



일반논문 (Regular Paper)

방송공학회논문지 제29권 제2호, 2024년 3월 (JBE Vol.29, No.2, March 2024)

<https://doi.org/10.5909/JBE.2024.29.2.180>

ISSN 2287-9137 (Online) ISSN 1226-7953 (Print)

ACELP-TCX 전이에서 3GPP EVS 코덱의 성능 개선

김 인 철^{a)†}

A Performance Improvement of the 3GPP EVS Codec in the Transition between the ACELP and the TCX

Rin Chul Kim^{a)†}

요 약

본 논문에서는 EVS에서 시간 영역 코덱과 TCX 사이의 전이에서 고 대역에 발생하는 스펙트럼 홀을 감소시킬 수 있는 기법을 제안한다. 먼저 TCX 구간에서도 고 대역 입력과 대역 확장 여기 신호를 끊임없이 생성한다. TCX 다음의 시간 영역 코덱에서 고 대역 복원 신호에도 cross-fading을 적용한다. 시간 영역 코덱 다음의 TCX에서 첫부분에 미리 확장 기법으로 고 대역 신호를 합성한다. 시뮬레이션 결과 수정된 기법은 EVS 표준에 비해 스펙트럼 홀을 크게 줄일 수 있음을 알 수 있었다.

Abstract

In this paper, the EVS codec is revised to alleviate the high-frequency (HF) spectral holes appearing in the transition between the time-domain codec and the TCX. In the revised codec, the HF input and bandwidth extended excitation signal are generated even in the TCX mode. The cross-fading is also applied to the HF reconstructed signal in the time-domain codec after the TCX. The time-domain HF reconstructed signal is synthesized from the result from the previous frame using the mirror extension. The simulation reveals the revised codec can effectively remove the HF spectral holes, as compared with the EVS codec.

Keyword : Enhanced voice service, High-frequency spectral hole, Codec transition

a) 서울시립대학교 전자전기컴퓨터공학부(dept. of ECE, Univ. of Seoul)

† Corresponding Author : 김인철(Rin Chul Kim)

E-mail: rin@uos.ac.kr

Tel: +82-2-6490-2336

ORCID: <https://orcid.org/0000-0002-7643-8717>

※ This work was supported by the 2020 Research Fund of the University of Seoul.

· Manuscript 30 January, 2024; Revised 26 February, 2024; Accepted 28 February, 2024.

1. 서론

최근 MPEG 3D 오디오 (3DA)^[1], Opus 코덱^[2], 3GPP의 enhanced voice service (EVS)^[3] 등이 오디오 코덱 표준으로 잘 알려져 있다. 이 중에서 EVS는 부호화 성능이 우수할 뿐만 아니라, 부호기와 복호기 모두 최적화된 코덱이 공개되어 있어^[4] 오디오 코덱의 성능 평가 및 개발의 기준으로 채택되고 있다. EVS 코덱은 algebraic code excited linear prediction (ACELP), transform coded excitation (TCX), high-quality (HQ) 등 3가지 코덱 중 하나를 선택하여 부호화한다. 이때 MDCT 기반 코덱인 TCX 혹은 HQ와 시간 영역 코덱인 ACELP 사이에 전이(transition)가 발생할 때 적절한 처리가 수반되지 않으면 매우 심한 부호화 손실이 발생한다.

3GPP의 AMR-WB^[5]에서는 ACELP에서 TCX로 전이될 때 TCX의 좌측 윈도우를 사각 윈도우로 적용하고, ACELP 가중 필터 및 합성 필터에서 얻어지는 zero impulse response (ZIR)를 제거하여 코덱 전이에 의한 영향을 최소화하였다. 한편, MPEG USAC^[6]에서는 ACELP와 인접한 TCX 프레임에 각각 128샘플만큼 추가 데이터를 전송하는 forward aliasing cancellation (FAC) 기법을 도입하였다. 3GPP EVS에서도 이와 유사한 방법으로 코덱이 전이될 때의 문제점을 해결하려고 노력하였다. 그러나 ACELP와 함께 시간 영역 대역 확장 (time-domain bandwidth extension; TBE) 기법으로 고 대역을 확장하는 시간 영역 코덱은 MDCT 기반 코덱과 고 대역에서 효과적으로 결합되지 않았다. 그 결과 ACELP와 MDCT 기반 코덱이 전이될 때 스펙트로그램 (spectrogram)을 보면 고 대역에 수직 막대 형태의 스펙트럼 홀 (spectral hole)이 발생한다. 이러한 현상은 특히 중고음 영역에 하모닉 구조가 강하지 않고 고 대역 에너지가 비교적 높은 신호들에서 크게 나타난다.

본 논문에서는 EVS에서 ACELP와 TBE를 채택하는 시간 영역 코덱과 MDCT 기반 코덱인 TCX가 전이될 때, 고 대역에서 스펙트럼 홀 발생을 감소시키는 방법을 제시한다. 먼저, 고 대역 신호 발생에 사용되는 대역 확장된 여기 신호 (bandwidth extended excitation signal)와 고 대

역 linear prediction (LP) 계수 추정을 위해 사용되는 고 대역 입력 신호를 TCX가 적용되는 프레임에도 계속 생성한다. 두 번째, TCX에서 시간 영역 코덱으로 전이되는 프레임에서 TBE로 산출되는 고 대역 복원 신호에 저 대역 복원 신호와 동일하게 cross-fading을 적용하여, 불완전하고 중복하여 복원된 고 대역 신호에 의한 영향을 최소화한다. 세 번째, TBE에서 TCX로 전이되는 프레임에서 처음 148샘플에 대해 미리 확장 기법으로 고 대역 신호를 합성하여 저 대역 신호와 합하여 복원된 신호를 산출한다. 본 논문에서 제시된 방법은 시뮬레이션을 통해 기존의 방법과 성능을 비교한다. 본 논문에서는, 별도로 표시되지 않으면, 편의상 신호의 표본화 주파수를 32KHz로 가정하고 시간 지연 및 동작 시간 등을 샘플 수로 표기하였다. 그리고 ACELP의 동작 주파수는 12.8KHz로 가정하였다. 필요하다면, 논문에 제시한 샘플 수에 적절한 비례 상수를 곱하여, 이와 다른 동작 주파수에서의 샘플 수를 얻을 수 있다.

II. 시간 영역 코덱

EVS^[3]의 시간 영역 코덱은 저 대역 부호화로 ACELP를 채택하고, 고 대역 확장을 위해 TBE 혹은 frequency-domain bandwidth extension (FBE) 기법을 사용한다. FBE는 MDCT를 이용하여 고 대역 신호를 확장하는 방법으로, 시간 영역 코덱과 MDCT 기반 코덱과의 전이가 발생할 때 적절하게 결합된다. 반면에 TBE는 코덱 간 전이가 발생할 때 효과적으로 결합되지 않는다. 본 절에서는 ACELP와 TBE를 채택하는 시간 영역 코덱을 간략히 살펴보고, 시간 영역 코덱과 TCX 사이의 전이에 대해 살펴볼 것이다. III절에서는 시간 영역 코덱과 TCX 사이에 전이가 발생할 때의 문제점을 분석하고 그 해결책을 제시한다.

1. 시간 영역 코덱 개요

그림 1에서는 TBE 복호기를 중심으로 시간 영역 코덱의 복호기의 구조를 제시하였다. 그림 1을 참조하면, ACELP 복

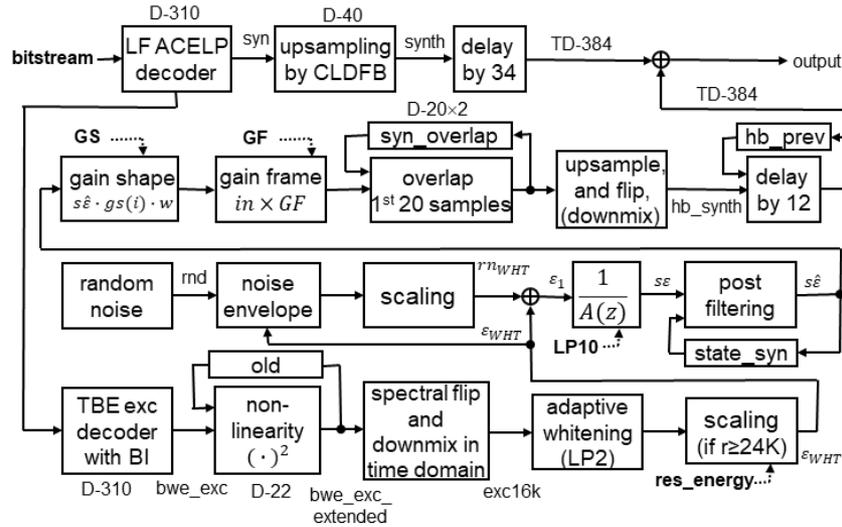


그림 1. 시간 영역 EVS 복호기의 블록선도
 Fig. 1. Block diagram of the time-domain EVS decoder

호기는 비트열 (bitstream)을 입력받아 저 대역 신호 syn을 복원한다. 이 신호는 complex low-delay filter bank (CLDFB)로 32KHz로 상향 표본화되고 34샘플 지연된 다음, 복원된 고 대역 신호와 합해진다. 이때 ACELP와 CLDFB의 시간 지연은 각각 310, 40샘플이므로 총 384샘플의 시간 지연이 발생한다.

ACELP 복호기에서는 여기 신호로부터 이중 선형 보간 (bilinear interpolation)으로 대역 확장된 여기 신호 bwe_exc를 얻는다. 대역 확장된 여기 신호는 비선형 연산자로 고주파 영역에 하모닉을 생성시킨 다음, 다운믹스를 통해 표본화 주파수를 16KHz로 변경한다. 다운믹스된 여기 신호는 2차 LP 분석으로 백색 잡음화를 수행하고, 별도로 생성된 잡음 신호와 합해져 고 대역 여기 신호를 합성한다. 고 대역 여기 신호는 수신된 LP 합성 필터로 합성되고, 후 처리 된다. 그런 다음, 수신된 이득 조정 (gain shape) 파라미터와 이득 프레임 (gain frame) 파라미터로 이득 조정이 되고 다시 32KHz로 상향 표본화된다. 이런 과정을 통해 얻은 고 대역 복원 신호는 총 시간 지연이 384샘플이 되도록 12샘플을 추가 지연시킨다. 그런 다음 상향 표본화된 저 대역 복원 신호와 합해져서 32KHz 표본화 주파수로 복원된 신호를 얻는다.

2. 시간 영역 코덱과 TCX 사이의 전이

본 절에서는 ACELP와 TBE로 구성되는 시간 영역 코덱과 TCX가 전이될 때의 동작을 살펴보면서 두 코덱이 전이될 때의 문제점을 분석한다. 그림 2에서는 첫 번째 프레임 $x_{old2}[n]$ 이 TCX, 두 번째 프레임 $x_{old}[n]$ 이 ACELP, 그리고 3번째 프레임 $x[n]$ 이 TCX로 부호화될 때 복호기에서의 타이밍 다이어그램을 제시한다. 그림 2에서 표본화 주파수가 32KHz이고, 가로축인 시간을 샘플 수로 나타내었다. 그림 2에서 첫 번째 줄은 입력 신호가 30샘플 지연된 다음 각 부호기에 입력되는 640샘플 길이의 프레임 3개를 나타내었다. 2~4번 줄은 시간 영역 코덱의 복원 신호, 5~7번 줄은 TCX로 복원된 신호를 각각 나타내었다. 마지막 8번 줄은 최종적인 복원 신호의 시간 위치를 나타내었다. 각 코덱에 입력되는 신호는 입력으로부터 30샘플 지연되는 것을 참조하면, 최종적인 복원 신호는 입력으로부터 384샘플 지연된다.

먼저 첫 번째 프레임 $x_{old2}[n]$ 은 TCX 코덱에서 부/복호화 과정을 거쳐 6번째 줄과 같은 신호를 얻는다. 이 신호는 아래쪽의 윈도우와 곱해지고, 이전 프레임 결과와 overlap-and-add (OLA) 과정을 통해 pcmbuffB로 표기된 블록만큼의 복원된 신호가 산출된다. 이때 pcmbuffB는 OLA

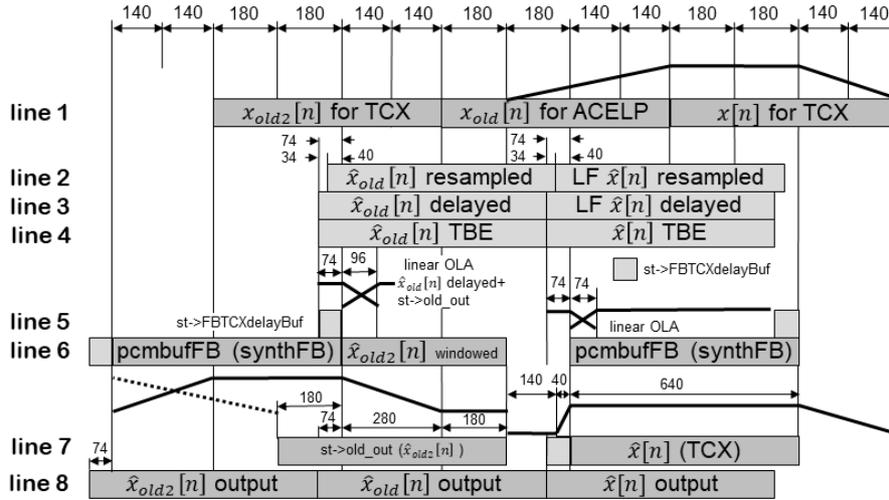


그림 2. 코덱이 TCX, ACELP, TCX 순서로 바뀔 때의 타이밍 다이어그램
 Fig. 2. Timing diagram when the codec is changed in the order TCX, ACELP, and TCX

로 인해 280샘플의 지연이 발생한다. 그리고 74샘플의 추가 지연을 통해 코덱의 출력 $\hat{x}_{old2}[n]$ output을 얻는다. 한편, TCX 출력의 윈도우된 마지막 부분은 7번째 줄의 st->old_out로 표기된 버퍼에 저장되어 다음 프레임 복원에 사용된다.

두 번째 프레임 $x_{old}[n]$ 은 ACELP와 TBE로 부호화된다. ACELP로 부/복호화된 신호는 CLDFB로 상향 표본화되면서 40샘플 추가 지연되어 두 번째 줄의 $\hat{x}_{old}[n]$ resampled 구간으로 표기된다. 이 신호는 다시 34샘플 추가 지연된 다음 TBE 출력과 합쳐져 복원된 $\hat{x}_{old}[n]$ 를 얻는다. 한편, TCX로 부호화될 때 TBE 동작에 필요한 중간 버퍼, 즉 그림 1의 old, state_syn, syn_overlap, hb_prev 등으로 표기된 버퍼의 값들은 모두 0으로 재설정된다. 따라서, TCX에서 TBE로 전이되면 TBE의 첫 번째 74샘플들은 재설정된 중간 버퍼 값들로 인해 불완전하게 복원된다. 이를 참조하여 첫 74샘플의 출력은 이전 프레임의 마지막 74샘플의 TCX 복원 신호로 대체된다. 다음 96샘플은 cross-fading 기법을 사용하여 윈도우된 TCX 복원 신호에서 ACELP와 TBE로 복원된 신호로 점진적인 전이가 이루어지게 한다.

세 번째 프레임 $x[n]$ 은 TCX로 부호화된다. 이전 프레임이 ACELP로 부호화되면 TCX의 좌측에 OLA 하여야 할 이전 프레임 결과가 없다. 따라서 복호기의 TCX 좌측에 윈

도우의 평편한 부분을 늘려 현재 프레임 결과만으로 충실히 복원된 신호를 얻을 수 있도록 한다. 즉, 그림 2의 오른쪽 6줄과 7줄 사이의 윈도우처럼, 평편한 부분이 좌측으로 확장되어 총 640샘플이 되도록 한다. 좌측 윈도우의 경사진 부분은 40샘플로 제한되고, 그 이전의 윈도우 값은 모두 0으로 처리한다. 우측 윈도우는 정상 윈도우와 동일하다. 이러한 방법을 통해 TCX로 복원된 프레임 pcmbuffFB는 OLA 없이 현재 프레임만으로 얻는다. 한편, 앞 부분의 복원된 신호를 얻기 위해서는 ACELP와 TBE로 복원된 신호 또한 필요하다. 이 신호는 TCX 복호기에서 pcmbuffFB와 동시에 출력되는 저 대역 복원 신호를 CLDFB로 상향 표본화하여 두 번째 줄의 오른쪽 LF \hat{x} resampled로 표기된 신호를 얻는다. 이 신호는 34샘플만큼 지연된다. 그런 다음, 처음 74 샘플은 ACELP와 TBE로 복원된 신호로 복원 신호를 얻고, 다음 74샘플은 cross-fading 기법으로 ACELP와 TBE로 복원된 신호에서 TCX로 복원된 pcmbuffFB로 점진적인 전이가 이루어지게 한다.

III. 수정된 전이 구간 처리

본 절에서는 II. 2절에서 설명한 시간 영역 코덱과 TCX

사이에 전이가 발생할 때의 문제점을 분석하고 그 해결책을 제시한다. 그림 1의 구조에서 TBE를 위한 대역 확장된 여기 신호 bwe_exc 는 이전 bwe_exc 로부터 피치만큼 이동하여 얻어진다. 그런데, EVS 코덱에서는 TCX로 부호화될 때 bwe_exc 는 0으로 재설정된다. 따라서 TCX 다음에 오는 몇 프레임의 TBE는 온전한 대역 확장된 여기 신호를 얻을 수 없다. 이러한 단점을 수정하기 위해 TCX로 부호화될 때에도 bwe_exc 를 생성하여 bwe_exc 가 끊김없이 유지되도록 한다. 한편, TCX는 ACELP와의 호환을 위해 ACELP와 동일한 표본화 주파수를 가지는 저 대역 여기 신호를 생성한다. 본 논문에서는 TCX에서 발생하는 저 대역 여기 신호를 시간 지연이 없는 이중 선형 보간으로 대역 확장된 여기 신호를 얻는다.

두 번째, 그림 1의 구조에서 고 대역 여기 신호 ϵ_1 는 수신된 LP 계수를 이용하여 LP 합성을 수행하여 복원된 고 대역 신호 $s\epsilon$ 을 얻는다. 한편 부호기 (encoder)에서는 LP 계수를 추정하여 전송하기 위해, 입력 신호를 CLDFB로 분석/합성하여 다운믹스된 고 대역 신호를 얻는다. LP 분석을 위해서는 540샘플의 다운믹스된 고 대역 신호가 필요하다. EVS 코덱에서는 TCX로 부호화될 때 이 다운믹스된 고 대역 신호를 생성하지 않으므로 TCX 다음의 TBE에서는 적절한 LP 분석이 이루어지지 않는다. 본 논문에서는 TCX로 부호화될 때도 다운믹스된 고 대역 신호를 얻도록 하여, TCX 다음에 오는 TBE에서도 적절한 LP 분석을 수행할 수 있도록 한다.

세 번째, TCX 다음에 오는 시간 영역 코덱에서 ACELP 뿐만 아니라 TBE도 내부 중간 버퍼들이 TCX 구간 동안 적절히 채워지지 않기 때문에, 첫 74샘플은 불완전한 복원 신호를 생성한다. 이 점을 고려하여 ACELP로 처리되는 저 대역 복원 신호에서 첫 74샘플은 이전 프레임의 TCX 복원 신호의 마지막 부분을 사용하고, 다음 96샘플은 cross-fading 기법으로 TCX 결과로부터 ACELP 결과로 점진적인 전이가 이루어지도록 한다. 반면에 TBE로 산출되는 고 대역 복원 신호는 EVS에서 아무런 처리 없이 단순히 합해진다. 본 논문에서는 TBE로 얻어지는 고 대역 복원 신호도 저 대역 복원 신호와 동일하게 처리한다. 즉, 첫 74샘플은 최종 결과에 합하지 않고, 다음 96샘플은 상승 윈도우가 곱해진 다음 최종 결과에 합하고, 나머지 샘플들은 최종 결과

에 그대로 합한다. 이 과정을 통해 TCX 다음에 오는 시간 영역 코덱 결과의 고 대역 부분도 적절히 복원할 수 있다.

네 번째, 시간 영역 코덱 다음에 오는 TCX에서는 앞부분의 복원 신호를 온전하게 처리하기 위해 시간 영역 코덱 결과가 필요하다. 저 대역 부분은 TCX에서 $pcmbuffB$ 와 함께 산출되는 저 대역 복원 신호를 CLDFB를 이용한 상향 표준화하고 34샘플 지연시켜 얻는다. 반면에 고대역 부분은 복원하지 않고 추후 작업을 진행한다. 이에 따라 복원 신호의 고 대역 부분에 스펙트럼 홀이 발생한다. 한편, EVS의 다른 부분에서는 미리 확장 기법, 즉 TBE에서 얻은 고 대역 복원 신호의 마지막 부분을 전치 (transposition) 시켜 고 대역 복원 신호로 사용한다. 본 논문에서도 이 방법을 이용하여 TBE 복원 값으로 사용한다.

본 절에서 제시한 네 가지 방법은 TCX와 시간 영역 코덱 사이의 전이에서 고 대역의 스펙트럼 홀을 감소시키는 데 기여할 것으로 기대한다.

IV. 모의 실험

본 절에서는 제III절에서 설명한 수정된 전이 구간 처리

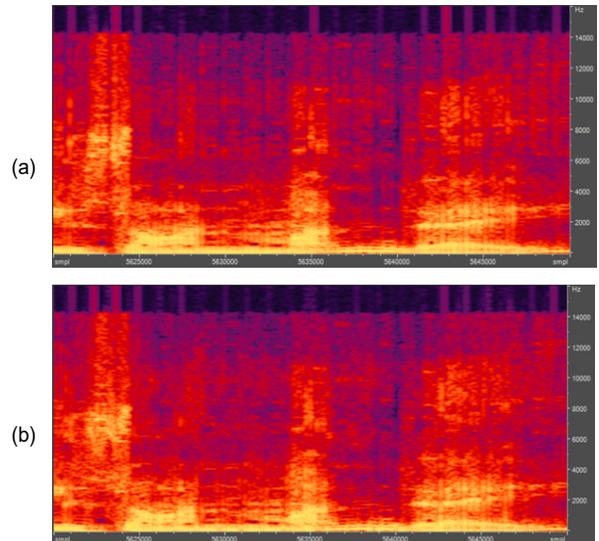


그림 3. (a) EVS, (b) 수정된 EVS로 복원된 lion 클립의 부분 스펙트로그램
Fig. 3. Spectrograms of a part of the lion clip reconstructed by (a) the EVS, (b) the revised EVS

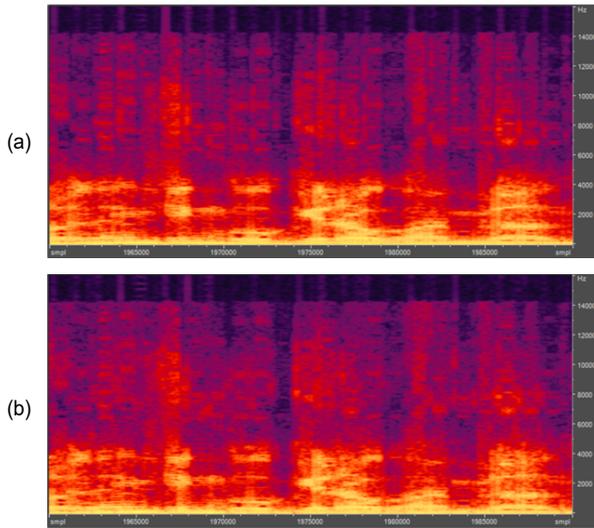


그림 4. (a) EVS, (b) 수정된 EVS로 복원된 wedding speech 클립의 부분 스펙트로그램

Fig. 4. Spectrograms of a part of the wedding speech clip reconstructed by (a) the EVS, (b) the revised EVS

를 채택하였을 때의 성능을 3GPP EVS의 성능과 비교한다. 실험에서 전송률은 13.2Kbps로 설정하고, 매 프레임마다 ACELP와 TBE로 구성된 시간 영역 코덱과 TCX가 전이되도록 하여 전이 구간의 특징이 주기적으로 나타나도록 하였다. 그림 3과 4에서는 EVS로 부호화된 결과와 수정된 EVS 결과를 비교하였다. 그림 3과 4에서는 각각 lion과 wedding speech라는 실험 샘플에 대한 복원 신호를 얻고 그 일부분의 스펙트로그램을 아도비 오디오션 3.0으로 도시한 것이다. 각 스펙트로그램은 가로 방향으로 30,000샘플 구간, 세로 방향으로 0부터 16KHz 주파수 구간을 도시하였다. 그림 3과 4를 살펴보면, EVS 복원 스펙트럼은 2 프레임 간격으로 고 대역에 수직 줄 형태의 스펙트럼 홀을 선명하게 볼 수 있다. 반면에 수정된 전이 구간 처리를 도입하였을 때 복원 신호의 스펙트럼은 고 대역의 스펙트럼 홀이

보이지 않음을 알 수 있다. 즉, 수정된 방법은 TCX와 시간 영역 코덱이 전이될 때 전이 구간의 성능을 크게 개선 시킬 수 있다.

V. 결론

본 논문에서는 EVS에서 시간 영역 코덱과 TCX 사이의 전이에서 고 대역에 발생하는 수직 막대 형태의 스펙트럼 홀을 감소시킬 수 있는 기법을 제안하였다. 제안한 기법은 비트열 측면에서는 EVS 표준과 동일하지만, 제III절에서 제시한 내용대로 복호기뿐만 아니라 부호기 동작도 수정해야 원하는 결과를 얻을 수 있다. 시뮬레이션 결과 수정된 기법은 EVS 표준에 비해 스펙트럼 홀을 크게 줄일 수 있음을 알 수 있었다. 본 논문에서 제안한 방법을 사용하면 코덱 간의 전이 중에도 견실한 성능을 보일 수 있어 차세대 오디오 코덱 개발에 기여할 것으로 기대한다.

참고 문헌 (Reference)

- [1] ISO/IEC 23008-3:201x(E), Information technology - High efficiency coding and media delivery in heterogeneous environments - Part 3: 3D audio (Phase 2), 2016.
- [2] Opus website. <http://opus-codec.org/>.
- [3] 3GPP TS 26.445 V16.1.1 3GPP Technical Specification Group Services and System Aspects; Codec for Enhanced Voice Services (EVS); Detailed Algorithmic Description, Oct. 2020.
- [4] 3GPP TS 26.443 V16.3.0 3GPP Technical Specification Group Services and System Aspects; Codec for Enhanced Voice Services (EVS); ANSI C code (floating-point), Dec. 2021.
- [5] 3GPP TS 26.290 V7.0.0 Extended Adaptive Multi-Rate - Wideband (AMR-WB+) codec; Transcoding functions, March 2007.
- [6] ISO/IEC FDIS 23003-3:2011(E), Information technology - MPEG audio technology - Part 3: Unified speech and audio coding, 2011.

저 자 소 개



김 인 철

- 1985년 : 서울대학교 제어계측공학과(공학사)
- 1987년 : 서울대학교 제어계측공학과(공학석사)
- 1992년 : 서울대학교 제어계측공학과(공학박사)
- 1992년 8월 ~ 1994년 3월 : 대우전자 영상 연구소
- 1994년 3월 ~ 1999년 8월 : 한성대학교 정보전산학부 조교수
- 1999년 9월 ~ 현재 : 서울시립대학교 전자전기컴퓨터공학부 교수
- ORCID : <https://orcid.org/0000-0002-7643-8717>
- 주관심분야 : 디지털 신호처리, 실시간 신호처리, 오디오 압축, 멀티미디어 통신 등