

파워 상호보완 윈도우를 이용한 지연 감소 IntMDCT

Low Delay IntMDCT Using Power Complementary Window

이상환, 이인성[†]

(Sang-hwan Lee and In-sung Lee[†])

충북대학교 전자통신공학과

(접수일자: 2013년 7월 29일; 채택일자: 2013년 9월 11일)

초 록: 본 논문은 파워 상호보완 윈도우를 이용한 지연 감소 알고리즘을 IntMDCT(Integer Modified Discrete Cosine Transform)에 적용하고자 제안한다. 기존에 사용하는 변환 방식인 MDCT는 정수 입력에 대해 실수 값을 출력하므로 데이터의 확장을 야기 시킨다. 이에 정수 대 정수 변환인 IntMDCT 기술이 등장했다. 하지만 IntMDCT는 MDCT(Modified Discrete Cosine Transform)처럼 알고리즘 지연을 갖는 문제점을 가지고 있다. 지연은 실시간 통신을 목표로 하는 환경에 있어 중요한 이슈로 자리 잡고 있다. 이러한 샘플의 지연을 최소화하기 위해 제안하는 알고리즘을 적용하고 성능을 평가한 결과, IntMDCT에서 발생하는 샘플 지연이 절반으로 감소하는 것을 확인 할 수 있었다.

핵심용어: 파워 상호보완 윈도우, 지연 감소 알고리즘, MDCT, IntMDCT

ABSTRACT: In this paper, we propose to apply low delay algorithm using power complementary window to Integer Modified Discrete Cosine Transform (IntMDCT). Conventional transform, the Modified Discrete Cosine Transform (MDCT) usually produces floating point values for integer input values. This causes the expansion of the data. To refine on this, IntMDCT that produces integer values even for integer input values have emerged. However, IntMDCT has a problem of the algorithm delay, such as MDCT. Delay has become a key issue in environments for the purpose of real-time communications. In order to reduce the delay, the proposed algorithm was applied and the results of the performance evaluation show that delay of IntMDCT has reduced by half existing delay.

Keywords: Power complementary window, Low delay algorithm, MDCT, IntMDCT

PACS numbers: 43.60.-c

1. 서 론

디지털 오디오는 좋은 음질과 유연성 때문에 아날로그 오디오를 대신하고 있다. 하지만 44.1 kHz 샘플링율, 16 bit/sample 사양의 CD 음질 디지털 오디오는 제한된 대역폭과 저장 공간 때문에 다루는데 한계가 있었다. 이러한 이유 때문에 오디오 압축 알고리즘에 대해 상당한 노력이 있어 왔다.^[1]

대부분의 오디오 압축 알고리즘은 MPEG-1 Layer III(MP3),^[2] MPEG-2/4 AAC^[3] 코덱과 같이 손실 압축을 기반으로 개발되어 왔다. 손실 압축은 사람이 들

을 수 없는 영역을 최소화하면서 사람이 들 수 있는 음질을 유지하는 압축 알고리즘이다.^[4] 손실 압축 알고리즘의 성능은 7~14:1 까지 개발되었다.

그러나 손실 압축은 본래의 음질을 손실하는 알고리즘이기 때문에 고성능의 응용프로그램에 적합한 알고리즘이 아니다. 이러한 이유 때문에 무손실 압축 알고리즘에 대한 기대가 증대되었다. 최근에는 광대역 네트워크와 저장 기술의 발전으로 96 kHz 샘플링율과 24 bit/sample의 사양, 고해상도, 무손실의 음질을 지원하는 응용 프로그램들이 증가하였다. 이처럼 환경의 발전 속에서도 여전히 많은 응용 프로그램들의 주요 관심사는 높은 압축률이다.^[1]

무손실 알고리즘을 사용하는 코덱으로는 MPEG에서 개발한 MPEG-4 Scalable to Lossless(SLS) 오디오

[†]Corresponding author: Insung Lee (inslee@chungbuk.ac.kr)
Department of Radio Communication Eng Chungbuk National University, Gaesindong, Heungdeokgu, Cheongju 361-763, Republic of Korea
(Tel: 82-43-261-3223, Fax: 82-43-271-4647)

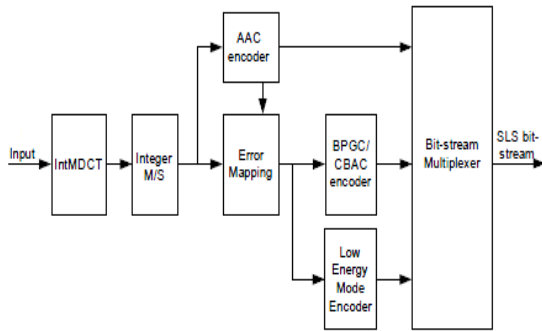


Fig. 1. Diagram of MPEG-4 audio SLS encoder.

코덱이 있다.^[1] Fig. 1의 SLS의 블록도 같이 SLS 코덱에서는 오디오 신호의 주파수 표현을 얻기 위해 필터뱅크 방식을 기반으로 한 Modified Discrete Cosine Transform(MDCT) 대신에 Integer Modified Discrete Cosine Transform(IntMDCT)를 사용한다. 손실 압축에서 사용하는 MDCT는 실수형의 계수를 생성하기 때문에 데이터 율을 증가시켜 고용량의 무손실 음원에 적합한 변환이 아닌 반면, IntMDCT는 정수형의 계수를 생성하여 데이터 율을 감소시키고, 완벽복원이 가능하여 무손실 음원에 적합한 알고리즘이다.^[5-8]

MDCT와 MDCT를 기반으로 한 IntMDCT를 사용하는 코덱에서는 데이터를 복원할 때 한 프레임 길이만큼의 샘플 지연이 발생한다. 알고리즘 지연은 커뮤니케이션 시스템에서 음질문제를 발생시키는 중요한 요인이다. 이러한 문제를 최소화하기 위해서 Constrained Energy Lapped Transform(CELT) 코덱에서는 MDCT에 샘플 지연을 절반으로 감소시키는 알고리즘을 적용하였다.^[9]

본 논문에서는 MDCT를 기반으로 한 IntMDCT에 파워 상호보완 윈도우를 이용한 지연 감소 알고리즘을 적용하여 무손실 환경에서도 샘플지연을 절반으로 감소하고자 제안한다.

본 논문의 2장에서는 MDCT, IntMDCT에 대한 이론과 샘플 지연을 감소하기 위한 방법에 대해 설명한다. 3장에서는 샘플 지연을 감소시킨 IntMDCT의 성능을 평가하고, 마지막으로 4장에서 결론을 맺는다.

II. 지연 감소 IntMDCT

본 논문에서는 IntMDCT에 샘플 지연을 감소시키

고자 CELT 코덱에서 사용한 알고리즘 적용을 제안한다.

먼저 IntMDCT에 지연 감소 알고리즘의 적용 가능한 점을 찾기 위해 MDCT와의 지연 특성의 차이점을 설명하고, 파워 상호보완 윈도우를 이용한 지연 감소 알고리즘에 대해 설명한다. 마지막으로 설명한 이론들을 이용하여 지연 감소 IntMDCT를 제안하고자 한다.

2.1 MDCT와 IntMDCT

N개의 데이터를 N개의 계수로 변환하고 중첩-합이 없는(Non-overlap-add) 구조인 Discrete Cosine Transform(DCT)의 연속되는 데이터의 양자화는 신호의 복원 과정에서 불연속점을 만든다.^[10] 이러한 불연속점은 음질을 손상시키는 요인이 된다. 이러한 단점을 개선하기 위하여 2N개의 데이터를 N개의 계수로 변환하는 중첩-합(Overlap-add) 구조인 MDCT가 등장하게 되었다.^[11]

MDCT 과정에서는 2N개의 데이터를 N개씩 중첩하면서 변환을 하게 되고 N개의 계수 값을 출력해 낸다. MDCT와 InverseMDCT(IMDCT)의 수식은 다음과 같다.

$$X_t(m) = \sqrt{\frac{2}{N}} \sum_{k=0}^{2N-1} w(k)x_t(k) \cos\left(\frac{\pi}{4N}(2k+1+N)(2m+1)\right), \quad (1)$$

$$m = 0, \dots, N-1$$

$$y_t(k) = w(k) \sqrt{\frac{2}{N}} \sum_{m=0}^{N-1} X_t(m) \cos\left(\frac{\pi}{4N}(2k+1+N)(2m+1)\right), \quad (2)$$

$$k = 0, \dots, 2N-1$$

식(1)과 식(2)의 $w(k)$ 는 사용하는 윈도우 함수를 의미하고 N은 한 프레임의 길이를 나타낸다. 사용하는 윈도우는 다음과 같은 특성을 보존해야 한다.

$$w(k)^2 + w(N+k)^2 = 1, \quad (3)$$

$$w(k) = w(2N-1-k). \quad (4)$$

MDCT / IMDCT 과정 후에 신호는 에일리어싱 (Aliasing)이 발생하게 되는데 이는 식(3)과 식(4)의 윈도우 특성을 이용하여 중첩-합을 하게 되면 에일리어싱이 제거되고 신호가 복원된다. 이 기술을 Time Domain Aliasing Cancellation(TDAC)라 부른다.^[11]

Fig. 2는 MDCT에서 중첩-합을 이용하여 신호를 복원하는 과정이다. 각 프레임은 MDCT / IMDCT 후에 에일리어싱이 발생한 신호이다. 프레임을 복원하기 위해 이전 프레임과 현재 프레임을 중첩-합 하게 되면 Fig. 2의 C, D의 데이터가 복원되고, 복원하기 위해 S_{EF} 샘플과 같이 N 만큼의 미래 샘플이 필요하므로 N의 샘플 지연이 발생한다.

데이터 율을 증가시키는 MDCT의 단점을 보완하기 위한 IntMDCT는 MDCT를 Widowing Time Domain Aliasing(WTDA)과 DCT-IV로 분해하여 각각 정수형의 데이터를 출력하여 데이터 율을 감소시키는 변형이다.^[8]

Fig. 3은 MDCT의 분해를 표현한 그림이다. WTDA

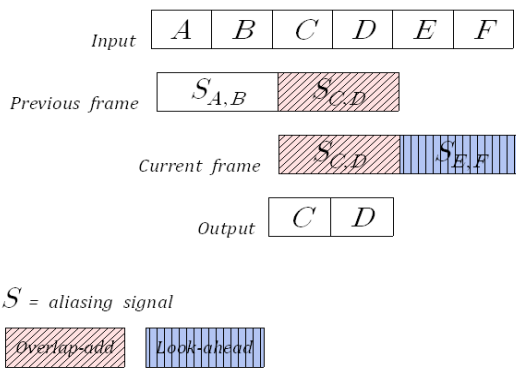


Fig. 2. Overlap-add process of N delay MDCT.

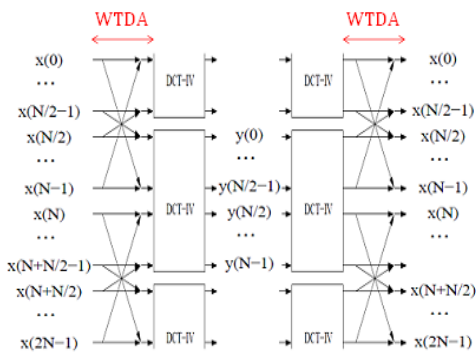


Fig. 3. Decomposition of MDCT and inverse MDCT into Givens rotations.

로 표시되어 있는 부분은 WTDA 신호를 발생하는 구간을 표현하고, DCT-IV 블록을 거친 $y(k)$ 의 데이터는 변환하여 출력된 계수 값을 표현한다. DCT-IV의 수식은 다음과 같다.

$$X_i(m) = \sqrt{\frac{2}{N}} \sum_{k=0}^{N-1} \tilde{x}_i(N-1-k) \cos\left(\frac{\pi}{4N}(2k+1)(2m+1)\right), \quad m = 0, \dots, N-1 \quad (5)$$

식(5)에서 $\tilde{x}_i(k)$ 는 WTDA 신호를 의미하고 N은 한 프레임의 길이를 나타낸다.

WTDA는 MDCT / IMDCT에서 발생하는 에일리어싱을 시간 축에서 먼저 발생시키는 과정이다. IntMDCT에서 WTDA 신호를 정수형으로 만들고 역변환 후에 완벽 복원 하기 위해 Givens rotations^[12]와 Lifting scheme^[13]을 사용한다. Givens rotations를 적용하기 전에 먼저 윈도우는 다음과 같이 계수로 표현된다.

$$a_k = \arctan \frac{w(k)}{w(N-1-k)}, \quad k = 0, \dots, \frac{N}{2} - 1. \quad (6)$$

식(6)을 이용하여 계수로 표현한 WTDA 신호를 만드는 행렬은 Givens rotations를 이용하여 다음과 같이 분해된 행렬로 표현된다.

$$\begin{pmatrix} \cos\alpha & -\sin\alpha \\ \sin\alpha & \cos\alpha \end{pmatrix} = \begin{pmatrix} 1 & \cos\alpha - 1 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ \sin\alpha & 1 \end{pmatrix} \begin{pmatrix} 1 & \cos\alpha - 1 \\ 0 & 1 \end{pmatrix}. \quad (7)$$

그리고 식(7)은 Fig. 4처럼 3단계의 Lifting step으로 표현된다. WTDA에 Lifting step을 적용하면 복원 할 때 오차가 발생하지 않고 완벽 복원이 가능하게 된다. DCT-IV에서도 Givens rotations와 Lifting step을 적용하여 정수형 계수를 출력하게 된다.^[8]

IntMDCT의 복원 과정에서도 MDCT와 같은 N의 샘플 지연이 발생한다. Fig. 5는 IntMDCT의 복원 과정이다. Fig. 5에서 이전 프레임과 현재 프레임으로

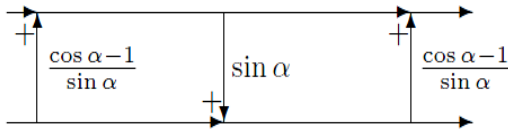


Fig. 4. Givens rotations by three lifting steps.

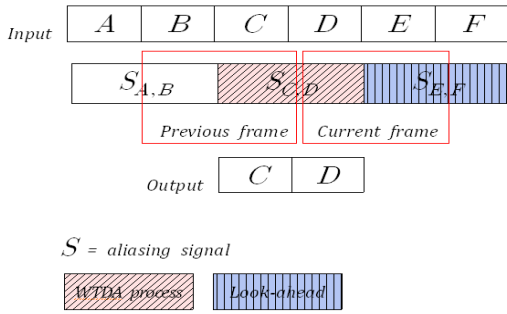


Fig. 5. WTDA process of N delay IntMDCT.

신호를 복원하기 위해 에일리어싱이 발생한 N/2 길이의 $S_{E,F}$ 샘플의 절반과, $S_{E,F}$ 샘플의 에일리어싱을 발생하기 위해 N/2 길이의 F 샘플이 관여 했으므로 전체 샘플지연 N이 발생한다.

2.2 CELT 코덱에 적용된 지연 감소 MDCT

CELT 코덱은 낮은 알고리즘 지연 특성을 갖는 음성 및 오디오 코덱으로, 음성 신호는 48 kbit/sec, 오디오 신호는 64 kbit/sec 에서 높은 성능을 내는 공개 소프트웨어 코덱이다. CELT에서는 샘플 지연을 N/2 으로 감소시킨 MDCT 알고리즘을 사용한다. CELT 의 샘플지연 감소를 위한 MDCT의 윈도우는 Vorbis 코덱^[14]에 사용하는 파워 상호보완 윈도우를 사용하며, 다음과 같은 수식으로 정의한다.^[9]

$$w(n) = \sin\left(\frac{\pi}{2} \sin^2\left(\frac{\pi(n+0.5)}{2L}\right)\right). \quad (8)$$

식(8)의 n은 프레임 값을 나타내고, L은 중첩함의 구간 길이이다. 이 윈도우를 이용하여 샘플 지연을 감소시키기 위해 윈도우의 양쪽 n/4 만큼은 0으로 설정되고, 가운데 n/2 만큼은 1로 설정 된다.

Fig 6은 지연 감소 알고리즘을 적용시켜 윈도우가 변형된 그림이다. 이 처럼 윈도우가 변형되면 중첩

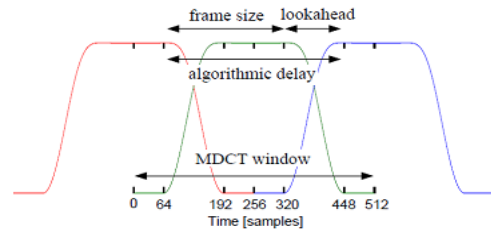


Fig. 6. Power complementary window of CELT.

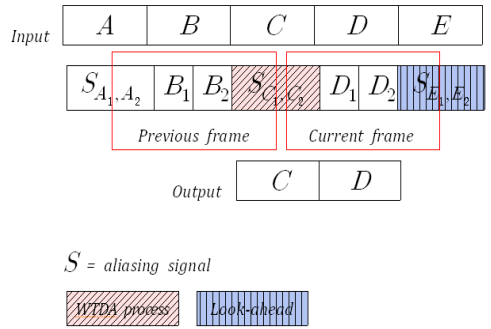


Fig. 7. WTDA process of N/2 delay IntMDCT.

합의 구간이 감소하게 되면서 샘플의 지연이 감소하게 된다.

2.3 지연 감소 IntMDCT

본 논문에서는 앞에서 설명한 IntMDCT의 지연 특성과 파워 상호보완 윈도우를 이용한 지연 감소 알고리즘을 적용한 지연 감소 IntMDCT를 제안한다.

먼저 파워 상호보완 윈도우를 이용한 지연 감소 알고리즘을 IntMDCT의 WTDA 과정에 적용하여 발생하는 에일리어싱 구간을 절반으로 감소하게 하였다. Fig. 7은 파워 상호보완 윈도우를 사용한 지연 감소 알고리즘을 적용한 후 예상되는 WTDA 신호와 복원 과정이다. 파워 상호보완 윈도우의 특성과 식(6), 식(7)이 적용된 Fig. 7의 WTDA 과정에서는 에일리어싱이 감소된 신호를 얻을 수 있을 것이다.

이처럼 에일리어싱 구간을 감소 시켜, 신호를 복원하기 위해 필요한 미래 샘플을 감소시키고 샘플 지연을 절반으로 감소하게 하였다. 감소된 에일리어싱의 신호를 복원하기 위해 이전 프레임과 현재 프레임을 이용하여 다시 WTDA 과정을 거치면서 Givens rotations와 Lifting step을 통해 완벽복원을 할 수 있게 하였다.

III. 성능 평가

본 논문에서 제안하는 지연 감소 알고리즘을 적용한 IntMDCT의 성능을 평가하기 위해 실험에 Mono 채널이며 16 bits/sample, 16 kHz로 샘플링 된 입력 신호를 사용하고, 200샘플을 한 프레임으로 설정 하였다. 실험은 파워 상호보완 윈도우를 적용한 IntMDCT의 과정과 결과를 확인하고, 기존 MDCT의 결과와 비교하는 것으로 진행하였다.

먼저 IntMDCT의 첫 번째 과정인 WTDA 신호를 발생시키기 위해 Fig. 8(a)의 실선과 같은 입력 신호를 사용하였다. 긴 점선으로 되어있는 부분은 파워 상호보완 윈도우이다. 파워 상호보완 윈도우와 식(6)과 식(7)을 이용하여 Fig. 8(a)의 0~199, 200~399 샘플마다 각각 Givens rotations과 Lifting step을 이용하여

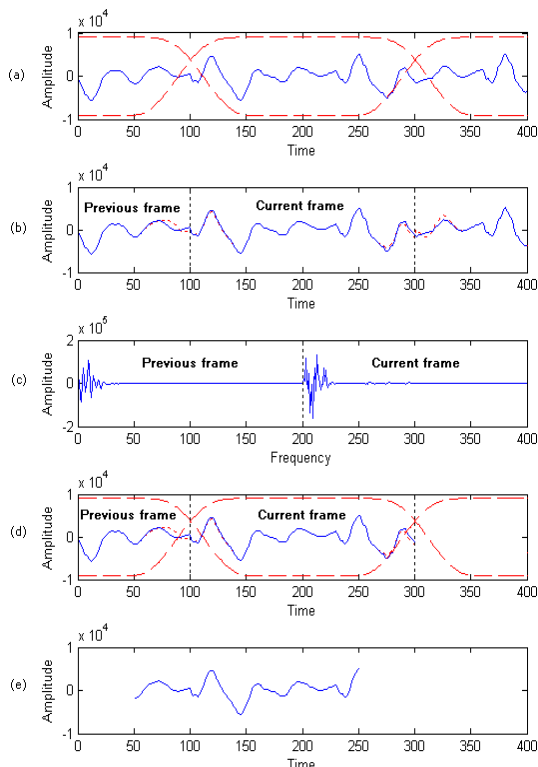


Fig. 8. Illustration of the low delay IntMDCT(N/2 delay). (a) The input signal and dashed lines power complementary window, (b) WTDA signal, (c) Coefficients of IntMDCT, (d) Coefficients of IntMDCT and dashed lines power complementary window, (e) Reconstructed signal.

정수형 WTDA 신호를 발생하는 과정을 진행 하였다.

Fig. 8(b)는 WTDA가 발생한 신호이다. 에일리어싱이 발생한 부분을 구분하기 위해 실선의 입력신호에 점선으로 에일리어싱을 표시하였다. 그리고 이전 프레임과 현재 프레임으로 각각 Givens rotations과 Lifting step을 이용하여 정수형으로 DCT-IV 하였다. 복원할 때 완벽하게 신호를 복원하기 위하여 이전 프레임에서 프레임 길이 보다 부족한 부분은 0의 값을 넣어서 변환 하였다.

Fig. 8(c)는 WTDA 신호를 DCT-IV한 결과이다. 이전 프레임을 이용하여 0~199 샘플의 변환 계수를 얻었고, 현재 프레임을 이용하여 200~399 샘플의 변환 계수를 얻었다.

Fig. 8(d)는 DCT-IV 계수를 역변환 한 계수이다. 신호를 확인한 결과 (b)의 WTDA신호의 이전 프레임과 현재 프레임 값이 동일한 것을 확인 하였다.

Fig. 8(e)는 Fig. 8(d)를 이용하여 프레임을 복원 한 것이다. 이 과정은 Fig. 7과 유사하다. (d)의 이전 프레임인 0~99까지의 샘플은 Fig. 7의 WTDA 신호에서 이전 프레임의 B_2 와 에일리어싱이 발생한 S_{C_1, C_2} 의 절반을 의미하고, (d)의 100~399까지의 샘플은 Fig. 7의 WTDA 신호에서 현재 프레임 값을 의미한다. 이 신호를 이용하여 WTDA 과정을 다시 거치게 되면 Fig. 8(a)의 입력신호에서 50~250까지의 샘플이 Fig. 8(e)처럼 완벽하게 복원 된다.

다음으로 파워 상호보완 윈도우를 적용한 기존의 MDCT의 결과와 IntMDCT의 결과를 비교 하였다.

먼저 Fig. 9(a)는 실수형의 MDCT 스펙트럼이고 Fig. 9(b)는 정수형의 IntMDCT의 스펙트럼이다. 확인 결과, 데이터의 형식은 다르지만 스펙트럼의 모양이 같음을 알 수 있었다. Fig. 10(a)는 MDCT의 복원 결과이고 Fig. 10(b)는 IntMDCT의 복원 결과이다. 두 신호를 비교 하였을 때 오차가 없이 결과가 같음을 확인 하였다.

실험 결과들로 IntMDCT에 파워 상호보완 윈도우를 적용하여 스펙트럼의 값은 정수형을 생성하고, 신호가 완벽하게 복원이 되는 것을 확인 하였다.

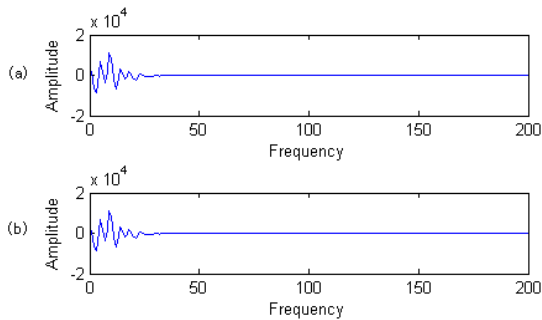


Fig. 9. Comparison between MDCT and IntMDCT spectrum using power complementary window. (a) Spectrum of MDCT, (b) Spectrum of IntMDCT.

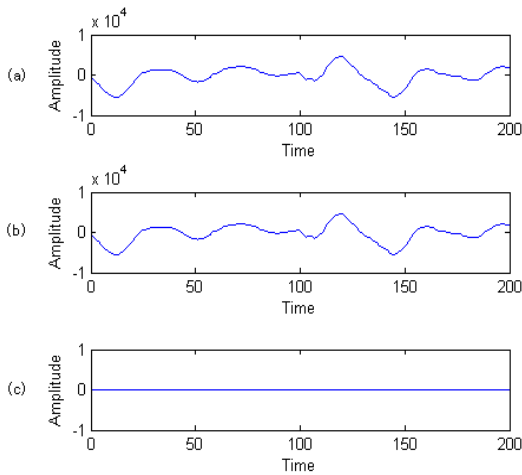


Fig. 10. Comparison between MDCT and IntMDCT reconstruction using power complementary window. (a) Reconstruction of MDCT, (b) Reconstruction of IntMDCT, (c) Error signal.

IV. 결 론

기존에 사용하는 MDCT는 변환 후에 데이터 확장 현상이 발생한다. 이러한 현상은 데이터의 용량이 큰 무손실 환경에서 문제점이 된다. 이 점을 보완하기 위해 정수형의 입력에 대해 정수형의 출력이 발생하는 IntMDCT를 사용한다. IntMDCT 과정에는 MDCT와 동일한 지연이 발생한다. 지연은 실시간 통신환경에서 품질을 결정하는 중요한 요소이다.

본 논문에서는 파워 상호보완 윈도우를 사용한 지연 감소 알고리즘을 IntMDCT에 적용하고 정수형의 스펙트럼을 생성하면서 샘플 지연을 절반으로 감소시키는 효과를 얻을 수 있었다.

감사의 글

본 논문은 2013년도 정부(교육과학기술부)의 재원으로 한국연구재단의 지원을 받아 수행된 기초연구사업임(No. 2011-0013830).

References

1. R. Geiger, T. Sporer, J. Koller, K. Brandenburg, "MPEG-4 Scalable to Lossless Audio Coding," 117th AES Convention, paper no. 6183 (2004).
2. ISO/IEC JTC1/SC29/WG11, "Information technology : coding of moving pictures and associated audio for digital storage media at up to about 1.5 Mbit/s - part3: audio," IS, paper no.11172 (1992).
3. ISO/IEC JTC1/SC29/WG11, "Coding of audiovisual objects, part 3. audio, subpart 4 time/frequency coding," IS, paper no.14496 (1999).
4. T. Painter and A. Spanias, "Perceptual coding of digital audio", P. IEEE, **88**, 451-513 (2000).
5. J. Koller, T. Sporer, K. Brandenburg, "Robust Coding of High Quality Audio Signals," 103rd AES Convention, paper no. 4621 (1997).
6. J. Koller, T. Sporer, K. Brandenburg, "Improving Lossless Audio Coding," AES 17th International Conference, paper no. 17 (1999).
7. M. Purat, T. Liebchen, P. Noll, "Lossless Transform Coding of Audio Signals," 102nd, AES Convention, paper no. 4414 (1997).
8. R. Geiger, T. Sporer, J. Koller, K. Brandenburg, "Audio coding based on Integer Transforms," 111th AES Convention, paper no. 5471 (2001).
9. J.-M. Valin, T. B. Terriberry, C. Montgomery, G. Maxwell, "A high-quality speech and audio codec with less than 10 ms delay," IEEE Trans, Audio Speech Lang Processing, **18**, 58-67 (2010).
10. K. R. Rao and P. Yip, *Discrete Cosine Transform: Algorithms, Advantages, Applications* (Academic Press, San Diego, 1990), pp.7-25.
11. J. P. Princen, A.B Bradley, "Analysis/Synthesis filter Bank Design Based on Time Domain Aliasing Cancellation," IEEE Trans. On ASSP, **34**, 1153-1161 (1986).
12. I. Daubechies, W. Sweldens, "Factoring wavelet transforms into lifting steps," J Fourier Anal Appl. **4**, 247-269, (1998).
13. P. P. Vaidyanathan, *Multirate Systems and Filter Banks* (DK, New Jersey, 1993), pp.290-295.
14. *Vorbis I specification*, http://www.xiph.org/vorbis/doc/Vorbis_I_spec.html (2004).

저자 약력

▶ 이 상 환(Sanghwan Lee)



2013년: 충북대학교 정보통신과(학사)
2013년 ~ 현재: 충북대학교 전파통신 공
학과(석사)
<관심 분야> 음성/ 오디오 신호처리

▶ 이 인 성 (Insung Lee)



1983년: 연세대학교 전자공학과(학사)
1985년: 연세대학교 전자공학과(석사)
1992년: Texas A&M University 전기 공학과
(박사)
1993년 ~ 1995년: 한국전자 통신연구원
이동통신 기술 연구단 선임연구원
1995년 ~ 현재: 충북대학교 전파통신공
학과 정교수
<관심 분야> 음성/ 영상 신호압축, 이동
통신, 적응필터