

RELP 방식을 이용한 디지털 음성 응답기

(A Digital Audio Response System Based on the RELP Algorithm)

金尚龍*, 殷鍾官*, 李愷洙*, 鮮于明勳**, 鄭信一***
 (S. R. Kim, C. K. Un, H. S. Lee, M. H. Sunwoo and S. I. Jung)

要 約

본 논문에서는 디지털 자동 음성 응답장치의 개발에 관하여 전반적인 사항을 기술하였다. 개발된 디지털 음성 응답 장치는 전화국에서 가입자가 전화번호를 문의하였을 때 자동 응답할 수 있도록 특별히 구성된 시스템이다. 본 시스템의 구현 algorithm으로는 pitch predictive loop(PPL)을 가지는 RELP(residual excited linear prediction)방식을 사용하였는데 system memory는 비교적 적은 반면 음질은 아주 좋은 것이 개발된 자동 응답기의 큰 장점이라 하겠다. Hardware는 bit-slice microprocessor를 사용한 음성 합성기와 controller 및 I/O로 이루어져 있는데 이들은 실시간 신호처리와 시스템의 적응성 및 신뢰성을 고려하여 설계하였다.

Abstract

This paper describes the overall procedure of the development of a digital audio response system. It has been developed specifically as an answering system to the inquiries of telephone numbers from subscribers. The system has been realized based on the residual excited linear prediction (RELP) algorithm that incorporates a pitch predictive loop. Its major advantage over other similar systems is that it produces high quality of synthetic speech, although its memory size is relatively small. The hardware which consists of a speech synthesizer, a controller and an I/O part has been constructed using 2900 series bit-slice microprocessors and an INTEL 8085 microprocessor. The system is capable of real time processing, reliable, and adaptable to other applications.

I. 序 論

최근에 컴퓨터나 로봇과 같은 복잡하고 많은 기능을 가진 시스템이 발달함에 따라 인간과 기계사이의 통신 능력을 보다 효율적으로 높일 수 있는 음성 통신에 대한 연구가 활발히 진행되고 있다. 인간과 기계와의 통신은 음성 인식부분과 합성부분으로 나누어지는데 음성 인식은 기계 즉, 컴퓨터가 인간의 음성을 인식하는 것이고 음성 합성은 인간의 명령에 의해 컴

*正會員, 韓國科學技術院 電氣 및 電子工學科
 (Dept. of Electrical Eng., KAIST)

**正會員, 韓國電氣通信研究所
 (Korea Electrotechnology Telecommunication
 Research Institute)

***正會員, 韓國放送公社 技術研究所
 (Korea Broadcasting System)

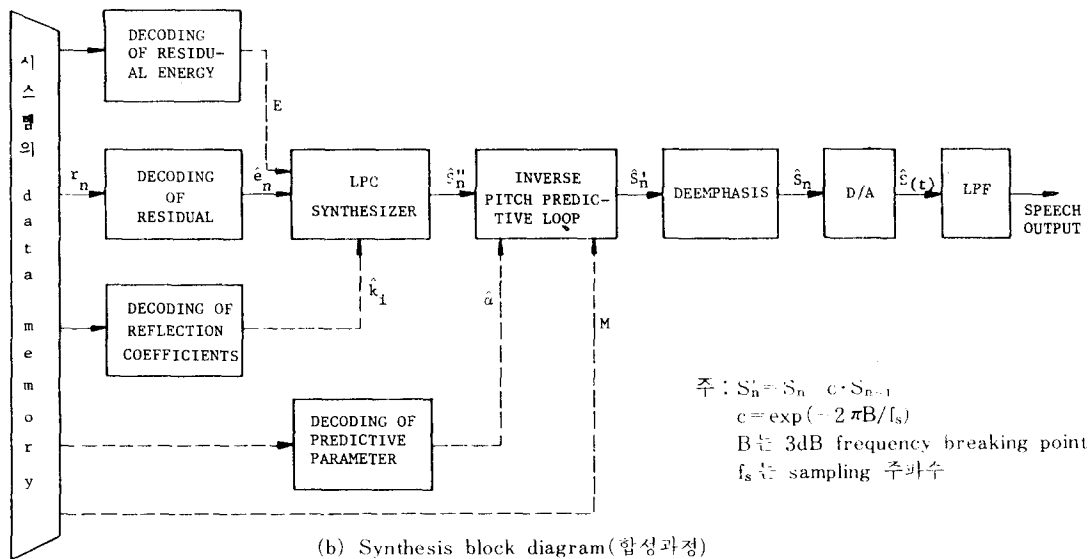
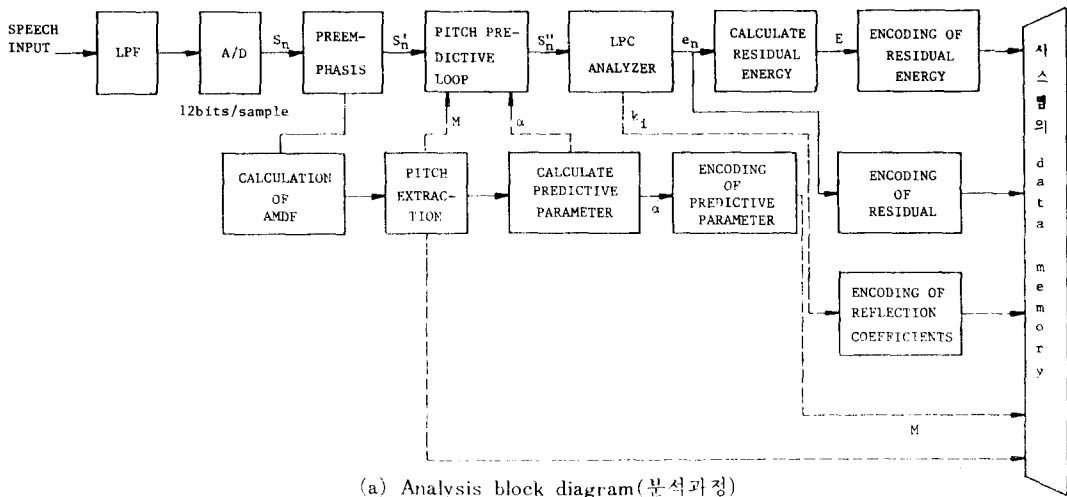
接受日字: 1984年 3月 29日

퓨터가 자동적으로 어휘 조합을 수행하여 인간에게 언어로 사실을 전하도록 하는 것이다.

1930년대 초에 analog 테이프 recoder를 사용한 음성 응답기가 처음 개발된 이래¹⁾, 최근에는 이 시스템이 디지털화되어 광범위한 분야에 걸쳐서 제작 응용되고 있다. 본 논문에서 연구 개발된 디지털 음성 응답기는 전화 가입자가 전화번호를 문의하였을 때 전화번호와 가입자 정보가 저장된 컴퓨터로부터 가입자 정보를 받아 이를 직접 음성으로서 응답할 수 있는 기능을 가진 특수 용도의 시스템이다.

이와 같은 음성 응답기에서는 음성을 디지털화하는데 있어서 음성의 intelligibility와 naturalness는 시스템의 성능 척도로서 중요한 고려 대상이 되며 어는 방

식을 채택하는가에 따라 이들 성능이 크게 달라지기 때문에 디지털화 방식을 택할 때는 여러 가지 면을 고려하여야 한다. 통상 음성의 디지털화 방식으로는 pulse code modulation(PCM) 또는 delta modulation(DM)과 같이 비교적 음질의 우수함이 보장되는 파형양자화 방법²⁾, linear predictive coding(LPC) vocoder 등 bit rate가 낮은 vocoding 방법³⁾, 또는 이 두 방식의 장점을 택한 residual excited linear prediction(RELP) vocoder⁴⁾ 등 hybrid coding으로 나눌 수 있다. 이 중 RELP vocoder 방식은 bit rate도 낮을 뿐 아니라 음질도 좋은 장점을 지니고 있어 본 연구에서 개발된 자동 음성 응답기는 RELP 방식을 채택하여 구현하였다.



주 : $S_n^i = S_n - c \cdot S_{n-1}$
 $c = \exp(-2\pi B/f_s)$
 B는 3dB frequency breaking point
 f_s 는 sampling 주파수

그림 1. RELP 방식에 의한 음성분석 및 합성과정
 Fig. 1. Speech analysis and synthesis by the RELP algorithm.

이 논문에서 연구된 과정은 음성 디지털화 방식, parameter coding 방법, 언어분석 및 조합, 디지털 여파기, microprogramming 및 컴퓨터 interface 등인데, 특히 입력 신호의 dynamic range를 줄여 음질을 보다 효율적으로 높일 수 있는 PPL을 갖는 REL P algorithm에 대해 연구되었다.

본 서론에 이어 제 II 장에서는 음성 해석 및 부호화 방식을, 제 III 장에서는 시스템의 컴퓨터 simulation 과 결과를, 제 IV 장에서는 시스템 구현에 관하여 자세히 기술하였으며 끝으로 V 장에서는 결론을 맺었다.

II. 음성 해석 및 부호화 방식:

본 디지털 음성 응답장치의 기본 algorithm은 4.8 - 16kbps의 전송 속도에서 좋은 음질의 음성을 만들어 낼 수 있는 REL P 방식을 채택하였다. 본 시스템에서는 원래의 REL P 방식을 개선하여 음성을 분석하기 전에 몇 단계의 preprocessing을 거치도록 하였으며, 음성 합성후에도 preprocessing의 역과정인 postprocessing을 거쳐 음성을 합성하도록 하였다. 사용한 REL P 방식의 상세한 block도는 그림 1에 도시하였다. Block도에 나타난 각 과정은 다음과 같다.

1. Preprocessing

Preprocessing은 우선 A/D 변환기를 거친 음성 신호를 preemphasis한다. 음성신호에 preemphasis를 행하면 음성신호의 energy가 저주파 대역에서 감소하고 고주파 대역에서는 증가되어 음성 spectrum의 dynamic range가 감소된다. 따라서 동일 bit수로 음성을 coding할 경우 음질상의 이득을 볼 수 있으며 역으로 동일 성능의 음질을 얻는데는 coding시 필요한 bit 수를 줄일 수 있다.^{[5][6][7]}

이외에도 preemphasis를 행함으로써 hardware implement시 적은 bit 수로 동일 성능을 얻을 수 있어 hardware cost를 감소시킬 수 있다.^[8]

두번째 과정으로 PPL을 구성하기 위하여는 pitch 주기(M)을 검출해야 한다. Pitch 추출 algorithm으로는 시간 영역 방법^[9], cepstrum 방법과 같은 spectrum 분석 방법^[10]과 autocorrelation 방법^{[11][12]} 등 여러가지 방법이 있는데 본 논문에서는 autocorrelation 방법의 일종인 average magnitude difference function (AMDF)^[13] 방법을 사용하였다.

AMDF 방법은 correlation 함수를 계산할 때 승산 (multiplication)을 하지 않고 가감의 계산으로만 하기 때문에 계산시간이 일반 autocorrelation 방식보다 빠른 반면 성능은 거의 같기 때문에 real time hardware

구현에 잇점이 있다.^{[14][15]}

다음으로 PPL을 사용하는 목적은 입력신호의 dynamic range를 줄여 residual 신호의 coding을 효율적으로 하여 bit 수를 감소시키는데 있다. 기본 원리는 임의의 한 sample을 pitch 주기만큼 이동시켜서 PPL parameter α 를 곱한후 원래의 sample에서 그 값을 빼는 것이다. Preemphasis된 음성 신호를 S'_n 이라 하면 PPL 출력 S''_n 는 다음과 같은 수식으로 표현된다.

$$S''_n = S'_n - \alpha \cdot S'_{n-M} \tag{1}$$

또는 Z-영역에서

$$S''(Z) = (1 - \alpha \cdot Z^{-M}) S'(Z) \tag{2}$$

로 주어지는데 여기에서 PPL parameter α 의 수식표현은 다음과 같다.

$$\alpha = \frac{\sum_{i=1}^N S'_i \cdot S'_{i-M}}{\sum_{i=1}^N |S'_{i-M}|^2} \tag{3}$$

여기서 N은 한 frame 당 sample 수(본 system에서는 180 sample임)이다. 위에서 설명한 원리에 의해 PPL을 구성하면 그림 2와 같다.

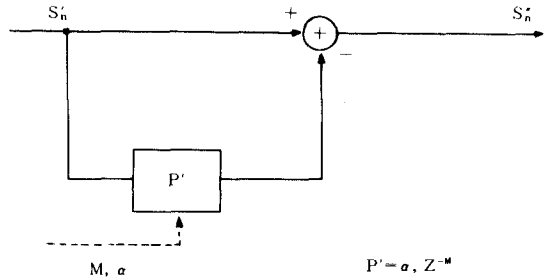


그림 2. 주기 예측 loop
Fig. 2. Pitch predictive loop.

2. 음성 분석

PPL을 통과한 신호는 LPC analyzer로 입력되며 여기서 성도 여파기의 예측 계수와 residual 신호를 얻게 된다. 이때 all-pole model 디지털 여파기의 전달함수는 Z-영역에서 다음과 같이 표현된다.^[16]

$$H(Z) = 1 / (1 - \sum_{i=1}^P a_i Z^{-i}) \tag{4}$$

여기에서 $|a_i|$ 는 여파기의 예측 계수이고 P는 예측 계수의 갯수이다.

LPC 분석 방법에는 LPC analyzer의 입력 음성 신호를 어떻게 해석하며 예측 오차를 어떻게 최소화하는가에 따라 autocorrelation (Toeplitz-form) 방법과 covariance (non-Toeplitz form) 방법이 있는데^[17] 본 논문에서는 합성 여파기의 안전성을 보장할 수 있는

autocorrelation 방법을 채택하였다.¹¹⁾ 여기서 window 함수는 Hamming window를 사용하였는데 window 길이는 180 sample로 하고 44 sample을 겹치도록 하였다. 그 이유는 해석하는 block 사이에서 LPC 계수의 급격한 변화를 방지하고, window 사이의 null에 있는 data sample의 소멸을 방지하기 위한 것이다. 음성 분석에 관한 자세한 내용은 문헌에 나와 있으므로 여기에서는 생략하기로 한다.¹³⁾¹⁴⁾¹⁵⁾¹⁶⁾¹⁷⁾¹⁸⁾¹⁹⁾²⁰⁾

3. 음성 합성 및 Postprocessing

음성 합성부에서는 분석부에서 구한 여파기 계수를 사용하여 합성 여파기를 구성하고, 여기에 residual 신호를 excitation 신호로 사용하여 음성을 합성한다. 이때 합성 여파기의 전달함수는 예측 여파기의 역 전달함수를 가진다.

반사계수를 사용하여 여파기를 구성하는 경우 여파기의 형태는 lattice 형이나 ladder 형이 되는데 여기서는 곱셈을 한번 행하는 lattice 형의 합성 여파기를 구성하였다.¹²⁾ 이와 같이 구성한 여파기는 안정성 여부의 판정이 용이함은 물론 계산 시간상의 이득을 볼 수 있다.

합성 여파기를 통과한 신호는 postprocessing 과정을 거쳐 최종 출력 신호가 되는데 postprocessing 과정은 preprocessing의 역과정으로서 inverse PPL과 deemphasis 회로를 거치게 된다. 마지막 과정으로 postprocessing을 거친 신호는 D/A 변환기와 저주파 대역 여파기를 거치면 합성된 음성신호를 얻을 수 있다.

4. Coding 및 Decoding

Residual 신호와 여타의 parameter들은 memory에 저장되기 전에 coding이 행하여 지는데 이때 발생하는 quantization error는 합성음성의 음질에 커다란 영향을 미치므로 적절한 양자화 방법의 사용은 매우 중요한 일이다.

Residual 신호의 coding 방법으로는 2bit adaptive PCM(APCM)을 사용하였는데 그 이유는 bit 율은 다소 높지만 합성 음성의 질이 보장되고 시스템 구성이 간단하기 때문이다.

APCM의 step size를 결정할 때에는 normalized residual energy(E)를 구하여 gain factor로 사용하였다.

그의 preprocessing과 분석부에서 구한 반사계수(k_i), normalized residual energy(E), PPL parameter(α) 그리고 pitch주기(M)등은 모두 PCM으로 coding하였다. 반사계수는 각 k_i 에 대하여 정보량의 경감에 따라 불균일한 bit 수를 할당하였고 각 계수의 최대치와 최소치 사이를 세부분으로 나누어 각 부분에 할당되는

양자화 level수를 달리하는 piecewise linear (PL) quantization 방식을 사용하였다.¹²²⁾

Normalized residual energy의 coding은 energy의 범위를 두 영역으로 나누어 PL 양자화를 하였고, PPL parameter는 linear quantization 방법을 사용하여 coding 하였으며 pitch 주기는 coding을 하지 않고 memory에 저장하였다.

III. 컴퓨터 Simulation과 결과

앞에서 설명한 디지털 자동 응답기의 simulation은 Data General사의 Nova 4/X minicomputer에 의해 수행되었다. 입력 음성 신호는 여성의 목소리를 사용하였으며 대역폭은 차단주파수가 3.2kHz인 저주파 대역 여파기로 filtering하여 제한하였고, 12bit A/D 변환기를 사용하여 6.8kHz로 sampling하였다.

Simulation시 window 길이는 20ms로 하였다. 따라서 normalized residual energy(E)를 제외한 모든 parameter들은 매 20ms마다 update하게 되며, residual energy는 residual의 quantization noise power를 줄이기 위하여 5ms마다 update하도록 하였다.

본 자동응답기에서 사용한 음성은 PPL parameter α 가 0.5에서 1.2사이에 분포하므로 6bit를 할당하여 linear quantization을 하였다. pitch 주기는 17sample에서 125sample 사이에 분포하므로 7bit를 할당하여 coding하지 않고 그대로 system memory에 저장하였다.

반사계수의 수를 결정할 때에는 입력신호의 주파수 대역이나 계산시간, coding시 요구되는 bit수 등이 먼저 고려되어야 한다. 본 논문에서는 대역폭을 3.2kHz로 제한하였고, 이때 음성신호의 spectrum은 보통 3개 정도의 formant를 갖는데 이 formant들은 pole에 의해 결정되며 이 pole들은 complex conjugate로 생기므로 적어도 6개 이상의 계수가 필요하다. 본 system에서는 계수의 수를 선정하기 위하여 simulation을 행한 결과 그림 3과 같은 합성 음성의 파형을 얻을 수 있었다. 그림 3을 살펴보면 계수의 수가 6개일 경우는 합성된 음성의 파형이 original 음성신호와 차이를 보임을 알 수 있다. 또한 계수의 수가 8개일 경우와 10개일 경우를 살펴보면 original 음성과 별 차이가 없음을 알 수 있다. 따라서 합성 음성의 파형이 original 음성에 가깝고 합성 filter의 구현에 직접 연관되는 계수의 수를 적게 함으로써 계산시간을 적게 하고 memory용량면에서의 이득은 물론 original 음성을 재현할 수 있도록 하기 위하여 계수의 수를 8개로 하였다.

각 반사계수를 coding하는데 사용되는 bit 수는 각각의 변화에 따른 합성 여파기의 민감도와 합성 음성의

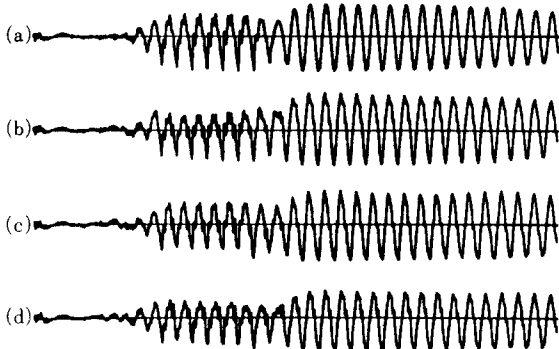


그림 3. 각각 다른 수의 reflection coefficient를 사용하여 합성한 음성들의 파형
(a) original (b) 6개 (c) 8개 (d) 10개
Fig. 3. Synthetic speech waveform with different reflection coefficient.

질을 고려하여 선택하였는데 일반적으로 낮은 차수의 계수들이 합성여파기에 큰 영향을 미치므로 많은 bit 수를 할당하였다. 반사계수들을 coding 하는데 사용된 PL quantizer를 구성하기 위하여 먼저 각 계수들의 분포를 구한 후에 이 분포를 토대로 k_i , max 값과 k_i , min 값을 정하고 영역의 경계치 R_1 과 R_2 값을 선택한 후 각 영역에 양자화 level수를 할당하였다. 정해진 양자화 level수와 경계치들은 표 1에 정리하였다.

표 1. 반사계수의 coding을 위한 PL 양자화의 구간과 level수 및 할당된 bit 수
Table 1. PL quantization of reflection coefficient.

반사계수	최소 k_i	R_2	R_1	최대 k_i	양자화 level 수			bit 수
					Q_1	Q_2	Q_3	
k_1	-0.8	0.2	0.6	0.96	24	12	28	6
k_2	-0.94	-0.5	0.3	0.84	4	16	12	5
k_3	-0.5	-0.3	0.5	0.7	4	24	4	5
k_4	-0.7	-0.5	0.2	0.4	3	11	2	4
k_5	0.7	-0.45	0.2	0.4	3	11	2	4
k_6	-0.5	-0.3	0.2	0.4	2	12	2	4
k_7	-0.4	-0.2	0.0	0.4	2	4	2	3
k_8	-0.4	-0.2	0.05	0.3	2	4	2	3

Normalized residual energy(E)는 범위를 두개의 영역으로 나누어 PL 양자화를 하였다. E는 0에서 250 사이에 분포하는데 0-63사이에는 분포밀도가 크고 64-250사이에는 밀도가 작은 것으로 밝혀졌다. 따라서 최소 값을 0, 최대값을 255로 하고, 영역의 경계치를 63으로 하고 각 영역의 양자화 level수를 32개로 하여

분포밀도가 큰 0-63사이를 세밀하게 양자화 하였다. 결과적으로 E를 coding 하는데 할당되는 bit 수는 6으로 하였다.

Residual coding에는 2bit APCM을 사용하여 불균일(non-uniform) 양자화를 행하였는데 APCM의 coding구간과 decoding level은 normalized residual energy에 임의의 계수 $|\beta_i|$, $|\gamma_i|$ 를 곱하여 사용하였다. Excitation신호로 사용되는 residual은 합성음성의 음질에 중요한 요인으로 작용하기 때문에 $|\beta_i|$ 와 $|\gamma_i|$ 의 최적치를 잘 택해야 한다. 최적치는 원래의 residual과 decoding된 residual간의 signal-to-quantization noise ratio(SQNR)와 합성음성의 음질에 의해 결정하였는데 결정된 최적치는 다음과 같다.

$$\beta_1 = -0.9816 \quad \beta_2 = 0.0 \quad \beta_3 = 0.9816$$

$$\gamma_1 = -1.51 \quad \gamma_2 = -0.4528 \quad \gamma_3 = 0.4528 \quad \gamma_4 = 1.51$$

위에서 설명한 방법에 의한 RELP system의 각 부분의 신호파형은 그림 4와 같으며 각 parameter들과 residual을 coding한 후 다시 decoding하여 합성한 음성의 파형과 coding을 하지 않은 값들을 사용하여 합성한 음성의 파형을 비교한 것이 그림 5에 나타나 있다. 그림에서도 알 수 있듯이 original 음성 파형과 최종 합성된 음성파형간에는 거의 차이가 없음을 알 수 있다.

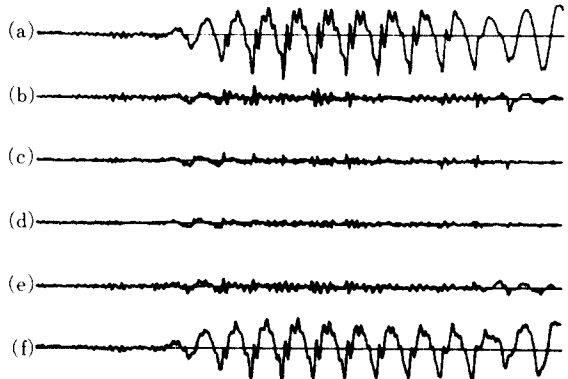


그림 4. 각 부분에서의 signal 파형
Fig. 4. Signal waveform in each part.

- (a) Input signal
- (b) Pre-processing output
- (c) Residual
- (d) Decoded residual
- (e) LPG synthesizer output
- (f) Post Processing output

본 digital 음성 자동응답기에 사용된 parameter와 방법들이 표 2에 요약되어 있다.

표 2. 개발된 digital audio response system에서 사용된 방법 및 coding bit 수

Table 2. Scheme and number of coding bit used in the developed digital audio response system.

• 1 초 길이의 음성을 합성하는 데 필요한 information 들을 저장 하는데 요구되는 memory size	Residual	13,600 bit/s
	반사계수	1,700 bit/s
	Residual energy	300 bit/s
	α	300 bit/s
	Pitch period	350 bit/s
	Total	17,150 bit/s
• Input bandwidth	3.2kHz	
• Input sampling rate	6.8kHz	
• Input quantization accuracy	12bits/sample	
• Window	Hamming window	
• 음성 해석 방법	Autocorrelation LPC method	
• Frame rate	50frames/s	
• 반사계수	8 개	
• Residual encoder	2bit APCM	
• Parameter encoder	PCM	

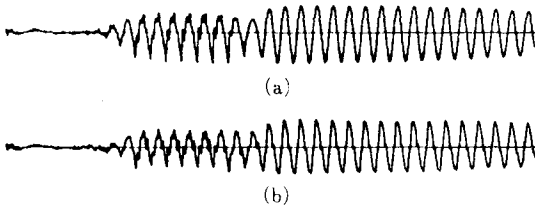
그림 5. Original 음성 파형과 최종 합성 음성 파형
(a) Original 음성파형 (b) 합성 음성파형

Fig. 5. Original speech waveform and synthetic output speech waveform.

IV. 음성 응답시스템의 구현

본 digital 음성 응답장치는 hardware와 software가 조화를 이루어 설계 개발되었으며 좋은 음질과 시스템 신뢰성을 높일 수 있도록 하였다.

1. Hardware 설계

Digital 음성 응답장치는 크게 나누어 컴퓨터 연결부를 포함한 I/O, code화 되어 data memory에 저장된 data를 memory로부터 음성합성기로 전송하는 controller와 controller로부터 전송되어 온 parameter를 합성하는 음성 합성기부분으로 나눌 수 있는데, 이들의 구성은 real time processing을 고려하여 bit-slice microprocessor인 2900series²³와 INTEL microprocessor를 사용하여 설계하였다. 이 시스템의 전체 block diagram은 그림 6에 도시하였는데 이를 기능별로 분리하여 자세히 설명하겠다.

먼저 operator의 명령에 의하여 host computer는

필요한 data(즉, 가입자 정보)를 access하여 controller로 전송한다. Controller는 USART(universal synchronous asynchronous receiver & transmitter)를 통하여 host computer와 연결되며 controller는 CPU와 memory 그리고 음성합성기와 연결되는 I/O로 구성된다. Memory는 operating program이 들어가는 2k PROM과 temporary memory(RAM) 그리고 미리 code화된 음성의 특성 parameter를 저장하는 data ROM으로 구성되어 있으며, I/O는 음성합성기의 상태를 나타내는 status register와 priority encoder 그리고 data buffer로 구성되어 있고 상태의 변화에 따라 그때 요구되는 data를 음성합성기로 전송하는 역할을 한다. 음성합성기의 기본구조는 저자들의 연구실에서 개발된 2-channel LPC vocoder²⁴와 비슷하며 CPU, microprogram control unit(MCU)와 memory로 구성되어 있다. CPU는 16bit arithmetic logic unit(ALU), multiplexer, status register로 구성되었고 MCU는 pipeline 기법을 사용할 수 있도록 sequencer와 micro instruction register 그리고 multiplexer로 구성하였다. Memory는 controller로부터 전송되어온 data의 보관과 음성합성의 중간 결과 및 최종 합성음성을 저장하는 RAM과 microprogram과 음성 합성시 필요한 decoding table을 저장하는 ROM으로 구성되어 있다.

따라서 이상의 내용을 정리하면 host computer로부터 data를 전달 받은 controller는 data ROM에서 특성 parameter를 차례로 access하여 음성 합성기로 전송하며 음성합성기를 거친 출력 신호는 D/A 변환기로 전송되고 이 신호는 deemphasis 회로와 저주파 대역 여

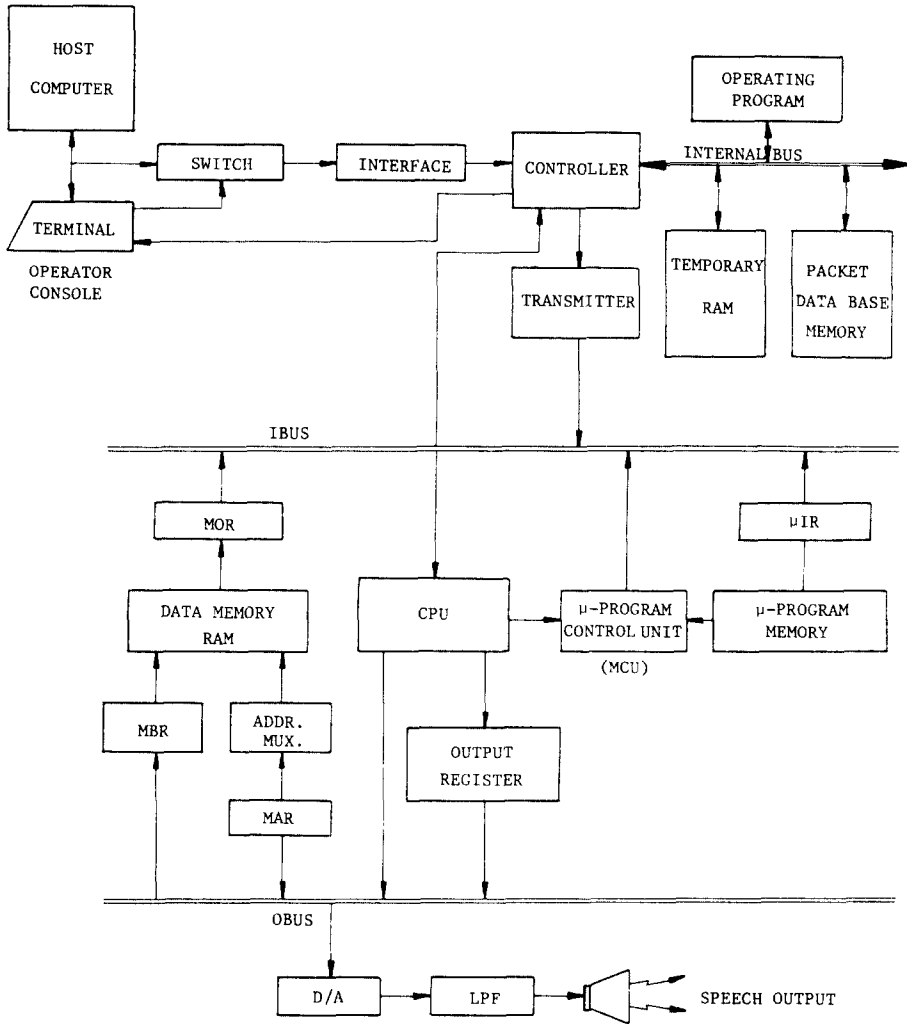


그림 6. Digital audio response system의 hardware block diagram
 Fig. 6. Hardware block diagram of digital audio response system.

표 3. Hardware 요약
 Table 3. Hardware specification.

(a) Synthesizer, I/O

Cycle time	204ns
기본 logic family	TTL
Microprogram	1 K × 52bits
Microprogram memory chip 수	13개
Data memory (ROM)	1 K × 16bits
Data memory (RAM)	1 K × 16bits
Microprocessor	4 AMD 2903 (4bit slice)
Microsequencer	1 AMD 2910 (12bit wide)
D/A conversion	12 bit accuracy
	147μs sampling interval
Total DIP Count	95

(b) Controller

X-tal frequency	6.144 Mhz
기본 logic family	TTL
Program size (ROM)	2 K × 8 bits
Data memory (RAM)	1 K × 8bits
Data memory (ROM)	32K × 8bits
CPU	INTEL 8085A
Total DIP Count	48

과기 및 speaker를 거쳐 합성음성으로 출력된다.
 이러한 일련의 과정을 거쳐 음성을 합성하는 본 시스템의 hardware 제원은 표 3에 요약되어 있다.

2. 시스템 Software

Digital 음성 응답장치에 사용된 시스템 software는 controller와 합성기에 내장된 program으로 나눌 수 있다. Controller에 내장된 program은 assembly 언어로서 초기 routine, main routine, parameter 전송 routine, interrupt service routine으로 나누어지는데 그 기능은 대략 다음과 같다.

표 4. 전화번호 구별표

Table 4. Classification of telephone number.

잘못된 data, data base에 기록되지 않는 전화번호	\$
대대표전화	CCCC*CC. CCC*CC.
보통의 전화	CCCb CCCC. CCb CCCC. CCCb CCCC -d d. CCb CCCC -d d. CCCb CCCC, d d d d. CCb CCCC, d d d d.

(주) 응답 형태별로 구별하면 1) 문의하신 번호는 번호부에 기재되어 있지 않습니다. 감사합니다. 2) 문의하신 번호는 대대표 CCCC에 CC입니다. 감사합니다. 3) 문의하신 번호는 CCC에 CCCC입니다. 감사합니다. 4) 문의하신 번호는 CCC에 CCCC부터 dd까지입니다. 감사합니다. 5) 문의하신 번호는 CC에 CCCC와 dddd입니다. 감사합니다. 등의 총 9종류가 위에 기록되어 있다.

b : blank

C : 전화번호

Host computer로부터 표 4에 나타난 9종류의 전화번호 format에 따라 가입자 정보를 받게 된다. 수신된 가입자 정보는 초기 routine에서 분석되어 error가 없을 경우 main routine으로 넘어간다. Main routine은 가입자 정보에 맞는 단어의 initial address와 end address 그리고 각 단어에 맞는 pause 기간을 mapping table에 기록한다. 음성의 naturality를 높이기 위하여 삽입하는 pause기간은 각 단어의 연결 부위에 넣게 되는데 여기서는 미리 coding된 random noise를 전송하였다. 이는 단순히 pause기간을 delay loop로 처리할 수도 있으나 음성합성시 단어의 연결부위가 부자연스럽고 해당 pause기간을 음성 합성기가 또 다시 check해야 하는 단점이 있기 때문이다. Pause 기간은 컴퓨터 simulation 결과로 얻은 것이며 표 5에 요약 정리하였다.

Main routine을 거쳐 mapping table이 완성되면 interrupt service routine은 합성기의 상태를 check한다. 이때 합성기가 ready상태에 있으면 parameter 전송

표 5. 각 단어에 대한 pause기간

Table 5. Pause period for each word.

단 어	Block 수	Pause기간(초)
0	3	0.113
1	3	0.113
2	4	0.15
3	3	0.113
4	3	0.113
5	3	0.113
6	5	0.19
7	3	0.113
8	3	0.113
9	4	0.15
입 니 다	10	0.4
에	6	0.23
부 터	6	0.23
까 지	6	0.23
과	6	0.23
와	6	0.23
대 대 표	6	0.23
문의하신 번호는 감사합니다.	10	0.4

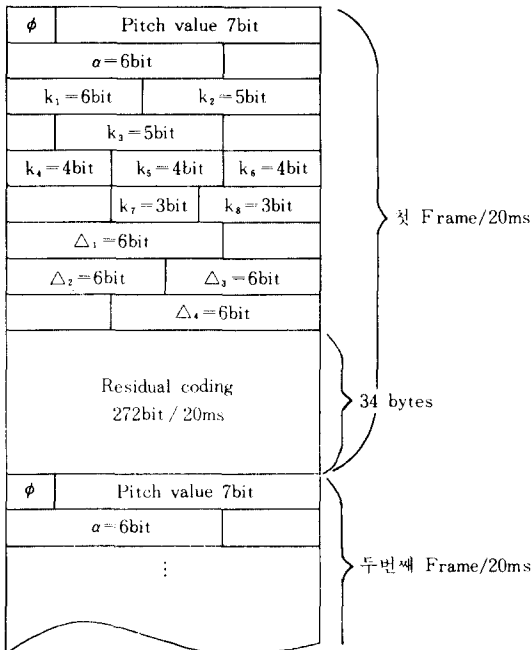
(주) 여기에서 1 block = 256 points

sampling frequency = 6.8kHz

routine과 interrupt service routine은 mapping table을 참고하여 그림 7에 도시한 data ROM의 data를 format에 맞게 masking하며 masking한 data는 frame interrupt, data interrupt, termination interrupt를 이용하여 전송한다. 이때 각종 interrupt 신호는 음성 합성시 필요한 timing정보와 data 정보(pitch값, PPL parameter, 반사계수, step size, residual signal)의 종류 그리고 sevice종료를 알리는 데 사용된다.

또한 음성합성기에 내장된 program은 microprogramming으로 초기 routine, main routine, interrupt service routine으로 나누어 지는데 main routine은 decoding routine, 합성여과기 routine, inverse PPL routine으로, 그리고 interrupt service routine에는 D/A, frame, data 전송, 종료 service routine으로 나누어진다. 이를 data 처리과정 순서로 살펴보면 다음과 같다.

Controller로부터 interrupt logic unit를 통하여 합성기로 입력된 data는 parameter별로 구분되어 decoding된다. 이때 decoding된 합성 여과기 계수와 inverse PPL parameter는 합성 여과기와 inverse PPL을 구성하는데 사용되며, decoding된 residual 신호는 구성된 합성 여과기와 inverse PPL을 거쳐 출력 음성



(주) Δ 는 5ms마다 update 되므로 20ms안에 4 개가 필요하다.
1 frame은 43 byte에 해당된다.

그림 7. PROM 에 저장된 data format
Fig. 7. Data format stored in PROM.

으로 buffer에 저장된다. 저장된 출력음성은 D/A service routine을 거쳐 최종출력 analog 음성신호로 변환된다.

위 과정에서 합성기는 한 frame의 analog 합성 음성을 내보내기 위하여 20ms동안 약48,000개의 micro-instruction을 수행하는데 이는 service 시간으로 환산할 경우 9.8ms에 해당한다. 따라서 본 시스템은 real time processing을 수행하는데 전혀 지장이 없음을 알 수 있다.

V. 結 論

Digital 음성 응답장치는 그 개발에 digital통신, digital 음성처리, 컴퓨터 및 microprocessor 응용 등 복합적인 최신 기술이 필요하고 또한 방대한 software를 요하는 장비이다. 개발된 digital 음성 응답장치는 PPL을 가지는 RELP algorithm을 사용하였으며 음질, 어휘의 다양성, 시스템 융통성면에서 우수하다. 특히 합성 음성의 음질을 보다 효율적으로 높여서 입력 음성을 재생할 수 있다는 것이 본 시스템의 큰 장점이라 하겠다.

謝 辭

본 논문을 상세히 검토하여 주시고 교정하여 주신

심사위원님께 감사드립니다.

參 考 文 獻

- [1] T.G. Homsby, Jr., *Voice Response Systems Modern Data*, pp. 46-50, Nov., 1972.
- [2] N.S. Jayant, *Waveform Quantization and Coding*. IEEE Press, New York, 1975.
- [3] J.D. Markel and A.H. Gray, Jr., *Linear Prediction of Speech*, Springer-Verlag, New York, 1976.
- [4] C.K. Un and D.J. Magill, "The residual excited linear prediction vocoder with transmission rate below 9.6kbps", *IEEE Trans. on Comm.*, vol. COM-23, no.12, Dec., 1975.
- [5] J. Makhoul and R. Viswanathan, "Adaptive processing for linear predictive speech compression systems," *J. Acoust. Soc. Am.*, vol.55, no.2, pp. 475-476, Feb., 1974.
- [6] A.H. Gray and J.D. Markel, "A spectral-flatness measure for studying the autocorrelation method of linear prediction of speech analysis," *IEEE Trans. on ASSP*, vol. ASSP-22, no.3, pp. 207-217 June, 1974.
- [7] J.D. Gibson, "Adaptive prediction in speech differential encoding systems," *Proc. on IEEE*, vol.68, no.4, pp. 448-527, Apr., 1980.
- [8] J.D. Markel and A.H. Gray, "Fixed-point truncation arithmetic implementation of a linear prediction autocorrelation vocoder", *IEEE Trans. on ASSP*, vol. ASSP-22, no.4, Aug, pp. 273-282, Aug, 1974.
- [9] B. Gold and L.R. Rabiner, "Parallel processing techniques for estimating pitch periods of speech in time domain," *J. Acoust. Soc. Am.*, vol.46, pp. 442-448, Aug., 1969.
- [10] A.M. Noll, "Cepstrum pitch determination," *J. Acoust. Soc. Am.*, vol.41, pp. 293-309, Feb., 1969.
- [11] J.D. Markel, "The SIFT algorithm for fundamental frequency estimation", *IEEE Trans. Audio Electroacoust.*, vol. AU-20,

- pp. 367-377, Dec., 1972.
- [12] C.K. Un and S.C. Yang, "A pitch extraction algorithm based on LPC inverse filtering and AMDF," *IEEE Trans. on ASSP*, vol. ASSP-25, no.6, Dec., 1977.
- [13] M.J. Ross, H.L. Shafter, A. Cohen, R. Freudberg, and H.J. Hanley, "Average magnitude difference function pitch extraction," *IEEE Trans. on ASSP*, vol. ASSP-22, pp. 353-362, Oct., 1974.
- [14] L.R. Rabiner, M.J. Cheng, A.E. Rosenberg, and C.A. McGonegal, "A comparative performance study of several pitch detection algorithms", *IEEE Trans. on ASSP*, vol. ASSP-24, pp. 399-417, Oct., 1976.
- [15] K.A. Oh and C.K. Un, *A Performance Comparison of Pitch Extraction Algorithms for Noisy Speech*. Proc. of IEEE Int. conf. on ASSP, Mar., 1984.
- [16] L.R. Labiner and R.W. Schafer, *Digital Processing of Speech Signals*. Prentice-Hall, pp. 98-106, 1978.
- [17] R.W. Schafer and L.R. Rabiner, "Digital representations of speech signal," *Proc. IEEE*, vol.63, pp. 662-667, Apr., 1975.
- [18] B.S. Atal and S.L. Hanauer, "Speech analysis and synthesis by linear prediction of the speech wave," *J. Acoust. Soc. Am.*, vol.50, pp. 637-655, Aug., 1971.
- [19] Hisashi Wakita, "Direct estimation of the vocal tract shape by inverse filtering of acoustic speech waveforms," *IEEE Trans. Audio Electroacoust.*, vol. AU-21, pp. 417-427, Oct., 1973.
- [20] E.A. Robinson, *Statistical Communication and Detection*. Hafner Publishing Co., New York, 1967.
- [21] F. Itakura and S. Saito, *On the Optimum Quantization of Feature Parameters in the PARCOR Speech Synthesizer*. IEEE Conf. on Speech Comm. and Process, pp. 434-437, Apr., 24-26, 1972.
- [22] C.K. Un and S.C. Yang, *Piecewise Linear Quantization of LPC Reflection Coefficients*. Proc. of IEEE Int. Conf. on ASSP, May, 1977.
- [23] The Am 2900 Family Data Book, Advanced Micro Device Inc., 1979.
- [24] C.K. Un et al., "Implementation of a two-channel LPC vocoder," *IEEE Trans. on Comm.* vol. COM-31, no.7, July, 1983.
-