http://dx.doi.org/10.7236/JIIBC.2013.13.6.55
JIIBC 2013-6-7

음성 합성용 저전력 고음질 부호기/복호기 설계 및 구현

Design and Implementation of the low power and high quality audio encoder/decoder for voice synthesis

박노경*, 박상봉**, 허정화**

Nho-kyung Park, Sang-bong Park, Jeong-hwa Heo

요 약 본 논문은 음성합성에서 사용되는 오디오 부호기/복호기 설계 및 구현을 기술한다. 설계된 회로는 원래 음성 샘플대신에 연속되는 음성 샘플의 차를 부호화하는 방식으로 압축율은 4:1 이다. FPGA를 이용해서 각각의 기능을 검증하고, 0.35㎛ 표준 CMOS 공정을 이용하여 칩으로 제작해서 성능을 측정하였다. 시스템 클럭 주파수는 16.384㎒ 를 사용한다. THD(Total Harmonic Distortion)+n은 주파수에 따라서 -40㏈에서 -80㏈ 값을 지니고, 전력 소모는 전원 전압 3.3V에서 80㎽로써, 고음질과 저전력 소모를 요구하는 모바일 응용에 적합하다.

Abstract In this paper, we describe design and implementation of audio encoder/decoder for voice synthesis. It uses the encoding of difference value of successive samples instead of the original sample value. and has the compression ratio of 4. The function is verified by using FPGA and the performance is measured by the fabricated chip using 0.35㎛ standard CMOS process. The system clock is 16.384㎒. The measured THD+n is from -40㏈ to -80㏈ with frequency variation and the power consumption is about 80㎽. It is suited for the mobile application of high audio quality and low power consumption.

Key Words : Audio Encoder/Decoder, Compression, ADPCM, Voice Synthesis

Ⅰ. 서 론

음성 합성은 인간의 소리인 음파를 기계가 자동적으로 생성하는 것으로, 사람의 소리를 녹음하여 음성 단위로 분할한 다음 부호를 붙여 필요한 음성 단위만을 소리로 만들어내는 기술이다. 음성 합성 기술은 디스플레이가 없는 정보통신 기기나 시각적인 핸디캡을 가진 사용자를 위한 효율적인 정보 제공 수단으로, 최근에는 지능형 로봇이나 전자 기기들의 송수신 수단으로 널리 사용되고 있다. 과거에는 음성 합성 방법론이 출현하면서 음성 출력 서비스 대부분이 미리 녹음된 안내 멘트를 재생해 주는 서비스가 주를 이루었으나, 최근에는 ARS (Automatic Response System), VMS(Voice Mailing System), UMS (Unified Messaging System) 서비스를 중심으로 서버형 음성합성 기술이 폭넓게 사용되면서 원격지 서버의 고성능 컴퓨팅 장치 및 거대 메모리를 활용하여 고품질의 합성음을 생성하여 서비스를 제공하고 있다.[1]

음성합성 방식은 알고리즘 상으로는 간단하지만, 실제 구현 시에는 여러 가지를 고려해야 한다. 소리 나는 장난감 인형처럼 한번 만들어져 칩 화되면 그 내용을 수정할

*정회원, 호서대학교 정보통신공학과
**정회원, 세명대학교 정보통신학과
접수일자 2013년 10월 8일, 수정완료 2013년 11월 9일
게재확정일자 2013년 12월 13일

Received: 8 October, 2013 / Revised: 9 November, 2013
Accepted: 13 December, 2013
*Corresponding Author: Nhokyung Park: nkpark@hoseo.edu
Dept. Information Communication Engineering, Hoseo University

수 없다. 또한 녹음되는 문장 수와 압축방법도 잘 고려해야 한다. 장치에서 사용할 문장 수가 많은 경우 압축하지 않고 저장해 놓으면 많은 저장 공간을 필요로 하므로 저장 공간 비용이 높아진다. 반면에 음성 데이터를 고압축 방법을 사용하여 압축을 하면 저장 공간은 줄일 수 있으나, 녹음된 내용을 재생해야 하는 경우 압축의 복원을 수행해야 한다. 일반적으로 고압축 방법은 압축을 풀 때에도 복잡한 연산을 수행하게 되므로 해당 칩이나 장치에 복잡한 코덱이 들어가고 연산량도 많아지므로 그 만큼 전력 사용량도 증가 하게 된다.[2]

본 논문에서는 파형 부호화 방식으로 음성 샘플을 부호화 하는 대신 음성 샘플과 예측된 샘플 사이의 차이를 부호화 하여 회로를 간단히 하여 전력 소모량을 줄이고 고음질을 얻도록 구현하였다. 본 논문의 Ⅱ장에서는 음성 합성 방식의 구조에 대해 설명하고, Ⅲ장에서는 제안된 알고리즘 및 구조에 대해 설명하고, Ⅳ장에서는 시뮬레이션 결과와 테스트 결과를 보여주고, 마지막으로 Ⅴ장에서 결론을 기술한다.

Ⅱ. 음성 합성 방식 구조

음성합성방식에는 3가지 방식이 있다. 파형부호화 방식과, 파라미터 부호화 방식, 혼합 부호화 방식이다. 파형부호화 방식은 음성파형을 그대로 부호화하는 방식이며, 이 파형부호화 방식 중에서 PCM(Pulse Code Modulation), ADPCM(Adaptive Differenctial Pulse Code Modulation)은 시간 도메인 형식이고, SBC(Sub Band Coding)같은 경우는 주파수 도메인 형식이다. 음성합성 방법에는 장・단점이 있지만, 최근 음성합성 녹음재생 방법에서는 파형부호화 방식이 많이 사용되고 있는데 다른 부호화 방식보다 간단하고, 전력 소모량이 적고 음질도 만족할 만큼 기술이 발전하였다.

1. SBC(Sub Band Coding) 방식

입력 신호인 아날로그 신호를 주파수 대역 상에서 다수개의 주파수 대역으로 분리하고, 그 후에 각 아날로그 신호에 대해 각각 PCM 또는 ADPCM을 사용해서 코딩하는 방식이다. 각 주파수 대역으로 분할되어 부호화된 정보는 상대적으로 ADPCM으로만 부호화 하는 방법보다 전송상 에러에 강인하기 때문에 부가정보에 대하여만 선택적으로 에러 정정부호를 사용하면 정보 전송 율이 크게 증가하지 않으면서 효율적으로 전송 선로상의 에러 문제점을 해결할 수 있다. 그러나 분할 대역의 수가 증가할수록 특성이 좋은 QMF(Quadrature Mirror Filter)를 사용해야 하므로 전송시간의 지연 및 계산 량의 증가로 인한, 연산량, 전력면에서 모바일 또는 임베디드 음성 합성 분야 응용에 한계가 있다.

2. ADPCM 방식

진폭 값으로 표현하지 않고, 이전단과의 차이점만 부호화를 할 경우, 정보를 줄일 수가 있으므로 부호화되는 비트의 수를 감소시킬 수 있는 알고리즘이다. ADPCM 방식은 음성신호가 상관성이 큰 특성을 이용한 방법으로 과거의 음성신호의 샘플을 기준으로 다음에 들어올 신호의 크기를 예측하는 방식이다. 기본적인 구조로는 적응 양자화기와 적응 예측기로 구성되어 있다. 적응 양자화기는 음성신호의 통계에 맞게 비 균일 특성을 갖도록 설계되며, 적응 예측기는 음성신호의 통계적 특성을 이용한 예측과정을 통하여 양자화기 입력신호를 효율적으로 줄이면서 계산 량이 많지 않도록 구현할 수 있다. 순방향 적응방식은 입력신호로부터 직접 변화를 예측하므로 역방향 적응방식보다 예측성능은 우수하지만 예측기에 대한 정보를 추가적으로 전송해야 하므로 비트의 효율성은 그만큼 감소된다. 이전단과 다음단의 차이점으로만 부호화를 하는 것이 아니라, 그 이전단의 값에서 다음 단에 들어올 신호의 크기를 예측하여 예측 표본 값과 실제 표본 값과의 차이를 부호화한다. PCM은 전체의 값을 저장하는 반면에 ADPCM은 샘플간의 차이값만 저장하는 방식을 보여주고 있기 때문에 저장 공간 면에서 더 효율적이다.[3]

Ⅲ. 제안된 알고리즘 및 구조

본 논문에서 사용된 ADPCM은 4:1 압축 기법으로 압축된 4비트의 음성이나 소리를 저장하고 있다가, 압축된 음성을 다시 복원 시키는 방식이다. 매 음성 샘플이 입력될 때마다 입력 신호의 예측치를 구해 이를 원 신호에서 빼주면 입력신호의 예측 오차를 구하여 이 예측 오차를 양자화 한다. 이 오차 신호를 적응적으로 양자화 하여 부호화기에서 출력한다. 이 값으로부터 다시 양자화 된 오차

신호를 규정된 테이블에 의해 구한다. 양자화 된 오차신호에 예측치를 더해 양자화 된 재생 신호를 만든다. 앞에서 구한 양자화 된 오차 신호와 양자화 된 재생 신호는 다음 입력 신호가 입력되었을 때 예측치를 계산하는 과정에서 쓰이며 동시에 예측 계수들이 업데이트된다. 그림 1은 본 논문에서 제안하는 부호화기와 복호화기 블록도이다.

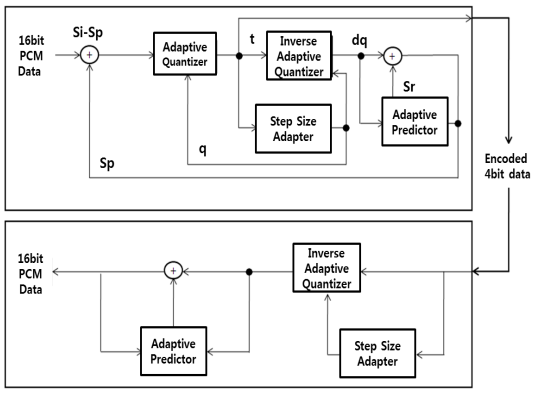

그림 1. 제안한 오디오 부호기/복호화기 블록도
Fig. 1. The block diagram of the proposed ADPCM

부호화기 블록의 음성압축은 16비트의 선형 PCM 데이터가 입력으로 들어가고, 입력 값과 이전의 입력 값과의 차이 값이 적응 양자화기를 거쳐 4비트의 압축된 값으로 나오게 되는 것이다. 압축된 4비트 값은 복호화기 블록에서 적응 양자화블록과 적응 예측기 블록을 거쳐서 다시 16비트의 선형 PCM값이 출력된다.

부호화기 블록은 4개의 세부 블록으로 구성되어 있다. 16비트 선형 PCM 입력 값 Si와 이전에 입력 값 Sp와의 오차 값이 적응 양자화기의 입력이다. 적응 양자화기에서는 16비트의 값을 양자화 과정을 거쳐 최종 4비트의 부호화된 값을 출력한다. 이 4비트의 값 t가 역 적응 양자화기의 입력으로 들어가서 스텝 사이즈 값과 다시 역 양자화를 실행하여 나오는 출력 값과 적응 예측기 블록에서 나오는 값이 더해져서 Sp의 값이 결정된다. 각각의 블록 기능은 다음과 같다.

가. 적응 양자화 블록

적응 양자화블록은 실제적인 양자화를 수행한다. 이 적응 양자화블록에서는 처음에 원샘플인 Si의 값에서 적응 예측기 블록에서 나온 이전 샘플인 Sp의 값을 빼줌으로서 차이 값인 difference 값을 구한다. difference 값이 양수인지 음수인지를 확인해서 부호화기에서 출력되는 압축된 4비트인 newsample의 부호비트가 결정된다. 양수이면 '0', 음수이면 '1'이 되고, difference 값에 절대 값을 취한다. newsample의 나머지 3비트는 difference값과 스텝사이즈 값의 크기를 비교한 다음 조건이 만족할 시에는 아래 식 (1)과 같은 공식을 수행하고, 만족하지 않는 경우에는 스텝사이즈 값을 우측으로 1비트 쉬프트 시킨 값으로 대신한다.

$$\text{newsample}\ [2:0] = 4 * (\text{difference}/\text{stepsize}) \qquad (1)$$

양자화 과정은 처음에 샘플 값이 양수인지 음수인지 확인해서 음수 일 때는 부호비트인 MSB를 0으로하고 양수일 때는 1로 하고 샘플값의 절대 값을 취한다. 절대 값을 취한 샘플 값과 스텝사이즈의 값과 비교를 해서 크면 1, 작으면 0을 3번째 비트인 BIT2값에 대입한다. BIT1의 값은 샘플 값과 스텝사이즈를 2로 나눈 값과의 크기 비교를 통해서 구한다. BIT1은 다시 스텝사이즈를 4로 나눈 값과 크기 비교를 통해서 구해서 마지막 BIT0까지 총 4비트의 부호화된 출력 값을 생성한다.

나. 적응 역양자화 블록

역양자화 블록은 양자화 된 신호를 다시 역양자화를 통하여 양자화 되기 이전의 형태로 바꾸어주는 역할을 한다. 양자화를 거치기 전의 신호를 가지고 예측을 하지 않고, 역양자화를 거친 신호로 연산을 하는 이유는 다음 신호에 대한 예측은 양자화의 결과로 나온 신호를 가지고 그 결과 값들의 변화가 어떠한지에 따라서 이루어지기 때문이다. 역양자화는 양자화와 같이 역양자화 테이블에 따라서 4비트 크기의 신호에서 16비트 크기의 값을 생성하고, 양자화 크기 값을 더하여 이전 양자화값의 보정을 위해서 뺀 것으로 이루어진다. 역양자화과정은 식 (2)로 나타낼 수 있다.

$$\text{difference} = 0$$
$$(\text{newsample} + 1/2) * \text{stepszie}/4 = \text{newsample} * \text{stepsize}/4 + \text{stepsize}/8 \qquad (2)$$

다. 스텝 사이즈 블록

스텝사이즈 블록은 부호화기 블록과 복호화기 블록에서

사용한다. 스텝사이즈 블록은 작거나 큰 샘플사이의 차이를 나타낸다. 초기 값은 리셋 기간에 설정한 값이다. 현재 스텝사이즈 값이 n이라 하면 다음의 스텝사이즈 값인 n + 1은 식 (3) 과 같다.

표 1. 스텝 사이즈 테이블 1
Table 1. Step Size Table 1

newsample	M(L(n))	newsample	M(L(n))
1111	+8	0111	+8
1110	+6	0110	+6
1101	+4	0101	+4
1100	+2	0100	+2
1011	-1	0011	-1
1010	-1	0010	-1
1001	-1	0001	-1
1000	-1	0000	-1

표 2. 스텝 사이즈 테이블 2
Table 2. Step Size Table 2

No.	StepsizeTable Value	No.	StepsizeTable Value
1	7	46	544
2	8	47	598
3	9	48	658
4	10	49	724
5	11	:	:
:	:	89	32767
45	494		

$$stepsize(n + 1) == stepsize(n) * 1.1M(L(n)) \quad (3)$$

위 식은 2개의 테이블표를 사용해야한다. 첫 번째는 표 1에서 보여주는 부호 절대 값방식이 쓰인 테이블표로서 인덱스의 값을 결정하는데 쓰인다. 표 2는 결정된 인덱스 값이 가리키는 값을 가지고 새로운 스텝사이즈의 값을 결정하게 된다.

라. 적응 예측기 블록

적응 예측기 블록은 이전의 적응 역양자화기에서 만들어낸 양자화 된 차등신호와 회귀신호로부터 다음신호를 예측한다. 회귀신호는 현재의 예측신호와 양자화 된 차등신호의 합으로 산출한다. 적응 예측기 블록에서는 식(4)와 같은 과정을 거친다. 여기서는 계산된 difference 값을 기반으로 예측된 값인 predictsample 값을 구한다.

```
if(newsample & 8)
    difference = | difference;
predictsample + = difference;                    (4)
```

predictsample 값이 스텝사이즈 테이블에 나와 있는 최대치 값인 32767값을 넘을 경우에는 32767 값을 사용하고 최하 값인 -32768을 넘어갈 경우에는 -32768 값을 사용한다. 인덱스는 스텝사이즈 테이블의 위치를 알려는 것이기 때문에 최대치가 88이고 최소치가 0인 이 범위를 초과하지 않도록 설계하였다.

그림 2는 제안한 부호기의 타이밍도이다. 시스템 메인 클럭은 6.384MHz이고, 리셋신호가 1일 때 활성화된다. sample_start는 입력이 들어 올 때마다 1로 활성화된다. 입력 값인 D_IN [15:0]은 병렬 값으로 들어온다. 부호기 블록의 모든 동작은 클록 사이클이 총 12번이 지났을 때 완료된다. 따라서 출력 값인 D_OUT[3:0]은 입력이 들어오고 나서 13번째 클럭에서 출력된다. 그림 3은 ADPCM 복호기 타이밍도이다. sample_start는 입력이 들어 올 때마다 1로 활성화가 된다. 입력 값인 D_IN [3:0]은 병렬 값으로 들어온다. 그리고 복호기 블록의 클록 사이클이 9 클럭이 필요하고 출력 값은 10번째 클럭에서 출력된다.

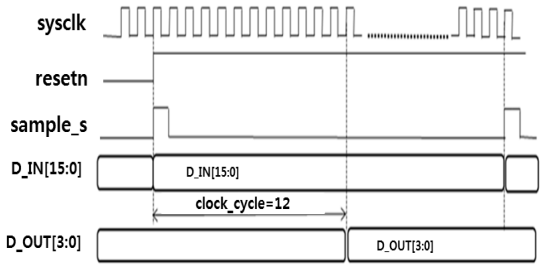

그림 2. 제안한 부호기 타이밍도
Fig. 2. The timing diagram of the proposed Encoder

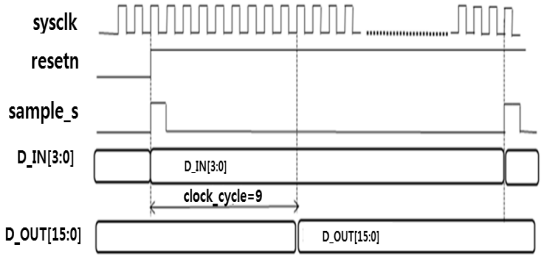

그림 3. 제안한 복호기 타이밍도
Fig. 3. The timing diagram of the proposed Decoder

Ⅳ. 시뮬레이션 결과 및 테스트 결과

Cooledit-Pro 툴을 이용해 디지털 값을 입력으로 넣어서 입력된 사인 파형에 대한 출력된 디지털 값과 비교 확인했다. Verilog HDL을 사용하여 설계된 디지털 블록에 대한 시뮬레이션 및 논리 합성 수행 후 0.35um 표준 CMOS 공정을 사용하여 칩으로 제작되었다.

1. 시스템 시뮬레이션 테스트 결과

여자 음성과 음악, 그리고 톤(Sine wave) 값을 넣어서 부호기/복호기 블록을 거쳐 나오는 결과 값과 원래 입력 값을 비교하였다. 그림 4는 샘플 레이트 16kHz 육성 음성 원본 파형이며, 그림 5는 복원 파형이다.

그림 6과 7은 샘플 레이트가 4kHz이고, 주파수가 1kHz 인 사인파에 대한 시뮬레이션 파형이다.

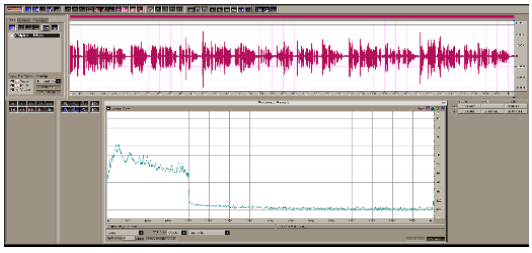

그림 4. 16k 음성 원본 파형
Fig. 4. The waveform of 16k voice source

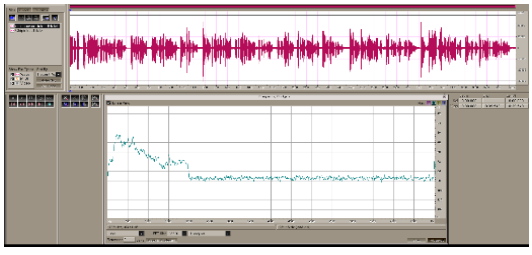

그림 5. 복원된 파형 (16k)
Fig. 5. The restored waveform of 16k

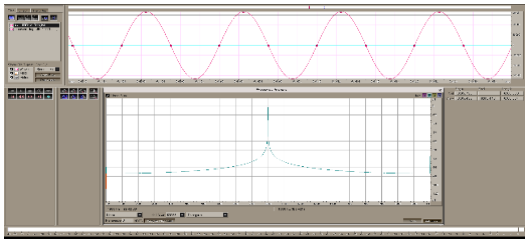

그림 6. 1kHz 사인파
Fig. 6. The waveform of 1kHz sine wave

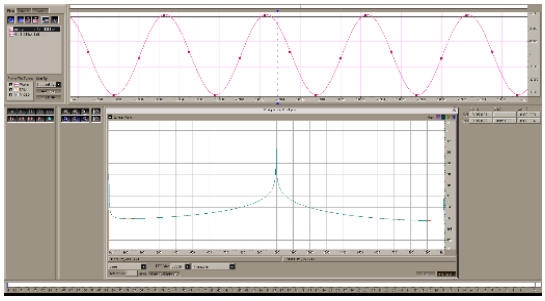

그림 7. 복원된 1kHz 사인파
Fig. 7. The restored waveform of 1kHz sine wave

2. FPGA 테스트 결과

본 논문에서는 FPGA를 이용하여 테스트 환경을 구성하고, Logic Analyzer를 이용하여 디지털 데이터 값을 확인하고, LRCLK와 BCLK를 사용해서 오디오 출력 단으로 음성 및 소리를 확인하였다. 입력으로는 아날로그 신호가 들어와서 FPGA에 있는 16비트 AD컨버터를 거쳐 디지털 값으로 변환이 되고, 이 16비트 디지털 값이 ADPCM에 들어가서 부호화와 복호화를 실행한다. 복호화를 거친 복원된 16비트 디지털값이 FPGA에 있는 16비트 DAC를 지나서 오디오 출력으로 재생된다. 그림 8은 FPGA를 이용한 테스트 환경과 Logic Analyzer를 이용한 출력 파형이다.

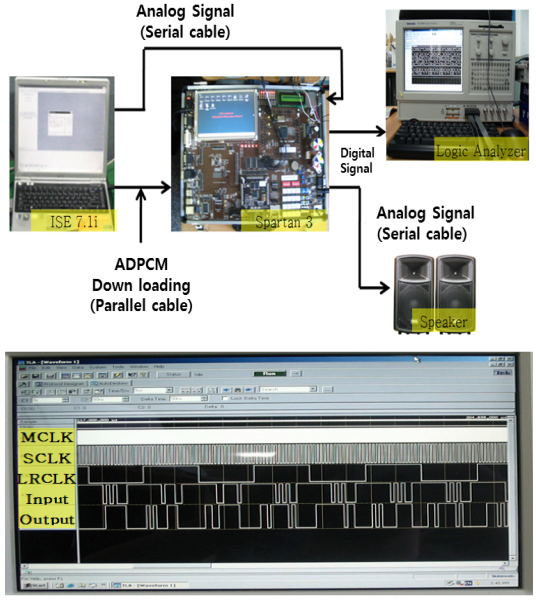

그림 8. FPGA 테스트 시스템 및 출력 파형
Fig. 8. FPGA test system and measured result

3. 레이아웃 및 칩 제작

그림 9는 레이아웃 그림과 제작된 칩 사진이다. 제작된 칩 면적은 4.35mm * 4.35mm이다. 전원 전압은 3.3V이고, 시스템 클럭은 16.384MHz 이다.

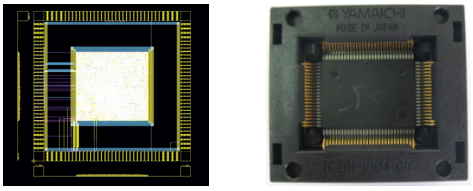

그림 9. 레이아웃 과 칩 사진
Fig. 9. Layout plot and fabricated chip

V. 결 론

본 논문에서는 모바일 환경에서 적용 가능한 음성 합성용 음성 복호화/부호화기를 구현하였다. 삼성 0.35um 표준 CMOS 2-poly-4-metal 공정을 이용하여 칩을 제작하였으며, 테스트 결과 원하는 클럭 주파수인 16.384 MHz에서 정상 동작함을 확인할 수 있었다. THD (Total Harmonic Distortion)+n 은 주파수에 따라서 -40 dB에서 -80dB 값을 지니고, 전력 소모는 전원 전압 3.3V에서 80㎽ 로 측정되었다. 기존 DPCM이나 SBC 음성 압축 방식보다 본 논문에서 제안하는 ADPCM 음성 압축 방식이 계산 량이 적고 복잡도나 칩의 면적이 작아 저전력, 고음질을 요구하는 모바일 분야의 음성 합성 IC에 적합하다고 사료된다.

References

[1] H. Han, “Variable Quad Rate ADPCM for Efficient Speech Transmission and Real Time Implementation on DSP”, Journal of Korean Institute of illuminating and Electrical Installation Engineers, vol. 18, No 1, pp. 129-136, January 2004.

[2] S. Y. Min, D. S. Na, “A Study on Implementation of Emotional Speech Synthesis System using Variable Prosody Model”, Journal of the Korea Academia-Industrial cooperation Society, v.14, no.8, pp.3992-3998, 2013.

[3] Bluetooth Audio Video Working Croup, “Bluetooth Specification: Advanced Audio Distribution Profile, Bluetooth SIG Inc. 2002 Hoc Networks”, IEEE Computer, Feb. 2004

[4] D.Hermann, R.L. Brenna, H.Sheikhzad, E.Cornu, “Low-Power implementation of the bluetooth subband audio codec”, IEEE, 2004.

[5] J. S. Choi, “Speech Synthesis Algorithm Applied to Methods of Multi-Cepstrum Extraction and Root Mean Square Amplitude”, Journal of Korean Institute of Information Technology, vol. 11, issue 6, pp. 157-162, June 2013.

[6] K. H. Han, “Coding Method of Variable Threshold Dual Rate ADPCM Speech Considering the Background Noise”, Journal of Korean Institute of illuminating and Electrical Installation Engineers, vol. 17, No 6, pp. 154-159, November 2003.

저자 소개

박 노 경(정회원)

- 1990년 2월 : 고려대 전자공학과 공학박사
- 1985년 3월 ~ 1989년 2월 : 삼성반도체 반도체설계연구소 연구원
- 1985년 3월 ~ 현재 : 호서대학교 정보통신공학과 교수

<관심분야 : SOC 설계, 임베디드 시스템 설계, 신호처리 >

박 상 봉(정회원)

- 1992년 2월 : 고려대 전자공학과 공학박사
- 1992년 3월 ~ 1992년 2월 : 삼성전자 선임 연구원
- 1999년 3월 ~ 현재 : 세명대학교 정보통신학부 부교수
- 2000년 7월 ~ 현재 : @lab(주) Digital 설계팀 기술고문

<관심분야 : RFID/USN 기술, ASIC 설계, 광센서, 멀티터치>

허 정 화(정회원)

- 2009년 2월 : 세명대학교 전산정보학과 이학 박사
- 2003년 3월 ~ 현재 : 세명대학교 시간강사

<관심분야 : ASIC 설계, 신호처리, ADC/ DAC, Multi-Touch>