

論文2003-40SP-6-9

TMS320C5416을 이용한 G.729A 보코더와 계산량 감소된 SOLA-B 알고리즘을 통합한 가변 전송율 보코더의 실시간 구현

(Real-time Implementation of Variable Transmission Bit Rate Vocoder Integrating G.729A Vocoder and Reduction of the Computational Amount SOLA-B Algorithm Using the TMS320C5416)

威明圭*, 裴明振*

(Myung-Kyu Ham, Myung-Jin Bae)

요약

본 논문에서는 8kbps의 전송율을 가진 ITU-T G.729A 보코더에 Henja가 제안한 SOLA-B (Synchronized Overlap Add) 알고리즘을 적용하여 가변 전송율의 보코더를 TMS320C5416에 실시간 구현하였다. 이 방법은 부호화 시 SOLA-B 알고리즘을 이용하여 음성의 속도를 빠르게 해주고, 복호화 시 다시 SOLA-B 알고리즘을 이용하여 음성의 속도를 느리게 해줌으로써 정상속도의 음성을 재생시켜준다. 이때 SOLA-B 알고리즘의 계산량을 줄이기 위해 상호 상관 함수가 수행되는 샘플의 간격을 3 샘플씩 건너뛰면서 처리하였다. 실시간 구현된 G.729A와 SOLA-B 알고리즘의 보코더는 8kbps 전송율일 때 인코더는 10.2MIPS이고 디코더에서는 2.8MIPS의 최대 복잡도를 나타내었다. 그리고, 6kbps 전송율일 때 인코더 18.5MIPS이고 디코더는 13.1MIPS의 최대 복잡도를 나타내었으며, 4kbps 전송율일 때 인코더 18.5MIPS 이고 디코더에서 13.1MIPS의 최대 복잡도를 나타내었다. 사용된 메모리는 program ROM 9.7kwords, table ROM 4.5kwords, RAM 5.1kwords 정도이다. 출력된 파형은 C simulator와 Bit Exact 한 출력 결과를 보여주었다. 또한, 실시간 구현된 가변 전송율 보코더의 음성 평가를 위해 MOS 테스트를 수행한 결과 4kbp의 전송율에서 MOS값이 3.69정도로 측정되었다.

Abstract

In this paper, we real-time implemented to the TMS320C5416 the vocoder of variable bit rate applied the SOLA-B algorithm by Henja to the ITU-T G.729A vocoder of 8kbps transmission rate. This proposed method using the SOLA-B algorithm is that it is reduced the duration of the speech in encoding and is played at the speed of normal by extending the duration of the speech in decoding. At this time, we handled that the interval of cross correlation function if skipped every 3 sample for decreasing the computational amount of SOLA-B algorithm. The real-time implemented vocoder of G.729A and SOLA-B algorithm is represented the complexity of maximum that is 10.2MIPS in encoder and 2.8MIPS in decoder of 8kbps transmission rate. Also, it is represented the complexity of maximum that is 18.5MIPS in encoder and 13.1MIPS in decoder of 6kbps, it is 18.5MIPS in encoder and 13.1MIPS in decoder of 4kbps. The used memory is about program ROM 9.7kwords, table ROM 4.5kwords, RAM 5.1kwords. The waveform of output is showed by the result of C simulator and Bit Exact. Also, for evaluation of speech quality of the vocoder of real-time implemented variable bit rate, it is estimated the MOS score of 3.69 in 4kbps.

Keyword : G.729A, SOLA-B, DSP, Variable bit rate, MOS score

* 正會員, 崇實大學交 情報通信工學科
(Dept. of Information & Communication Engineering,

SoongSil University)

接受日字:2003年6月5日, 수정완료일:2003年11月22日

I. 서론

CELP(Code Excited Linear Prediction)구조의 보코더들은 낮은 전송율에서 양호한 음질을 얻을 수 있으며 ITU-T 국제표준화 기구를 통해 다양한 응용분야에서 표준화가 이루어지고 있다. 특히 PCS 및 전화기 라인 상에서의 인터넷을 통한 화상회의를 위하여 낮은 전송율에서 고음질을 가지는 코덱이 많은 주목을 받고 있다. 이러한 CELP 계열 보코더들 중에서 G.729A는 멀티미디어 통신 환경하의 음성 전송 표준 보코더로 개발되었다^[1].

음성속도 변환 알고리즘인 SOLA-B는 음질이 우수하여 MP3 플레이어나 Voice Recorder의 외국어 학습을 위한 응용 분야에서 사용되어지고 있다.

본 논문에서 사용한 TMS320C5416은 TI사의 고정 소수점 DSP로서 40bit의 ALU, 개선된 멀티버스 등을 지원한다. 이러한 고정 소수점 DSP를 사용하여 G.729A 음성 보코더에 계산량이 줄어든 SOLA-B 알고리즘을 적용하여 가변 전송율의 보코더를 실시간 구현하였다.

II. 음성속도 변환 방식

음성속도 변환(Time Scale Modification) 알고리즘은 시간 영역에서만 처리를 하여 계산시간이 적게 소모되는 장점을 가지고 있다^[2]. 또한, 이 방식은 음성신호를 부드럽게 연결하므로 속도 변환된 음성의 음질이 좋다^[3]. 그러나, 기존의 SOLA 알고리즘은 압축모드에서는 클릭킹 현상이 발생하고 확장모드에서는 잔향 현상이 발생한다. 이러한 현상들을 개선하기 위해 수정 변형된 SOLA-B 알고리즘은 Henja에 의해서 제안되었다^[3].

<그림 1>은 SOLA-B 알고리즘을 사용하여 음성을 압축하는 것을 보여주고 있다. 여기서, Sa는 분석을 위한 분석의 분할 단위를 나타내며, Ss는 합성을 위한 합성의 분할 단위를 나타내고 있다^[4].

음성속도 변환 알고리즘의 과정은 크게 세 단계로 나누어진다. 첫째는 윈도우를 씌워 프레임을 생성하는 윈도우링 과정이고, 두 번째는 상호상관 함수(Cross-correlation)를 사용하여 동기화 시켜주는 과정이며, 마지막으로 세 번째 과정은 오버랩(Overlap)되는 프레임을 더해(Add)주는 과정이다.

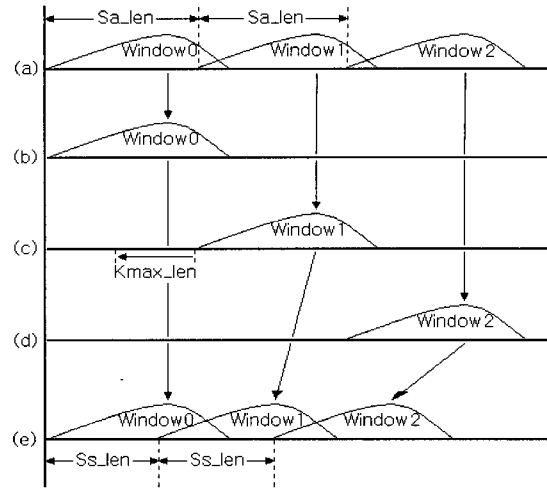


그림 1. SOLA-B 알고리즘을 이용한 압축모드 속도변환 (a) 원음성 신호, (b) 윈도우 0의 신호, (c) 윈도우 1의 신호, (d) 윈도우 2의 신호, (e) 출력된 신호

Fig. 1. Compression mode Time Scale Modification using the SOLA-B algorithm. (a) Original speech signal, (b) Signal of window 0, (c) Signal of window 1, (d) Signal of window 2, (e) Output signal.

1. 윈도우(Windowing)

원음성 신호에 대해 윈도우를 씌워 프레임 단위로 분할하는 과정이다. 분할을 하기 위한 윈도우에는 Hamming 윈도우를 사용하였다. 아래의 식 (1)은 윈도우가 곱해져 일정한 크기로 분해된 프레임을 나타내는 식이다.

$$S_{analysis}(n) = W_{analysis}(m-n)S(n) \tag{1}$$

- $S_{analysis}(n)$: 분할된 프레임 신호
- $W_{analysis}(n)$: 분석 윈도우 함수
- m : m번째 윈도우
- $S(n)$: 원음성 신호

2. 동기화(Synchronization)

윈도우링 과정에 의해 분해된 프레임들에 대해 상호상관 함수를 이용하여 동기를 맞추어 주는 과정이다. 상호상관 함수는 일반적으로 많이 쓰이는 식으로 표현이 된다. 상호상관 함수에서 검색의 범위인 Kmax_len는 100에서 200 사이를 사용한다. 본 논문에서는 남, 여의 다양한 음성시료를 사용하여 실험한 결과 120의 값이 가장 적합하여 이 값을 사용하였다. 동

기화 과정에서는 상호 상관 함수의 계산 결과가 가장 큰 값의 인덱스를 다음의 합성 과정에 넘겨준다.

3. 합성(Overlap and Add)

동기화 과정에서 얻은 위치 값을 이용하여 프레임의 오버랩되는 부분을 더해주는 과정이다. 오버랩시키는 부분은 가중치를 주고, 더해주면 합성음이 완성된다. 따라서, 음성의 속도를 빠르거나 느리게 합성하기 위해서는 프레임의 간격들을 넓거나 좁게 재배열하면 된다.

III. 실험의 실시간 구현

제안한 방법의 성능을 평가하기 위해 범용 16bit fixed-point DSP를 이용하여 실시간 구현하였다. G.729A와 SOLA-B 알고리즘의 실시간 구현에 사용된 보드는 TI사의 TMS320C5416 EVM 보드를 사용하였다^[6]. 가변 전송율 G.729A 보코더의 개발순서는 다음과 같이 실시하였다. 보드에 G.729A Fixed-point C 소스와 SOLA-B Fixed-point C 소스를 먼저 DSP 상에서 비실시간 동작되도록 구현하였다. 그리고, 비실시간 동작된 소스에 대해 루틴별로 MIPS를 측정하고 함수들의 통합 및 불필요한 메모리할당을 제거하여 C 레벨에서 최적화를 하였다. 핸드 어셈블 작업에 들어가기 전에 5416은 하버드 구조를 가지고 있기 때문에 효율적으로 메모리를 관리하고 수행최적화를 이루기 위해서는 변수들의 메모리 할당과 파이프라인 충돌에 주의하여야 한다. 따라서, 메모리 할당과 함수간 인자전달방법에 대해 명백한 사전정의를 하였다. 이러한 과정을 기본으로 하여 본 논문에서는 프로그램의 전 과정이 어셈블리어로 구현되었다. 마지막으로 음성의 입력과 출력은 Pingpong 방식을 사용하였다. 이 방식은 두 개의 버퍼를 두고 ping 버퍼가 full 되었으면 이 버퍼에 대해 압축을 수행하고 압축하는 동안 pong 버퍼에 새로운 데이터를 받는 방식으로 동작한다.

음성 시료로는 ITU-T에서 제공하는 테스트 벡터를 사용하였으며, 그 외에 다양한 음질 평가를 위해 남성과 여성에 대한 한국어와 영어의 2가지 언어를 사용하였다.

<그림 2>는 본 논문에서 제안한 부호화단의 알고리즘을 나타낸 블록도이다. 그리고, 각 블록에 대한 번호(①...③)의 설명은 다음과 같다.

1. 부호화단(Encoder)

- ① 원음성 신호가 입력되었을 때 전송율에 따른 흐름을 결정하는 블록이다. 여기서, SOLA-B 알고리즘을 사용하면 다양한 전송율이 구현 가능하나 사용상의 편의성과 알고리즘의 복잡도를 줄이기 위하여 8kbps, 6kbps, 4kbps의 3가지 전송율만을 고려하였다. 속도 변화율 $a(S_s/S_a)$ 의 값이 1일 경우에는 G.729A의 기본 전송율인 8kbps를 유지하기 위해 그대로 G.729A의 부호화단에서 처리하게 된다. 그러나, a 의 값이 1보다 작을 경우에는 6kbps와 4kbps의 전송율을 위해 아래의 블록에서 SOLA-B 알고리즘을 적용하게 된다.

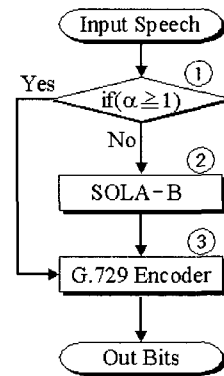


그림 2. 제안한 알고리즘의 부호화단 블록도
Fig. 2. Encoder block diagram of proposed algorithm.

- ② 이 블록에서는 음성의 길이를 줄여주기 위해 SOLA-B 알고리즘의 압축모드를 적용하는 블록이다. 전체 SOLA-B 알고리즘의 계산량 중에서 상호 상관 함수가 차지하는 비율은 75%정도이다. 따라서, 계산량을 줄이기 위해 상호상관 함수의 검색 범위인 K_{max_len} 안에서 이동하는 샘플 수를 1샘플이 아닌 2샘플 이상으로 사용하였다. 그리고, 4kbps의 전송율을 위해서는 $a=0.5$ 의 값이 필요하고, 6kbps의 전송율을 위해서는 $a=0.75$ 의 값이 필요하다. 이렇게 a 의 값에 대한 배율만큼 음성의 길이가 변경된다.
- ③ G.729A의 부호화 시 프레임의 크기는 80샘플을 사용하고 이것을 다시 40샘플의 서브프레임으로 나누어 처리를 수행한다. 그리고 LPC 분석을 위한 윈도우의 크기는 240샘플을 사용하며, Auto correlation 방법을 사용하여 LPC 계수값을 얻을 수 있다. 이렇게 얻어진 10차의 계수값들은 다시 LSP로 양자

화가 이루어진다. LSP 양자화는 Switched MA (Moving Average) 예측오차에 대해서 2-Stage의 벡터 양자화를 이용하고 있다. 각 Stage에 대한 비트 할당을 <표 1>에 나타내었다.

1st stage에서는 10차의 벡터 양자화로 128 (7bit)개의 테이블을 형성하고 있다. 그리고 2nd stage에서는 5차의 스플릿 벡터 양자화로 32(5bit)개의 테이블을 형성한다. 이렇게 하여 1개의 프레임에 대해 80bit의 파라미터들이 생성된다.

표 1. LSP 양자화의 비트할당
Table 1. Bit allocation of LSP quantizer.

1st stage	10th order	7bit
2nd stage	5th low order	5bit
	5th high order	5bit
Total		17bit

2. 복호화단(Decoder)

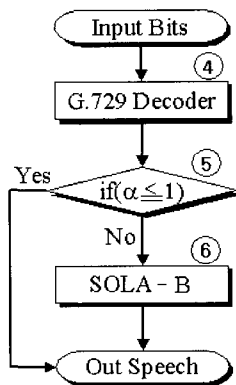


그림 3. 제안한 알고리즘의 복호화단 블록도
Fig. 3. Decoder block diagram of proposed algorithm.

- ④ 저장된 파라미터들을 다시 복호화 시키는 블록이다. 80bit의 파라미터를 입력으로 하여 복호화되고 80샘플로 음성을 출력한다.
- ⑤ 음성속도 변화율 α 의 값이 1일 경우에는 G.729A의 복호화단에서 처리하게 된다. 그러나, α 의 값이 1보다 클 경우에는 6kbps와 4kbps의 전송율을 위해 아래의 블록에서 SOLA-B 알고리즘을 적용하게 된다.
- ⑥ G.729A로 복호화된 음성은 속도가 빨라진 상태로 변화되어 있으므로 이를 다시 복원하기 위해서

SOLA-B 알고리즘을 이용하여 음성의 속도를 느리게 해준다. 이 블록에서는 부호화단과는 반대의 개념으로 전송율이 4kbps인 경우에는 $\alpha=2$ 가 필요하고, 6kbps인 경우에는 $\alpha=1.25$ 가 필요하게 된다.

이렇게 처리된 음성은 정상 속도를 가진 음성으로 출력되어진다.

IV. 실험 결과

<그림 4>는 원 음성신호의 파형을 나타낸 것이다. 그리고 <그림 5>는 원 음성신호를 4kbps의 전송율로 부호화하고 복호화한 최종 출력파형을 나타내고 있으며, SOLA-B 알고리즘 적용 시 상호상관 함수에서 계산되어지는 샘플의 수를 하나씩 처리한 Fixed-point C로 시뮬레이션 결과 파형이다. <그림 6>은 <그림 5>

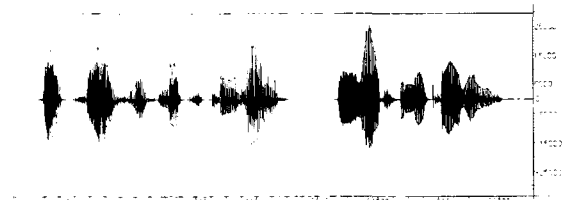


그림 4. 원 음성신호 파형
Fig. 4. Original speech signal waveform.

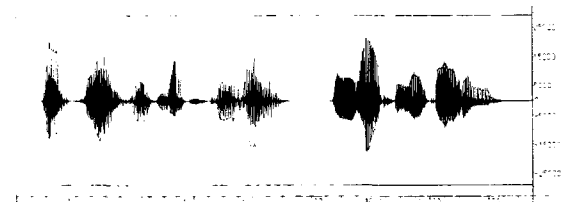


그림 5. 1개의 샘플씩 처리된 C 시뮬레이션 출력파형
Fig. 5. The output waveform of C simulation which 1 sample is processed.

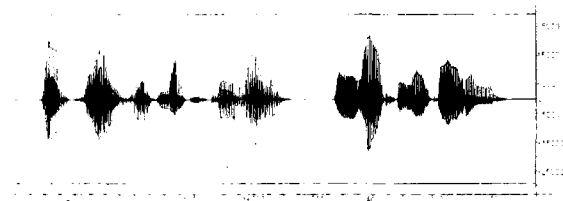


그림 6. 하나의 샘플씩 처리된 TMS320C5416 출력파형
Fig. 6. The output waveform of TMS320C5416 which one sample is processed.

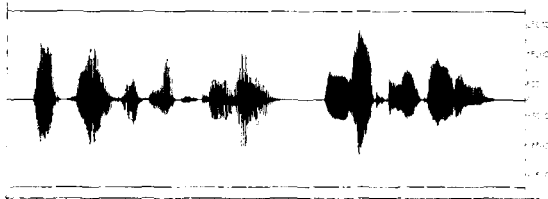


그림 7. 3개의 샘플씩 처리된 C 시뮬레이션 출력파형
Fig. 7. The output waveform of C simulation which 3 sample is processed.

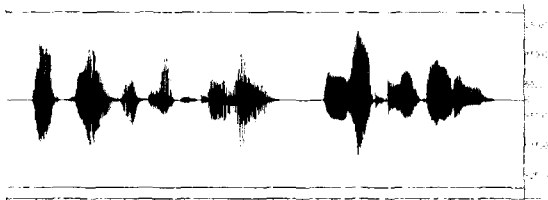


그림 8. 3개의 샘플씩 처리된 TMS320C5416 출력파형
Fig. 8. The output waveform of TMS320C5416 which 3 sample is processed.

와 처리 과정은 동일하며 TMS320C5416으로 처리된 파형을 나타내고 있다. 또한, <그림 7>은 4kbps의 전송율에 대해 상호상관 함수에서 샘플을 3개씩 건너뛰면서 처리한 Fixed-point C로 시뮬레이션한 결과를 보여주는 파형이다. <그림 8>은 <그림 7>에 대해서 TMS320C5416을 이용하여 실시간 처리한 결과를 보여주는 파형이다.

G.729A에 SOLA-B 알고리즘을 적용하여 실시간 구현된 가변 전송율 보코더의 MOS 테스트 결과를 <표 2>에 나타내었다. 테스트 결과 8kbps의 전송율에서는 평균 3.91정도의 MOS 값이 측정되었다. 6kbps, 4kbps에서는 상호상관 함수의 처리되는 샘플에 따라 아래 <표 2>와 같은 MOS 결과 값을 얻었다.

표 2. MOS 점수
Table 2. MOS score.

	6kbps	4kbps
1 샘플씩 처리	3.80	3.64
2 샘플씩 처리	3.78	3.60
3 샘플씩 처리	3.75	3.56
4 샘플씩 처리	3.71	3.51

최대 복잡도는 <표 3>에서와 같이 1샘플씩 처리 했을 경우보다 3샘플씩 처리하였을 경우 평균 4-5MIPS

정도의 차이가 있었다.

표 3. 최대 복잡도(MIPS)
Table 3. Maximum Complexity(MIPS).

	6kbps		4kbps	
	부호화	복호화	부호화	복호화
1 샘플씩 처리	21.2	13.8	23.5	16.1
2 샘플씩 처리	19.1	11.7	21.1	13.7
3 샘플씩 처리	17.3	9.9	18.5	11.1
4 샘플씩 처리	15.3	7.9	16.1	8.7

논문의 응용 범위는 음성 압축방법에 모두 적용 가능하다. 이러한 보코더들로는 G.723.1, G.726, AMR, IMBE, Dod-CELP, QCELP, G.728, EVRC, VSELP 등이 있다. 본 논문의 알고리즘을 적용할 경우 위와 같은 보코더들이 기존의 전송율에서 1/2배와 1/4배로 압축이 가능하다.

V. 결 론

본 논문은 TI사의 TMS320C5416 DSP 칩을 이용하여 G.729A에 음성속도 변환 알고리즘인 SOLA -B 알고리즘을 적용하여 가변율을 가지는 보코더를 실시간 구현하였다. ITU-T에서 제공하는 테스트 벡터들에 대해 어셈블리어로 구현되어진 G.729A 가변 보코더와 C-simulation된 결과는 Bit Exact한 결과임을 확인하였다. 그리고, 실시간 구현된 G.729A와 SOLA-B 알고리즘의 가변 전송율 보코더는 8kbps 전송율일 때 인코더는 10.2MIPS이고, 디코더에서는 2.8MIPS의 최대 복잡도를 나타내었다. SOLA-B 알고리즘에서 상호상관 함수를 3개의 샘플씩 처리하였을 경우 6kbps 전송율일 때 인코더 17.3MIPS, 디코더는 9.9MIPS의 최대 복잡도를 나타내었으며, 4kbps 전송율일 때 인코더 18.5MIPS, 디코더에서 11.1MIPS의 최대 복잡도를 나타내었다. 또한, MOS 테스트를 통한 음질 평가 결과 8kbps에서는 3.91, 6kbps에서는 3.75, 4kbps에서는 3.56 정도의 음질을 가졌다.

참 고 문 헌

[1] Coding of speech at 8kbit/s using conjugate-structure algebraic-code-excited linear- predic-

- tion (CS-ACELP), ITU-T Recommendation G.729.
- [2] E.Hardam "High quality time scale modification of speech signals using fast synchronized overlap add algorithms," IEEE. Processings of the 1990 International Conference on Acoustics, Speech, and Signal Processing, vol. 1, 409-412, Feb. 1990.
- [3] D.J.Hejna Jr., "Real-Time Time-Scale Modification of Speech via the Synchronized Overlap-Add Algorithm." Master's thesis, Department of Electrical Engineering and Computer Science, MIT, Feb. 1990.
- [4] 박규식, 오승록, 김선영, "G.729 음성 복호화기와 듀얼 SOLA 알고리즘을 통합한 최적의 음성 속도 변환 시스템." 한국음향학회지, 21(3), 293-303, April 2002
- [5] 안도건, 유승균, 최용수, 이재성, 강태익, 박성현, "16비트 고정 소수점 DSP를 이용한 다채널 G.729A 음성 복호화기의 실시간 구현." 한국음향학회지, 19(4), 45-51, May 2000

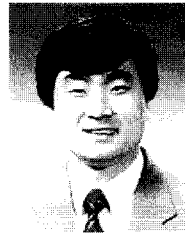
저 자 소 개



咸 明 圭(正會員)

1997년 2월 : 송실대학교 정보통신공학과 졸업(학사). 1999년 2월 : 송실대학교 대학원 정보통신공학과 졸업(석사). 1999년 9월~현재 : 송실대학교 대학원 정보통신공학과 박사과정. <주관심분야 : 음

성합성, 음성코딩, 음성인식, DSP Processor>



裴 明 振(正會員)

1981년 2월 : 송실대학교 전자공학과(공학사). 1983년 2월 : 서울대학교 대학원 전자공학과(공학석사). 1987년 8월 : 서울대학교 대학원 전자공학과(공학박사). 1992년 9월~현재 : 송실대학교 정보통신전

자 공학부 정교수. <주관심분야 : 음성코딩, 음성합성, 화자인식, 디지털 신호처리, 디지털통신>