

선분 근사법과 곡선의 적합성을 이용한 피아노 음의 모델링

Modeling of Piano Sound Using Method of Line-Segment Approximation and Curve Fitting

임 훈*, 정 의 필*
(Hun Lim*, Ui Pil Chong*)

* 이 연구는 정보통신부에서 지원하는 대학기초연구 지원사업에 의해 연구되었습니다.

요 약

본 논문에서는 피아노 음을 FFT(Fast Fourier Transform)를 이용하여 주파수 영역으로 변환한 후, 크기(magnitude)와 위상(phase)에 대한 특성들을 분석한다. 이 분석한 결과로 크기와 위상을 모델링하여 파라미터를 생성한다. 크기의 특성에서 기본주파수와 고조파 부분은 다른 부분에 비해 크기가 매우 크다. 그래서 이 부분의 모델링은 오차를 줄이기 위해 곡선의 적합성(Curve Fitting) 방법을 이용하였고, 노이즈 부분의 모델링은 기본주파수 전후의 노이즈가 피아노 음색 특성에 중요한 역할을 하고 불규칙하므로 선분 근사법을 이용하였으며, 크기가 작고 완만하게 변하는 나머지 노이즈 부분은 곡선의 적합성을 이용하였다. 위상은 크기에 적용한 방법과 동일하게 적용하였다. 생성된 파라미터로 크기와 위상을 만들어 역변환 FFT를 하면 피아노 원음에 매우 근접한 음을 얻을 수 있다.

핵심용어: 컴퓨터 음악, 음원 모델링, 피아노 음

ABSTRACT

In this paper, we will discuss the characteristics of the magnitude and the phase of the piano sound in frequency domain by using the FFT(Fast Fourier Transform). The method deciding the parameters representing those sounds through the mathematical model is described. We used the curve fitting method for the modeling of the harmonic part of the sound including the fundamental frequency in order to minimize the errors between original sounds and modeled sounds. Furthermore, we used the line segment approximation method for the modeling of the noise part around fundamental frequency. We also applied the same method for the phase model and could get the modeled sound to be similar to the original sound using the parameters. Therefore the high compression ratio comparing the modeled sound to the original sound is achieved.

Key words: computer music, sound source modeling, piano sound

투고분야: 음악음향 및 심리음향(8.1)

I. 서 론

악기음을 생성하고 합성하는 방법은 가산합성(additive synthesis), 감산합성(subtractive synthesis), FM(Frequency Modulation), waveshaping에 의한 합성방식, 샘플링에 의한 합성방식, SMS(Spectral Modeling Synthesis) 방식들이 있다[1~4].

본 논문에서는 악기 원음을 주파수영역에서 모델링하여 원음에 근접한 음을 만들 수 있는 SMS 방식을 사용하였다. 기존의 SMS 방식은 주파수 영역에서 정현파 성분을 추출하여 모델링을 하고, 나머지 노이즈 부분은 선분 근사법(Line-Segment Approximation)을 이용하여 모델링 하는

방식[3,4]이고, 본 연구는 주파수 영역에서 음에 대한 특성을 분석하여 곡선의 적합성과 선분 근사법을 이용하여 모델링 한다. 기존의 방식과 비교할 때 파라미터의 수가 적어 모든 피아노 음에 대한 데이터베이스화가 가능하고, 정현파 성분을 추출하는 과정이 없어서 구현하기가 쉽다는 장점이 있다.

본 논문의 구성은 다음과 같다. 2절은 SMS 방식에 대한 간단한 소개와 피아노 원음의 주파수영역에서의 특성을 분석하여 기술하였고, 음의 크기와 위상에 대한 모델링을 하기 위해 이용한 곡선의 적합성 방법과 선분 근사법에 대한 내용, 그리고 적용하는 방법에 대해 소개하고, 3절은 결과를 분석하며, 4절에서는 결론을 내렸다.

II. 본 론

2.1. SMS(Spectral Modeling Synthesis)방식

* 울산대학교 컴퓨터공학과

접수일자: 2000년 1월 7일

피아노의 음을 DAT(Digital Audio Tape)로 녹음하면 아날로그의 음을 디지털 신호로 변환할 수 있으며, 그 변환된 신호인 피아노 음, $x(n)$ 를 식 (1)의 FFT를 이용하여 주파수영역의 데이터 $X(k)$ 로 변환한다.

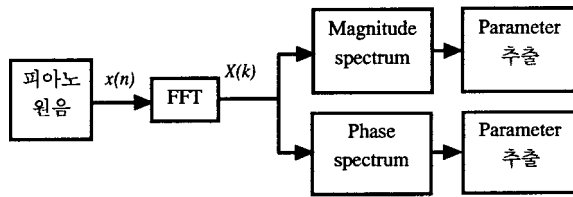
$$X(k) = \sum_{n=0}^{N-1} x(n)e^{-j2\pi kn/N}, \quad 0 \leq k \leq N-1. \quad (1)$$

주파수영역의 데이터 $X(k)$ 는 식 (2)와 같이 크기(magnitude) $|X(k)|$ 와 위상(phase) $\angle X(k)$ 으로 표현할 수 있다.

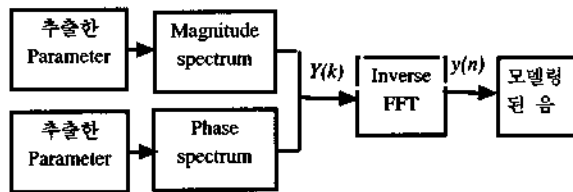
$$X(k) = |X(k)|e^{j\angle X(k)} \quad (2)$$

그림 1(a)에서 피아노 음의 크기와 위상에 대한 모델링을 통하여 각각 파라미터를 추출한다.

그림 1(b)에서는 추출된 파라미터로 모델링 되어진 주파수영역의 데이터로 나타낼 수 있으며, 이 데이터 $Y(k)$ 를 식 (3)의 Inverse FFT를 이용하여 시간영역의 데이터 $y(n)$ 으로 변환할 수 있다. 여기서 얻어진 데이터 $y(n)$ 은 피아노 원음에 매우 근접한 음이 된다.



(a) Analysis 부분의 Block diagram



(b) Synthesis 부분의 Block diagram

그림 1. SMS(Spectral Modeling Synthesis)방식에 대한 Block diagram

Fig. 1. Block diagram for SMS; (a) Analysis part, (b) Synthesis part.

$$y(n) = \frac{1}{N} \sum_{k=0}^{N-1} Y(k)e^{j2\pi kn/N}, \quad 0 \leq n \leq N-1. \quad (3)$$

2.2. 피아노 원음의 주파수영역에서의 특성 분석

모든 피아노 원음은 주파수 영역에서의 크기와 위상에 대한 공통적인 특성을 가지고 있다. 여기서는 피아노 A4(라) 음을 예로 들어 그림과 함께 특성을 간략히 설명하고, 이 특성들에 의해 부분적으로 모델링할 방법을 선택하게 된다.

2.2.1. 크기에 대한 특성분석

그림 2(a)는 샘플링 주파수(sampling frequency)가 44100Hz인 피아노 A4 원음에 대한 크기를 보여주는데, 여기서 22050Hz를 중심으로 대칭이고, 5000Hz 이상인 부분의 크기는 거의 0에 가깝다.

그림 2(b)는 주파수가 0~5000Hz까지의 그림 2(a)의 일부분의 크기를 보여주는데, 피아노 A4음의 기본주파수(fundamental frequency)는 440Hz인 것을 알 수 있다. 그리고 기본주파수의 배수가 되는 고조파(harmonic)가 존재하는데, 이 부분에 대한 크기가 다른 부분에 비해 매우 크므로 모델링을 할 때 오차를 작게 하여야 한다.

그림 2(c)는 기본주파수와 기본주파수 전후에 있는 노이즈 부분의 크기를 보여 주는데 다른 부분의 노이즈에 비해 이 부분의 노이즈가 피아노 음색 특성에 중요한 역할을 하므로 모델링을 잘 해 주어야 한다.

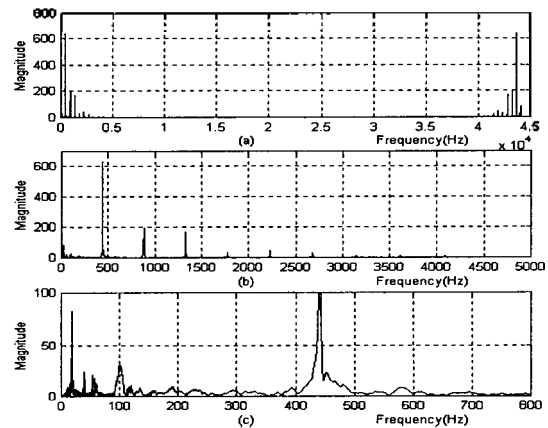


그림 2. 피아노 A4 원음에 대한 크기
Fig. 2. Magnitude for A4 original sound of piano.

2.2.2. 위상에 대한 특성분석

그림 3(a)는 피아노 A4 원음에 대한 위상을 나타내며, 여기서 22050Hz를 중심으로 대칭인 부분의 위상은 서로 부호만 다르다.

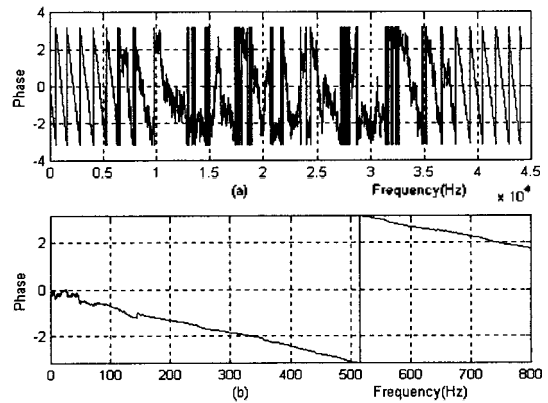


그림 3. 피아노 A4 원음에 대한 위상
Fig. 3. Phase for A4 original sound of piano.

그림 3(b)는 그림 2(c)와 같은 주파수 영역의 위상을 나타내며, 0~200Hz 부분은 위상의 분포가 비선형이고, 크기의 노이즈 중요부분과 관련되어 음색에 영향을 주므로 모델링시에 충분히 고려되어야 하며, 나머지 부분은 거의 선형에 가깝다.

2.3. 크기와 위상에 대한 모델링

크기와 위상을 모델링 하기 위해 변화가 완만한 부분에는 곡선의 적합성 방법과 급격하게 변화하는 부분에는 선분 근사법을 사용하였으며, 이 두 가지 방법에 대해 간략히 설명하고, 크기와 위상에 적용한 내용에 대해 설명한다.

2.3.1. Curve Fitting(곡선의 적합성)

음의 모델링에 곡선의 적합성을 이용한다는 것은 여러 개의 데이터가 주어졌을 경우 데이터의 특성 또는 경향을 잘 대표할 수 있는 점근 함수를 찾아내는 방법을 말한다. 여기서 적합시킬 수 있는 함수의 선정이 중요하며, 본 연구에서는 직교 다항식의 조건을 만족시키는 Gram-Schmidt 다항식[5]을 이용하였다.

$$p_{m,n}(u) = \sum_{k=0}^m (-1)^k \binom{m}{k} \binom{m+k}{k} \binom{u}{k} \binom{n}{k} \quad (4)$$

단, m 은 직교 다항식의 차수이며, 데이터의 개수는 $(n+1)$ 개이고, $u=0, 1, \dots, n$ 이 된다. 식 (4)를 전개하면 다음과 같이 쓸 수 있다.

$$p_{0,n}(u) = 1$$

$$p_{1,n}(u) = 1 - \frac{2u}{n}$$

$$p_{2,n}(u) = 1 - \frac{6u}{n} + \frac{6u(u-1)}{n(n-1)}$$

$$p_{3,n}(u) = 1 - \frac{12u}{n} + \frac{30u(u-1)}{n(n-1)} - \frac{20u(u-1)(u-2)}{n(n-1)(n-2)}$$

... ..

$$g(u) = c_0 p_{0,n}(u) + c_1 p_{1,n}(u) + \dots + c_r p_{r,n}(u) \quad (5)$$

이 다항식은 차수를 높여주면 원래 곡선과의 오차는 작아지지만, 계산량이 많아진다. 실험결과 식 (5)와 같은 7차인 다항식과 5차 다항식을 이용하여 계수를 구하고, 이것을 파라미터로 사용하였다.

그림 4(a)는 기본주파수의 크기가 최고인 점 이전의 50개의 데이터를 보여주고, 그림 4(b)는 점근함수, 7차 다항식을 이용하여 모델링한 결과이다.

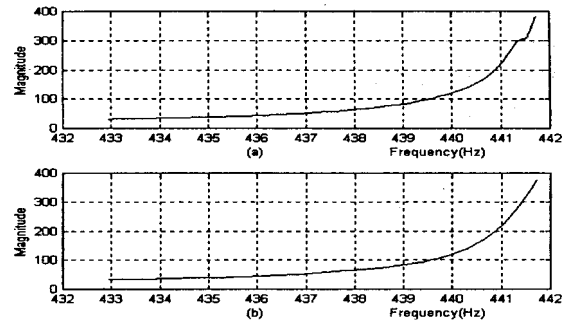


그림 4. 7차 다항식을 이용한 모델링의 예
(a) 기본주파수의 앞부분에 있는 50개의 데이터
(b) 모델링을 한 결과
Fig. 4. Modeling example using the 7th polynomial;
(a) fifty data of the previous fundamental frequency,
(b) modeling result.

2.3.2. Line-Segment Approximation(선분 근사법)

선분 근사법(Line-Segment Approximation)이란 어떤 복잡하고 불규칙하게 변화하는 데이터에 대해 최고점들이나 최저점들을 찾아 서로간에 직선으로 연결하여 복잡한 데이터에 대해 많은 변화를 주지는 않으면서 근접한 데이터로 변환하는 방법을 말한다.

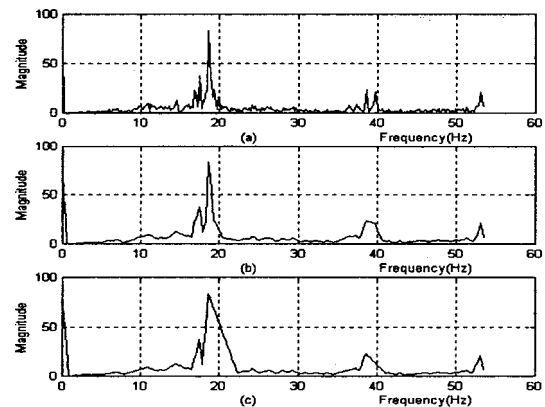


그림 5. 선분 근사법을 이용한 모델링의 예
(a) 기본주파수 앞에 있는 노이즈의 일부분(300개)
(b) 최고점들을 찾아 선형으로 연결
(c) 최고점들과 최저점들을 찾아 선형으로 연결
Fig. 5. Modeling example using Line-Segment approximation;
(a) original,
(b) connecting the each maximum point,
(c) connecting the each maximum & minimum point.

복잡한 노이즈 부분(그림 5(a))의 모델링은 오차를 줄이기 위하여 곡선의 적합성 대신 선분 근사법을 적용하였다. 그림 5(a)에서 크기가 커지다가 작아지는 최고점들을 찾아 서로 직선으로 연결하면 그림 5(b)와 같이 되고, 파라미터의 수를 더 줄이기 위해 그림 5(b)에서 크기가

커지다가 작아지는 최고점들과 크기가 작아지다가 커지는 최저점들을 찾아 서로 직선으로 연결하여 그림 5(c)와 같이 노이즈 부분을 모델링하였다.

2.3.3. 크기에 대한 모델링

기본주파수와 고조파들(8개)의 부분에서 가장 높은 점을 찾아 그 점을 파라미터로 사용하고, 그 점을 제외한 전후의 50개의 데이터를 가지고 곡선의 적합성을 이용하여 그 계수들을 파라미터로 사용한다. 기본주파수 전후의 노이즈 부분은 선분 근사법을 이용하여 최고점 및 최저점들을 파라미터로 사용하고, 나머지 노이즈 부분은 두 부분으로 나누어 곡선의 적합성을 이용하여 그 계수들을 파라미터로 사용하였다.

이 방법은 모든 피아노 음에 적용이 가능하고, 이 파라미터들을 이용하여 크기에 대해 모델링을 하였다. 그리고 5000Hz정도(모델링을 한 마지막 고조파 부분)부터 22050Hz까지의 데이터는 아주 작은 값이므로 모두 0으로 처리하고, 0~22050Hz의 모델링을 한 데이터를 가지고 대칭적으로 사용해서 22051~44100Hz까지의 크기에 대한 모델링을 할 수 있다. 그림 6(a)은 크기에 대한 모델링 결과를 보여주고 있다.

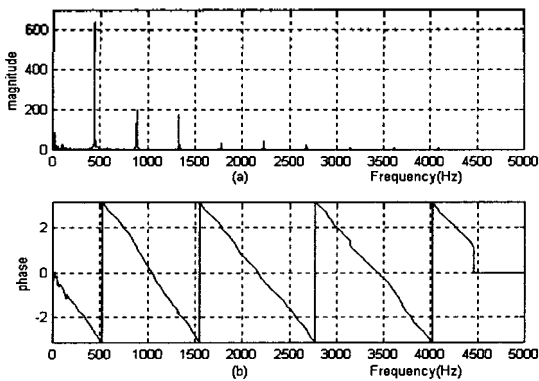


그림 6. 크기와 위상에 대한 모델링 결과의 일부분
(a) 크기(0~5000Hz) (b) 위상(0~5000Hz)
Fig. 6. Modeling result of magnitude & phase;
(a) magnitude, (b) phase.

2.3.4. 위상에 대한 모델링

크기에 대한 모델링 결과를 바탕으로 크기에서 오차가 적어야 하는 부분은 위상에서도 오차가 적어야 하므로 위상에서 0~200Hz까지는 선분 근사법을 사용하여 파라미터를 구하고, 나머지 부분에서는 크기와 같은 범위에서 곡선의 적합성을 똑같이 사용하여 파라미터를 구한다. 이 파라미터를 이용하여 위상에 대해 모델링을 할 수 있다.

그림 6(b)은 위상에 대한 모델링 결과를 보여주고 있다.

III. 실험 결과

곡선의 적합성 방법에 있어서 다항식의 차수에 선택이 중요하다. 그래서 3차, 5차, 7차, 9차일 때의 계산량과 오차(RMSE)를 비교하여 그림 7과 그림 8에 나타내었다.

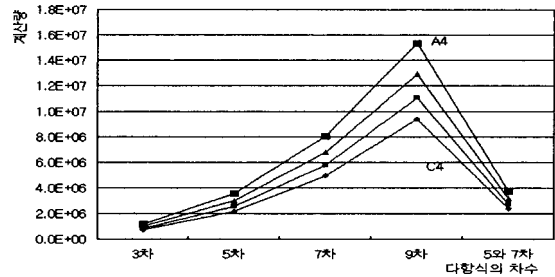


그림 7. 곡선의 적합성에 사용한 다항식의 차수에 따른 계산량
Fig. 7. Computational complexity of curve fitting method depend on the change of polynomial order.

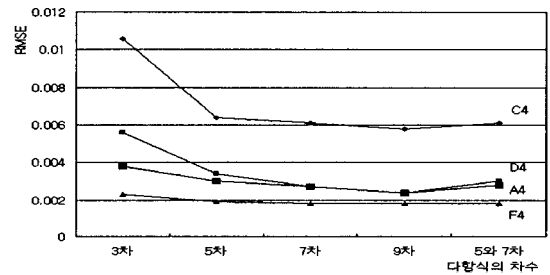


그림 8. 곡선의 적합성에 사용한 다항식의 차수에 따른 오차 (RMSE)
Fig. 8. RMSE of curve fitting method depend on the change of polynomial order.

그림 7, 8에서 볼 수 있듯이 다항식의 차수가 높아지면 계산량은 많아지고 오차는 줄어든다. 그래서 크기가 큰 부분(기본주파수와 고조파 전후의 50개 데이터)에는 오차를 줄이기 위해 7차 다항식을 적용하였고, 크기가 작은 부분(노이즈)에는 계산량을 줄이기 위해 5차 다항식을 적용하였다. 다항식에서 얻은 계수가 크기와 위상에 대한 파라미터가 된다.

모델링한 크기와 위상을 식 (2)에 적용시키면 주파수 영역의 데이터로 나타낼 수 있으며, 이 데이터를 식 (3)의 Inverse FFT를 이용하여 시간영역의 데이터로 변환할 수 있다. 그림 9(a)는 시간영역에서의 피아노 A4 원음이고, 데이터의 수는 246600개(샘플링 주파수가 44100Hz이므로 약 5.6초의 데이터)이다. 그림 9(b)는 모델링을 한 결과이다.

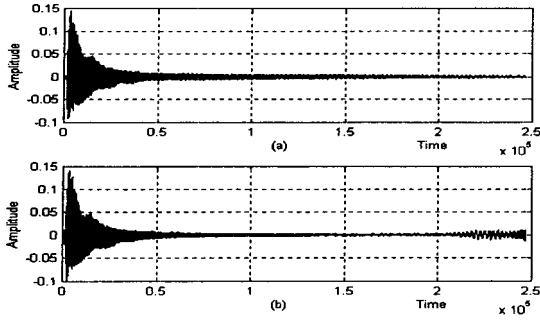


그림 9. 시간영역에서의 A4음
 (a) A4 원음 (b) 모델링한 결과
 Fig. 9. A4 sound of time domain;
 (a) A4 original sound, (b) modeling result.

그림 9의 결과를 두 주기만 자세히 비교해 보면 그림 10과 같이 매우 비슷한 것을 볼 수 있다.

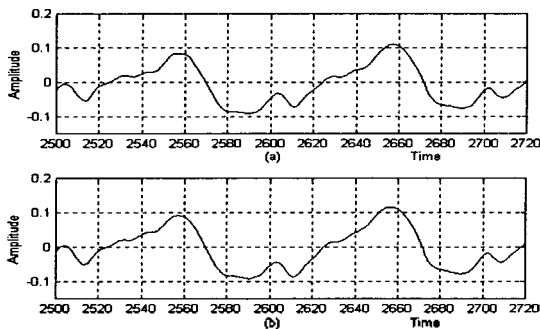


그림 10. 시간영역에서의 A4음의 일부분(두 주기)
 (a) A4 원음 (b) 모델링한 결과
 Fig. 10. Two periodic of A4 sound in time domain;
 (a) original, (b) modeling result.

피아노 원음(C4(도)음, D4(레)음, F4(파)음, A4(라)음)과 모델링 결과를 얻는데 필요한 파라미터의 데이터 수를 비교해 보면, 표 1과 같다.

표 1. 원음 데이터와 모델링한 데이터의 압축률과 오차
 Table 1. Compression ratio and RMSE between original sound and modeling sound.

	원음의 데이터 수 (Sampling 개수)	Parameter의 수 (magnitude+phase)	압축률	RMSE
C4	246501	1156(730+426)	213:1	0.0061
D4	246501	1256(824+432)	196:1	0.0030
F4	246600	1482(968+514)	166:1	0.0018
A4	246600	1354(878+476)	182:1	0.0028

피아노 원음과 모델링한 결과 사이의 시간영역에서의 데이터들의 오차를 식 (6)의 RMSE(Root Mean Squared Error)를 이용해서 구한 다음 표 1에 나타내었다.

$$RMSE = \sqrt{\frac{1}{N} \sum_{n=1}^N (x(n) - y(n))^2} \quad (6)$$

N : 데이터의 전체 개수

x(n) : 피아노 원음의 n번째 데이터

y(n) : 모델링한 결과의 n번째 데이터

그림 11(a)는 피아노 A4음이고, 그림 11(b)는 간단하게 만든 2초 길이의 envelop이며, 그림 11(c)는 그림 11(a), (b)를 이용하여 만든 2초 길이의 A4음이다. 다른 음들과 모델링하여 만들어진 음에도 그림 11과 같이 적용을 하여 연속적으로 이어주면 그림 12와 같은 결과를 얻을 수 있다.

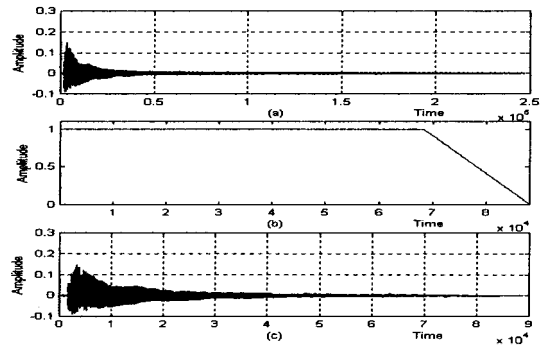


그림 11. 음을 envelop를 사용하여 2초 음으로 만들
 (a) 6초 길이의 음 (b) envelop (c) 2초 길이의 음
 Fig. 11. Change of sound duration;
 (a) 6 seconds length of the sound, (b) envelop,
 (c) 2 seconds length of the sound.

그림 12는 C4음, D4음, F4음, A4음에 대한 원음과 모델링하여 얻어진 음을 순서대로 연결하여 서로 비교해 본 결과이다.

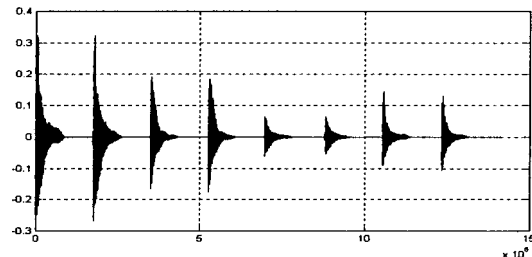


그림 12. 원음과 모델링한 음의 비교(C4, D4, F4, A4)
 Fig. 12. Comparison of original sound and modeling sound.

IV. 결 론

피아노 원음에 대한 주파수영역에서의 크기와 위상에 대한 특성을 분석하고, 피아노 음색에 중요한 노이즈 부분까지 모델링하여 원음에 근접한 음을 생성할 수가 있다. 피아노 원음 4개를 예로서 모델링한 결과 평균 압축률 190:1, RMSE는 0.0034의 평균값을 나타내었다. 모델링된 음을 실제 여러 사람들에게 들려줌으로서 원음과의 차이를 거의 느낄 수 없었다. 이 방법을 이용하면 많은 데이터를 가지는 피아노 음을 적은 파라미터를 사용하여 음을 생성할 수가 있으며, 이러한 음들을 데이터베이스화 하면 피아노 작곡 및 연주에 이용할 수가 있다.

참 고 문 헌

1. 권민도, 장호근 외, "FM 합성방식을 이용한 악기음 합성용 DSP 설계," 한국음향학회지 14권 5호, pp. 63-73, 1995.
2. 주준열, 이문형 외, "FM 방식의 디지털 악기음 합성을 위한 소프트웨어 시뮬레이터 및 파라미터 추출 알고리즘 개발," 전자공학회논문지 31권 3호, pp. 24-37, 1994.
3. X. Serra and J. Smith, "Spectral Modeling Synthesis : A Sound Analysis/Synthesis System based on a Deterministic plus Stochastic Decomposition," ComputerMusic Journal 14(4):12- 24, 1990.
4. X. Serra "A System for Sound Analysis/Transformation/Synthesis based on a Deterministic plus Stochastic Decomposition," Ph.D. Dissertation, Stanford University, 1989.
5. 김창근, Matlab을 이용한 수치해석, 교우사, pp. 251-286, 1997.

▲정 의 필(Uipil Chong)

1951년 9월 10일생



1978년 2월: 울산대학교 전기공학과 (공학사)

1980년 2월: 고려대학교 전기공학과 (공학석사)

1985년 6월: Oregon State University (공학석사)

1997년 1월: Polytechnic University (공학박사)

1997년 3월~현재: 울산대학교 컴퓨터정보통신공학부 교수

※ 주관심분야: 신호처리, 데이터압축, 컴퓨터음향

▲임 훈(Hun Lim)

1972년 6월 3일생



1998년 2월: 울산대학교 컴퓨터공학과 (공학사)

2000년 2월: 울산대학교 컴퓨터공학과 (공학석사)

2000년 1월~현재: 삼창기업(주) 제어기술연구소 연구원

※ 주관심분야: 디지털 신호처리, 디지털 신호압축, 컴퓨터음악