International Journal of Multimedia and Ubiquitous Engineering
Vol.9, No.8 (2014), pp.347-354
http://dx.doi.org/10.14257/ijmue.2014.9.8.32

A Sound Analysis and Synthesis System for Generating an Instrumental Piri Song

Myeongsu Kang and Jong-Myon Kim*

School of Electrical Engineering, University of Ulsan, 93 Daehak-ro, Ulsan 680-749, South Korea
{ilmareboy, jmkim07}@ulsan.ac.kr

Abstract

This paper proposes spectral modeling of the piri instrument based on its acoustic characteristics. Since the harmonics of piri sounds are evenly distributed to the high frequency band (up to 20 kHz), spectral modeling synthesis is suitable for generating piri sounds. In this study, sinusoids of the piri are only considered with formant models (or resonant models) to reduce the number of synthetic parameters using cepstral analysis. In addition, noise components are recreated by exploiting a line-segmentation approximation. Experimental results indicated that single notes of the piri and a synthesized instrumental piri song have similar spectra to originals.

Keywords: *Cepstral analysis; formant model; line-segmentation approximation; spectral modeling synthesis*

1. Introduction

Spectral modeling synthesis (SMS) reproduces an input signal in frequency domain, and it is based on perceptual mechanisms of the listener [1]. The first spectral modeling is a sinusoidal model that is also called additive synthesis. It is rooted in Fourier’s theorem which states that any periodic waveform can be modeled as a sum of sinusoids at various amplitudes and harmonic frequencies. Despite the fact that the sinusoidal model (or additive synthesis) is appropriate for synthesizing vowel sounds or string and wind instruments that have periodic characteristics [2], it is not suitable to model transient part of sounds, such as attacks of most percussion instrument, the bow noise in string instruments, or the breath noise in wind instruments [3]. In 1990s, *Serra* introduced SMS that musical sounds can be modeled as a sum of sinusoids plus a noise residual [4, 5]. This paper synthesizes single notes and an instrumental song played by the *piri*, which is one of representative Korean traditional wind instrument. In SMS, sinusoidal components are generated by using additive synthesis, subtractive synthesis, and formant synthesis. Additive synthesis and subtractive synthesis requiring high computation complexity limit their usage in real-time audio applications [6]. Consequently, the proposed paper is concerned only with formant models (or resonant models) to generate sinusoids of the *piri* sounds using cepstral analysis. Likewise, a simple line-segment approximation is utilized to model noise components and one way to carry out the line-segment approximation is to set through the magnitude spectrum and find local maxima in the predefined sections.

The rest of this paper is organized as follows. Section 2 introduces the target instrument, and Section 3 analyzes its acoustic characteristics. Then, analysis/synthesis procedures are described in Section 4. Synthetic sounds are shown in Section 5 and Section 6 finally concludes this paper.

---

* Corresponding author.

ISSN: 1975-0080 IJMUE

2. Target Instrument: Korean Traditional Wind Instrument, *Piri*

This paper synthesizes the *piri* sounds based on SMS by utilizing acoustic characteristics of the *piri*. The *piri* is one of the most popular wood wind instruments among double-reed recorders in Korean traditional musical instruments, and it is the generic term of recorders which are composed of a bamboo bar with some holes. There are several different kinds of the *piri* such as the *hyang piri* and *tang piri*, and Figure 1 shows the structure of them. As shown in Figure 1, they are very similar to each other in their appearance, and the principle of sounding is also similar. They have eight finger holes in front and one finger hole in the back side. Then, it is possible to adjust intonation on the *piri* by covering/uncoverting finger holes [7].

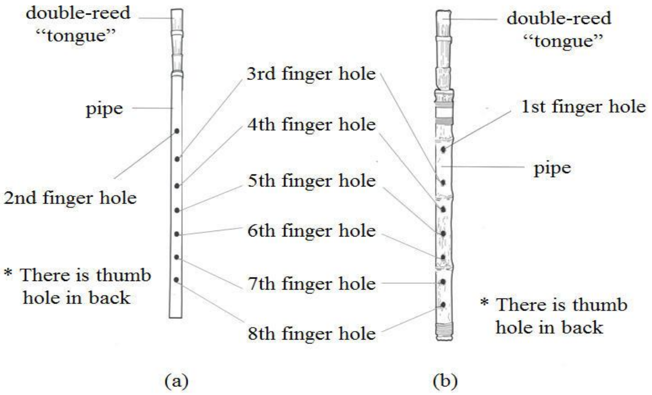
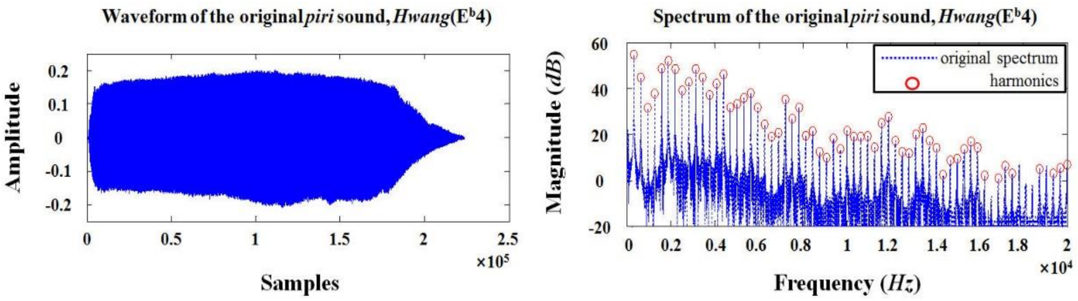


Figure 1. Structure of the Korean Traditional Wind Instrument, *piri* (a) *hyang piri,* (b) *tang piri*

3. Acoustic Characteristics of the *Piri*

Single *piri* notes were sampled at 44.1 kHz with 16-bit quantizastion. Specific terms for notes produced by various combinations of covered and uncovered holes are provided by Korean musical theory. This paper utilizes the following single notes in order to analyze acoustic characteristics of the *piri* sounds: *Hwang* (E^b4), *Tae* (F4), *Jung* (A^b4), *Im* (B^b4), and *Nam* (C5). To analyze harmonic structure and frequency characteristics of the *piri*, this study used single notes sustained for 0.5 seconds.

Figure 2(a) shows a waveform of the note *Hwang* (E^b4, 313 *Hz*), which has a very fast attack. In addition, the sound has no decay and a very long sustain because constant vibrational energy is delivered to produce the *piri* sound. Thus, the sound only fades once the breath input stops. Moreover, the *piri* generates regular harmonics to 20 *kHz*. Figure 2(b) shows the spectrum and the harmonic structure of the note *Hwang*. As shown in Figure 2(b), intervals between two adjacent harmonics are almost the same, and harmonic frequencies are equally spaced based on the width of the fundamental frequency. Table 1 describes fundamental and prominent frequencies for the five single reference notes.



(a) (b)

Figure 2. Acoustic Characteristic of the Note *Hwang* (E^{b}4, 313 *Hz*) (a) Waveform, (b) Spectral envelope

Table 1. Fundamental and Prominent Frequencies for the Five Single Reference Notes

Notes	Fundamental frequency		Prominent frequency (*Hz*)
	Frequency (*Hz*)	Magnitude (*dB*)	
Hwang **(E^{b}4)**	313	54.56	313
Tae **(F4)**	347	53.96	348
Jung **(A^{b}4)**	417	53.54	3576
Im **(B^{b}4)**	465	52.11	3250
Nam **(C5)**	525	54.20	3679

For notes *Jung*, *Im*, and *Nam*, prominent frequencies differ from fundamental frequencies and lie in the range from 3.2 *kHz* to 3.8 *kHz*. However, prominent frequencies of notes *Hwang* and *Tae* are the same as their fundamental frequencies. Figure 3 shows the cepstral envelope of the single reference *piri* notes, where all of the single notes have a formant (or a resonance) around 3.5 *kHz*, including notes *Hwang* and *Tae*. Thus, we observe that frequency components around 3.5 *kHz* affect the tone of the *piri*.

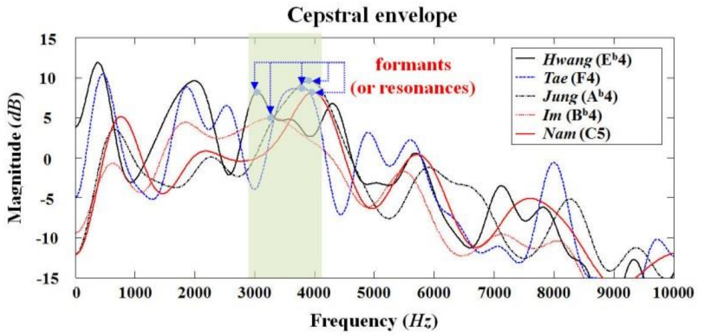


Figure 3. Cepstral Envelope of the Reference Single Notes of the *piri*

4. Analysis and Synthesis for Generating *Piri* Sounds

4.1. Analysis of *piri* Sounds

As described in Section 3, harmonics of the *piri* are evenly distributed to the high frequency band, and formants (or resonances) are also shown in the high frequency band, as illustrated in Figures 2 and 3. Thus, the sinusoidal components of *piri* sounds are generated by utilizing formants (or resonances) extracted from the cepstral envelope. To extract synthetic parameters, we first analyze the reference *piri* sounds using a frame-based process. Consequently, short-time Fourier transform (STFT) should be carried out using a hamming window with an overlap of 50 % and synthetic parameter extraction performed as follows.

- ***Step* 1**. ***Classification of an input frame***: In the chosen instrumental *piri* song, magnitude variation occurs in its waveform, which may result in generating low-quality synthetic sounds. To address this drawback, it is necessary to extract accurate synthetic parameters for the given frame. In this study, each input frame is partitioned into silence, attack, and sustain regions in order to identify precise parameters, where an input frame is considered a silence region when the root mean square (RMS) value of the frame is less than or equal to a predefined threshold. However, the first input frame whose RMS value exceeds the threshold after the previous frame (which was included in the silence region) is segmented into the attack region. In this study, the threshold value is set to 0.0025 by quantitatively measuring RMS values of a number of frames, including various fundamental frequencies, as shown in Figure 4. For the RMS variation in Figure 4, we analyzed 9.6 seconds of instrumental *piri* song at 20 *ms* intervals and observed that RMS values are very low where the fundamental frequencies change rapidly.

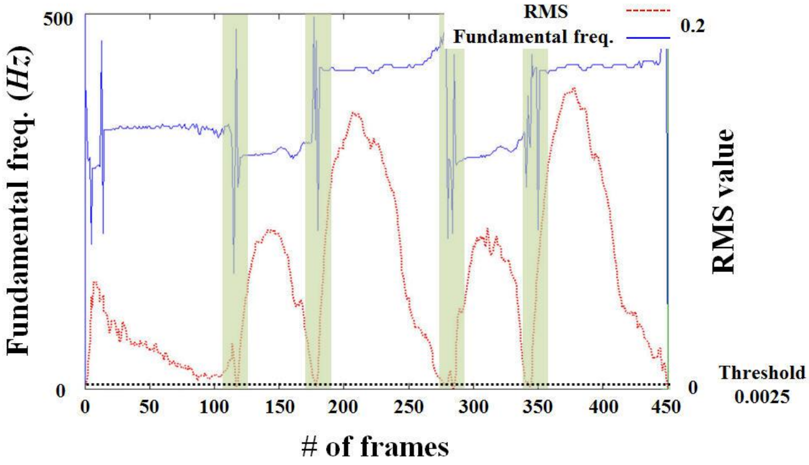


Figure 4. RMS Variation for a Number of Frames Including Various Fundamental Frequencies

- ***Step* 2. *Cepstral analysis***
  - To synthesize the instrumental *piri* song, we extract synthetic parameters for input frames involving either attack or sustain regions. In this study, once an input frame is determined to be in the silence region, the frame is considered to contain only noise components, and a line-segment approximation is employed to produce it. For the frame involving the attack region, the length of the input frame should be readjusted to have a length three times the fundamental frequency of the frame, which is enough to capture synthetic information according to analytical characteristics of the *piri*. Finally, we calculate cepstral coefficients of the frame.
  - If the previous frame was in the attack region and its fundamental frequency is the same as that of the current frame, then the current frame is classified into the sustain region, and it is only necessary to compute cepstral coefficients for the current frame without readjusting its frame length.

  - If the previous frame was in the sustain region and the RMS value of the current frame is greater than 0.0025, then we consider the current frame to also be in the sustain region. However, the fundamental frequency of the current frame may differ from that of the previous frame due to magnitude variation of the waveform. Thus, this study sets the fundamental frequency of the current frame to that of the previous frame if the two adjacent frames are in the sustain region.
- ***Step* 3. *Resonance extraction***
  - It is necessary to obtain the cepstral envelope by exploiting cepstral coefficients with an appropriate analysis window, and the length of the analysis window is set to the same size of the input frame in this study.
  - We then extract resonances (or formants) from the cepstral envelope.
    - First, cepstral peaks are detected from the cepstral envelope, and we then extract peaks that satisfy the following two conditions: 1) the frequency interval between two adjacent peaks should be greater than or equal to 0.8 times the fundamental frequency; 2) we recognize peaks as synthetic resonances when their magnitudes are greater than the maximum peak magnitude of 50 *dB*.
    - The bandwidths of extracted cepstral peaks can be measured at half-power points.

4.2. Synthesis of the *piri* Sounds

The sinusoids are synthesized with extracted parameters such as resonant frequencies, resonant magnitudes, and bandwidths. It is then possible to obtain noise components (or the residual signal) by subtracting the synthesized sinusoids from the original *piri* sound. To model the residual signal, we first obtain the local maximum values of every 100 samples in the original residual and then apply a line-segment approximation to its log-magnitude spectrum.

Synthesis is also a frame-based process. The *piri* sounds are generated with synthetic parameters through the analysis process. Figure 5 illustrates a digital resonator generating the sinusoids, where two parameters are used to specify the input-output characteristics of the resonator: the resonant (or formant) frequency f_r and the resonance bandwidth ω_{BW} .

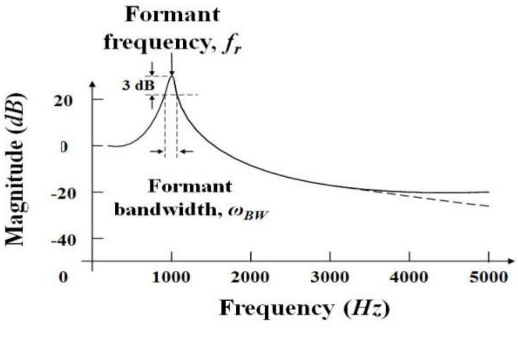


Figure 5. Frequency response of the resonator (f_r = 1000 *Hz*, ω_{BW} =50 *Hz*)

Samples of the output of the digital resonator $y(n)$ are then computed from the input signal $x(n)$ using the formula (1) below:

$$y(n) = Ax(n) + By(n-1) + Cy(n-2), \tag{1}$$

where the constants *A*, *B*, and *C* are related to the resonant frequency f_r and the bandwidth ω_{BW} by the impulse-invariant transformation as follows:

$$A = 1 - B - C. \tag{2}$$

The digital resonator is a second-order difference equation, and the transfer function of the digital resonator is given by the formula (3) below:

$$T(z) = A / 1 - Bz^{-1} - Cz^{-2}, \tag{3}$$

where $z = e^{j\omega}$ [8].

To produce the sinusoids, a parallel formant synthesizer is used in this study. The parallel formant synthesizer sums the outputs of the simultaneously excited formant resonators. In this case, a superposition of resonators at adjacent resonant frequencies can occur; thus, attenuation around resonances will not be well described. To overcome this drawback, a digital band-pass filter (BSF) is connected to the output of each formant resonator. The center frequency and bandwidth are specified by each resonant frequency and the difference between two adjacent resonant frequencies, respectively. For example, if the first and second resonant frequencies are 250 *Hz* and 550 *Hz*, the center frequency, bandwidth, and upper and lower cutoff frequencies of the BSF should be 250 *Hz* (the first resonant frequency), 300 *Hz* (the difference between two adjacent resonant frequencies), 100 *Hz* (the first resonant frequency – (bandwidth/2)), and 400 *Hz* (the first resonant frequency + (bandwidth/2)), respectively. Figure 6 shows a block diagram that synthesizes the single reference notes of the *piri*. Consequently, the synthesis system includes digital resonators, digital band-pass filters, and the line-segment approximation.

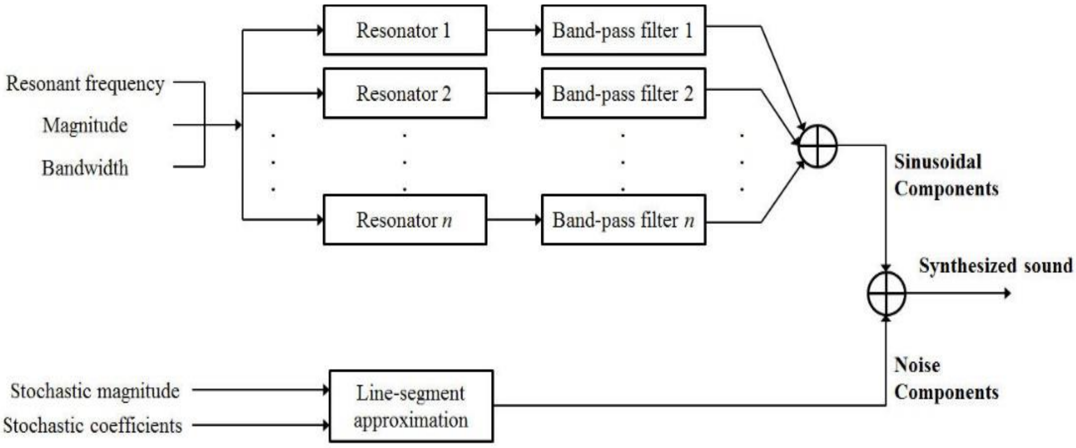


Figure 6. A Block Diagram for Generating the Single Reference Notes of the *piri* with Extracted Synthetic Parameters

4.3. Synthetic Results of the Instrumental *piri* Song

In this section, the proposed spectral modeling synthesis technique recreates the single reference notes and an instrumental *piri* song. Figure 7 depicts a synthetic result of the instrumental *piri* song, which is similar to the original in its shapes and spectra. Some magnitudes of the synthesized sound are greater or less than those of the original because frames overlap in the analysis and synthesis processes, but this does not decrease the quality of the synthesized sound. More synthetic results are available at http://eucs.ulsan.ac.kr/IJMUE/2014-02.

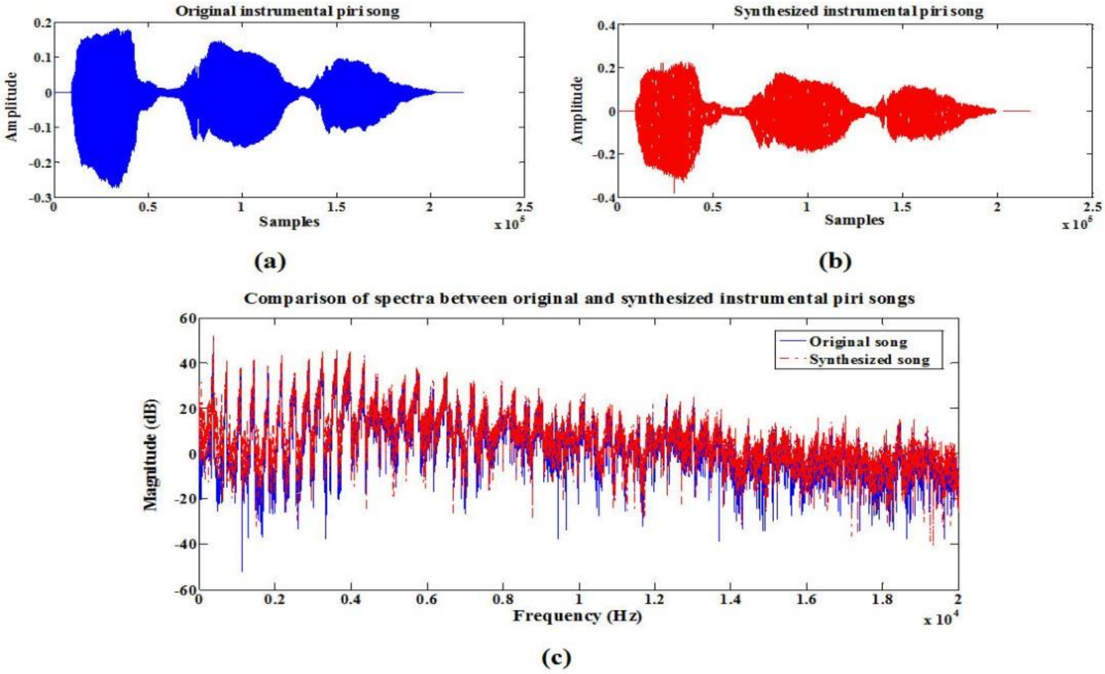


Figure 7. Comparison of the Synthetic Results of the Instrumental *piri* Song

(a) Waveform of the 5.6-second long original instrumental *piri* song, (b) Waveform of the produced instrumental song, (c) Spectral comparison between the original (solid) and synthesized (dotted) instrumental *piri* songs.

5. Conclusions

This paper proposed a spectral modeling synthesis technique for the *piri* instrument based on acoustic characteristics. The *piri* had a very short attack and no decay, and its harmonics are evenly distributed to 20 *kHz*. These acoustic characteristics of the *piri* were suited to spectral modeling synthesis, which generates sinusoids using formants obtained from the cepstral analysis as well as noise components by line-segment approximation. The synthetic result using the proposed spectral modeling synthesis technique was very similar to the original instrumental *piri* song.

Acknowledgements

This work was supported by a National Research Foundation of Korea (NRF) grant funded by the Korean government (MEST) (No. NRF-2013R1A2A2A05004566).

References

[1] J. Bonada, A. Loscos, P. Cano and X. Serra, “Spectral Approach to the Modeling of the Singing Voice,” Proceedings of the 11th Audio Engineering Society Convention, (**2001**) November 30, New York, USA.

[2] X. Serra and J. O. Smith, “Residual Minimization in a Musical Signal Model Based on a Deterministic plus Stochastic Decomposition,” Journal of Acoustical Society of America, vol. 95, (**1994**), pp. 2958-2959.

[3] Spectral Audio Signal Processing, (**2014**) February, http://ccrma.stanford.edu/~jos/sasp.

[4] X. Serra and J. O. Smith, “Spectral Modeling Synthesis: A Sound Analysis/Synthesis System Based on a Deterministic plus Stochastic Decomposition,” Computer Music Journal, vol. 14, (**1990**), pp. 12-24.

[5] X. Serra and J. Bonada, “Sound Transformations Based on the SMS High Level Attributes,” Proceedings of International Conference on Digital Audio Effects (DAFX98), (**1998**) November 19, Barcelona, Spain

[6] P. –Y. Tsai, T. –M. Wang and A. Su, “GPU-Based Spectral Model Synthesis for Real-Time Sound Rendering,” Proceedings of International Conference on Digital Audio Effects (DAFx-10), (**2010**) September 6-10, Graz, Austria
[7] National Gugak Center, (**2014**) February, http://www.gugak.go.kr:9001/eng/index.jsp.
[8] H. K. Dennis, “Software for a Cascade/Parallel Formant Synthesizer,” Journal of Acoustical Society of America, vol. 67, (**1980**), pp. 971-995.

Authors

Myeongsu Kang, He received the B.S. and M.S. degrees in computer engineering and information technology from the University of Ulsan, Ulsan, South Korea, in 2008 and 2010, respectively. His current research interests include high-performance multimedia signal processing, application-specific system-on-a-chip design, and fault diagnosis of machinery.

Jong-Myon Kim, He received the B.S. degree in electrical engineering from Myongji University, Yongin, South Korea, in 1995, and the M.S. degree from the University of Florida, Gainesville, and the Ph.D. degree from the Georgia Institute of Technology, Atlanta, in 2000 and 2005, respectively, both in electrical and computer engineering. He is currently an Associate Professor of electrical engineering with the University of Ulsan, Ulsan, South Korea. His current research interests include multimedia processing, multimedia-specific processor architectures, parallel processing, and embedded systems.