

캡스트럼 분석을 이용한 해금의 스펙트럼 모델링

Spectral Modeling of Haegeum Using Cepstral Analysis

홍연우*, 강명수*, 조상진**, 김종면*, 이정철*, 정의필*

(Yeonwoo Hong*, Myeongsu Kang*, Sangjin Cho**, Jongmyon Kim*, Jungchul Lee*, Uipil Chong*)

*울산대학교 컴퓨터정보통신공학부, **울산대학교 전기전자정보시스템공학부

(접수일자: 2010년 2월 8일; 채택일자: 2010년 3월 23일)

본 논문에서는 해금 소리의 시간에 따른 변화를 사실적으로 묘사하기 위해 캡스트럼 분석을 이용한 전통 악기 해금의 스펙트럼 모델링을 제안한다. 정확한 캡스트럼 분석 결과를 얻기 위해 프레임 사이즈는 입력 신호의 3주기로 하였고 포먼트 추출에 더 많은 캡스트럼 계수를 활용하였다. 정현파 성분 합성 과정에서 대역통과 필터의 차단주파수를 공명점 별로 유동적으로 조절하고 노이즈 성분에 남아있는 피크 성분들을 제거하는 과정을 추가하여 성능을 향상시켰다. 음 높이의 변화를 판단하기 위해 입력 프레임을 묵음구간, 어택구간, 지속구간으로 분류하였고 기본주파수에 따라 프레임 사이즈를 가변적으로 조절하였으며 지속구간에서의 기본주파수 검출 오류를 수정함으로써 정확도를 향상시켰다. 해금 연주 전문가의 청취테스트를 통해 원음과 합성음이 96~100 % 유사하다는 평가 결과를 얻었다.

핵심용어: 해금 소리 합성, 스펙트럼 모델링, 캡스트럼 분석, 캡스트럼 포락선, 포먼트 합성

투고분야: 음악음향 및 음향심리 분야 (8.6)

This paper proposes a spectral modeling of Korean traditional instrument, Haegeum, using cepstral analysis to naturally describe Haegeum sounds varying with time. To get a precise result of cepstral analysis, we set the frame size to 3 periods of input signal and more cepstral coefficients are used to extract formants. The performance is enhanced by flexibly controlling the cutoff frequency of bandpass filter depending on the resonances in the synthesis process of sinusoidal components and the deleting peaks remained in the residual signal. To detect the change of pitch, we divide the input frames into silence, attack, and sustain region and determine which region the current frame is involved in. Then, the proposed method readjusts the frame size according to the fundamental frequency in the case of the current frame is in attack region and corrects the extraction errors of the fundamental frequency for the frames in sustain region. With these processes, the synthesized sounds are much more similar to the originals. The evaluation result through the listening test by a Haegeum player says that the synthesized sounds are almost similar to originals (96~100 % similar to the original sounds).

Keywords: Sound synthesis of Haegeum, Spectral modeling, Cepstral Analysis, Cepstral envelope, Formant synthesis

ASK subject classification: Signal Musical Acoustics and Psychoacoustics (8.6)

1. 서론

스펙트럼 모델링은 입력 신호를 주파수 영역에서 재현하는 기술로 사람의 청각 기능에 주목하여 청자의 인지 메커니즘을 흉내내는 것이다 [1]. 최초의 스펙트럼 모델은 임의의 사운드는 정현파의 합으로 표현된다는 푸리에 이론에 바탕을 둔 정현파 모델이다. 가산 합성법이라고도 불리는 정현파 모델은 모음이나 현악기, 관악기와 같

은 주기성이 있는 사운드를 합성하는 데는 효과적이지만, 바람소리나 플룻에서의 숨소리 같은 노이즈 사운드나 타격음과 같은 천이 신호 (transient signal)에는 상대적으로 비효율적이다 [2]. 1930년대 Homer Dudley는 음성 전송 시 요구되는 대역폭을 줄이기 위해 보코더 (vocoder)를 개발하였는데, 이는 정현파 모델이나 노이즈 모델로 볼 수 있다 [3]. 1960년대에는 STFT (Short Time Fourier Transform)를 기초로 하여 위상을 쉽게 계산할 수 있는 위상 보코더가 소개되었으나 후에 정현파 모델로 대체되었다 [4]. 정현파 모델은 축소된 위상 보코더 할 수 있으며 보코더에서와 같이 유니폼 (uniform)

책임저자: 이 정 철 (jungcle@ulsan.ac.kr)

680-749 울산광역시 남구 대학로 93 울산대학교 컴퓨터정보통신공학부
(전화 : 052-259-1269; 팩스 052-259-1687)

필터 뱅크에서 순간 진폭과 위상을 추출하는 대신 STFT 스펙트럼에서 피크를 추출한다. 이러한 STFT 기반의 정현파 모델링을 악기 음 합성에 응용한 것은 Smith와 Serra에 의해 시작되었다.

80년대 중반에, Smith는 피치가 바뀌고 불협화음의 소리를 표현하기 위해 PARSHL이라는 프로그램을 개발하였다 [5]. 비슷한 시기에 Quatieri와 McAulay는 음성 분석을 위해 PARSHL과 유사한 기술을 개발했다 [6, 7]. 하지만 PARSHL 프로그램은 많은 악기 사운드의 어택(attack)과 같은 노이즈 신호를 표현하기에는 어려움이 있었다. Serra는 사운드를 정현파 성분과 노이즈 성분의 합으로 해석하여 분석, 변환, 합성하는 SMS (Spectral Modeling Synthesis)라는 스펙트럼 모델링 기술을 소개하였다 [8-10]. SMS는 후에 정현파 + 노이즈 모델에 천이 신호를 포함하는 정현파 + 노이즈 + 천이 모델로 개선되었다 [11-13].

음 합성 기술을 국악기에 적용한 연구로는 산조가야금의 물리적 모델링 [14-17], 태평소의 음향분석을 통한 팔랑 특성 추출 [18], 디지털 피리 구현을 위한 스펙트럼 모델링 [19, 20], 피리의 FM합성 [21, 22], 해금의 음향 분석과 포만트 합성 [23-26] 등이 있다. FM 합성 방식을 제외하면 물리적 모델링과 스펙트럼 모델링을 위한 것으로 특히 본 논문과 관련된 선행 연구는 [23-26] 정도이다.

본 논문에서는 선행 연구를 개선한 캡스트럼 분석을 이용한 해금 소리의 스펙트럼 모델링을 제안한다. 분석과 합성과정은 프레임 기반으로 처리되며 정확한 캡스트럼 분석 결과를 얻기 위해 프레임 사이즈는 입력 신호의 3주기로 하고 충분한 캡스트럼 계수를 활용한다. 기본주파수에 따라 프레임 사이즈를 조정하고 입력 프레임이 묶음 구간, 어택구간, 지속구간 중 어디에 속하는지를 판단하여 파라미터 추출과정을 달리함으로써 연주곡을 효율적으로 합성할 수 있다. 또한 청취테스트를 통해 합성음의 음질을 평가한다.

II. 기존 연구

스펙트럼 모델링 합성법은 시변 스펙트럼을 부분적으로 선형적이며 진폭과 주파수 포락선으로 표현할 수 있는 정현파 성분과 시변 필터링된 노이즈 성분으로 모델링하는 기술이다. 정현파 성분 (sinusoidal component, deterministic component)은 시스템의 주 진동 모드에 해당하

며 정현파가 아닌 에너지 성분과 여기신호에 의해 발생하는 에너지 성분은 노이즈 성분에 해당한다. 관악기의 경우 관대 내부의 진동에 의해 정현파 성분이 생성되며 노이즈 성분은 연주자가 좁은 구멍으로 공기를 불어넣을 때 생성된다.찰현악기의 경우 현의 주 진동 모드에 의한 결과로 정현파 성분이 생성되며 활과 현의 마찰과 공명기의 비선형적 현상에 의해 노이즈가 생성된다. 정현파는 스펙트럼 상에서 피크를 형성하게 되고 크기와 주파수의 함수로 표현된다. 노이즈의 경우 주파수에 대한 시그널 파워 (signal power)의 기댓값인 파워 스펙트럼 밀도 함수 (power spectral density)로 묘사할 수 있다. 정현파 성분과 노이즈 성분의 합으로 다양한 사운드를 합성할 수 있도록 한 기존의 스펙트럼 모델링 기술은 SMS라고도 칭한다. SMS의 경우 정현파 성분 합성을 위해 가산 합성법을 활용하는데, 가산합성 방식은 각 배음 성분의 주파수, 위상, 크기 정보를 모두 필요로 하므로 유효 배음의 수가 많을수록 원음에 거의 흡사한 소리를 만들어 낼 수는 있지만 합성시 요구되는 파라미터 수는 비례적으로 많아질 수밖에 없다.

전통 국악기 해금의 스펙트럼 모델링을 위해 캡스트럼 포락선을 이용한 포만트 합성법 [26]은 기존의 SMS에서 정현파 성분 합성을 위해 사용한 가산 합성법 대신 포만트 합성법을 적용하여 합성 과정에 필요한 파라미터의 수를 줄이고자 하였다. 캡스트럼 포락선에서 포만트 정보를 추출하고 IIT (Impulse Invariant Transform)에 의한 공명기를 정현파 성분의 합성에 활용하였다. 원음과 포만트 합성으로 재생된 정현파 성분과의 차이를 통해 노이즈 성분을 구하고 선형 보간법을 활용하여 노이즈 성분을 합성하였다. 이러한 과정을 통해 원음과 유사한 스펙트럼을 재현하였으나 녹음된 해금 단위음에서 지속 구간만을 파라미터 추출에 사용하였다. 즉, [26]의 방법에서는 입력 신호를 STFT 분석의 한 프레임이라 가정하여 분석하고 합성한 결과로, 해금 소리의 시간에 따른 변화를 사실적으로 묘사하기에는 제한적이었다.

III. 제안하는 모델

본 논문에서는 해금 소리의 시간에 따른 변화를 사실적으로 묘사하기 위해 캡스트럼 분석을 이용한 해금의 스펙트럼 모델링을 제안한다. 음 합성에 필요한 파라미터를 추출하기 위한 분석 과정은 그림 1과 같다. 분석과정은 프레임 기반으로 처리되며 입력 프레임의 길이는 입력

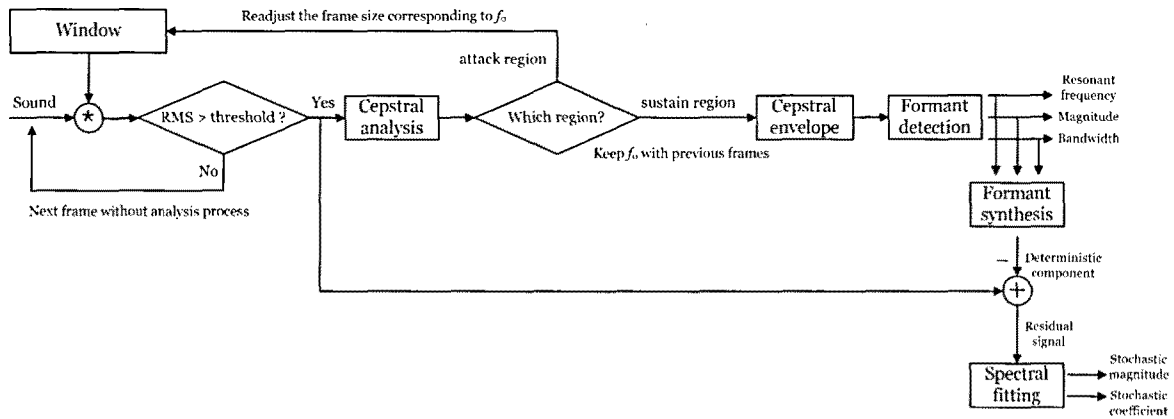


그림 1. 분석 과정 블럭도
Fig. 1. Block diagram of analysis process.

신호의 3주기로 하였다. 캡스트럼 분석을 통해 포만트 정보를 추출하게 되는데, 정확한 포만트 추출을 위해서는 스펙트럼 피크를 정확하게 묘사하는 포락선을 구할 필요가 있다. 캡스트럼 포락선을 구하는 과정에서 입력 샘플의 수가 많아지면 전체적인 에너지의 평균을 나타내게 되어 상대적으로 피크 주변의 크기가 커지게 된다. 다양한 입력 샘플 수에 대한 실험을 통해 3주기 정도의 입력 신호를 이용할 경우 피크가 정확하게 묘사됨을 확인할 수 있었다. [26]에서는 입력 신호를 STFT 분석에서의 한 프레임이라 보고 각 단위에서 500~1000 샘플 정도를 추출하여 활용하였는데, 이는 각 단위의 3주기 정도의 구간에 해당한다. 분석에는 해밍 윈도우 (hamming window)를 사용하였으며 홉 사이즈 (hop size)는 50%로 하였다. 이는 분석과 합성 전체 과정에 동일하게 적용하였다.

입력 신호 $x(n)$ 에 대한 캡스트럼 c_r 는 식 (1)로 정의된다.

$$c_r = \frac{1}{2\pi} \int_{-\pi}^{\pi} \log[X(e^{j\omega})] e^{j\omega r} d\omega \quad (1)$$

캡스트럼 계수는 $x(n)$ 의 스펙트럼 $X(e^{j\omega})$ 의 로그 값의 역푸리에 변환이며 이는 신호의 주기성에 관한 정보를 쉽게 표현하는 장점이 있어 주기를 결정하는 데 주로 사용된다. $x(n)$ 의 캡스트럼 포락선은 길이 n_c 의 사각 윈도우 (rectangular window)와 c_r 의 곱을 푸리에 변환하여 구할 수 있다. 이때 윈도우 크기 n_c 는 포락선의 조밀함과 완만함을 결정하게 되며 보통은 기본주파수와 관계되는 두 번째 피크를 기준으로 윈도우 크기를 정하지만 n_c 가 클수록 조밀한 포락선을 구할 수 있다. 본 논문에서는 세

밀한 포락선을 구하기 위하여 캡스트럼 계수에서 입력 샘플 수만큼의 계수를 취하였다.

추출된 조밀한 캡스트럼 포락선은 많은 피크를 포함하고 있기 때문에 적절한 공명점을 찾아야 한다. 해금 음은 15000 Hz 이후에 분포하는 배음 성분의 크기가 상대적으로 작은 특징이 있으므로 15000 Hz 이내에서 공명점을 추출하였고 인접한 공명점 간의 거리가 기본주파수의 0.8배 보다 크며 인접한 공명점보다 크기가 더 클 경우 유효 공명점으로 판단하였다. 이웃한 공명점간의 거리는 최소한 기본주파수 정도이지만 정확히 기본주파수와 같은 값이 되지 않는 좀 더 작은 값인 기본주파수의 0.8배를 기준으로 하여 공명점 추출에 활용하였다. 각 공명점의 대역폭은 에너지가 공명주파수의 절반이 되는, -3 dB가 되는 지점을 기준으로 결정하였다. 추출된 공명점에 대한 정보는 합성을 위한 파라미터로 활용되며 각 공명점의 주파수, 크기, 대역폭이 해당된다.

추출한 공명점 정보를 통해 포만트 합성된 정현파 성분과 원음과의 차로 잔여신호 (residual signal)를 구할 수 있는데 이러한 감산은 시간영역에서 이루어 질 수도 있고 주파수 영역에서 이루어 질 수도 있다. 본 논문에서는 시간 영역에서 감산을 수행하였다. 잔여신호는 이론적으로 노이즈의 형태를 보이게 되지만 원음과 정현파 성분과의 차이로 얻어진 잔여신호에는 피크 성분이 일부 남아있을 수 있다. 따라서 이를 보완해 줄 필요가 있으므로 잔여신호의 스펙트럼에서 그 크기가 현저히 크게 나타나는 몇몇 피크를 제거하는 과정이 필요하다. 그림 2의 원음과 추출된 노이즈 성분의 스펙트럼 비교에서 노이즈 스펙트럼에 존재하는 몇몇 피크를 볼 수 있는데 이들은 원음의 스펙트럼에서 나타나는 배음 성분의 위치와는 다르며 원

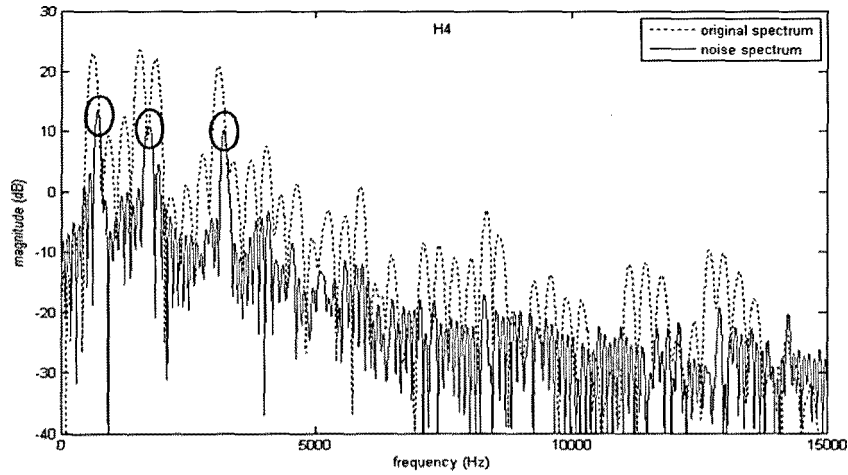


그림 2. 황 (黃)음의 스펙트럼과 추출된 노이즈 성분의 스펙트럼
 Fig. 2. Spectra of Hwang (dotted) and extracted noise component (solid).

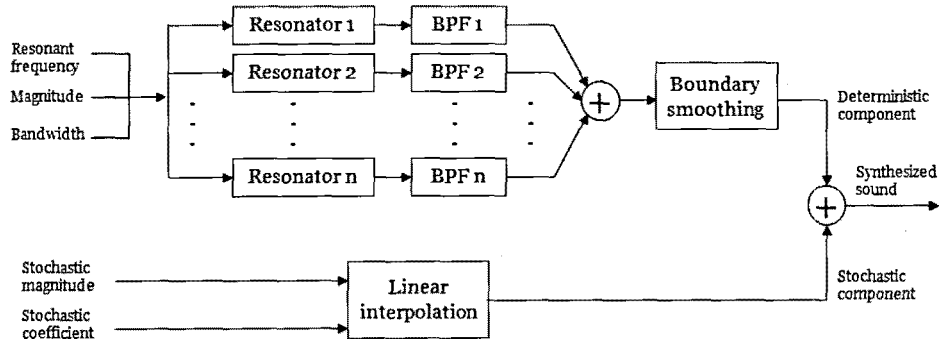


그림 3. 합성 과정 블록도
 Fig. 3. Block diagram of synthesis process.

음의 스펙트럼에서 피크 사이의 감쇠부분에 위치함을 확인할 수 있다. 따라서 배음 성분과는 다르며 원음과 포맷트 합성음과의 감산과정에서 생겨나는 결과라고 할 수 있다. 이들은 에너지가 큰 저주파 대역에서 주로 나타나며 합성과정에 그대로 적용할 경우 합성음에 영향을 미치게 된다. 따라서 노이즈 스펙트럼을 근사화하기 이전에 이를 제거하는 과정을 추가하였다. 저주파 대역에 주로 분포하는 점에 주목하여 5000 Hz 이상의 노이즈 스펙트럼 중 최대값을 나타내는 피크의 크기를 기준 값으로 정하고 기준 값 이상의 크기를 나타내는 모든 피크의 크기를 기준 값으로 조정하였다. 잔여신호에 남아있는 피크를 제거한 노이즈 성분은 주파수 영역에서 선형보간법 (linear interpolation)을 위한 파라미터 추출 과정을 거쳐지게 되는데, 본 논문에서는 노이즈 스펙트럼을 100샘플 단위로 나누어 각 구간에서 최대값을 구하였다. 각 구간의 최대값과 그 위치 정보는 노이즈 성분 합성 과정에 활용된다.

분석 과정을 통해 추출된 파라미터를 이용하여 해금음을 합성하는 과정은 그림 3과 같으며 프레임 단위로 처리된다. 정현파 성분 합성에는 IIT에 의한 필터를 공명기로 사용하였으며 인접한 공명점 간의 중첩으로 인해 공명점 부근의 감쇠영역에서 크기가 커지는 현상을 방지하기 위해 공명기의 출력에 대역 통과 필터를 적용하였다. [26]에서는 인접한 공명점 사이의 중간 지점을 차단 주파수로 결정하였으나, 본 논문에서는 각 공명점의 중심주파수와 대역폭이 다르다는 사실에 주목하여 인접한 공명점에 대한 필터의 주파수 응답 곡선과 해당 공명점에 대한 필터의 주파수 응답 곡선의 교점을 차단주파수로 결정하였다. 각 공명점을 재현하기 위해 공명기와 대역 통과 필터링이 적용된 결과는 모두 더해져 정현파 성분에 해당하는 스펙트럼을 만들어 낸다.

노이즈 성분 합성에는 분석과정에서 추출된 구간별 최대값과 그 위치 정보를 이용한다. 각 구간별 최대값을 이웃한 구간의 최대값과 직선으로 연결하는 선형보간법을

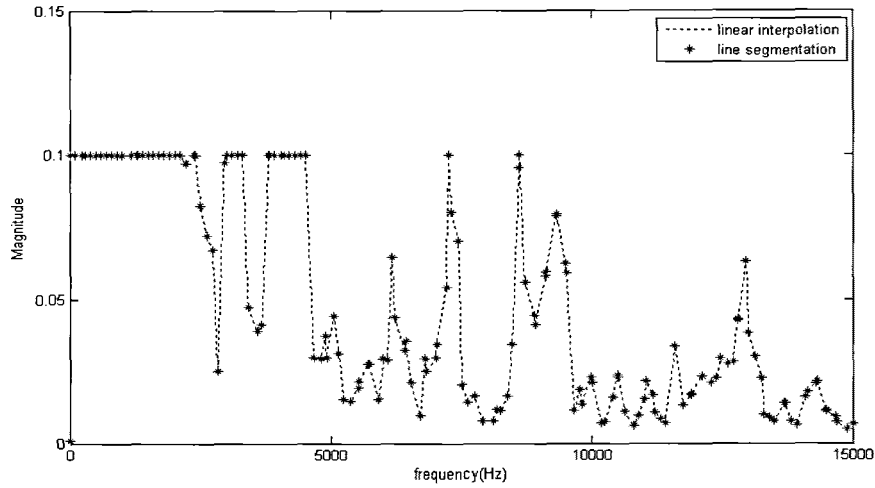


그림 4. 분석 과정에서 추출된 구간별 최대값과 선형보간법으로 재현된 노이즈 성분
 Fig. 4. Synthesized noise spectrum by linear interpolation and local maximum extracted in the synthesis process.

표 1. 황종운지법 단위음 합성을 위한 프레임 사이즈
 Table 1. Frame sizes to synthesize sounds of Hwangjong position of Haegeum.

	기본주파수 (Hz)	프레임사이즈 (samples)
중(仲)	209	689
임(林)	229	629
남(南)	257	561
황(黃)	310	465
태(太)	340	423
중(仲)	414	349
임(林)	457	315
남(南)	511	281

통해 스펙트럼으로 재현된다. 그림 4는 분석과정에서 파라미터화된 각 구간별 최대값과 합성과정에서 선형보간법으로 재현된 노이즈 성분을 비교하여 보여준다. 재현된 노이즈 성분의 저주파 대역에서 값이 0.1인 구간이 많은 이유는 잔여신호에 남아있는 배음과는 무관한 피크 성분들을 제거한 후 파라미터화 하였기 때문이다. 즉, 5000 Hz 이상의 스펙트럼에서 최대값을 기준 값으로 하여 그 이상의 피크를 모두 기준 값으로 조정하였기 때문에 상대적으로 크기가 크게 나타났던 5000 Hz 미만의 스펙트럼은 기준 값인 0.1로 조정된 것이다.

IV. 음 합성 결과

본 논문에서는 황종운지법 8개 단위음과 해금 연주곡 '천년만세'의 도입부를 제안한 합성법으로 합성하였다. 캡스트럼 분석에서 정확한 포먼트 정보를 추출하기 위해

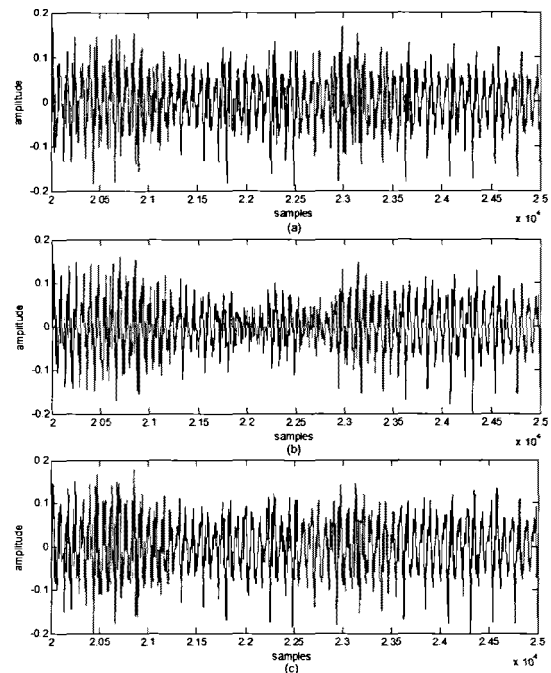


그림 5. 황(黃)음의 일부 구간에 대한 파형 비교 : (a) 원음 (b) [26]의 방법을 이용한 합성음 (c) 기본주파수 조정 후의 합성음

Fig. 5. A waveform comparison of Hwang sound : (a) the original sound (b) the synthesized sound using the method in [26] (c) the synthesized sound using fundamental frequency readjust.

서는 3주기 정도의 입력신호가 적절함을 앞에서 밝힌 바 있다. 그러므로 입력 사운드의 기본주파수에 따라 프레임 사이즈가 결정되어야 한다. 각 단위음의 경우 기본주파수에 따라 프레임 사이즈를 표 1과 같이 조절하여 합성하였는데, 그 결과 그림 5 (b)과 같이 일부 구간에서 그 크기가 갑작스럽게 작아지는 현상이 발생하였다. 이는

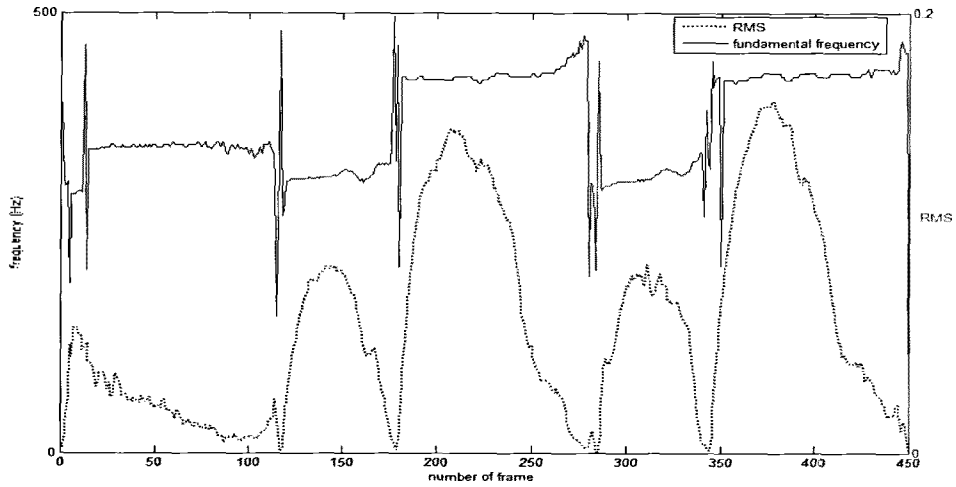


그림 6. 해금 연주곡 '천년만세'의 프레임별 RMS와 기본주파수 변화
 Fig. 6. RMS values and fundamental frequencies of each frame of the tune of 'Cheon-nyeon-man-se' only played by the Haegeum.

캡스트럼 계수를 통해 정확한 기본주파수를 추출하지 못하여 생겨난 결과이다.

기본주파수는 공명점 추출 과정에서 추출한 공명점이 유효한지를 판단하는 기준이 되는데 기본주파수가 너무 큰 값으로 추출되면 유효 공명점 간의 간격이 커지게 되며 그 사이에 존재하는 일부 유효 공명점을 찾아내지 못하는 결과를 초래한다. 따라서 에너지가 큰 공명점들을 재현하지 못하게 되어 합성음의 파형의 크기가 작아진다. 즉 기본주파수를 잘못 판단하게 되면 분석과정에서 스펙트럼 파라미터가 제대로 추출되지 못하므로 합성 결과 또한 원음과는 차이를 보이게 된다. 이러한 문제를 해결하기 위하여 이전 프레임과 기본주파수를 비교하여 조절해주는 과정을 추가하였다. 입력 프레임은 음 높이의 변화가 거의 없는 단위음에 속한다는 것에 주목하여 추출한 기본주파수가 이전 프레임의 기본주파수보다 일정 크기 이상 차이가 나면 이전 프레임의 기본주파수로 바꿔주었다. 이러한 기본주파수 조절 결과를 그림 5 (c)에 보였는데, 그림 5 (b)에서의 파형의 감소 문제가 향상되었음을 확인할 수 있다.

연주곡의 경우 음이 계속적으로 변하여 음 높이가 일정하지 않으므로 일정한 프레임 사이즈를 유지하게 되면 정확한 합성 결과를 기대하기 힘들다. 이러한 문제점을 해결하기 위해서는 음 높이가 바뀌는 구간, 즉 기본주파수가 바뀌는 구간을 인식하여 기본주파수에 따라 프레임 사이즈를 가변적으로 조절해야 한다. 연주곡의 파형을 살펴보면 음이 바뀌는 구간에서 그 크기가 작아지는 것을 확인할 수 있는데 이를 이용하여 묵음 구간을 판단하였

다. 연주곡을 20 ms 단위로 나누어 스펙트럼의 RMS를 조사한 결과 그림 6과 같이 음이 바뀌는 4개의 구간에서 그 값이 확연히 작아지는 것을 알 수 있었다. 따라서 각 프레임에서 스펙트럼의 RMS 값이 임계값 (threshold, 본 논문의 경우 0.00025)보다 작으면 묵음 구간으로 판단하고 합성 과정을 수행하지 않도록 하였다. 또한 묵음 구간 이후에 스펙트럼의 RMS 값이 임계값보다 큰 프레임은 새로운 음이 시작되는 어택 (attack) 구간이므로 기본주파수를 추출하여 프레임 사이즈를 재조정 하였다. 또한 단위음 전체구간 합성 과정에서와 마찬가지로 같은 음이 유지되는 동안에는 기본주파수를 일정하게 유지해 줄 필요가 있으므로 묵음 구간이나 어택 구간이 아니면 지속 (sustain) 구간으로 간주하고 이전 프레임과 비교하여 기본주파수를 조정하였다. 그 결과를 그림 7 (c)에 나타내었는데 일정한 프레임 사이즈를 적용하여 합성한 그림 7 (b)에서와 같은 갑작스런 파형의 크기 변화 부분이 조정되어 원음과 더욱 유사한 결과를 얻을 수 있었다.

합성된 해금 음의 음질을 평가하기 위하여 해금 연주 전문가에게 원음과 합성음에 대한 비교 평가를 의뢰하였다. 전문가는 원음과 합성음을 각각 듣고 그 유사도를 5 단계로 나누어 음질을 평가하였다. 5단계의 유사도는 각각 0~50 %, 51~70 %, 71~85 %, 86~95 %, 96~100%이며 단위음 전체 구간을 합성한 황종운지법 8개음과 연주곡에 대해 이루어졌다. 그 결과 모두 96~100 % 유사하다고 평가하였고 합성음에 대한 청취 결과 모든 음이 원음과 매우 유사하다는 소감을 밝혔다.

해금의 스펙트럼 모델링을 위한 선행 연구인 [26]은 원

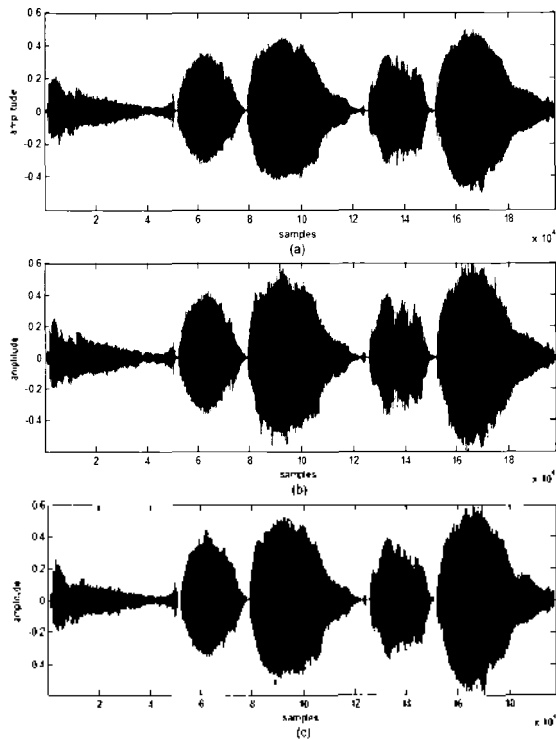


그림 7. 해금 연주곡 '천년만세'의 파형 비교 : (a) 원음 (b) [26]의 방법을 이용한 합성 결과 (c) 제안한 방법을 이용한 합성 결과

Fig. 7. A waveform comparison of the tune of 'Cheon-yeon-man-se' : (a) the original tune (b) the synthesized tune using the method in [26] (c) the synthesized tune using the proposed method.

음과 유사한 스펙트럼을 재현하였으나 해금 단위음에서 10~20 ms 정도의 지속구간만을 입력 샘플로 활용하였기 때문에 합성 결과를 청취하기 힘들 뿐 아니라 해금 소리의 시간에 따른 변화를 사실적으로 묘사하기에는 제한적이었다. 본 논문에서는 이를 개선하여 프레임 기반의 단위음 전체 구간을 합성하였으며 유사한 방법으로 연주곡을 합성하였다. 본 논문에서 수행한 해금 음 합성은 모두 연주법이 포함되지 않은 기본음에 대한 분석 결과를 토대로 이루어졌다. 이러한 결과는 단위음의 지속구간을 대상으로 한 분석 및 합성 방법을 적용하여 연주법과 꾸밈음 등이 포함되지 않은 단위음 전체 구간이나 연주곡 또한 분석하고 합성할 수 있음을 증명한다.

V. 결론

본 논문에서는 해금 소리의 시간에 따른 변화를 사실적으로 묘사하기 위해 캡스트럼 분석을 이용한 전통 악기 해금의 스펙트럼 모델링을 제안하였다. 캡스트럼 포락선이 피크 성분을 정확히 묘사할 수 있도록 프레임 사이즈

를 3주기로 결정하였으며 포먼트 추출에 더 많은 캡스트럼 계수를 활용하였다. 정현파 성분 합성 과정에서 대역 통과 필터의 차단주파수를 공명점 별로 유동적으로 조절하여 성능을 향상시켰으며 노이즈 성분에 남아있는 피크 성분들을 제거하였다. 연주곡 합성에서 입력 프레임을 묵음구간, 어택구간, 지속구간으로 판단하는 과정을 통해 음 높이의 변화를 인지하였고 기본주파수에 따라 프레임 사이즈를 가변적으로 조절하여 정확도를 향상시켰다. 합성 결과는 해금 연주 전문가의 청취테스트를 통해 원음과 합성음이 96~100% 유사하다는 평가를 받았다. 향후 분석과정에서 추출한 파라미터를 효과적으로 줄일 수 있는 방법에 대한 연구가 수행되어야 하며 이를 통해 실시간 연주가 가능한 디지털 해금을 개발하는데 한 걸음 더 다가갈 수 있을 것이다.

감사의 글

이 논문은 2010년도 정부 (교육과학기술부)의 재원으로 한국과학재단의 지원을 받아 수행된 연구임 (No. R01-2008-000-20493-0).

참고 문헌

1. Bonada, J., Loscos, A., Cano, P., Serra, X., "Spectral Approach to the Modeling of the Singing Voice", in *Proc. of the 111th AES Convention*, 2001.
2. J. O. Smith, *Spectral Audio Signal Processing*, <http://ccrma.stanford.edu/~jos/sasp/>, Online Book, 2007.
3. R. W. Schafer and J. D. Markel, eds., *Speech Analysis*, New York: IEEE Press, 1979.
4. J. L. Flanagan and R. M. Golden, "Phase vocoder," *Bell System Technical Journal*, vol. 45, pp. 1493-1509, 1966.
5. X. Serra and J. O. Smith, "PARSHL: An Analysis/Synthesis Program for Non-Harmonic Sounds based on a Sinusoidal Representation," in *Proc. of the 1987 International Computer Music Conference*, Computer Music Association, 1987.
6. McAulay, R.J. and T.F. Quatieri, "Magnitude-only Reconstruction using a Sinusoidal Speech Model," in *Proc. of the 1984 IEEE International Conference on Acoustics, Speech and Signal Processing*, New-York: IEEE Press, 1984.
7. McAulay, R.J. and T.F. Quatieri, 1986, "Speech Analysis/Synthesis based on a Sinusoidal Representation," *IEEE Trans. on Acoust., Speech and Signal Processing* vol. 34, no. 4, pp. 744-754, 1986.
8. X. Serra and J. Smith, "Spectral Modeling Synthesis: A Sound Analysis/Synthesis System based on a Deterministic plus Stochastic Decomposition," *Computer Music Journal*, vol. 14, no. 4, pp. 12-24, 1990.
9. X. Serra and J. O. Smith, "Residual Minimization in a Musical Signal Model based on a Deterministic plus Stochastic Decomposition," *J. Acoust. Soc. Am.*, vol. 95, no. 5-2, pp.

- 2958-2959, 1994.
10. Serra, X, Bonada, J, "Sound Transformations Based on the SMS High Level Attributes," in *Proc. of International Conference on Digital Audio Effects (DAFX98)*, 1998.
 11. Verma, T. S., T. H. Y. Meng., "Time Scale Modification Using a Sines+Transients+Noise Signal Model," in *Proc. of International Conference on Digital Audio Effects (DAFX98)*, 1998.
 12. Tony S. Verma and Teresa H. Y. Meng, "An analysis/synthesis tool for transient signals," in *Proc. 16th International Congress on Acoustics/135th Meeting of the Acoustical Society of America*, vol. 1, pp. 77-78, 1998.
 13. Verma, T. S., T. H. Y. Meng., "Extending Spectral Modeling Synthesis with Transient Modeling Synthesis", *Computer Music Journal*, vol. 24, no. 2, pp. 47-59, 2000.
 14. 조상진, 정의필, "산조가야금의 물리적 모델링," *한국음향학회지*, 23권, 7호, 521-531쪽, 2004.
 15. 조상진, 최진규, 정의필, "안죽과 몸통의 임펄스 응답을 이용한 가야금 사운드 합성," *한국신호처리시스템학회논문지*, 7권, 3호, 102-107쪽, 2006.
 16. 조상진, 정의필, "개선된 산조 가야금의 물리적 모델링을 이용한 오른손 주법의 음 합성," *한국음향학회지*, 25권, 8호, 325-332쪽, 2006.
 17. 조상진, 정의필, "안죽이 있는 악기의 개선된 현의 모델 개발," *한국음향학회지*, 26권, 7호, 328-333쪽, 2007.
 18. 변중배, 조상진, 홍연우, 정의필, "태평소의 음향분석을 통한 팔랑 특성 추출," *한국음향학회지*, 27권 1호, 12-17쪽, 2008.
 19. 강영수, 홍연우, 조상진, 정의필, "디지털 피리 구현을 위한 개선된 스펙트럼 모델링 합성 알고리즘," *한국공학예술학회논문지*, 1권, 1호, 5-11쪽, 2009.
 20. Kieu Huu Thu, Sangjin Cho, Yeonwoo Hong, Myeongsu Kang, Uipil Chong, "Synthesis of Piri Based on the Modified SMS," *The 10th Western Pacific Acoustics Conference*, CD Proceeding, pp. 45, 2009.
 21. 변중배, 조상진, 정의필, "주파수 영역에서의 상관함수를 이용한 피리의 FM합성 파라미터 추출," *한국디지털아트미디어학회 학술발표대회 논문집*, 4권, 1호, 85-88쪽, 2006.
 22. 변중배, 조상진, 정의필, "비대칭 FM합성방식을 이용한 피리 소리의 합성," *한국신호처리시스템학회 하계학술대회 논문집*, 7권, 1호, 37-40쪽, 2006.
 23. 홍연우, 조상진, 변중배, 정의필, "운지법에 따른 해금 소리의 베음 구조 분석," *한국신호처리시스템학회 하계학술대회논문집*, 8(1), pp. 58-61, 2007.
 24. 홍연우, 조상진, 최명환, 정의필, "해금 소리 합성을 위한 스펙트럼 파라미터 추출," *한국디지털아트미디어학회 학술발표대회 논문집*, 5권, 1호, 7-10쪽, 2007.
 25. Yeonwoo Hong, Sangjin Cho, Myeongsu Kang, Hyungseob Han, Uipil Chong, "Spectrum modeling of Haegum using formal extracted from cepstral envelope," *The 10th Western Pacific Acoustics Conference*, CD Proceeding, pp. 44, 2009.
 26. 홍연우, 조상진, 김중면, 정의필, "캡스트럼 포락선을 이용한 해금 소리의 포먼트 합성," *한국음향학회지*, 28권, 6호, 526-533쪽, 2009.

저자 약력

- **홍 연 우 (Yeonwoo Hong)**
한국음향학회지 제28권 제6호 참조
- **강 명 수 (Myeongsu Kang)**
한국음향학회지 제28권 제2호 참조
- **조 상 진 (Sangjin Cho)**
한국음향학회지 제23권 제7호 참조
- **김 중 면 (Jongmyon Kim)**
한국음향학회지 제28권 제6호 참조
- **정 의 필 (Uipil Chong)**
한국음향학회지 제23권 제7호 참조
- **이 정 철 (Jungchul Lee)**
한국음향학회지 제28권 제6호 참조