A new pitch tracking algorithm for the spectral modeling synthesis

Hyungseob Han, Sangjin Cho*
School of Electrical Engineering, University of Ulsan, Ulsan, Republic of Korea
sjcho75@ilsan.ac.kr

Abstract—**The modeling methods for generating musical sounds are largely categorized into two major sections: physical modeling and spectrum modeling. Spectrum modeling is to parameterize a sound recognizing at the human auditory cognitive mechanism in frequency domain. This paper describes a new pitch tracking algorithm for the spectral modeling synthesis (SMS) technique for Korean traditional percussion instrument, Jing, showing that the pitch tracking technique in conventional SMS cannot represent characteristics of percussion instruments. In contrast, the proposed can guarantee accurate sinusoids and well-approximated residual signals. The peak-picking algorithm improved the conventional SMS by extracting accurate magnitudes of peaks changing irregularly for each frame. The pitch-tracking algorithm enables a pitch continuation to be performed successfully, as well as, softening phenomena and beat phenomena can be interpreted. Compared to the conventional algorithm, the proposed algorithm reduced sinusoidal components in the residual signal significantly, so not only it can synthesize successfully Jing sounds but be also applied to analyze general percussion instruments.**

I. INTRODUCTION

Sound synthesis methods are generally classified into sampling, modulation, filtering, and modeling methods[1]. The modeling methods are largely classified into physical modeling and spectrum modeling. While physical modeling is a method to synthesize sounds based on the principles of sound generation in musical instruments[2], spectrum modeling is a method to synthesize a sound recognizing at basilar membrane of the ear in frequency domain, modeling the human auditory cognitive mechanism[3].

The beginning of spectrum modeling is the additive synthesis proposed by Moorer in 1977[4]. In 1978, Moorer[5] improved his previous additive synthesis using a nonrectangular window and the digital phase vocoder proposed by Portnoff[6]. In 1987, Smith and Serra introduced the PARSHL program that is the improved version of Moorer's phase vocoder, and can represent inharmonic and pitch changing sounds[7]. However, PARSHL had a drawback that it is difficult to represent noise-like sounds, such as the attack of instrument sounds. In 1990, Serra and Smith proposed spectral modeling synthesis (SMS) that synthesizes sounds separating a sound into a deterministic (sinusoidal) and a stochastic (residual) component[3].

SMS finds the slowly varying sinusoidal components in a signal using peak-picking algorithm in the frequency domain and determines significant sinusoidal components using pitch-tracking algorithm. Subsequently, synthesized sinusoids by additive synthesis are subtracted from the original signal to obtain a residual signal composed of transients and noise. Therefore, the precise peak-picking and pitch-tracking algorithm can guarantee accurate sinusoids and well-approximated residual signal. Otherwise, unexpected sinusoidal component will be contained in the residual signal[8]. This paper describes modified peak-picking and pitch-tracking algorithm and evaluates this method by comparing it to conventional SMS for the Korean percussion instrument called Jing.

II. SOUND ANALYSIS

The Jing is a large gong used in traditional Korean music. This is usually made from brass and is struck by a hammer that is layered with soft cloth to smoothen the texture of the sound produced. It is typically played at the onset of ceremonies and special occasions. Jing sounds used in this paper are recorded in an anechoic chamber by professional musicians in compliance with Korean traditional percussion instrument playing styles. Recording devices are as follows: microphones (measurement: BEHRINGER ECM-8000, recording: AKG C1000s), DAT recorder (TASCAM DA-P1) with a sampling frequency of 48KHz, and 16bit quatization. Softwares for sound analysis are Mathworks MATLAB 7.6 and Sony SoundForge 8.0.

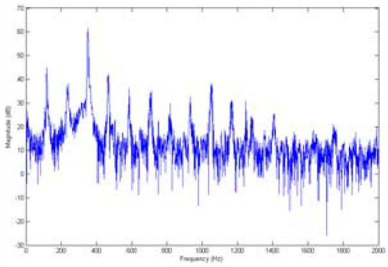

Fig. 1. Spectral representation of the Jing sound

* Corresponding author: sangjin.cho@gmail.com
This research was supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education, Science and Technology (2012R1A1B6002600).

Fig. 1 depicts the spectrum of range up to 2 kHz of the recorded sound. Here, the remarkable thing is that the harmonic is observed unlike other percussion instruments. For investigating harmonic relationships in detail, the harmonic partials generated by different attacks are shown in Table 1. It is known that each partial is set to the nearest multiple of fundamental frequency regardless of the kind of attack.

TABLE I.
Fundamental Frequency of Jing Sounds

The number of partials	Strong attack			Weak attack		
	Sound 1	Sound 2	Sound 3	Sound 1	Sound 2	Sound 3
1st	117	118	117	120	120	120
2nd	239	237	239	238	238	238
3rd	351	352	351	357	357	357
4th	469	466	469	475	476	475

Flat plates generally show hardening phenomena but most curved plates exhibit softening. Softening behavior is that the frequency of plates of shells is initially formed lower and gradually rising, depending on the curvature. P. L. Grossman reported that the frequency of vibration is 10% less than its small-amplitude value[9]. These phenomena occur when nonlinear vibrations at large-amplitude return to linear vibration at small-amplitude[10, 11]. Because the Jing is a kind of curved plates, we can expect that softening effect is observed.

III. PROPOSED METHOD

A. Conventional SMS for the Jing

In order to analyze Jing sounds using SMS, analysis parameters are set as follows: Blackman-Harris window, window size of 0.06 sec, 40 peaks to pick, 6 sinusoids, 16 times zero-padding. Fig. 2 shows the results of pitch tracking for 6 sinusoids selected among 40 peaks.

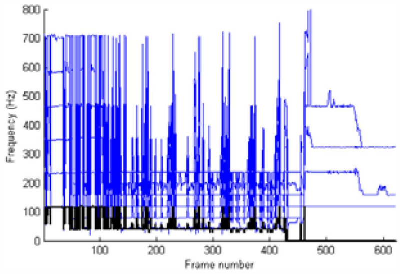

Fig. 2. Pitch tracking of the Jing sound using conventional SMS

As illustrated in Fig. 2, the number of sinusoids and those values in frame are detected and measured but pitch tracking for inter-frame are performed incorrectly. According to the conventional SMS, in the procedure of peak-picking, the peak values and the locations of them are computed and stored and then pitch-tracking is performed computing the distance of locations of peaks for the inter-frame[3]. This pitch-tracking algorithm is suitable that the magnitude of peaks has monotonic envelops in each frame. However, envelop patterns of Jing's partials are irregular so that the unexpected results of pitch-tracking were shown as Fig. 2. Therefore, this paper proposes a modified peak-picking and pitch-tracking algorithm in accordance with characteristics of Jing sounds.

B. Proposed Method

The procedure for the proposed peak-picking and pitch-tracking algorithm has four steps, as follows:

Step 1. Compute the STFT, $x_m(e^{jw})$ for input sounds x, the number of frames m.

Step 2. Decompose a signal into sub-band by predefined band size.

Step 3. Find the spectral peak in each sub-band, and store the values and locations of each peak.

Step 4. If values of peaks are less than the predefined threshold value, the peaks are discarded, which are not used for parameters.

IV. EXPERIMENTAL RESULTS

In this paper, the size of sub-band is determined 100Hz and the threshold value is set 0.3dB according to analysis of the spectral characteristics of the Jing. Also the other parameters are set the same as the configuration described in section 2.2. Fig. 3 described the results of the proposed algorithm applied to the sound of Fig. 2. Compared to the results of Fig. 2, the proposed algorithm performs more accurate pitch-tracking.

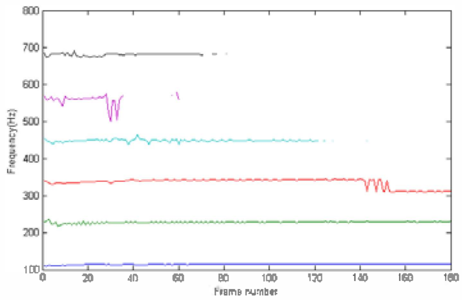

Fig. 3. Results of the proposed algorithm

In order to observe the softening effect of the Jing, the proposed algorithm is applied to the sound minimizing the window size and increasing zero padding parameters simultaneously for high resolution. Fig. 4 depicts changes of frequency of the third partials which have the highest energy distribution among partials. The slope of frequency of third peak generally increases. This represents the softening behavior that the conventional SMS cannot observe.

The unexpected frequency changes in 4, 5, 9 and 30 frames can be interpreted as beat phenomenon. In acoustics, beat phenomenon is defined as an interference between two sounds having slightly different frequencies and shows periodical changes of magnitudes between the two frequencies[12]. In Fig. 5(a), this interference occurs at both approximately

700 and 6,500 samples simultaneously. Fig. 5(b) shows that two different frequencies near 350 Hz are interfered each other.

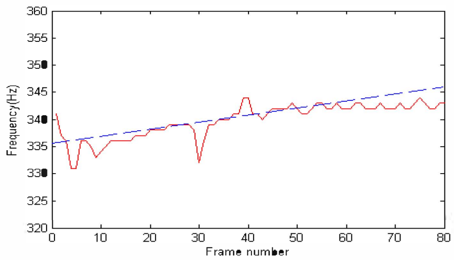

Fig. 4. Softening effect: tracking (solid line) and trend (dashed line) of the third partial

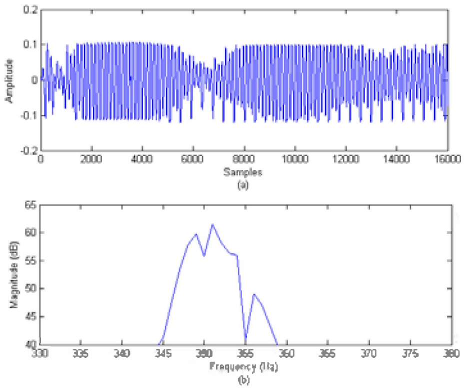

Fig. 5. Beat phenomenon in the (a) time domain and (b) frequency domain

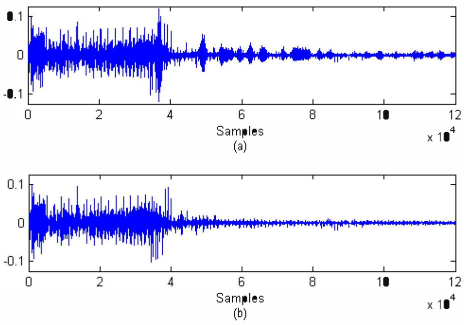

Fig. 6. Residual signals from (a) the conventional SMS and (b) the proposed SMS

The locations of falling amplitude, around 700 and 6,500 samples, correspond to the 4th and 5th frame and the time interval of these locations is approximately 0.122 sec (about 8 Hz). This can explain the unexpected frequency changes in Fig. 4. In addition, if we consider the parameters of beat phenomena to be extracted, the window size should be increased enough to include two falling amplitudes, at least over 0.112 sec.

In order evaluate the proposed algorithm by comparing it to conventional SMS, the residual signals of them are represented in Fig. 6. A fair number of sinusoidal components are included in residual signal in Fig. 6(a) especially from 5,000 samples, whereas the proposed algorithm significantly reduced sinusoidal components in residual signal in Fig. 6(b).

V. CONCLUSIONS

This paper described an effective method for the precise peak-picking and pitch-tracking algorithm that can guarantee accurate sinusoids and well-approximated residual signals. The proposed method improved the conventional SMS by extracting accurate magnitudes of peaks changing irregularly for each frame. The proposed algorithm allows pitch-tracking and for Jing sounds to be performed, which the conventional SMS cannot do, as well as, softening phenomena and beat phenomena can be interpreted. The proposed algorithm reduced sinusoidal components in the residual signal significantly, so not only it can synthesize successfully Jing sounds but be also applied to analyze general percussion instruments.

In the proposed method, big window size should be used for parameter extraction of beat phenomena, whereas window size should be as small as possible for representing a softening effect. To determine the both window size are left to future works.

REFERENCES

[1] C. Roads, *The Computer Music Tutorial*, The MIT press, London, 1996

[2] J. O. Smith, “Physical Modeling using Digital Waveguides,” *Computer Music Journal*, vol. 16, no. 4, pp. 74–87, 1992

[3] X. Serra, J. O. 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

[4] J. A. Moorer, “The Heterodyne Filter as a Tool for Analysis of Transient Waveforms,” Memo AIM-208. Stanford Artificial Intelligence Laboratory. Computer Science Dept., Stanford University, 1973

[5] J. A. Moorer, “The Use of the Phase Vocoder in Computer Music Applications,” *Jounal of the Audio Engineering Society*, vol. 26, no. 1/2, pp. 42–25, 1978

[6] M. R. Portnoff, “Implementation of the Digital Phase Vocoder Using the Fast Fourier Transform,” *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 24, no. 3 pp. 243–248, 1976

[7] X. Serra, J. O. Smith, “PARSHL: an analysis/synthesis program for non-harmonic sounds based on a sinusoidal representation,” *Proc. Int. Computer Music Conf.,* San Francisco: Computer Music Association, 1987

[8] Y. Hong, S. Cho, M. Kang, H. Han, U. Chong, “Spectrum modeling of Haegum using format extracted from cepstral envelope,” *The 10th Western Pacific Acoustics Conference*, pp. 44, 2009

[9] P. L. Grossman, B. Koplik, Y-Y. Yu, “Nonlinear Vibrations of Shallow Spherical Shells,” *Journal of Applied Mechanics*, vol. 36, pp. 451–458, 1969

[10] N. H. Fletcher, “Nonlinear frequency shifts in quasispherical-cap shells: Pitch glide in Chinese gongs,” *Journal of the Acoustical Society of America*, vol. 78, no. 6, pp. 2069–2073, 1985

[11] T. D. Rossing, N. H. Fletcher, “Nonlinear vibrations in plates and gongs,” *Journal of the Acoustical Society of America*, vol. 73, no. 1, pp. 345–351, 1983

[12] M. E. McIntyre, R.T. Schumacher, J. Woodhouse, “On the oscillations of musical instruments,” *Journal of the Acoustical Society of America,* vol. 74, no. 7, pp. 1325–1345, 1983