얼마나 여기 신호를 얼마나 잘 모델링 하였는지의 척도로 사용이 가능하기 때문이다. 구체적으로 적응 코드북 이득이 크면 검색 공간의 제한폭을 늘려 계산량을 줄이고, 적응 코드북 이득이 작으면 검색 공간을 덜 제한해 고정 코드북 벡터를 정확히 검색함으로써 음질 저하를 막았다.

고정 코드북 이득도 AMR에서 EVRC로의 상호부호화와 마찬가지로 간단한 선형 보간과 스케일 조정으로 변환이 가능하였으며, 그에 따른 음질 저하도 없었음을 확인하였다.

5. 프레임 에너지 변환

프레임 에너지 변환은 음성이 아닌 배경 잡음 구간을 담고 있는 제 1 프레임형을 위한 상호부호화에서만 수행된다.

매 부프레임당 하나씩, 한 프레임에 총 3개의 프레임 에너지가 전송되는 EVRC 음성부호화기와 달리 AMR 음성부호화기는 프레임당 하나의 프레임 에너지만을 필요로한다. 본 논문에서는 EVRC 음성부호화기에서 얻은 3개의 프레임 에너지의 산술 평균을 취한 후 스케일을 적절히 조절하여 AMR 음성부호화기를 위한 프레임 에너지를 얻는 방법으로 프레임 에너지를 변환하였다. 모의 실험 결과 이와 같은 변환으로 인하여 음질의 저하는 발생하지 않았다. 그러나 입력 음성에 잡음이 존재하는 경우와 전송 과정의 오류가 존재하는 상

황하에서의 성능은 추후 실험으로 검증해보아야 할 것이다.

V. 성능 평가

본 장에서는 제안된 파라미터 직접 변환 방식의 상호부호화 알고리즘의 성능을 다양한 주/객관적 지표를 이용하여 평가한다. 성능 평가는 계산량, 지연 시간, 객관적/ 주관적 음질 평가로 이루어져있다. 성능 평가를 위하여 제안된 파라미터 직접 변환 방식의 상호부호화 알고리즘을 C 언어를 이용하여 고정 소수점 형태로 구현하였다. 그리고 입력 음성으로는 남자 4명, 여자 4명에 의해 발음된 8초 길이의 한국어 음성 16 가지가 모든 모의 실험에서 공통적으로 사용되었다. 각 음성 파일은 8kHz로 표본화 되었으며 13 비트 선형 PCM 형식이다. 전송 과정의 오류는 없다고 가정하였다.

1. 계산량

Weighted Million Operations Per Second (WMOPS)를 측정하여 제안된 알고리즘의 계산량을 평가하였다. <표 3>은 제안된 파라미터 직접 변환 방식의 계산량과 Tandem 방식의 계산량을 비교한 것이다. <표 3>의 결과를 보면 제안된 파라미터 직접 변환 방식의 계산량이 Tandem 방식에 비하여 최저 42%에서 최고 58% 적음을 확인할 수 있다. 이러한 계산량 감소는 부가적인 선형 예측 분석의 생략, 피치 지연 검색 범위의 제한, 효율적인 고속 고정 코드북 벡터 검색 알고리즘의 도입 등으로 가능했다.

표 3. 계산량 비교

Table 3. Comparison of computational complexity.

Tandem	13.62	14.29	13.20	13.22
Proposed	7.48	8.22	5.58	6.01
감소율	45%	42%	58%	55%

2. 지연 시간

음성 통신 시스템의 총 지연 시간은 알고리즘 (Algorithmic), 처리(Processing), 그리고 전송 (Transmission) 지연 시간을 합하여 계산이 가능하다. 그러나 전송 지연은 통신망의 구조와 상태에 따라 달라지기 때문에 본 논문에서는 고려하지 않고 알고리즘 지연과

처리 지연 시간의 합을 총 지연 시간으로 가정한다. Tandem 방식의 상호부호화 알고리즘(D_{AB}^{td}, D_{BA}^{td})과 제안된 파라미터 직접 변환 방식의 상호부호화 알고리즘(D_{AB}^{di}, D_{BA}^{di})의 총 지연 시간을 정리하면 다음과 같다.

$$D_{AB}^{td} = 35ms + \alpha_A + \beta_A + \alpha_B + \beta_B \quad (5)$$

$$D_{AB}^{di} = 25ms + \alpha_A + P_{AB} + \beta_B \quad (6)$$

$$D_{BA}^{td} = 35ms + \alpha_B + \beta_B + \alpha_A + \beta_A \quad (7)$$

$$D_{BA}^{di} = 30ms + \alpha_B + P_{BA} + \beta_A \quad (8)$$

위의 식에서 A는 AMR 음성부호화기를, B는 EVRC 음성부호화기를 의미하며, AB는 AMR에서 EVRC로의, 그리고 BA는 EVRC에서 AMR로의 상호부호화를 뜻한다. 그리고 α_m 과 β_m ($m=A$ 혹은 B)은 각각 양 음성부호화기의 부호화기와 복호화기의 처리 지연이며, P_{AB} 와 P_{BA} 는 각각 제안된 상호부호화 알고리즘에서 AB와 BA 방향의 상호부호화를 수행하는 데에 필요한 처리 지연이다. Tandem 방식에서의 알고리즘 지연 35ms는 한 프레임의 길이 20ms에 AMR과 EVRC 음성부호화기의 예견 구간 길이 10ms와 5ms를 더해 얻은 값이다. 이에 반해 파라미터 직접 변환 방식은 부가적인 선형 예측 분석없이 간단한 선형 보간으로 LSP를 변환하므로 AMR에서 EVRC 음성부호화기로의 상호부호화의 경우 EVRC 음성부호화기에서 요구되는 10ms의 예견 구간이, 그리고 EVRC에서 AMR 음성부호화기로의 상호부호화의 경우 AMR 음성부호화기에서 요구되는 5ms의 예견 구간이 필요 없게 되어 알고리즘 지연이 각각 25ms와 30ms로 짧아지게 된다. 게다가 상호부호화에 필요한 처리 지연도 제안된 파라미터 직접 변환 방식의 상호부호화 알고리즘이 $P_{AB} \ll \beta_A + \alpha_B$, $P_{BA} \ll \beta_B + \alpha_A$ 로 짧기 때문에 결과적으로 제안된 알고리즘의 총 지연 시간이 기존의 Tandem 방식에 비하여 많이 짧아지게 된다. <표 4>는 실제로 Tandem 방식과 제안된 파라미터 직접 변환 방식으로 상호부호화를 수행하여 그 처리 지연을 측정하여 비교한 것이다.

앞에서 언급된 8초 분량의 시험 음성을 처리한 시간을 측정하여 평균을 취한 것이며, 모든 실험을 수행한 컴퓨터의 환경은 인텔 펜티엄4 1.8GHz, 512MB 주메모리, 그리고 마이크로소프트 윈도우 XP 운영체제이다. 표를 보면 실제로 제안된 파라미터 직접 변환 방식의

처리 지연이 Tandem 방식보다 짧다는 것을 확인할 수 있다.

표 4. 처리 지연 시간 비교

Table 4. Comparison of processing delay.

Algorithm	AMR → EVRC		EVRC → AMR	
	Male	Female	Male	Female
Tandem	3.42	3.10	3.48	3.20
proposed	3.44	3.09	3.47	3.11

3. 객관적 음질 평가

여러 가지 객관적 음질 평가 모델 중 우리는 ITU P.862 Perceptual Evaluation of Speech Quality (PESQ)^[12]를 사용하여 제안된 상호부호화 알고리즘의 객관적 음질을 평가하였다. <표 5>는 Tandem 방식과 제안된 파라미터 직접 변환 방식의 PESQ 점수를 정리한 것이다. <표 5>를 보면 Tandem 방식과 제안된 파라미터 직접 변환 방식의 PESQ 점수 차가 PESQ의 오차 범위 내일 정도로 미미하고 어떤 경우에는(AMR에서 EVRC로의 상호부호화에서 남성 화자가 발음한 음성을 입력으로 사용한 경우) 제안된 방식의 음질이 더 뛰어나게 나온 것을 알 수 있다. 이 결과는 제안된 파라미터 직접 변환 방식의 상호부호화 알고리즘이 Tandem 방식과 전체적으로 동등한 음질을 구현함을 보여주고 있다.

표 5. 객관적 음질 평가 결과

Table 5. Comparison of PESQ score.

Algorithm	AMR → EVRC		EVRC → AMR	
	Male	Female	Male	Female
Tandem	3.42	3.10	3.48	3.20
proposed	3.44	3.09	3.47	3.11

4. 주관적 음질 평가

주관적 음질 평가를 위하여 선호도 평가(ABX Test)를 실시하였다. 10명의 훈련되지 않은 청자가 실험에 참가하였으며 청자들은 각각 Tandem 방식과 파라미터 직접 변환 방식으로 상호부호화 된 음성을 연달아 들은 후 어떤 결과 음성의 품질이 뛰어난지, 혹은 두 음성의 품질을 구분할 수 없는지 평가하였다. 실험에 사용된 입력 음성은 앞에서 밝힌 바와 같으며 실험시 Meridian의 566DAC와 Senheiser의 HD600 헤드폰을 사용하였다. <표 6>은 주관적 음질 평가의 결과를 정

리한 것이다. <표 6>의 결과를 보면 청자들이 Tandem 방식과 파라미터 직접 변환 방식으로 상호부호화된 음성의 품질을 거의 구분하지 못했다는 것을 확인할 수 있다. 이는 앞 절의 객관적 음질 평가 결과와도 일치하는 것으로, 제안된 파라미터 직접 변환 방식의 상호부호화 알고리즘이 Tandem 방식과 동등한 음질을 구현한다는 사실을 입증하고 있다.

표 6. 주관적 음질 평가 결과
Table 6. Results of ABX test.

Tandem	30%	32%	35%	27%
proposed	40%	33%	32%	30%
No Preference	30%	35%	33%	43%

VI. 결 론

본 논문은 이동 통신망에서 널리 쓰이고 있는 AMR과 EVRC 음성부호화를 위한 새로운 파라미터 직접 변환 방식의 상호부호화 알고리즘을 제안하였다. 제안된 알고리즘은 C 언어로 구현되어 그 성능이 여러 가지 주,객관적 지표를 통하여 평가되었다. 앞 장에서 제시된 성능 평가 결과는 제안된 알고리즘이 Tandem 방식에 비하여 적은 계산량과 짧은 지연 시간으로도 동등한 수준의 음질을 구현함을 보여주고 있다.

각종 통신망의 통합이라는 최근의 추세 속에서 가장 기초적인 통신 수단이자 매체라고 할 수 있는 음성 통신의 효율적인 운용을 위한 상호부호화 기술의 개발은 매우 중요한 의미를 지니게 되었다. 제안된 알고리즘은 비록 AMR과 EVRC 음성부호화기 고유의 성능을 부분적으로 이용하는 하였으나 기본적으로 CELP 계열의 음성부호화기들에 공통적으로 적용할 수 있는 구조를 지니고 있다. 그러므로 본 알고리즘을 G.729A, G.723.1 등 VoIP 망에 쓰이는 음성부호화기나 Selectable Mode Vocoder(SMV) 등 새로이 개발된 이동 통신용의 음성부호화기에 적용해보는 것은 새로운 도전이 될 것이다. 그리고 본 논문에서는 고려되지 않은 입력 음성에 잡음이 첨가된 경우와 전송 과정에 오류가 발생한 경우 등도 새로이 고려되어야 할 환경 들이다. 이러한 환경에 효율적으로 대응하기 위해서는 기존 음성부호화기가 이와 같은 상황을 위하여 채택하고 있는 잡음 억제,

오류 은닉 알고리즘이 상호부호화에 끼치는 영향을 살펴 봐야 할 것이다.

참 고 문 헌

- [1] 3GPP TS 26.071 V5.0.0, *AMR Speech CODEC: General Description*, Jun., 2002.
- [2] 3GPP TS 26.090 V5.0.0, *Adaptive Multi-Rate (AMR) speech codec: Transcoding functions*, Jun., 2002.
- [3] 3GPP TS 26.093 V5.0.0, *Adaptive Multi-Rate (AMR) speech codec: Source controlled rate operation*, Jun., 2002.
- [4] 3GPP TS 26.094 V5.0.0, *Adaptive Multi-Rate (AMR); Voice Activity Detector(VAD)*, Jun., 2002.
- [5] 3GPP TS 26.092 V5.0.0, *Adaptive Multi-Rate (AMR); Comfort noise aspects*, Jun., 2002.
- [6] 3GPP TS 26.091 V5.0.0, *Adaptive Multi-Rate (AMR); Error concealment of lost frames*, Jun., 2002.
- [7] 3GPP TS 26.073 V5.0.0, *ANSI-C code for Adaptive Multi-Rate(AMR) speech codec*, Jun., 2002.
- [8] TIA/EIA/IS-127, *Enhanced variable rate codec, speech service option 3 for wideband spread spectrum digital systems*, 1997.
- [9] ITU-T Rec. G.711, *Pulse code modulation of voice frequencies*, 1988.
- [10] ITU-T Rec. G.723.1, *Dual-rate Speech Codec for Multimedia Communications Transmitting at 5.3 and 6.3 kbit/s*, 1996.
- [11] ITU-T Rec. G.729 Annex A, *Reduced Complexity 8 kbit/s CS-ACELP Speech Codec*, 1996.
- [12] ITU-T Rec. P.862, *Perceptual evaluation of speech quality(PESQ), an objective method for end-to-end speech quality assessment of narrowband telephone networks and speech codecs*, 2000.
- [13] W.B. Kleijn, "Analysis-by-Synthesis Speech Coding Based on Relaxed Waveform-Matching

- Constraints," *Ph.D. dissertation, Delft University of Technology*, 1991.
- [14] W.B. Kleijn, P. Kroon, "The RCELP Speech-Coding Algorithm," *European Trans. On Telecom.*, Vol. 5, No. 5, pp. 573~582, 1994.
- [15] Manfred R. Schroeder, Bishnu S. Atal, "Code Excited Linear Prediction(CELP): High-Quality Speech at Very Low Bit Rates," *Proc. Of ICASSP*, pp. 937~940, 1985.
- [16] F.K. Soong, B.H. Juang, "Line Spectrum Pair (LSP) and speech data compression," *Proc. of ICASSP*, pp. 1.10.1~1.10.4, 1984.
- [17] Hong-Goo Kang, Hong-Kook Kim, R.V. Cox, "Improving transcoding capability of speech coders in clean and frame erased channel environments," *Proc. of IEEE Workshop on Speech Coding*, pp. 78~80, Jan., 2000.
- [18] A.M. Kondo, *Digital Speech: Coding for Low Bit Rate Communications Systems*, John Wiley & Sons Ltd, 1994.
- [19] K.K. Paliwal, B.S. Atal, "Efficient vector quantization of LPC parameters at 24 bit/frame," *IEEE Trans. on Speech and Audio Processing*, Vol. 1, No. 1, pp. 3~14, 1993.
- [20] Yongmiao Hong, "Testing for independence between two covariance stationary time series," *Biometrika*, Vol. 83, No. 3, pp. 615~625, 1996.
- [21] E.J. Hannan, *Multiple Time Series*, John Wiley and Sons Inc., 1970.
- [22] 이선일, AMR과 EVRC 음성부호화기를 위한 새로운 Tandemless 방식의 상호 부호화 알고리즘, 석사 학위 논문, KAIST, 2002.
- [23] 이선일, 유창동, "AMR과 EVRC 음성 부호화기간의 비탄덤 방식을 이용한 상호 부호화," 한국음향학회지, 제 21권 제 6호, pp. 531~542, 2002.
- [24] Kyung Tae Kim, Sung Kyo Jung, Young Cheol Park, Yong Soo Choi, Dae Hee Youn, "An efficient transcoding algorithm for G.723.1 and EVRC speech coders," *Proc. of 54th IEEE VTC*, Vol. 3, pp. 1561~1564, 2001.
- [25] 윤성완, 정성교, 박영철, 윤대회, "8kbps G.729A에서 5.3 kbps G.723.1로의 상호 부호화 알고리즘," 신호처리 합동 학술대회 논문집 제 13권 제 1호, pp. 832~826, 2000.
- [26] Qualcomm Inc., *CELP-based to CELP-based vocoder packet translation*, US Patent No. US6260009B1, Jul., 2001.

 저 자 소 개

李 善 一(學生會員)

2001년 2월 연세대학교 기계전자공학부 전자공학전공 학사, 2002년 8월 한국과학기술원 전자전산학과 전기 및 전자공학전공 석사, 2002년 9월~현재: 한국과학기술원 전자전산학과 전기 및 전자공학전공 박사과정, <주관심분야 : 음성부호화, 디지털 신호처리>

柳 昌 東(正會員)

1986년 California Institute of Technology 학사, 1988년 Cornell University 석사, 1996년 MIT 박사, 1997년 1월~1999년 3월: KT Senior Researcher, 1999년 3월~현재: 한국과학기술원 전자전산학과 조교수, <주관심분야 : 음성신호처리, 디지털 신호처리>