

# OakDSP Core 를 이용한 GSM-EFR 음성 부호화기의 실시간 구현

최민석<sup>o</sup>, 변경진, 김경수  
한국전자통신연구원

## Real-time Implementation of a GSM-EFR Speech Coder using a OakDSP Core

Min Seok Choi<sup>o</sup>, Kyung Jin Byun, Kyung Soo Kim  
Electronics and Telecommunications Research Institute (ETRI)

E-mail : mschoi@etri.re.kr

### 요약

본 논문에서는 DSP Group사의 16 비트 고정 소수점 DSP인 OakDSP Core를 사용하여 유럽의 이동통신에서 표준으로 사용되고 있는 음성 부호화기 알고리즘인 GSM-EFR (Global System for Mobile communications - Enhanced Full Rate)을 실시간으로 구현하였다. 구현된 GSM-EFR 음성 부호화기의 계산량은 약 24 MIPS가 소요되며, 7.06 K 워드의 코드 메모리와 12.19 K 워드의 데이터 메모리를 사용하였다. 구현된 음성 부호화기는 ETSI에서 제공하는 시험 벡터 샘플을 모두 통과하였으며, 객관적 평가 틀을 이용하여 지각 평가를 수행한 결과, 32 kbps ADPCM과 비슷한 음질을 보였다. 본 논문에서 실시간으로 구현된 GSM-EFR 음성 부호화기는 IMT2000 비동기 방식의 음성 부호화기 표준인 GSM-AMR의 최상위 전송률 모드로서, 앞으로 IMT-2000 비동기식 단말기용 모델 ASIC에 탑재할 GSM-AMR 음성 부호화기의 구현을 위한 기본 구조로 이용될 예정이다.

### I. 서론

지금까지 제안된 음성 부호화 방식들 중에서 우수한

음질을 갖는다고 알려져 있는 CELP (Code-Excited Linear Prediction)[1]는 표준 음성 부호화 알고리즘의 주류를 이루고 있으며 G.723.1[2], G.729[3], GSM-EFR[4] 등과 같은 최근의 음성 부호화기 국제 표준에서도 CELP 기반의 음성 부호화 알고리즘인 ACELP (Algebraic CELP)[5]를 사용하고 있다. 초기의 유럽의 이동통신 표준안인 GSM 시스템에서는 음성 부호화 방식으로 13kbps RPE-LTP (Regular Pulse Excitation-Long Term Prediction)을 사용하였으나, 최근에는 열악한 채널 환경이나 잡음 환경에서 보다 나은 음질을 제공하기 위해 1996년 GSM-EFR (Enhanced Full Rate speech codec) 음성 부호화 표준안 (GSM 06.60)[6]을 채택하였다. 이 방식은 ITU-T G.729 표준안과 유사하며 ACELP 음성 부호화 방식에 기반을 두고 있는데, 음성 부호화를 위해 12.2 kbps, 채널 부호화를 위해 10.6 kbps의 비트율을 할당하여 전체 채널 전송률은 22.8 kbps가 된다. 또한, GSM-EFR 음성 부호화 방식은 차세대 이동통신 표준안으로서, IMT-2000에서 채택된 AMR (Adaptive Multi-Rate) 음성 부호화 방식의 기초가 되고 있다.

본 논문에서는 IMT2000 비동기 방식 단말기에서 사용되는 GSM-AMR 음성 부호화기의 개발의 일환으로서 최상위 전송률을 가지는 GSM-EFR 음성 부호화 방식을

DSP Group 사의 OakDSP EVM (Evaluation Module) 보드 상에서 실시간으로 구현하였다.

본 논문의 구성은 다음과 같다. II장에서 GSM-EFR 음성 부호화 알고리즘의 구조를 간단히 살펴본 후, III장에서는 GSM-EFR 음성 부호화 알고리즘의 실시간 구현 및 성능평가에 대해 상세히 설명하고, 마지막으로 IV장에서는 결론을 맺는다.

## II. GSM-EFR 음성 부호화 알고리즘

GSM-EFR 음성 부호화기는 20 ms의 분석 프레임을 사용하며, 한 프레임은 5 ms의 서브프레임으로 나뉘어 분석된다. 음성 부호화 과정에서 분석 프레임별로 할당되는 비트수는 표 1과 같다.

표 1. GSM-EFR 음성 부호화기의 프레임별 비트 할당

Parameter	1 <sup>st</sup> &3 <sup>rd</sup> subframe	2 <sup>nd</sup> &4 <sup>th</sup> subframe	Total per frame
2 LSP sets			38
Pitch delay	9	6	30
Pitch gain	4	4	16
Algebraic code	35	35	16
Codebook gain	5	5	20
Total (bit)			244

GSM-EFR 음성 부호화 알고리즘의 부호화 과정은 그림 1과 같다. 고역통과필터를 통과한 음성 신호에 2개의 상이한 30 ms의 비대칭 창함수를 적용하여 한 프레임 당 2 세트의 10차 선형 예측 계수를 구한 후 LSP (Line Spectral Pair) 계수로 변환한다. 1차 MA(Moving Average) 예측을 통한 LSP 잔여 벡터는 분할 행렬 양자화 (Split Matrix Quantization (SMQ))를 사용하여 양자화된다. 적응 코드북 탐색을 위한 피치 예측 과정은 제한된 지연 영역인  $\left[17\frac{3}{6}, 143\right]$  사이에서 개루프 탐색과 페루프 탐색을 수행함으로써 이루어진다. 가중 필터를 거친 음성 신호로부터 매 10ms 마다 예측된 개루프 지연을 이용하여 매 서브프레임마다 페루프 피치 분석이 이루어지고, 피치 지연과 이득이 양자화된다. 위의 과정을 거친 잔차 신호를 모델링하기 위해 매 서브프레임마다 대수 코드북(algebraic codebook)을 사용하게 된다. 이 코

드북에서는 각 여기벡터 들은  $\pm 1$ 의 크기를 가지는 10개의 non-zero 펄스열로 구성된다. 따라서, 각 서브프레임에서 40개의 펄스 위치는 두 개의 펄스열을 가지는 5개의 트랙으로 나뉘어 모델링된다. 최적의 펄스 위치는 적응 코드북 탐색 과정을 거쳐 나온 잔차 신호와 대수 코드북 사이의 상관이 최대가 되는 값을 찾음으로써 구해진다. 이때, 대수 코드북의 구조에서는 5개의 검색 루프를 통해 연속적으로 최적의 펄스 위치를 찾게 되므로 계산량의 감소와 우수한 음질을 얻을 수 있게 된다. 고정 코드북의 이득 값은 MA 예측을 이용하여 양자화된다. GSM-EFR 음성 부호화기의 복호화 과정에서는 전송된 파라미터들로부터 LP 파라미터, 피치, 코드북 인덱스, 그리고 이득 값들을 추출하여, 이 파라미터들을 이용하여 서브프레임 단위로 음성신호를 합성하게 된다. 또한, 복원된 합성음의 음질을 향상시키기 위해 후처리 필터링을 거치게 된다.

## III. GSM-EFR 음성 부호화기의 실시간 구현

본 연구에서 사용된 OakDSP Core는 DSP Group, Inc.의 16 비트 고정 소수점 DSP로서 40 MHz의 동작 주파수와 40 MIPS의 계산 능력을 가지며, 16 비트 데이터와 프로그램 버스를 지닌다. 또한, 저전력으로 설계되어 있어 통신과 전자 응용분야에서 다양하게 사용되고 있다. OakDSP Core는 일반 범용 DSP와는 달리 IP 형태로 제공되기 때문에 향후 IMT2000 이동통신 방식 단말기의 ASIC 구현에 용이하다. OakDSP Core의 주요 특징은 다음과 같다.

- 40 MHz 동작 주파수 (40 MIPS)
- Single cycle multiply - accumulate instruction
- 36 bit ALU
- 4 개의 36 비트 누산기
- Single cycle exponent evaluation
- Max/Min single cycle instruction with pointer latching and modification
- 4 K word on-core memory, 6K word on-chip memory
- 64 k word maximum program memory space

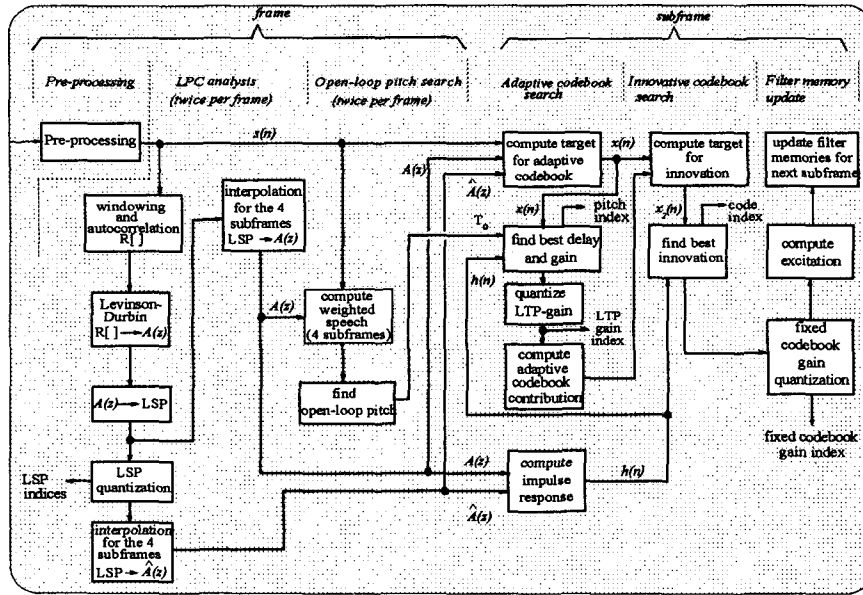


그림 1. GSM-EFR 부호화기의 기본 블록도

GSM-EFR의 실시간 구현은 표준안(GSM 06.60)과 함께 제공되는 고정 소수점 C 프로그램(GSM 06.53)을 바탕으로 OakDSP 어셈블리 언어로 구현하였다. C의 각 모듈을 어셈블리 코드로 구현한 후, C의 결과와 구현된 결과가 일치하는지를 검증하기 위해 시험 벡터를 C에 인가하여 구한 입/출력 데이터를 구현된 어셈블리 코드에 인가하여 결과가 일치함을 확인하였다. GSM-EFR 음성 부호화기를 OakDSP EVM 보드 상에서 실시간 구현함에 있어서 표준 알고리즘을 그대로 구현하기 위해서 알고리즘 상의 최적화는 고려되지 않았으며, OakDSP Core의 구조적 특성을 이용하여 계산량을 감소시키는데 집중하였다. 표 2는 구현된 GSM-EFR 음성 부호화기의 주요 모듈의 계산량을 나타낸 것이다.

표 2. 구현된 주요 모듈의 계산량

부호화기 전체 (22.187 MIPS)		
전처리부	0.313 MIPS	1.3%
LPC 분석 및 양자화	5.396 MIPS	22.4%
피치 검색 (개루프 검색)	1.714 MIPS	7.1%
적응코드북 및 이득양자화	3.308 MIPS	13.8%
고정코드북 및 이득양자화	11.068 MIPS	46.1%
기타 모듈	0.388 MIPS	1.6%
복호화기 전체 (1.844 MIPS)		
파라미터 추출부	1.049 MIPS	4.4%
후처리 필터	0.760 MIPS	3.2%
기타 모듈	0.035 MIPS	0.1%

전체	24.031 MIPS	100%
----	-------------	------

구현된 음성 부호화기는 ETSI에서 제공하는 모든 테스트 벡터[7]에 대해서 C 프로그램의 수행 결과와 완전히 일치된 결과를 보였다. 표 3에서는 GSM-EFR의 실시간 구현에 사용된 메모리 양을 나타내었다.

표 3. GSM-EFR 구현 결과

	부호화기	복호화기	전체
Program ROM	5204 words	1856 words	7060 words
Data RAM	6394 words	823 words	7217 words
Data ROM	4842 words	127 words	4969 words

실시간으로 구현한 GSM-EFR 음성 부호화기의 객관적 성능 평가를 위해 G.721 (32 kbps ADPCM)과 EVRC (Enhanced Variable Rate Codec) 음성 부호화기의 모의실험 결과를 비교하여 표 4에 나타내었다. 객관적 평가 톨로 널리 사용되는 것으로는 SNR과 segSNR, CD (Cepstral Distance), BSD(Bark Spectral Distance) 등이 있으나, 최근의 음성 부호화기 들은 파형 부호화기와는 달리 파형 매칭 보다는 지각적 성능에 중점을 두고 개발되었기 때문에 SNR과 같은 객관적 톨에 의한 평가는 적당하지 않다. 따라서 본 논문에서는 최근에 제안된 지각 평가 방법인 MNB (Measuring Normalizing Block) 알고리즘[8]을 사용하여 그림 2와 같이 객관적 평가를 수행하였다. 표

4에서 보면, 코딩 SNR이 실제 주관적인 평가와는 달리 상당히 저하됨을 알 수 있다. 그러므로 이러한 SNR 측정법은 최근의 음성 부호화기들의 객관적 평가 툴로 사용하기에 적합하지 않다. MNB 알고리즘을 사용한 지각 평가에서는 GSM-EFR이 EVRC에 비해 좋은 성능을 보이며, ADPCM과도 비슷한 성능을 보이는 것을 알 수 있다. 이를 통해 EVRC 음성 부호화기에 비해 GSM-EFR이 좀 더 지각적 특성을 잘 반영하는 것을 알 수 있다.

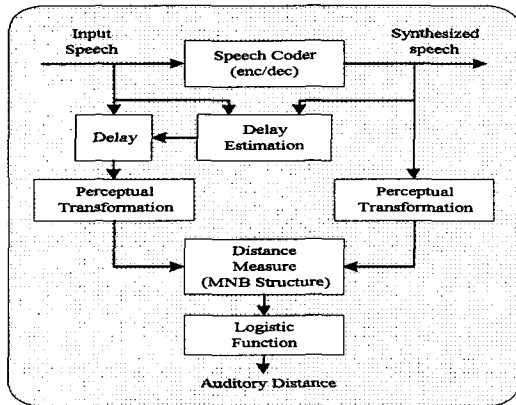


그림 3. MNB 알고리즘의 기본 블록도

표 4. Perceptual Measure (PM) 및 SNR 측정 결과

	PM (0~1.0)			SNR (dB)		
	G.721	EFR	EVRC	G.721	EFR	EVRC
Male1	0.814	0.832	0.701	16.115	4.142	-1.121
Male2	0.824	0.828	0.689	16.575	4.125	-1.960
Male3	0.829	0.831	0.676	18.397	3.359	-1.568
Female1	0.838	0.837	0.738	21.815	10.487	-3.450
Female2	0.794	0.815	0.630	23.213	7.233	-1.983
Female3	0.819	0.816	0.692	21.384	7.907	-3.046
Average	0.820	0.827	0.690	19.583	6.209	-2.188

#### IV. 결론

본 논문에서는 비동기 방식의 IMT2000 이동통신 단말기에서 사용되는 GSM-AMR 음성 부호화기의 최상위 전송률 모드인 GSM-EFR 음성 부호화기를 OakDSP Core에서 실시간으로 구현하였다. 구현된 GSM-EFR 음성 부호화기에 소요된 계산량은 부호화기와 복호화기에 각각 22.187 MIPS와 1.844 MIPS이었으며, 필요한 메모리 크기는 코드메모리가 7.06 Kwords, 데이터메모리가 12.19

Kwords였다. 실험 결과, 구현된 GSM-EFR 음성 부호화기는 ETSI에서 제공하는 시험 벡터 샘플을 모두 통과하였으며, MNB 알고리즘을 이용한 32 kbps ADPCM, EVRC 음성 부호화기와의 지각 평가 비교를 통해 GSM-EFR 음성 부호화기의 우수한 성능을 확인하였다. 개발된 GSM-EFR 음성 부호화기는 IMT-2000 비동기 방식 음성 부호화기의 표준인 GSM-AMR 음성 부호화기의 개발의 기본적인 구조로 사용될 예정이다.

#### 참고 문헌

- [1] M. R. Schroeder, and B. S. Atal, "Code-Excited Linear Prediction (CELP): High Quality Speech at Very Low Bit Rates," *IEEE Proc. Int. Conf. Acoust. Speech and Signal Proc.*, pp.25.1.1-25.1.4, 1985
- [2] ITU-T Rec. G.723.1, *General Aspects of Digital Transmission Systems: Dual Rate Speech Coder For Multimedia Communications Transmission at 5.3 and 6.3 kbit/s*, March 1996.
- [3] ITU-T Recommendation G.729, *Coding of Speech at 8 kbit/s Using Conjugate-Structure Algebraic-Code-Excited Linear Prediction (CS-ACELP)*, March, 1996.
- [4] ETSI Draft prETS 200 726, *Digital Cellular Telecommunications System; Enhanced Full Rate (EFR) Speech Transcoding (GSM 06.60 Version 5.0.0)*, March, 1996.
- [5] J. P. Adoul, P. Mabilieu, M. Delprat, and S. Morissette, "Fast CELP Coding Based on Algebraic Codes," *IEEE Proc. Int. Conf. Acoust. Speech and Signal Proc.*, pp.1957-1960, 1987.
- [6] GSM 06.60 (ETSI EN 300 726), *EFR Speech Transcoding*, ETSI, 1998
- [7] GSM 06.54 (ETSI EN 300 725), *Test sequences for the GSM EFR speech codec*
- [8] Stephen Voran, "Objective Estimation of Perceived Speech Quality - Part I: Development of the Measuring Normalizing Block Technique," *IEEE Transactions on Speech and Audio Processing*, Vol.7, No.4, July 1999