

## TMS320C5416을 이용한 SOLA-B 알고리즘과 G.729A 보코더의 음질 향상된 가변 전송률 보코더의 실시간 구현\*

Real-time Implementation of Variable Transmission Bit Rate Vocoder Improved Speech Quality in SOLA-B Algorithm & G.729A Vocoder Using on the TMS320C5416

함 명 규\*\* · 배 명 진\*\*  
Myung-kyu Ham · Myung-jin Bae

### ABSTRACT

In this paper, we implemented the vocoder of variable rate by applying the SOLA-B algorithm to the G.729A to the TMS320C5416 in real-time. This 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. But the method applied to the existed G.729A and SOLA-B algorithm is caused the loss of speech quality in G.729A which is not reflected about length variation of speech. Therefore the proposed method is encoded according as it is modified the structure of LSP quantization table about the length of speech is reduced by using the SOLA-B algorithm. The vocoder of variable rate by applying the G.729A and SOLA-B algorithm is represented the maximum complexity of 10.2MIPS about encoder and 2.8MIPS about decoder in 8kbps transmission rate. Also it is evaluated 17.3MIPS about encoder, 9.9MIPS about decoder in 6kbps and 18.5MIPS about encoder, 11.1MIPS about decoder in 4kbps according to the transmission rate. The used memory is about program ROM 9.7kwords, table ROM 4.69kwords, RAM 5.2kwords. The waveform of output is showed by the result of C simulator and Bit Exact. Also, the result of MOS test for evaluation of speech quality of the vocoder of variable rate which is implemented in real-time, it is estimated about 3.68 in 4kbps.

**Keywords:** TMS320C5416, SOLA-B, G.729A, LSP Quantization Table

### 1. 서 론

현재까지 발표된 음성부호화기 중 가장 많은 연구가 이루어지고 있는 방식은 CELP(Code Excited Linear Prediction) 구조이다. 이러한 구조의 보코더들은 낮은 전송률에서 양호한 음질을 얻을 수 있으며 ITU-T 국제표준화 기구를 통해 다양한 응용분야에서 표준화가 이루어

---

\* 본 논문은 숭실대학교 연구지원으로 이루어졌습니다.  
\*\* 숭실대학교 정보통신공학과

지고 있다. 특히 PCS 및 전화기 라인상에서의 인터넷을 통한 화상회의를 위하여 낮은 전송율에서 고품질을 가지는 코덱이 많은 주목을 받고 있다. 이러한 CELP 계열 보코더들 중에서 G.729A는 멀티미디어 통신 환경하의 음성 전송 표준 보코더로 개발되었다[1].

G.729A는 8 kbps의 전송율을 갖는 구조로 현재 별정 통신으로 상용화되는 인터넷폰과 그 외의 이동 통신용 보코더로 사용되어지고 있으며, 낮은 전송율에 비해서 우수한 음질을 제공하고 있다. 그리고, 음성속도 변환 알고리즘인 SOLA-B는 음질이 우수하여 Voice Recorder의 외국어 학습을 위한 응용 분야에서 사용되어지고 있다.

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

## 2. 음성속도 변환 방식

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

그림 1은 SOLA-B 알고리즘을 사용하여 음성을 압축하는 것을 보여주고 있다. 여기서,  $S_a$ 는 분석을 위한 분석의 분할 단위를 나타내며,  $S_s$ 는 합성을 위한 합성의 분할 단위를 나타내고 있다[4].

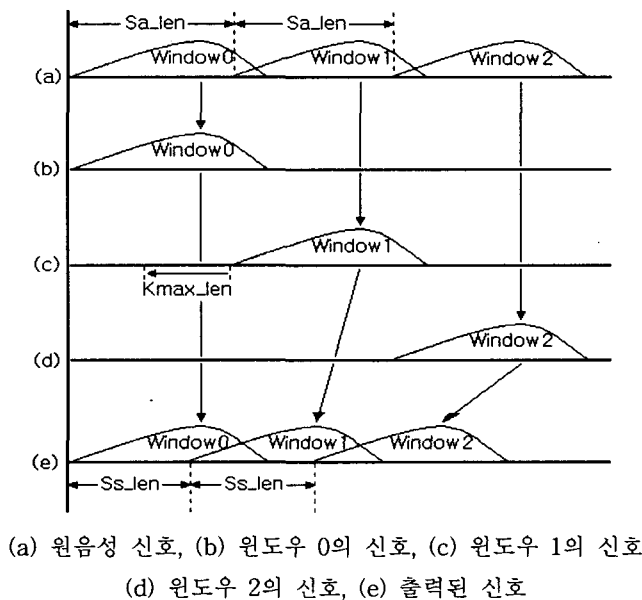


그림 1. SOLA-B 알고리즘을 이용한 압축모드 속도변환

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

### 2.1. 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.

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

$S_{analysis}(n)$ : 분할된 프레임 신호

$W_{analysis}(n)$ : 분석 윈도우 함수

$m$ : m번째 윈도우

$S(n)$ : 원음성 신호

### 2.2. 동기화(Synchronization)

윈도우 과정에 의해 분해된 프레임들에 대해 상호 상관 함수를 이용하여 동기를 맞추어 주는 과정이다. 상호 상관 함수는 일반적으로 많이 쓰이는 식으로 표현이 된다. 상호 상관 함수에서 검색의 범위인  $K_{max\_len}$ 는 100에서 200 사이를 사용하였다. 본 논문에서는 남, 여의 다양한 음성시료를 사용하여 실험한 결과 120의 값이 가장 적합하여 이 값을 사용하였다. 동기화 과정에서는 상호 상관 함수의 계산 결과가 가장 큰 값의 인덱스를 다음의 합성 과정에 넘겨준다.

### 2.2. 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.

## 3. 실험의 실시간 구현

제안한 방법의 성능을 평가하기 위해 범용 16 bit fixed-point DSP를 이용하여 실시간 구현하였다. G.729A와 SOLA-B 알고리즘의 실시간 구현에 사용된 보드는 TI사의 TMS320C5416 EVM 보드를 사용하였다[5]. 가변 전송률 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 측

정하고 함수들의 통합 및 불필요한 메모리 할당을 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 full되었으면 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 2 가지 언어를 사용하였다.

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

### 3.1. 부호화단(Encoder)

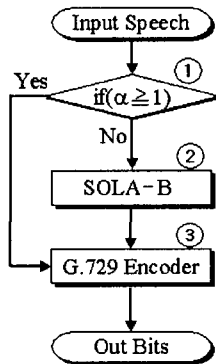


그림 2. 제안한 알고리즘의 부호화단 블록도

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

② 이 블록에서는 음성의 길이를 줄여주기 위해 SOLA-B 알고리즘의 압축모드를 적용하는 블록이다. 4 kbps의 전송률을 위해서는  $\alpha=0.5$ 의 값이 필요하고, 6 kbps의 전송률을 위해서는  $\alpha=0.75$ 의 값이 필요하다. 이렇게  $\alpha$ 의 값에 대한 배율만큼 음성의 길이가 변경된다.

③ G.729A의 부호화 시 프레임의 크기는 80 샘플을 사용하고, 이것을 다시 40 샘플의 서브프레임으로 나누어 처리를 수행한다. 그리고, LPC 분석을 위한 윈도우의 크기는 240 샘플을 사용하며, Auto correlation 방법을 사용하여 LPC 계수값을 얻을 수 있다. 이렇게 얻어진 10 차의 계수값들은 다시 LSP로 양자화가 이루어진다. LSP 양자화는 Switched MA(Moving

Average) 예측오차에 대해서 2-Stage의 벡터 양자화를 이용하고 있다. 각 Stage에 대한 비트 할당을 표 1에 나타내었다.

표 1. LSP 양자화의 비트할당

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

1st stage에서는 10 차의 벡터 양자화로 128(7 bit) 개의 테이블을 형성하고 있다. 그리고, 2nd stage에서는 5 차의 스플릿 벡터 양자화로 32(5 bit) 개의 테이블을 형성한다. 그러나, 본 논문에서는 음성을 압축하여, G.729A에서 부호화하므로, 기존의 LSP 양자화 테이블에서 정확한 LSP 정보를 보낼 수 없다는 단점을 가지고 있다. 이를 보완하여 음질을 향상시키기 위해서 1st stage의 LSP 양자화 테이블에 1 bit를 더 할당하여 256 개의 양자화 테이블을 재구성하였다.

LSP 양자화 테이블의 구성은 기존 G.729A에 존재하는 128 개의 테이블 내용에 대해 각각 4 kbps와 6 kbps에 대해 SOLA-B 알고리즘으로 압축한 음성 신호에서 새로 훈련시킨 형태의 LSP 양자화 테이블을 재구성하였다.

벡터 테이블을 새로 구성하기 위해 영어, 한국어에 대해 각각 100씩의 남성과 여성으로 구성된 음성 데이터베이스에서 SOLA-B 알고리즘을 적용해  $\alpha=0.5$ 와  $\alpha=0.75$ 로 압축하였다. 이렇게 압축된 음성 신호는 표준 LBG 방법을 사용하여 벡터 테이블을 구성하였다[6]. 또한, 8 kbps일 때는 기존의 LSP 양자화 테이블을 그대로 사용하고, 6 kbps, 4 kbps일 때는 추가된 LSP 양자화 테이블을 사용하므로 호환성은 그대로 유지된다.

### 3.2. 복호화단(Decoder)

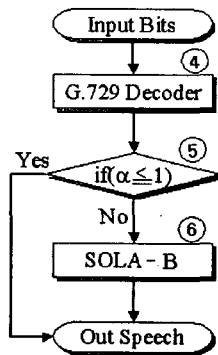


그림 3. 제안한 알고리즘의 복호화단 블록도

④ 저장된 파라미터들을 다시 복호화시키는 블록이다. 80 bit의 파라미터를 입력으로 하여 복호화되고, 80 샘플로 음성을 출력한다.

⑤ 음성속도 변화율  $\alpha$ 의 값이 1일 경우에는 G.729A의 복호화단에서 처리하게 된다. 그러나,  $\alpha$ 의 값이 1보다 클 경우에는 6 kbps와 4 kbps의 전송률을 위해 아래의 블록에서 SOLA-B 알고리즘을 적용하게 된다.

⑥ G.729A로 복호화된 음성은 속도가 빨라진 상태로 변환되어 있으므로 이를 다시 복원하기 위해서 SOLA-B 알고리즘을 이용하여 음성의 속도를 느리게 해준다. 이 블록에서는 부호화단과는 반대의 개념으로 전송률이 4 kbps인 경우에는  $\alpha=2$ 가 필요하고, 6 kbps인 경우에는  $\alpha=1.25$ 가 필요하게 된다. 이렇게 처리된 음성은 정상 속도를 가진 음성으로 출력되어진다.

#### 4. 실험 결과

그림 4는 원음성 신호의 파형을 나타낸 것이다. 그리고, 그림 5는 2 배로 빨라진 음성 신호를 G.729A의 부호화와 복호화를 거쳐 SOLA-B 알고리즘을 통해 음성속도를 2 배로 느리게 하여 정상적인 음성 속도로 복원된 최종 출력 음성신호의 Fixed-point C로 시뮬레이션한 파형을 나타낸 것이다. 그림 6은 2 배로 빨라진 음성 신호를 G.729A의 부호화와 복호화를 거쳐 SOLA-B 알고리즘을 통해 음성속도를 2 배로 느리게 하여 정상적인 음성 속도로 복원된 최종 출력 음성신호의 TMS320C5416으로 처리된 파형을 나타낸 것이다.

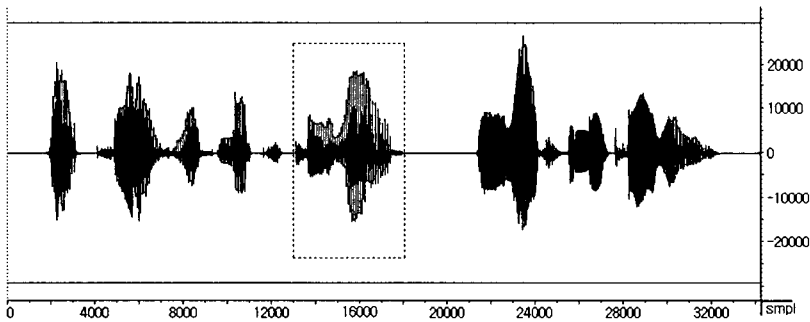


그림 4. 원 음성신호 파형

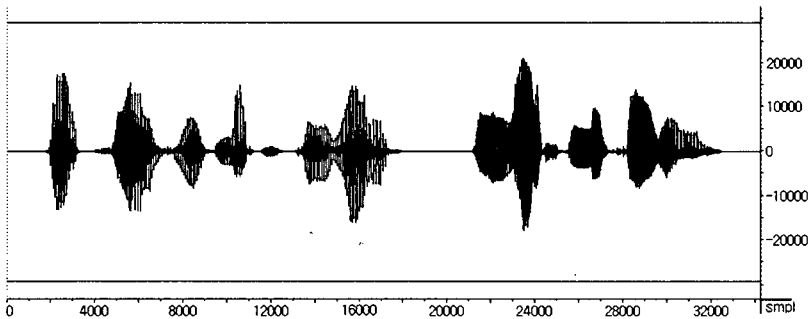
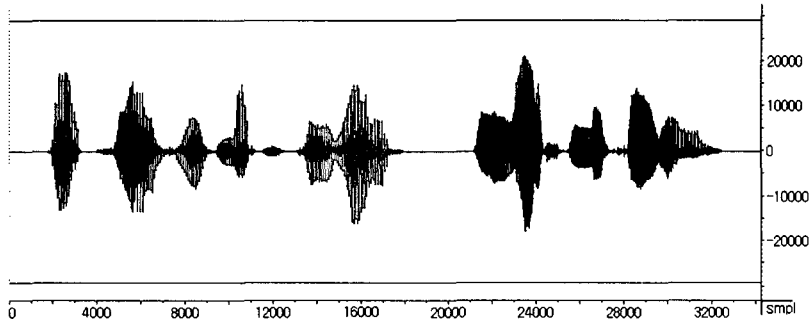


그림 5. SOLA-B의 C 시뮬레이션 출력 파형( $\alpha=2$ )

그림 6. SOLA-B의 TMS320C5416 출력 파형( $\alpha=2$ )

G.729A는 부호화기와 복호화기로 구성되어 있으며 부호화기의 경우 크게 전처리단, LP계수 분석단, 개회로 피치 검색단, 적응 코드북 검색단, 고정 코드북 검색단으로 구분할 수 있다. 표 2는 G.729A 부호화기의 기능별 최대 복잡도를 나타내었다. 그리고, 복호화기의 경우 크게 LP 계수와 코드북들을 복호화하는 복호화단, 후필터단, 후처리단으로 구분할 수 있으며, 각각에 대한 최대 복잡도를 표 3에 나타내었다. 또한, SOLA-B 알고리즘의 경우 크게 동기화단, 합성단으로 구분되며 두 단계에 대한 최대 복잡도를 표 4에 나타내었다.

표 2. G.729A 부호화기의 최대 복잡도(MIPS)

Pre processing	LP analysis	Open loop Pitch search	Adaptive codebook search	Fixed codebook search
0.12	1.836	1.02	1.122	5.304

표 3. G.729A 복호화기의 최대 복잡도(MIPS)

Decoder	Post filter	Post processing
1.291	1.024	0.337

표 4. SOLA-B algorithm의 최대 복잡도(MIPS)

Synchronization	Overlap and add
5.325	1.427

G.729A와 SOLA-B 알고리즘을 LSP 양자화 테이블의 확장 없이 처리하여 실험한 결과가 표 5와 같으며, 음질 향상을 위해 LSP 양자화 테이블을 확장하여 MOS 테스트를 수행한 결과가 표 6과 같다. LSP 양자화 테이블을 확장하여 처리하였을 경우 8 kbps에서는 3.91, 6 kbps에서는 3.83, 4 kbps에서는 3.68 정도의 음질이 향상되었다.

표 5. 처리 전의 MOS 점수

	8 kbps	6 kbps	4 kbps
음성 1(남)	3.90	3.76	3.54
음성 2(남)	3.93	3.78	3.59
음성 3(여)	3.91	3.74	3.57
음성 4(여)	3.89	3.72	3.52
평 균	3.91	3.75	3.56

표 6. 처리후의 MOS 점수

	8 kbps	6 kbps	4 kbps
음성 1(남)	3.90	3.81	3.69
음성 2(남)	3.93	3.87	3.73
음성 3(여)	3.91	3.83	3.71
음성 4(여)	3.89	3.79	3.6
평 균	3.91	3.83	3.68

아래의 그림들은 그림 4의 점선으로 그어진 상자의 음성에 대한 스펙트럼을 나타낸 것이다. 그림 7은 원음성 신호에 대한 스펙트럼을 나타낸다. 그리고, 그림 8은 기존의 LSP 양자화 테이블을 사용하였을 경우에 대한 스펙트럼이다. 또한 그림 9는 LSP 양자화 테이블을 확장하였을 경우에 대한 스펙트럼을 나타내고 있다. 그림과 같이 LSP 양자화 테이블을 확장하였을 경우 왜곡이 적어짐을 볼 수 있다.

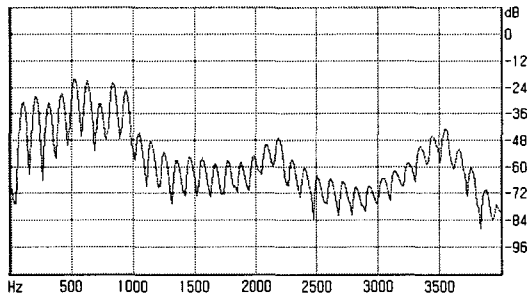


그림 7. 원음성 신호의 스펙트럼

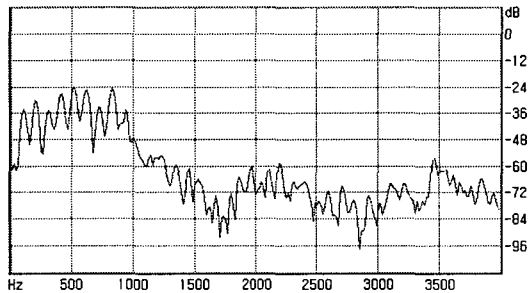


그림 8. 처리 전의 스펙트럼



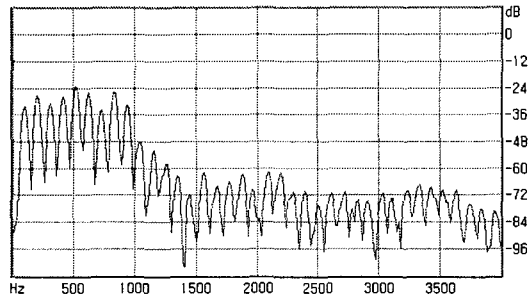


그림 9. 처리 후의 스펙트럼

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 전송률에서 1/2 배와 1/4 배로 압축이 가능하다.

## 5. 결론

본 논문은 TI사의 TMS320C5416 DSP 칩을 이용하여 LSP 양자화 테이블을 확장한 G.729A 보코더에 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 8 kbps 전송률일 때 인코더는 10.2 MIPS이고, 디코더에서는 2.8 MIPS의 최대 복잡도를 나타내었다. 6 kbps 전송률일 때 인코더 17.3 MIPS이고, 디코더는 9.9 MIPS의 최대 복잡도를 나타내었으며, 4 kbps 전송률일 때 인코더 18.5 MIPS이고, 디코더에서 11.1 MIPS의 최대 복잡도를 나타내었다. 또한, MOS 테스트를 통한 음질 평가 결과 8 kbps에서는 3.91, 6 kbps에서는 3.83, 4 kbps에서는 3.68 정도의 음질을 가졌다.

## 참고 문헌

- [1] Coding of speech at 8kbit/s using conjugate-structure algebraic-code-excited linear-prediction (CS-ACELP), ITU-T Recommendation G.729.
- [2] Hardam, E. 1990. "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.
- [3] Hejna, D. J. Jr. 1990. *Real-time Time-scale modification of speech via the synchronized overlap-add algorithm*. Master's thesis, Department of Electrical Engineering and Computer Science, MIT.

- [4] 박규식, 오승록, 김선영. 2002. "G.729 음성 복호화기와 듀얼 SOLA 알고리즘을 통합한 최적의 음성 속도 변환 시스템." *한국음향학회지*, 21(3), 293-303.
- [5] 안도건, 유승균, 최용수, 이재성, 강태익, 박성현. 2000. "16비트 고정 소수점 DSP를 이용한 다채널 G.729A 음성 복호화기의 실시간 구현." *한국음향학회지*, 19(4), 45-51.
- [6] Linde, Y., A. Buzo & R. M. Gray. 1980. "An algorithm for vector quantizer design." *IEEE Trans. on Communications*, 28(1).

접수일자: 2003. 7. 17.

게재결정: 2003. 8. 31.

▲ 함명규

서울특별시 동작구 상도5동 (우: 156-743)

숭실대학교 공과대학 정보통신공학과

Tel: +82-2-824-0906 (O), +82-2-6245-7897 (H)

Fax: +82-2-2-561-7560

E-mail: mkham@mindtell.co.kr

▲ 배명진

서울특별시 동작구 상도5동 (우: 156-743)

숭실대학교 공과대학 정보통신공학과

Tel: +82-2-820-0902

E-mail: mjbae@ssu.ac.kr