

論文95-32B-7-8

FM 방식을 이용한 디지털 악기음 합성기의 구현 (Realization of Digital Music Synthesizer Using a Frequency Modulation)

朱世哲*, 金鎮範*, 金基斗**

(Se-Chul Chu, Jin-Bum Kim, and Ki-Doo Kim)

요약

본 논문에서는 유전 알고리즘(GA : Genetic Algorithm)에 기초한 FM 악기음 합성을 범용 디지털 신호 처리기(DSP)에 적용하여 실시간 악기음 합성기를 구현하였다. 특히 FM 합성시 여러개의 반송파와 한개의 변조파로 악기음을 모델링하여 다양하고 복잡한 악기음을 실제로 가깝게 합성하였다. 또한 FM 방식의 악기음을 합성하기 위한 파라미터를 추출하기 위하여 PCM 데이터를 기준으로 주파수 영역에서 FFT를 이용한 스펙트럼을 분석하고, 이에 유전 알고리즘을 적용하여 여러개의 반송파와 한개의 변조파를 갖는 FM 합성을 위한 최적 파라미터 추출 방법을 사용하였다. 제시한 파라미터 추출 알고리즘을 사용하여 몇 가지 악기에 해당하는 파라미터를 추출하고, 추출된 파라미터를 실시간 악기음 합성기에 적용하여 악기음을 합성하였다. 합성음을 직접 듣는 주관적인 방법 외에, 평가를 보다 객관적으로 행하기 위하여 파라미터 추출시 기준으로 사용한 악기음과 이것으로부터 추출된 파라미터에 의해 합성된 악기음을 시간 영역 및 주파수 영역에서 비교 분석하여 파라미터 추출 알고리즘 및 DSP를 이용한 실시간 악기음 합성기의 타당성을 입증한다.

Abstract

In this paper, we realize a real time digital FM synthesizer based on genetic algorithm using a general purpose digital signal processor. Especially, we synthesize diverse music sounds nicely using a synthesis model consisting of a single modulator and multiple carriers. Also we present genetic algorithm-based technique which determines optimal parameters for reconstruction through FM synthesis of a sound after analyzing the spectrum of PCM data as a standard music sound using FFT. Using the suggested parameter extraction algorithm, we extract parameters of several instruments and then synthesize digital FM sounds. To verify the validity of the parameter extraction algorithm as well as realization of a real time digital music synthesizer, the evaluation is first done by listening the sound directly as subjective test. Secondly, to evaluate the synthesized sound objectively with an engineering sense, we compare the synthesized sound with an original one in a time domain and a frequency domain.

* 準會員, ** 正會員, 國民大學校 電子工學科

(Dept. of Elec. Eng., Kook-Min Univ.)

※ 이 논문은 1993년도 한국 학술진흥재단의 공모과제 연구비에 의하여 연구되었음.

接受日字: 1994年12月17日, 수정완료일: 1995年6月26일

I. 서론

악기음을 생성하고 합성하는 방법은 디지털 신호처리 기술의 한 응용 분야로서, 지금까지 악기음을 합성하기 위하여 사용되어 온 방식은 가산합성(additive

synthesis), 감산합성 (subtractive synthesis), PCM (pulse code modulation) 및 FM (frequency modulation) 방식들이 있다^[1-5]. FM 방식은 1973년 미국 Stanford 대학의 John Chowning에 의해 처음 연구되었고^[3-5], 1980년대 중반 야마하 (Yamaha) 및 Digital Keyboards 회사에서 상업용 악기음 IC를 개발함으로써 많이 보급된 방식이다^[5]. 본 방식은 주파수를 변조시키므로써 다양한 스펙트럼을 만들어 실제에 가까운 악기음을 생성하는데, 다른 방식에 비하여 계산량이 적게 요구되며, 회로 구성이 간단하고 여러 악기음을 만드는데 필요한 파라미터의 수가 적어 메모리를 적게 차지하는 등 효율적이다. 하지만, FM 방식은 각 악기에 해당하는 정확한 파라미터를 요구하기 때문에 파라미터를 추출하기 위해서 시간 영역 및 주파수 영역에서 많은 방법들이 연구되어 왔다^[6,7].

기존의 FM 방식은 악기음을 반송파와 변조파를 각각 하나씩 갖는 FM 식으로 모델링하여 합성하였고, 이에 대한 파라미터 추출 알고리즘도 경험적인 요소들이 많았다^[8]. 본 논문에서는 참고문헌 [8]을 실시간 처리를 위하여 신호처리를 이용하여 악기음을 합성한 참고문헌 [12]의 결과를 개선하여 기존의 방법보다 실제음에 보다 가깝고 다양한 음을 합성할 수 있도록 악기음을 여러개의 반송파와 하나의 변조파를 갖는 FM 식으로 모델링하였다. 또한 파라미터 추출 알고리즘도 자연 생태계의 적자 생존의 원리를 모방한 적응적 탐색기법의 일종인 유전 알고리즘 (GA)을 이용하여 보다 체계적으로 최적의 파라미터를 추출할 수 있다. GA는 대부분의 최적화 기법과 달리 문제와 관련된 사전지식이 필요하지 않고 탐색점이 여러개로 구성되어 매우 효과적이다^[9].

II. FM 방식을 이용한 악기음 합성이론

FM 방식을 이용한 악기음 합성방식은 변조파로 반송파의 주파수를 변화시켜 고조파 성분 뿐만 아니라 비고조파 성분을 포함하는 파형을 생성하므로써 다양한 악기음을 만드는데 사용된다. 반송파와 변조파를 각각 한개씩 갖는 일반적인 FM 악기음 합성식은 다음과 같다^[3-5]

$$X(t) = W(t) \sin(2\pi f_c t + I(t) \sin(2\pi f_m t)) \quad (1)$$

여기서, $W(t)$ 는 시변 진폭, f_c 는 반송 주파수, f_m 은 변조 주파수, $I(t)$ 는 시변 변조지수를 의미한다. 식 (1)은 Bessel 함수를 이용하여 다음과 같이 나타낼 수 있다^[10]

$$X(t) = W(t) \sum_{k=-\infty}^{\infty} J_k(I(t)) \sin(2\pi f_c t + 2\pi k f_m t) \quad (2)$$

특히 f_c 가 f_m 의 정수배인 경우 ($f_c = n f_m$, n 은 정수), 식 (1)과 (2)는 다음과 같이 표현된다^[11].

$$X(t) = W(t) \sin(2\pi n f_m t + I(t) \sin(2\pi f_m t)) \quad (3a)$$

$$= W(t) \sum_{k=-\infty}^{\infty} J_k(I(t)) \sin(2\pi(n+k)f_m t) \quad (3b)$$

$$= W(t) \sum_{k=1}^{\infty} a_k(I(t)) \sin(2\pi k f_m t) \quad (3c)$$

여기서 $a_k(I(t)) = J_{(k-n)}(I(t)) - J_{-(k+n)}(I(t))$ 이다.

한편 한개의 반송파와 한개의 변조파를 사용하는 대신 여러개의 반송파와 한개의 변조파로 악기음을 모델링한다면 식 (3)에서 표현하기 어려운 보다 복잡한 악기음의 스펙트럼을 합성하여 실제 악기음에 가까운 음을 낼 수 있다. 이때 반송파의 갯수를 N_{cars} 라 하면 식 (3a)와 (3c)는 다음과 같이 변형된다.

$$X(t) = \sum_{j=1}^{N_{cars}} W_j(t) \sin(2\pi n_j f_m t + I_j(t) \sin(2\pi f_m t)) \quad (4a)$$

$$= \sum_{j=1}^{N_{cars}} W_j(t) \sum_{k=1}^{\infty} a_{k,j}(I_j(t)) \sin(2\pi k f_m t) \quad (4b)$$

여기서 $a_{k,j}(I_j(t)) = J_{(k-n_j)}(I_j(t)) - J_{-(k+n_j)}(I_j(t))$ 이며 j 번째 반송파의 k 번째 고조파의 크기를 나타내고, $W_j(t)$ 는 j 번째 반송파의 시변 진폭, n_j 는 j 번째 반송파의 주파수 (f_{c_j})와 변조 주파수 (f_m)의 비, 그리고 $I_j(t)$ 는 j 번째 반송파의 시변 변조지수이다. 식 (4b)는 다시 아래와 같이 표현할 수 있다.

$$X(t) = \sum_{k=1}^{\infty} b_k(t) \sin(2\pi k f_m t) \quad (5)$$

여기서 $b_k(t) = \sum_{j=1}^{N_{cars}} W_j(t) a_{k,j}(I_j(t))$ 이며, 합성된 악기음의 k 번째 고조파의 크기를 나타낸다. 이때 한개의 반송파를 이용할 경우에는 복잡한 악기음의 스펙트럼을 조절하기 위하여 변조지수값이 시간에 대한 함수이다. 하지만, 여러개의 반송파를 사용하는 경우에는 각 반송파에 따른 변조지수가 일정한 상수이어도 실제에 가까운 복잡한 스펙트럼을 갖는 악기음을 합성할 수 있다. 그러

므로 식 (4a), (4b), (5)는 각각 다음과 같이 다시 쓸 수 있다.

$$X(t) = \sum_{j=1}^{N_{cars}} W_j(t) \sin(2\pi n_j f_m t + I_j \sin(2\pi f_m t)) \quad (6a)$$

$$= \sum_{j=1}^{N_{cars}} W_j(t) \sum_{k=1}^{\infty} a_{k,j}(I_j) \sin(2\pi k f_m t) \quad (6b)$$

$$= \sum_{k=1}^{\infty} b_k(t) \sin(2\pi k f_m t) \quad (6c)$$

여기서

$$b_k(t) = \sum_{j=1}^{N_{cars}} W_j(t) a_{k,j}(I_j), \quad a_{k,j}(I_j) = J_{(k-n_j)(I_j)} - J_{-(k+n_j)(I_j)} \quad (6d)$$

즉, 식 (6a)에서 볼 수 있는 바와 같이 변조 주파수 f_m 과 각 반송파의 시간에 따른 파형의 변화를 나타내는 $W_j(t)$, 각 반송파의 변조지수 I_j , 그리고 각 반송파와 변조파의 주파수 비 n_j 를 추출하면 FM 방식을 이용하여 악기음을 합성할 수 있다. 그림 1은 여러개의 반송파와 한개의 변조파를 이용한 FM 악기음 합성의 블럭 다이어그램이다.

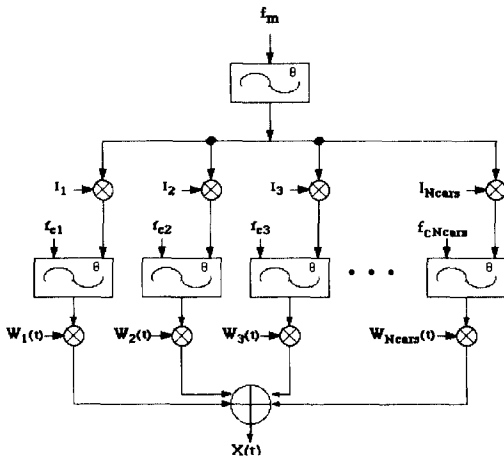


그림 1. 여러개의 반송파와 한개의 변조파를 이용한 FM 악기음 합성 모델

Fig. 1. FM synthesis model consisting of a single modulator and multiple carriers.

III. 유전 알고리즘

유전 알고리즘 (GA)은 자연 생태계의 적자 생존의 원리를 모방한 이론으로서 적응적 탐색기법의 일종이다¹⁹⁾. 기존의 탐색 방법이 탐색 공간에 대한 지식이

필요하고 국부적인 최적점의 탐색인데 반해, GA는 문제에 관련된 사전 지식이 아닌 적합도 함수 (fitness function)만을 이용하여 탐색을 수행한다. GA는 다음과 같은 몇가지 특징을 가진다. 첫째, 적자생존과 자연선택의 개념을 탐색에 도입한다 (population search). 즉, 탐색점 (searching point)이 한개가 아니라 여러개이고, 이들 중 가장 적합한 것이 다음 단계에 영향을 미친다. 둘째, 조합 행위자 (mating operator)를 이용해서 한 세대 (generation)내의 탐색점을 조합하여 더 나은 탐색점을 만들어 내며, 생물계에서 일어나는 유전 현상을 모델링한 유전 행위자를 사용하여 탐색 효율을 향상시킨다. 이 과정은 확률적 근거에 의해 진행된다. 셋째, 탐색 과정은 모두 적합도 함수에 따라 일어난다. 넷째, GA는 파라미터 그 자체를 사용하는 것이 아니라, 파라미터의 세트를 유한한 길이의 유전자 열 (string)로 코딩하여 사용한다.

GA는 재생산 (reproduction), 교배 (crossover), 돌연변이 (mutation)의 세가지 주된 기능자를 갖는다. 재생산은 자연계의 적자생존의 원리와 마찬가지로 주어진 유전자 열중에서 적합도에 따라 적합도가 높은 유전자 열은 똑같은 유전자 열을 하나 더 만들어 내고, 대신 적합도가 낮은 유전자 열을 도태시켜서 전체적으로 적합도가 높은 유전자 열들로 집단을 구성하게 한다. 교배는 이렇게 구성된 유전자 열들에서 두 유전자 열을 임의의 위치에서 서로 치환하는 과정이다. 돌연변이는 선택된 유전자 열중에서 임의의 인자를 바꾸는 과정이다. 그림 2와 3은 각각 교배와 돌연변이의 과정을 보여준다.

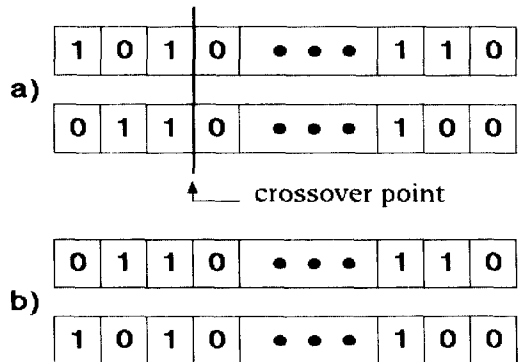


그림 2. 교배 과정 a)교배 전 b)교배 후
Fig. 2. Crossover process a) before crossover b) after crossover.

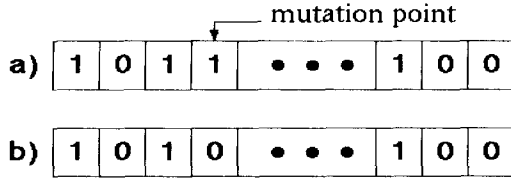


그림 3. 돌연변이 과정 a) 돌연변이 전 b) 돌연변이 후
 Fig. 3. Mutation process a) before mutation b) after mutation.

GA의 일반적인 전개는 다음과 같이 설명될 수 있고, 그림 4는 GA의 전개 순서도를 보여준다.

- 단계 1 : 파라미터를 2진수 1과 0으로 코딩하여 집단 (population)을 초기화한다.
- 단계 2 : 집단 내의 각 유전자 열에 해당하는 적합도 함수 값을 계산한다.
- 단계 3 : 적합도 함수 값을 근거로 하여 유전자 열을 복제 또는 삭제하여 교배에 적합한 유전자 열을 선택한다.
- 단계 4 : 교배와 돌연변이를 통하여 새로운 집단을 구성한다.
- 단계 5 : 원하는 적합도가 얻어질 때까지 단계 2 ~ 단계 5를 반복한다.

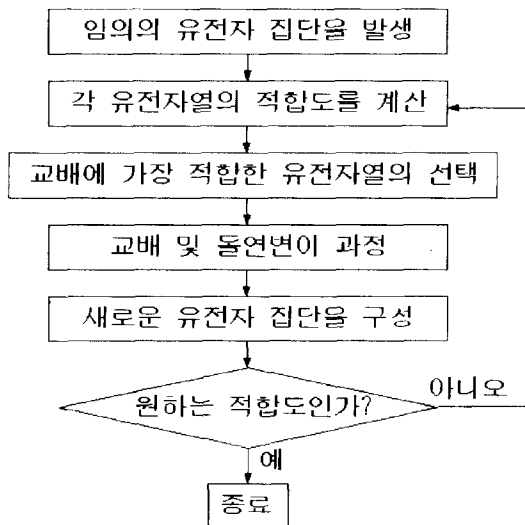


그림 4. 유전 알고리즘의 순서도
 Fig. 4. The flowchart of genetic algorithm.

IV. 유전 알고리즘을 이용한 FM 악기음 파라미터 추출 알고리즘

본 논문의 파라미터 추출 알고리즘은 경험을 바탕으로 한 기존의 파라미터를 추출하는 방법 [8] 대신 자연 생태계의 적자 생존의 원리를 모방한 적응적 탐색기법의 일종인 유전 알고리즘 (GA)을 이용하여 보다 체계적으로 최적의 파라미터를 추출할 수 있다. 이 방법은 일반성과 최적성을 보장할 수 없는 기존의 FM 파라미터 추출 알고리즘과 대조된다.

식 (6)에서 보인 바와 같이 각 반송파의 시변진폭과 변조지수, 변조파와의 주파수 비, 그리고 변조 주파수를 이용하면 실제의 악기음과 같은 주파수 특성을 갖는 악기음을 합성할 수 있다. 여러개의 반송파와 한개의 변조파를 갖는 FM 합성식으로 모델링된 악기음의 파라미터를 추출하기 위하여 기준음으로 사용한 PCM 데이터의 스펙트럼과 식 (6d)를 이용하여 다음과 같이 식 (7)을 구성할 수 있다.

$$A W \approx B \tag{7}$$

여기서

$$A = \begin{pmatrix} a_{1,1} & \dots & a_{1,Nbars} \\ a_{2,1} & \dots & a_{2,Nbars} \\ \vdots & & \vdots \\ a_{Nbars,1} & \dots & a_{Nbars,Nbars} \end{pmatrix}, W = \begin{pmatrix} w_{1,1} & \dots & w_{1,Nframes} \\ w_{2,1} & \dots & w_{2,Nframes} \\ \vdots & & \vdots \\ w_{Nbars,1} & \dots & w_{Nbars,Nframes} \end{pmatrix},$$

$$B = \begin{pmatrix} b_{1,1} & \dots & b_{1,Nframes} \\ b_{2,1} & \dots & b_{2,Nframes} \\ \vdots & & \vdots \\ b_{Nbars,1} & \dots & b_{Nbars,Nframes} \end{pmatrix}$$

행렬 A의 원소 $a_{k,r}$ 는 j번째 반송파에 해당하는 k번째 고조파의 크기를 나타내고, 행렬 W의 원소 $w_{j,r}$ 는 r번째 시간 프레임 (frame)에서 j번째 반송파의 진폭을 나타낸다. 그리고 행렬 B의 원소 $b_{k,r}$ 는 기준음으로 사용한 PCM 데이터의 r번째 프레임에서 k번째 고조파의 크기를 나타낸다. 또한, Nframes는 총 시간 프레임의 수를 나타내고 Nbars는 기준음의 고조파 성분의 갯수로서, 고조파의 수는 식 (6)에서와 같이 무한개이지만 실제로 파라미터 추출 알고리즘에 적용하기 위하여 유한한 갯수의 고조파만을 사용하였다. 합성음의 r

번째 프레임에서 j번째 반송파의 k번째 고조파의 크기는 $a_{k,j} * w_{j,r}$ 으로 표현할 수 있고, 식 (7)로부터 기준음의 r번째 프레임에서 k번째 고조파의 크기, $b_{k,r}$ 는 각 반송파의 r번째 프레임에서 k번째 고조파 성분의 합으로 근사화할 수 있다. 이는 다음과 같이 나타낼 수 있다.

$$\sum_{j=1}^{Ncars} a_{k,j} w_{j,r} \approx b_{k,r}, \quad 1 \leq k \leq Nhars, 1 \leq r \leq Nframes \quad (8)$$

최적의 파라미터를 찾는 GA는 각 파라미터를 하나의 유전자 열로 코딩하므로, 반송파와 변조파의 주파수비 n_j 와 각 반송파의 변조지수 I_j 의 비트 수를 결정해야 한다. 이때 I_j 는 대개 일정한 범위안에 있으므로 범위를 설정하여 보다 정확하게 변조지수를 표현할 수 있도록 변조지수의 비트 수를 결정한다. 또한 사용될 반송파의 갯수를 결정하여야 한다. 이때 반송파의 수는 실제 악기음의 고조파의 수보다 항상 작아야 한다. 사용되는 반송파의 수가 실제 악기음의 고조파의 수보다 크거나 같다면, FM 방식으로 악기음을 합성하기보다 가산합성 방식을 이용하는 것이 더욱 적합할 것이다. 그림 5는 n_j 와 I_j 의 코딩 예를 보여주는데, 본 논문에서는 변조지수를 7비트로 코딩하고 범위를 0에서 3까지로 하여 분해능이 0.023이 되게 하였고, 반송파와 변조파의 주파수비는 4비트로 코딩하였다.

n_1	I_1	n_2	I_2	...	n_{Ncars}	I_{Ncars}
-------	-------	-------	-------	-----	-------------	-------------

그림 5. n_j 와 I_j 의 코딩 ($j = 1, 2, \dots, Ncars$)
 Fig. 5. The coding of n_j and I_j ($j = 1, 2, \dots, Ncars$).

기준음으로 사용한 PCM 데이터를 프레임별로 단시간 (short-time) FFT를 수행한 후, 각 프레임에서의 고조파 성분들을 구하여 행렬 B를 구성한다. 주파수 분석을 수행하면서 변조 주파수 즉, 기본 주파수를 구할 수 있다. GA를 이용하여 n_j 와 I_j 를 코딩한 후 값들을 변화시키면서 행렬 A를 구성하고, 이렇게 매 세 대마다 구성된 행렬 A와 이미 구성된 행렬 B를 이용하여 행렬 W를 구한다. 만약, 파라미터 추출시 사용한 반송파의 갯수가 기준음의 고조파 갯수와 같다면, 쉽게 구할 수 있겠지만 파라미터 추출시 사용한 반송파의 갯수는 기준음의 고조파 갯수보다 항상 작으므로 식

(9)를 최소화시키는 행렬 W를 최소 자승법을 사용하여 구하였다.

$$\frac{\sum_{r=1}^{Nframes} \sum_{k=1}^{Nhars} \left(\sum_{j=1}^{Ncars} a_{k,j} w_{j,r} - b_{k,r} \right)^2}{\sum_{k=1}^{Nhars} \sum_{r=1}^{Nframes} b^2_{k,r}} \quad (9)$$

이 오차값을 적합도 함수로 하고 적합도 함수값이 일정한 범위안에 들어오면 이때 사용된 n_j 와 I_j 그리고 행렬 W를 구성하는 각 반송파의 시변 진폭을 여러 개의 반송파와 한개의 변조파를 갖는 FM 합성을 위한 최적 파라미터로 결정한다.

V. FM 방식을 이용한 디지털 악기음 합성 알고리즘

악기음을 합성하는데는 PCM 방식과 FM 방식이 가장 많이 사용되고 있다. PCM 방식은 비록 양질의 악기음을 합성할 수 있지만, 필요한 모든 악기의 샘플값들을 가지고 있어야 하므로 많은 메모리가 필요하고 메모리의 양을 줄이기 위하여는 ADPCM (Adaptive Delta PCM)과 같은 복잡한 알고리즘과 하드웨어가 요구된다. 이에 반하여 FM 방식은 음질은 다소 PCM 방식으로 합성한 악기음보다 떨어지지만 각 악기에 해당하는 몇개의 파라미터만 가지고 있으면 악기음을 합성할 수 있으므로 현재 PC용 사운드 카드 등에서 음원합성 방법으로 많이 사용되고 있다. 또한, 가산합성 방식은 하나의 정현파 발생기가 하나의 배음을 생성하고 음색을 조절하기 위해서는 각 정현파 발생기의 엔빌로프를 변화시켜야 하므로 많은 파라미터가 필요하고 계산량이 많아지게 된다. 즉 정현파 발생기의 엔빌로프가 가장 간단한 ADSR (Attack Decay Sustain Release) 방식이라 해도, 한 개의 배음에 대해 최소 8개의 파라미터를 조정해 주어야 한다. 정현파를 발생시킬 경우 12.8 MIPS (Million Instructions Per Sec)의 성능을 갖는 TMS320C25를 이용하여 23개의 인스트럭션 사이클 (instruction cycle)이 필요하며 1.79 μ sec의 시간이 소요되었고, 엔빌로프를 발생하기 위해서는 24개의 인스트럭션 사이클이 필요하여 1.87 μ sec의 시간이 소요되었다^[12]. 본 논문에서 사용한 FM 방식인 3개의 반송파와 하나의 변조파가 사용될 경우 반송파와 변조파를 발생시키기 위해 정현파 발생기가 4개 필요한 반면 가산합성 방식의 경우 예를들어

10배음까지 포함한 음을 생성하기 위해서는 최소한 10개의 정현파 발생기가 필요하므로 FM 방식보다 훨씬 많은 정현파 발생기가 요구된다. 마찬가지로 엔빌로프 발생기의 경우도 본 논문에서 사용한 FM 방식의 경우 3개가 요구되는 반면 가산합성 방식의 경우 10개가 필요하게 되어 단지 정현파와 엔빌로프를 발생시키기 위해서도 약 3 배의 계산량과 이에따른 시간이 요구된다. 여러개의 반송파를 포함하는 FM 방식을 이용한 악기음 합성식인 식 (6)을 구현하기 위하여 그림 6에 보인 바와 같은 오퍼레이터 (operator)를 정의한다. 정의된 오퍼레이터의 동작 알고리즘의 주된 특징은 ASIC 설계에 대비하여 하드웨어 구현이 용이하도록 하였다. 한 예로 테이블참조 (table look-up) 방식을 이용하여 sine 값을 계산하였으며, 모든 데이터를 고정 소숫점 (fixed point)으로 표현하였다¹⁸⁾.

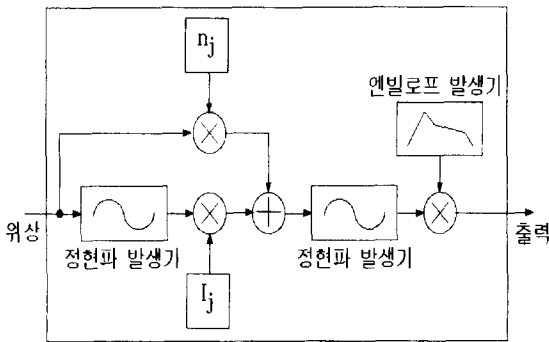


그림 6. 오퍼레이터의 내부 구조
Fig. 6. Internal structure of an operator.

정현파 발생기, 엔빌로프 발생기 등으로 이루어진 오퍼레이터는 외부와 위상 입력단 및 출력단으로 연결되며, 샘플링 시간 단위로 동작하도록 구성하였다. 여러개의 반송파를 갖는 FM 방식의 악기음 합성은 반송파 갯수의 오퍼레이터를 병렬로 연결하여 구현할 수 있다. 이때 연결된 각 오퍼레이터는 한개의 반송파에 대한 합성기로 작용하고, 이들의 출력을 모두 합하면 식 (6a)와 같은 악기음의 합성이 가능하다.

1. 정현파 발생기

정현파 발생기는 식 (6)의 반송파 및 변조파의 정현파를 만들어내기 위하여 사용된다. 정현파 발생을 위하여 테이블참조 방식을 이용하였다. 테이블참조 방식은 미리 계산된 정현파의 값을 테이블 형태로 저장해 놓

고 입력된 위상에 상응하는 정현파의 값을 구하는 방식이다. 그림 7은 본 방식에 의한 정현파 발생기를 보여준다.

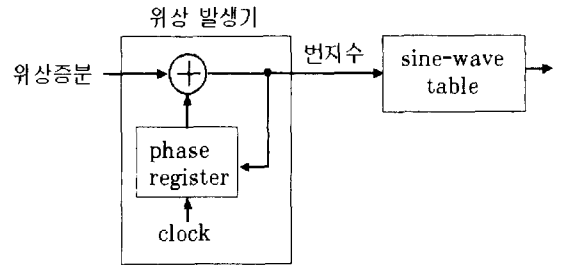


그림 7. 테이블참조 방식에 의한 정현파 발생기
Fig. 7. Sine wave generator using table look-up method.

그림 7에서 보는 바와 같이 테이블참조 방식을 이용해 정현파를 발생시키기 위해서는 위상을 발생시키는 역할을 하는 위상 발생기가 필요하다. 위상 ϕ 는 음의 높이를 결정하는 값 $\omega (= 2\pi f)$ 와 샘플링 주파수 f_s 에 의해 변화하는 값 t 의 곱으로 표현된다. 한 음이 유지되는 동안 ω 의 값은 일정하므로 위상값의 계산을 식 (10)과 같이 위상증분 $\Delta\phi$ 를 사용하여 표현하면 매 샘플링 시간마다 한번의 덧셈만으로 위상값을 얻을 수 있다.

$$\phi \leq \phi + \Delta\phi, \quad \Delta\phi = 2\pi f / f_s \quad (10)$$

참조 테이블의 입력으로 사용되는 위상값을 $0 \leq \phi < 2\pi$ 의 범위안에 있도록 하기 위해, 위상증분의 계산을 식 (10)과 같이 수행하는 대신 참조 테이블의 데이터 갯수에 따라 식 (11)과 같이 조정한다면 2π 의 범위를 벗어나는 위상을 자동으로 범위안에 들어오게 할 수 있다.

$$\Delta\phi = 2^n * f / f_s \quad (11)$$

여기서 n 은 테이블의 데이터를 저장하는 번지수의 비트수를 나타낸다. 예를 들어 참조 테이블의 총 데이터 갯수가 1,024개 ($n=10$)라면, 라디안 (radian)으로 0인 위상은 번지수 0으로, 라디안으로 π 인 위상은 번지수 512로 각기 대응된다. 계산상으로 3π 처럼 위상값이 참조 테이블의 번지수 범위를 벗어날 경우에도, 위상값의 하위 n 개의 비트만을 취하면 위상 값 π 를 지정하는 번지수와 동일한 512가 된다.

2. 엔빌로프 발생기

엔빌로프 발생기는 매 샘플링 순간마다 각 반송파의 진폭을 발생한다. 본 논문에서 사용한 엔빌로프 발생기는 엔빌로프를 프레임별로 구간을 나누어 각 구간을 직선으로 근사화한 방식을 사용하였다. 이 방식은 각 프레임에서의 진폭의 기울기만 파라미터로 제공하여 매 샘플링 순간마다 해당 프레임의 진폭의 기울기에 해당하는 증감분을 이전의 진폭에 더하면 한번의 덧셈만으로 진폭을 계산할 수 있다. 그림 8은 각 프레임을 직선으로 근사화시킨 시변진폭을 보여준다.

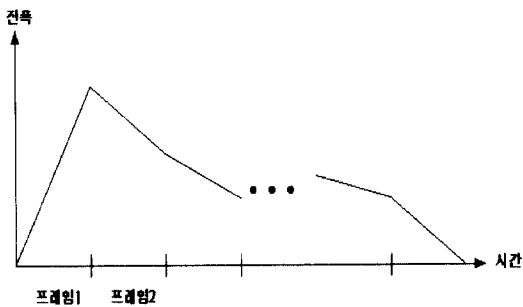


그림 8. 각 프레임마다 직선으로 근사화시킨 시변진폭

Fig. 8. Time-varying amplitude approximated by line segments according to the time frame.

VI. DSP를 이용한 FM 악기음 합성기의 구현

본 논문에서는 여러개의 반송파를 포함하는 FM 방식에 의한 악기음 합성을 실시간으로 구현하기 위하여 Ariel사의 DSP-16 보드(board)를 사용하였다^[13,14]. DSP-16 보드는 Texas Instrument사의 TMS320C25를 사용한 디지털 신호처리 전용 보드이다. DSP-16 보드는 그림 9에서 나타낸 바와 같이 악기음 합성에 필요한 DSP (TMS320C25), Program/Data RAM, DAC를 내장하고 있다. DSP의 처리 속도는 12.8 MIPS이고 한 채널당 샘플링 속도는 50 KHz이다.

그림 10에서 보는 바와 같이 FM 방식을 이용한 악기음 합성시 필요한 파라미터들은 PC에서 DSP-16 보드로 전송된다. 또한 FM 방식을 이용한 악기음 합성 프로그램이 프로그램 메모리에 다운로드 (download)

되고, 파라미터는 DSP를 통해 데이터 메모리에 저장된다.

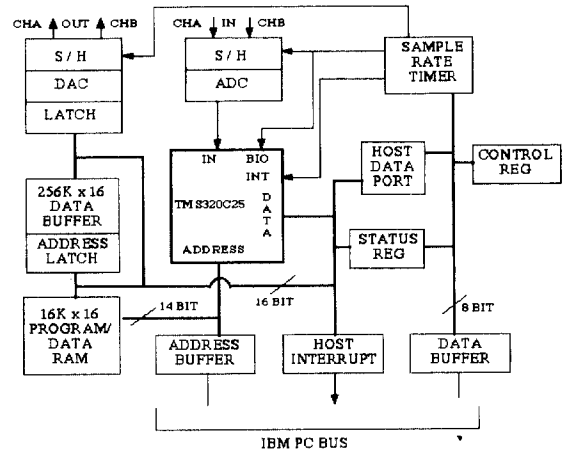


그림 9. DSP-16 하드웨어의 블록 다이어그램
Fig. 9. The block diagram of DSP-16 hardware.

한 음을 합성할 때 요구되는 데이터 즉, 음정과 박자가 PC의 MIDI 파일로부터 전송되면 DSP는 매 샘플링마다 V장에서 제시한 알고리즘을 통하여 계산된 악기음을 DAC를 거쳐 스피커 (speaker)로 출력시킨다. 즉, 미리 저장된 악기의 파라미터를 이용하여, MIDI 파일로부터 음정 및 박자에 대한 정보가 전송될 때마다 실시간으로 악기음을 합성하므로써 DSP가 음원 모듈의 역할을 하도록 구현하였다.

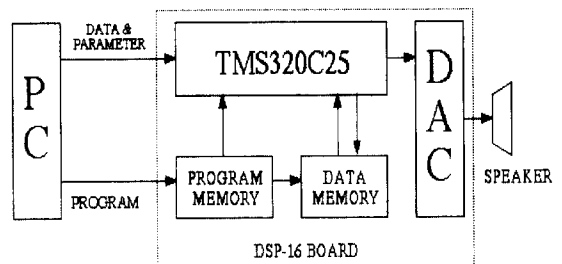


그림 10. PC와 DSP-16과의 인터페이스
Fig. 10. Interface between personal computer and DSP-16.

VII. 파라미터 추출 및 악기음 합성

본 논문에서의 대상 악기는 오보에와 호른이며, 로랜드 (Roland)사의 음원 모듈인 사운드 캠퍼스와 매킨

토시 컴퓨터에서의 악기음 데이터를 기준으로 하였다. 이들은 데이터 분해능 (resolution)이 8비트이고 샘플링 주파수가 22 KHz인 PCM 데이터이다. 2개의 악기음 데이터를 가지고 IV장에서 논의한 파라미터 추출 알고리즘을 이용하여 파라미터를 추출하고 추출한 파라미터를 VI장에서 논의한 DSP를 이용한 실시간 악기음 합성기에 적용하여 악기음을 합성하였다. 먼저, 악기음 합성시 필요한 파라미터를 추출하기 위하여 반송파의 갯수를 결정하여야 한다. 반송파의 갯수가 많아지면 실제 악기음을 보다 잘 표현할 수 있지만, 합성시에 많은 계산량을 요구하기 때문에 실시간 처리가 어려워진다. 따라서, 반송파의 갯수는 실시간 처리 여부와 음질을 모두 고려하여 결정해야 한다. 그림 11은 반송파 갯수에 따른 적합도 함수 값을 보여준다. 반송파의 갯수가 증가할수록 실제 음과의 오차를 표현하는 적합도 함수가 감소하여 음질이 향상되며 두 악기 모두 3개 이상에서는 음질의 차이가 거의 없음을 알 수 있다. 따라서, 본 논문에서는 실시간 처리를 위하여 두 악기 모두 3개의 반송파를 이용하여 악기음 합성에 필요한 파라미터를 추출하였다.

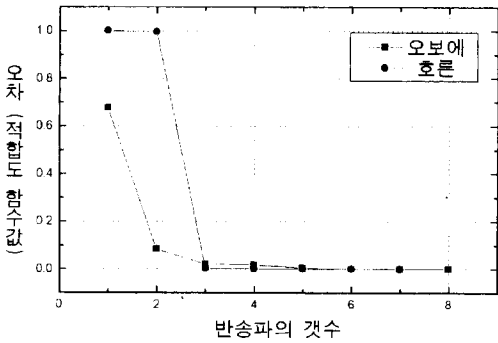


그림 11. 반송파 갯수에 따른 오차
Fig. 11. Error with increasing numbers of carriers.

두개의 기준 악기음 데이터로부터 IV장에서 논의한 알고리즘에 의하여 추출된 3개의 반송파에 대한 파라미터를 그림 12와 그림 13에 나타냈다. 그림 12는 오보에의 각 반송파의 크기 (CAR1, CAR2, CAR3)와 변조지수 (I1, I2, I3) 및 반송 주파수와 변조 주파수의 비 (N1, N2, N3)를 나타내고, 마찬가지로 그림 13은 호른에 대한 파라미터를 보여준다.

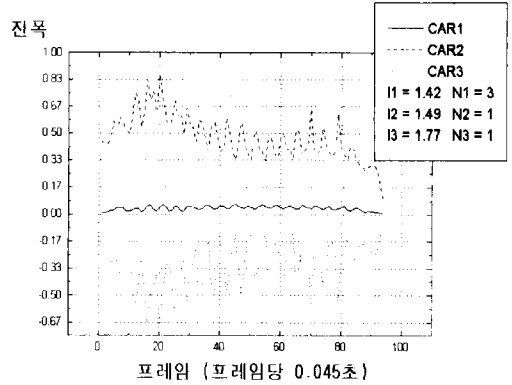


그림 12. 오보에의 파라미터 값
Fig. 12. The parameter values of oboe.

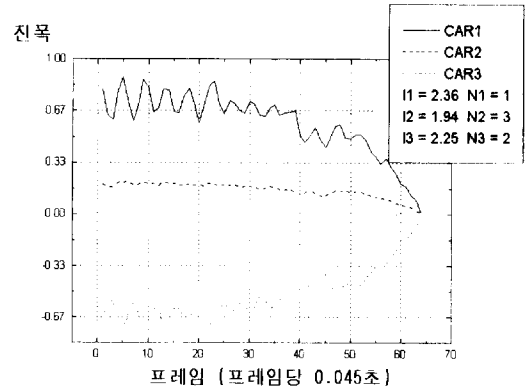


그림 13. 호른의 파라미터 값
Fig. 13. The parameter values of horn.

추출된 파라미터를 VI장에서 논의된 DSP를 이용한 실시간 악기음 합성기에 적용하여 악기음을 합성하였다. 악기음 합성 프로그램을 구성하는 모든 인스트럭션의 사이클 수는 1038 개로 12.8 MIPS인 TMS320C25를 이용하여 한 음을 합성하는데 걸리는 시간은 약 0.081 ms이므로 샘플링 주파수를 11 kHz로 하여 실시간 악기음 합성을 할 수 있었다. 합성된 악기음의 평가는 먼저 주관적 평가 (subjective test)로 행해졌다. 합성된 악기음은 PCM 방식에 의한 음질 보다는 이미 알려진 바와 같이 약간 떨어지나 각 악기의 독특한 특성을 쉽게 확인할 수 있었다. 평가를 보다 객관적으로 행하기 위하여 본 논문에서 합성한 악기음

과 기존의 방식¹⁸⁾으로 추출한 파라미터로써 FM 합성된 악기음을 각각 주파수 영역에서 오차를 구하여 비교하였다. IV장에서 파라미터 추출시 언급된 적합도 합수를 나타내는 식 (9)의 $\sum a_{k,j} \omega_{j,r}$ 에 합성된 오보에 악기음의 k번째 고조파의 크기를 대입하여 비교한 결과 본 논문에서 합성한 악기음의 기준음에 대한 오차제곱값은 약 0.13이었고 기존의 방법으로 합성한 악기음의 경우 오차제곱값은 약 0.31로 약 2배 이상 향상됨을 알 수 있었다. 이때 4096 FFT가 사용되었다.

그림 14 ~ 그림 21은 기준으로 사용된 악기음과 이로부터 추출된 파라미터로 합성한 악기음의 시간 영역에서의 파형 및 주파수 영역에서의 스펙트럼을 나타낸다. 그림 14와 15는 오보에의 기준 악기음과 합성음의 파형 비교로서 그림 14는 파라미터를 추출하기 위해 기준이 된 PCM 데이터의 파형을 보여주며, 그림 15는 그림 14의 파형으로부터 추출된 파라미터에 의해 합성된 파형을 보여준다. 합성된 악기음의 파형은 시간에 따라 엔빌로프가 조금씩 감소하면서 진폭에 약간의 오실레이션이 발생하는 반면, 기준 악기음은 일정한 크기를 유지하는 것을 볼 수 있다. 그림 16과 17은 오보에의 기준 악기음과 합성음의 스펙트럼 비교로서 그림 16은 기준이 된 PCM 데이터의 스펙트럼을 보여주며, 그림 17은 추출된 파라미터에 의해 합성된 악기음의 스펙트럼을 보여준다. 합성된 악기음이 3 KHz 이상의 주파수에서 실제의 악기음보다 고조파 성분들이 다소 작은 것을 볼 수 있다. 그림 18과 19는 호른의 기준 악기음과 합성된 악기음의 파형을 보여주는데 파형이 거의 유사함을 볼 수 있다. 그림 20과 21은 호른의 기준 악기음과 합성된 악기음의 스펙트럼을 나타낸다. 합성된 악기음 스펙트럼의 두번째 고조파가 기준 악기음의 스펙트럼보다 작은 것을 볼 수 있다.

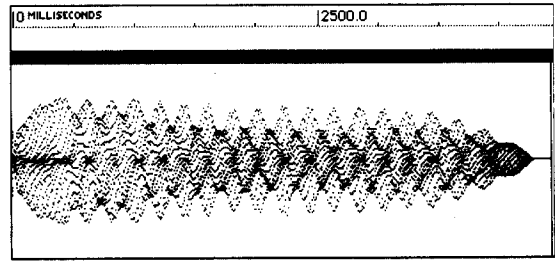


그림 15. 합성된 오보에 악기음의 파형
Fig. 15. The synthesized waveform of oboe.

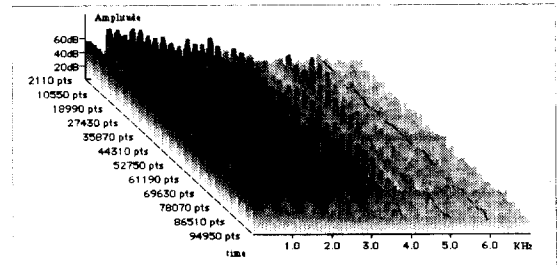


그림 16. 기준이 된 오보에 악기음의 스펙트럼
Fig. 16. The standard spectrum of oboe.

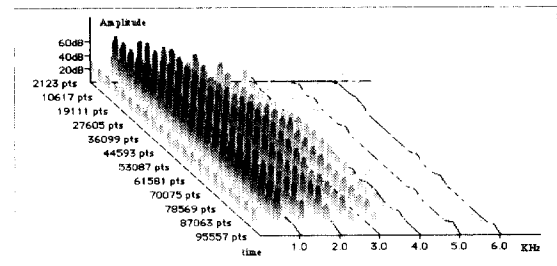


그림 17. 합성된 오보에 악기음의 스펙트럼
Fig. 17. The synthesized spectrum of oboe.

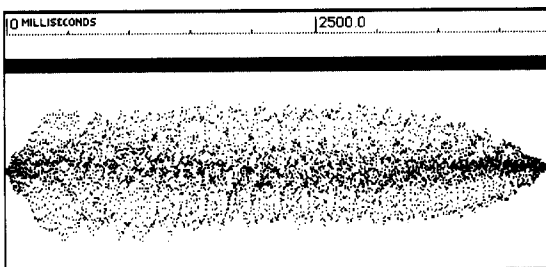


그림 14. 기준이 된 오보에 악기음의 파형
Fig. 14. The standard waveform of oboe.

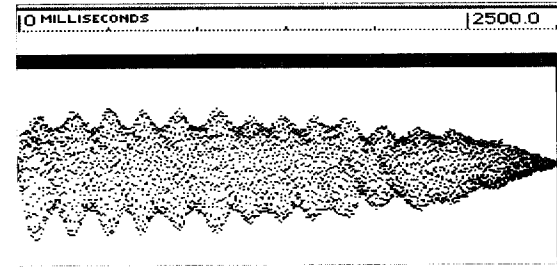


그림 18. 기준이 된 호른 악기음의 파형
Fig. 18. The standard waveform of horn.

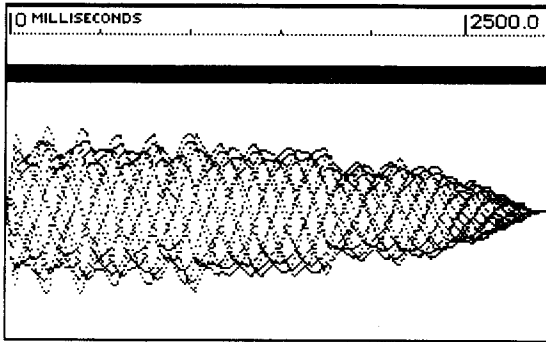


그림 19. 합성된 호른 악기음의 파형
Fig. 19. The synthesized waveform of horn.

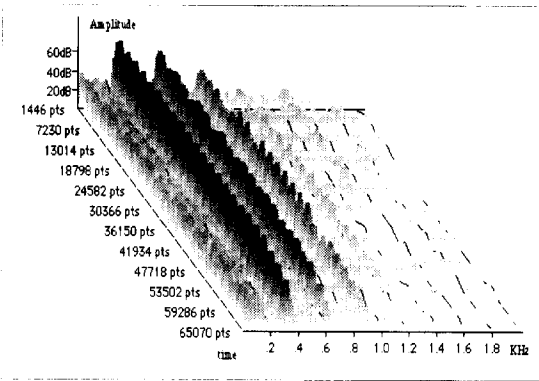


그림 20. 기준이 된 호른 악기음의 스펙트럼
Fig. 20. The standard spectrum of horn.

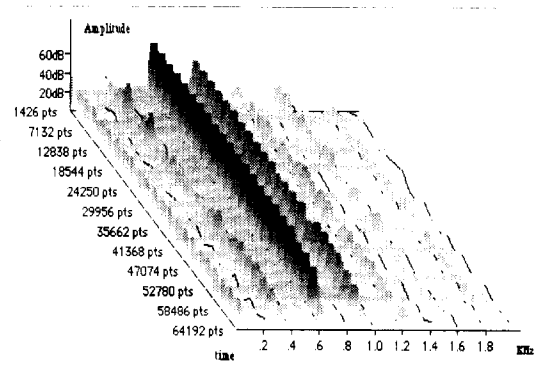


그림 21. 합성된 호른 악기음의 스펙트럼
Fig. 21. The synthesized spectrum of horn.

VIII. 결론

본 논문에서는 FM을 이용한 악기음 합성방식의 반

송파와 변조파를 각각 하나씩 이용하여 악기음을 모델링하는 기존의 방식보다 다양하고 복잡한 실제 악기음에 보다 가깝게 악기음을 합성하기 위하여 여러개의 반송파와 하나의 변조파를 갖는 개선된 FM 식으로 악기음을 모델링하고 이를 실시간으로 합성하기 위해 범용 디지털 신호처리기 (DSP)를 이용하여 악기음 합성기를 구현하였다. 특히 합성기를 구성하는 정현파 발생기, 엔빌로프 발생기 등의 구현 알고리즘은 ASIC 설계에 대비하여 하드웨어 구현이 용이하게 구현하였다. FM 방식을 이용한 악기음 합성은 비교적 계산량이 적게 요구되며 회로 구성이 간단하고, 여러 악기음을 합성하는데 필요한 파라미터의 수가 적어 메모리를 적게 차지하는 효율적인 방식이지만 각 악기에 해당하는 정확한 파라미터를 얻을 수 없다면 원하는 악기음을 얻을 수 없다. 따라서 각 악기에 해당하는 최적의 파라미터들을 추출하기 위하여 PCM 데이터를 기준으로 주파수 영역에서 스펙트럼을 분석하고, 이에 유전 알고리즘을 적용하여 여러개의 반송파와 한개의 변조파를 갖는 FM 악기음 합성을 위한 최적 파라미터 추출 방법을 사용하였다. 파라미터 추출 알고리즘은 기존의 경험을 바탕으로 한 방법보다 체계적으로 최적의 파라미터를 추출할 수 있다. 파라미터 추출 알고리즘을 이용하여 2가지 악기에 대한 파라미터를 추출하였고, 추출한 파라미터를 DSP를 이용한 FM 방식의 실시간 악기음 합성기에 적용하여 악기음을 합성하였다. 합성음을 직접 듣는 방법 외에 파라미터 추출시 기준으로 사용한 악기음과 이로부터 추출한 파라미터로 합성된 악기음을 시간 영역 및 주파수 영역에서 비교 분석하므로써 파라미터 추출 알고리즘 및 DSP를 이용한 실시간 악기음 합성기의 유효성을 확인하였다.

참고 문헌

- [1] T.D. Rossing, *The Science of Sound*, Addison-Wesley Publishing Company, 1989.
- [2] 新井 純, *The Theory of Synthesizer Operation*, Rittor Music, 1988.
- [3] C. Roads and S. Strawn, *Foundation of Computer Music*, MIT press, 1985.
- [4] J.A. Moorer, "Signal processing aspects of computer music: A survey," Pro-

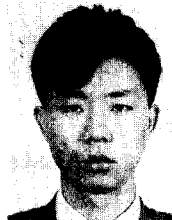
- ceeding of The IEEE, August 1977.
- [5] R.J. Higgins, *Digital Signal Processing in VLSI*, Prentice-Hall, 1990.
 - [6] J. Justice, "Analytic signal processing in music computation," *IEEE Trans. Acoust., Speech, and Signal Process.* vol. 27, no. 6, pp. 670-684, 1979.
 - [7] J. Beauchamp, "Synthesis by amplitude and brightness matching of analyzed musical instrument tones," *Journal of the audio engineering society*, vol. 30, no. 6, pp. 396-406, 1982.
 - [8] 주준열, 이문형, 김기두, "FM 방식의 디지털 악기음 합성을 위한 소프트웨어 시뮬레이터 및 파라미터 추출 알고리즘 개발," *대한전자공학회 논문지*, 제31권, 제3호, 1994년 3월
 - [9] D. E. Goldberg, *Genetic Algorithms in Search, Optimization, and Machine Learning*, Addison-Wesley, 1989.
 - [10] M. LeBrun, "A derivation of the spectrum of FM with a complex modulating wave," *Computer Music J.* 1, no. 4, 1977.
 - [11] A. Horner, J. Beauchamp, and L. Haken, "Genetic algorithm and their application to FM matching synthesis," *Computer Music J.* 17, no. 4, 1993.
 - [12] 백창기, 주세철, 이문형, 김기두, "디지털 신호 처리기 (DSP)를 이용한 FM 악기음 합성기의 구현," *제6회 신호처리합동학술대회 논문집*, vol. 6, no. 1, pp. 3-7, 1993
 - [13] *DSP-16 Data Acquisition Processor Manual*, Ariel Co., 1986.
 - [14] *TMS320C2X User's Guide*, 1991.

— 저 자 소 개 —



朱世哲(準會員)

1971年 2月 3日生. 1994年 2月 국민대학교 전자공학과 졸업(학사). 1994年 3月~현재 국민대학교 대학원 전자공학과 석사과정. 주관심 분야는 디지털 신호처리, GPS 시스템 등임.



金鎮範(準會員)

1971年 12月 4日生. 1994年 2月 국민대학교 전자공학과 졸업(학사). 1994年 3月 ~ 현재 국민대학교 대학원 전자공학과 석사과정. 주관심 분야는 디지털 신호처리, 멀티미디어 등임.

金基斗(正會員) 第 31卷 B編 第 3號 參照.

현재 국민대학교 전자공학과 부교수