

다이나믹 윈도우 스위칭기법을 적용한 AC-3 오디오 필터뱅크의 성능향상에 관한 연구

김준성*, 강현철**, 장병건***, 변윤식*

시립 인천대학교 전자공학과*, 시립 인천대학교 정보통신공학과**

시립 인천대학교 전기공학과***

A Study on the Dynamic Window Switching MDCT for Enhanced AC-3 Audio Filterbank

Junsung Kim*, Hyunchul Kang**, Byong-Kun Chang***, Youn-Shik Byun*

Dept. of Electronics Eng., Univ. of Incheon*, Dept. of Info. & Telecom. Eng., Univ of Incheon**.

Dept. of Electrical Eng., Univ. of Incheon

E-mail : byun@lion.inchon.ac.kr

Abstract

This paper presents a technique to enhance TDAC in the AC-3 algorithm. To reduce block boundary noise without decreasing the performance of transform coding, new special window adopted. They improves the defect of the AC-3 algorithm that could not properly cancel aliasing in the transient period. In addition, a fast MDCT calculation algorithm based on a fast Fourier Transform, is adopted.

I. 서론

주파수 해상도가 높아질 수록 밴드폭이 좁아지고 밴드 수가 많아지므로, 많은 대역분할은 밴드들 중에서 신호가 들어있는 것만을 코딩할 수 있어 코딩 이득을 높이는 장점이 있는 반면 시간축 해상도가 낮아지므로, 프레임 내에서 천이 발생하는 신호의 경우 별도의 특별한 처리가 필요한 단점이 있다. AC-3 디지털 오디오 부호화 알고리즘은 MDCT에 기반을 둔 변환 부호화 방식으로, 93.75Hz 정도의 높은 주파수 해상도를 유지하며, 천이구간에서는 블록사이즈 변환으로 시간축 해상도를 높이는 방법을 사용한다. 이러한 방법은 신호의 갑작스런 변화에 의해 발생하는 pre-echo를 최소화시키거나, 천이구간에서의 TDAC를 안정적으로 처리하기 못하는 구조를 가지고 있다[1]. 본 연구에서는 이러한 AC-3 알고리즘의 불완전한 부분을 보완하며, fast algorithm으로 분석과 합성 과정을 설계하고 성능비교를 한다.

II. AC-3 알고리즘의 TDAC 변환

2.1 Time Domain Aliasing Cancellation 변환

일반적으로 신호 분석시 중첩, 가산을 행하는 시스템에서는 인접한 시간축 윈도우간의 중첩에 의해서 aliasing 이 일어나게 되는데, TDAC 방법을 사용하면 분석단에서 발생한 aliasing을 효율적으로 제거함으로써 신호를 완전 복원해 낼 수 있다[2][3][4]. TDAC subband의 복원 과정은 식(1)과 같다.

$$\hat{x}_{m0}(r) = x_{m0} \{ f(r+M)h(r+M) + f(r)h(r) \} + x_{m0(M-1-r)} \{ f(r)h(r+M) - f(r+M)h(r) \} \quad \text{식(1)}$$

식(1)의 첫 번째 항은 입력신호성분을 나타내고 두 번째 항은 time alias의해 발생된 성분을 나타낸다. 분석과 합성 윈도우가 같고 대칭적일 때 완전복원 조건은 식(2)와 같으며, aliasing cancellation 조건은 식(3)과 같다.

$$f(r)^2 + f(r+M)^2 = 1 \quad \text{식(2)}$$

$$f(r)f(r+M) - f(r+M)f(r) = 0 \quad \text{식(3)}$$

2.2 Modified Discrete Cosine Transform(MDCT)

MDCT는 linear orthogonal lapped transform으로 TDAC에 기초한 변환방식으로 50% 중첩한 데이터의 변환후 샘플의 수가 변환하기 전의 샘플수와 같은 임계

표본화율 보장하며, fast algorithm으로의 구현이 가능하다[5]. 식(4),식(5)에 기본 변환과 역변환식이 나타나 있다.

$$X(m) = \sum_{k=0}^{n-1} f(k)x(k) \quad \text{식(4)}$$

$$\cos\left(\frac{\pi}{2n}\left(2k+1+\frac{n}{2}\right)(2m+1)\right)$$

$$\text{for } m = 0 \dots \frac{n}{2} - 1$$

$$y(p) = f(p) \frac{1}{n} \sum_{m=0}^{\frac{n}{2}-1} X(m) \quad \text{식(5)}$$

$$\cos\left(\frac{\pi}{2n}\left(2p+1+\frac{n}{2}\right)(2m+1)\right)$$

$$\text{for } p = 0 \dots n-1$$

where $f(x)$ is a window $f(x) = \sin\left(\pi \frac{x}{N}\right)$

2.3 고속 MDCT의 구현

MDCT는 FFT로 구현이 가능하며, 단순한 구현방법으로는 길이 N 의 변환을 위해 odd 변환을 하는 FFT는 $2N$ length가 요구되어진다. 그러나 MDCT는 odd-time, odd-frequency Discrete Fourier Transform ($O^2 DFT$)에 의해 fast 구현이 가능하며 식(6),식(7)에 나타내어진다[10].

$$X(m) = R(O^2 DFT_{\text{odd}(k)}(m)) \quad \text{식(6)}$$

$$= R\left(\sum_{k=0}^{N-1} f\left(k-\frac{n}{4}\right)x\left(k-\frac{n}{4}\right)e^{-\frac{j\pi}{2n}(2k+1)(2m+1)}\right)$$

고속알고리즘의 표현식은 다음과 같다.

$$W = O^2 DFT(\text{odd}(f(k-\frac{n}{4})x(k-\frac{n}{4}))) \quad \text{식(7)}$$

$$= X(m)$$

$$W_{2k} = R(P_k)$$

$$W_{\frac{n}{2}-2k} = T(P_k)$$

$$W_{2k+1} = -W_{n-2(k+1)}$$

where

$$P_k = 2 e^{-j\frac{2\pi}{n}(k+\frac{1}{8})}$$

$$\sum_{r=0}^{N/4-1} ((x(2r) - jx(\frac{n}{2}+2r)) e^{-j\frac{2\pi}{n}(r+\frac{1}{8})}) e^{-j\frac{2\pi}{n/4}rk}$$

윗식의 시그마 이후 부분은 바로 $N/4$ point FFT를 나타낸다. 그러므로 MDCT는 적절한 pre- 그리고 post-rotation과 $n/4$ point FFT로 구현되며, IMDCT 역

시 비슷한 방법으로 구현된다.

2.4 AC-3의 MDCT와 블록-스위칭

변환 부호화 방식을 사용하는 오디오 코딩에서 분석 블록의 크기 결정은 가장 기본이며, 중요한 문제이다. AC-3 알고리즘은 주파수와 시간축 해상도를 모두 고려하여 trade-off시키는 방법을 사용하여 두 개의(512,256) 블록의 스위칭 기법을 사용한다. 기본적으로 512-point 블록으로 하며, 구간 내에서 천이가 발생하는 경우 발생하는 pre-echo 성분을 제거하기 위해 256-point 2개를 사용하여, 각각을 변환한 후 인터리빙(interleaving)하여, 512-point MDCT한 블록과 같은 $N(256)$ 을 유지시키지만, 시간축 해상도를 높이는 방법을 사용한다.

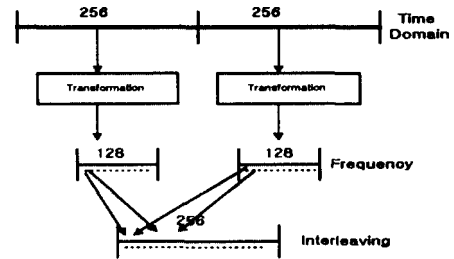


Fig 1. Shorter Block Transformation in AC-3 audio

III. 가변 Block-Size Windowing

3.1 적용 Block-Size MDCT를 위한 Window

AC-3 알고리즘은 512 samples MDCT(Normal blocks)의 결과로 256개의 주파수 도메인 변환 계수를 나타내며, 천이구간에서 512 samples 윈도우가 쪼개진 오디오 블록을 2개의 256 블록으로 나눈 후 각각의 절반 블록을 변환하여, 각각 $0 \sim f_c/2$ 의 주파수를 나타낸 후 계수들간의 인터리빙(interleave)을 수행하여 하나의 블록(single-block)을 만든다. 그 결과 큰 블록을 512 samples MDCT를 수행한 것과 같은 샘플수를 만들면서, 시간축 해상도를 높여 pre-echo를 제거시킨다. 그러나 AC-3 알고리즘의 변환 이전 시간축 계수들은 중첩, 가산을 행한 샘플들이므로, 위와 같은 방법으로 pre-echo는 제거되나, 천이구간에서 aliasing과 block-effect가 발생된다. 따라서 TDAC를 위해 큰 블록뿐만 아니라 작은 블록 그리고 블록 스위칭이 일어나는 부분에 특별한 윈도우가 요구되어지며, 이러한 부분은 AC-3 알고리즘의 불완전한 부분으로 사료된다.

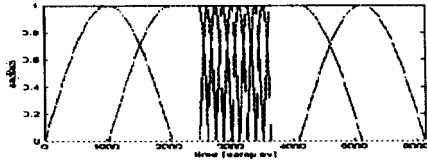


Fig2. Window Transitions for Input Signals Requiring High Temporal Resolution

본 연구에서는 sine 윈도우를 사용하였으며, 윈도우 $f(x)$ 는 50%의 중첩을 한 이웃 윈도우와 TDAC가 되도록 long, long-short, short-long, short의 4가지 윈도우를 사용하였으며, 수식(2),(3)의 조건을 만족한다. 그림4는 구현된 4가지의 윈도우를 보여준다.

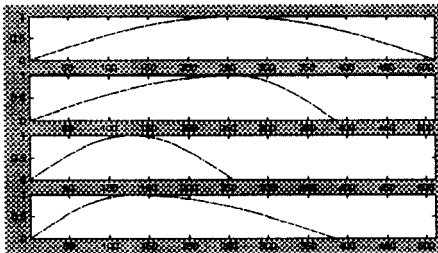


Fig 4. Window for adaptive block-size MDCT

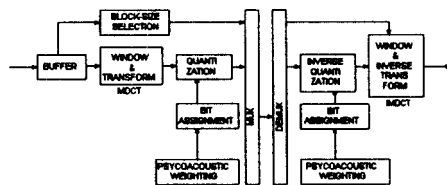


Fig5 Adaptive MDCT audio system

3.2 적응 Block-Size 선택 오디오 시스템

3.2.1 적응 블록사이즈 선택 알고리즘

AC-3 알고리즘들은 높은 주파수 해상도를 유지하여, 높은 코딩 이득을 얻으며, 시간축 권이구간에서 시간축 해상도를 높여야 하므로, 스위칭 알고리즘은 주파수 보단, 시간축의 정보를 더 잘 나타내야 한다. 영교차율, 차분, 에너지 엔트로피의 세가지 스위칭 기준을 가지고 실험하였으며, 세가지 스위칭 기준의 실험 결과가 그림 5에 나타나 있다.

3.2.2 양자화기 설계

사람의 청각특성을 반영한 임계 대역과 유사한 24 대역으로 분할 하였으며, 각각의 주파수 도메인 변환 계수들은 표3과 같이 묶었다. 심리음향 모델은 MPEG의 psychoacoustic Model-2를 적용적으로 할당하여 사용하였다.

표3. Coefficient Clustering-Frequency Domain

Band	First bin - Last bin	Band	First bin - Last bin
1	0 - 1	13	25 - 28
2	2 - 3	14	29 - 33
3	4 - 5	15	34 - 38
4	5 - 6	16	39 - 45
5	6 - 7	17	46 - 53
6	8 - 9	18	54 - 64
7	10 -11	19	65 - 77
8	12 -13	20	78 - 91
9	14 -15	21	109 - 128
10	16 -18	22	129 - 149
11	19 -21	23	150 - 170
12	22 -24	24	171 - 255

IV. 모의실험 결과

본 연구에서는 TDAC 변환을 효율적으로 하는 MDCT와 fast-MDCT의 error energy와 연산량을 비교하였으며, AC-3의 TDAC 변환에 제안된 윈도우를 적용하고, 3가지의 스위칭 기준에 의한 성능 평가하였다.

4.1 fast algorithm error energy

그림1은 cosine 함수를 overlap-add 한 후 512 point MDCT/IMDCT한 재생신호와 원신호간의 error energy를 나타낸다. 맨 위 그림은 입력신호를 두 번째는 변환된 신호를 세 번째 그림은 direct-MDCT에 의한 재생신호, 네 번째 그림은 fast-MDCT에 의한 재생신호와 원신호간의 차이를 나타낸다.

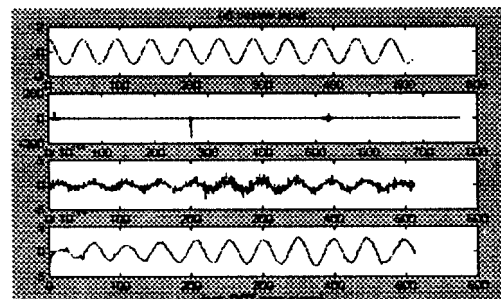


Fig6 Comparison of error signal between fast and direct transform case of cosine input

위 그림에 나타나듯이 fast algorithm의 성능은 direct form에 거의 차이가 없는 성능을 나타냄을 알 수 있다.

4.2 fast algorithm의 연산량 비교

AC-3 알고리즘 신호의 분석과 합성구간이 전체 연산량의 절반정도를 차지하며, 그 분석과 합성구간은 MDCT와 IMDCT가 대부분을 차지한다. 특히 AC-3

HDTV와 같은 방송용으로 사용되기 때문에 알고리즘의 실시간 구현을 위해서는 그 복잡도가 최소가 되어야 한다. 길이 N인 신호에 대해 direct MDCT/IMDCT와 fast MDCT/MDCT의 수식적인 연산량과 실험결과를 표-1에 나타내었으며 N은 512point 일 때이다.

표1. fast MDCT/IMDCT 연산량 비교

입력신호	Direct		Fast	
	MDCT	IMDCT	MDCT	IMDCT
cos	1.6000	3.0400	0.0600	0.0600
pop	1.4900	3.0400	0.0600	0.0800
ramp	1.5500	3.0400	0.0300	0.0400

위 표의 실험 결과는 다음의 수식에서 예상되는 결과와 거의 일치함을 보여준다. direct-MDCT /IMDCT는 각각 $(N^2)*1.5$, $(N^2) *3$ 의 real multiply-adds 그리고 fast-MDCT /IMDCT는 각각 $N*\log_2(N)+3.5*N$, $N*\log_2(N)+5 *N$ real multiply-adds가 소요되며, 위 실험 결과는 pentium-200 IBM PC에서 각각의 소요 CPU time을 측정 한 것이다.

4.3 각 스위칭 기준에 대한 성능 평가

44.1kHz로 표본화된 drum 신호에 대한 성능평가 그림 5에 나타나 있다. 본 실험결과 시변특성을 가장 잘 나타내는 영교차출법이 가장 좋은 성능을 보였으며, 이는 4-band QMF 필터의 가장 높은 주파수대역의 성분율 가지고, 120이상의 ZCR이 나타나는 구간을 천이 구간으로 인정하였다[9].

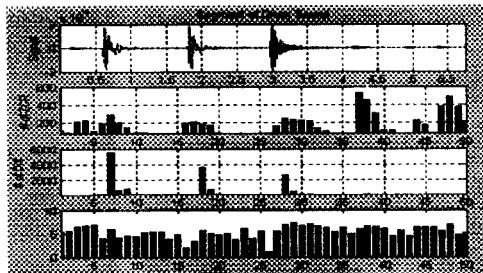


Fig5 Result of Switching Criterion for drum

4.4 음질평가

단일채널에 대해 512샘플마다 snr을 계산하였으며, 표 2에 나타나 있다.

V. 결 론

본 연구에서는 모든 디지털 오디오 압축 알고리즘의 목표인 저복잡도, 고품질 오디오 코덱의 구현에 초점을 맞추어, AC-3 오디오 알고리즘의 TDAC 변환, 역변환 부를 FFT를 사용한 고속 알고리즘을 적용한 경우, direct-MDCT/IMDCT를 한 경우에 비해 MDCT는

표2. 객관적인 음질평가 (averaged SNR)

음원	Polyphase	MDCT(B.S)	MDCT(W.S)
Drum	26.61	23.1	23.2
Piano	33.28	31.5	31.7
Violin	25.65	23.0	23.5
Popsong	23.37	30.5	30.7
평균	27.2275	27.025	27.275

N=512 일때 평균 1/36 정도의, IMDCT의 경우는 1/50 정도의 처리시간이 감소하였으며, 원신호와 복원신호간의 error energy의 경우는 거의 차이를 나타내지 않은 결과를 보였다. 본 연구에선 모의 실험을 위해 mono 신호의 경우 128kbps에서 CD quality를 유지하는 오디오 코덱을 설계 하였으며, 시간축 천이구간에서 TDAC를 고려하는 WINDOW의 도입으로 aliasing에 의해 일어나는 음질저하를 막을 수 있었으나, 주관적인 큰 음질의 차이는 느끼지 못하였다. 이는 심리음향 모델링과 양자화기의 최적화로 보다 저 전송율 오디오 코덱의 설계를 했을 때 보다 좋은 성능향상이 나타날 것으로 기대되어진다.

참 고 문 헌

- [1].J.C McKinney, R.Hopkins " Digital Audio Compression Standard(A/52). ATSC 1995
- [2].A.W.Johnson, "Frequency Domain Coding of Speech Waveforms at Low Bit Rates;Masters Thesis, RMIT, Melbourne, 1988.
- [3].J.P Princen and A. Bradley," Anlysis /Synthesis Systems in the Presence of Quantization". Proc. Int. Conf. on Acoust., Speech and Sig. Proc., pp1314-4, vol. 2, 1989
- [4].G. Smart and A. Bradley, "Efficient Filter Bank Design Based on Time Domain Aliasing Cancellation", submitted to IEEE Transactions on Signal Processing.
- [5].Shina.D. and Tewfik A.H., "Low Bit Rate Transparent Audio Compression Using Adaptive Wavelets", IEEE Trans. on Sign. Proc., Vol. 41, No. 12 Dec. 1993.
- [6].N.S Javantetal., Digital Coding of Waveforms. EnglewoodCliffs, Prentice Hall 1984.
- [9].신유철, 김준성, 상현철, 장병건, 변윤식, 웨이브릿 변환과 MDCT의 스위칭 기법을 이용한 오디오 부호화, 신호처리 학술대회 1997.
- [10].K. Sayood, Introduction to Data compression, Morgan Kaufmann Publishers, 1996.

본 연구의 일부는 한국 과학 재단 지정 인천대학교 멀티미디어 연구센터의 지원에 의한 것입니다.