# A Modified Spectral Modeling Synthesis Algorithm for Whale Sound

**Pranab Kumar Dhar†, Mohammad Ibrahim Khan††, Kaushik Deb†, and Jong-Myon Kim*††**

†Chittagong University of Engineering and Technology, Chittagong-4349, Bangladesh
†† University of Ulsan, San 29, Mugeo 2-Dong, Nam-Ku, Ulsan, 680-749, Korea

**Summary**
Spectral modeling synthesis (SMS) is a powerful tool for musical sound modeling. This technique considers a sound as a combination of a deterministic plus a stochastic component that makes possible for a synthesized sound to attain all the perceptual characteristics of the original sound. In this paper, we propose a modified SMS which includes a new transformation technique to synthesize the deterministic component of the whale sound from a base sound. This transformation process utilizes the synthesis parameter set (magnitude, frequency, and phase of each sinusoid) of base sound and the difference parameter set which is calculated by subtracting the base sound parameter set from the desired sound parameter set to synthesize the target sound. In addition, this transformation technique reduces the data size of the parameter set. Analysis and simulation results indicate that the proposed SMS well synthesized whale sound which is similar to original one in both time and frequency domain.

***Key words:***
*Spectral modeling synthesis, whale sound, short time Fourier transform*

## 1. Introduction

The whale sound is one of the most complex, non-human, acoustic displays in the animal kingdom. They use sound to attract mates, repel rivals, communicate within a social group or between groups, navigate, or find food [1]. Different species of whales produce distinct sounds, such as songs, moans, clicks, and roars. Some of the sounds produced are not only particular to a species but are also unique in certain areas.
The spectral modeling synthesis (SMS) extracts the synthesis parameters out of real sounds using analysis procedures, being able to reproduce and modify actual sounds. This approach is based on modeling sounds as stable sinusoids (partials) plus noise (residual components) to analyze sounds and generate new sounds. The analysis procedure detects partials by utilizing the time-varying spectral characteristics of a sound, and represents them with time-varying sinusoids [2], [3], [4]. These partials are then subtracted from the original sound where the remaining residual is represented as a time-varying filtered white noise component. The synthesis procedure is a combination of additive synthesis for the sinusoidal part and subtractive synthesis for the noise part [5], [6].
In this paper, we propose modified SMS that includes a new transformation technique to synthesize deterministic component of the whale sounds from one base sound. In this transformation process, the parameters such as magnitude, frequency and phase of each sinusoid of sounds are extracted first. Then subtraction between the parameters of base sound and the target sounds are performed to calculate the difference parameters. Finally, the difference parameters are added with the parameters of the base sound to generate the parameter set for desired sound. This parameter set is used as the components for synthesis process. Simulation results show that the proposed SMS provides good synthesized whale sound which is similar to original one in both time and frequency domain.
The rest of this paper is organized as follows. Section 2 describes the background information regarding the deterministic plus stochastic model. Section 3 presents an overview of the SMS analysis and synthesis process, magnitude and phase spectra computation, peak detection, and peak continuation process. Section 4 introduces our proposed SMS which includes a new transformation process for whale sound synthesis. Section 5 summarizes and discusses the experimental results of the synthesized whale sound using the proposed method. Finally, section 6 concludes this paper.

## 2. Background Information

### 2.1 Deterministic plus Stochastic Model

A sound model assumes certain characteristics of the sound waveform or the sound generation mechanism. Sounds produced by musical instruments, any physical system, or any human voice can be modeled as the sum of sinusoid plus noise residual components. The sinusoidal or deterministic component normally corresponds to the main

Manuscript received September 5, 2010
Manuscript revised September 20, 2010

modes of vibration of the system. The residual component comprises the energy produced by the excitation mechanism not transformed by the system into stationary vibrations plus any other energy component that is not sinusoidal in nature.

A deterministic signal is traditionally defined as anything that is not noise. A stochastic or noise signal is fully described by its power spectral density which gives the expected signal power versus frequency. When a signal is assumed stochastic, it is not necessary to preserve the instantaneous phase. This model considers a waveform signal *s(t)* as the sum of a series of sinusoids plus a residual *e(t),* which is defined as

$$s(t) = \sum_{r=1}^{R} A_r(t)\cos[\theta_r(t)] + e(t) \quad (1)$$

where $R$ is the number of sinusoids, $A_r(t)$ and $\theta_r(t)$ is the instantaneous amplitude and phase of the $r^{th}$ sinusoid, respectively, and $e(t)$ is the noise component at time $t$ (in seconds).

The model assumes that the sinusoids are stable partials of the sound, and each one has a slowly changing amplitude and frequency. The instantaneous phase is taken to be the integral part of the instantaneous frequency $\omega_r(t)$ and therefore satisfies

$$\theta_r(t) = \int_0^t \omega_r(\tau)d\tau \quad (2)$$

where $\omega_r(t)$ is the frequency in radians and $r$ is the sinusoidal number.

By assuming that $e(t)$ is a stochastic signal, it can be described as a filtered white noise,

$$e(t) = \int_0^t h(t,\tau)u(\tau)d\tau \quad (3)$$

where $u(t)$ is the white noise and $h(t,\tau)$ is the response of a time varying filter to an impulse at time $t$. Thus, the residual signal is modeled by the convolution of white noise with time varying frequency-shaping filter [5], [7].

## 3. An Overview of SMS Analysis and Synthesis Process

The deterministic plus stochastic model supports many possible implementations. Both analysis and synthesis models are the frame-based process with the computation done one frame at a time. Figure 1 shows a block diagram for the SMS analysis process. We have analyzed the sound by multiplying it with an appropriate analysis window. Its spectrum is obtained by fast Fourier transform (FFT) and then the prominent spectral peaks are detected and incorporated into the existing partial trajectories by the mean of a peak continuation algorithm. It detects the magnitude, frequency, and phase of the partials presented in the original sound (the deterministic components). When the sound is pseudo harmonic, a pitch detection step can improve the analysis by utilizing the fundamental frequency information in the peak continuation algorithm as well as by selecting the size of the analysis window [5], [6].

The stochastic component of the current frame is calculated by generating the deterministic signal with additive synthesis and then subtracting it from the original waveform in time domain. The stochastic representation is then obtained by performing a spectral fitting of the residual signal.

Figure 2 shows a block diagram of the SMS synthesis process. The deterministic component (sinusoidal component) is calculated from the frequency, magnitude, and phase trajectories. The result of the synthesized stochastic signal is a noise signal by time varying spectral shape obtained in the analysis (i.e., subtractive synthesis). It can be implemented by a convolution in time domain or by a complex spectrum for every spectral envelope of the residual and an inverse-FFT in frequency domain. We then add the deterministic component with stochastic one using an overlap add method [8], [9] in time domain for each frame to obtain the synthesized whale sound.

### 3.1 Magnitude and phase spectra computation

The computation of the magnitude and phase spectra of the current frame is the first step in the analysis. By analyzing the spectra the sinusoid are tracked and decided whether a part of the signal is considered as deterministic or noise. The computation of the spectra is carried out by the short time Fourier transform (STFT). Figure 3 shows the magnitude and phase spectrum for the first frame of a whale sound.

### 3.2 Peak Detection:

Once the spectrum of the current frame is computed, the next step is to detect its prominent magnitude peaks. A peak is defined as a local maximum in the magnitude spectrum. A sinusoid that is stable both in amplitude and in frequency has a well defined frequency representation. Figure 4 shows the peak detection of the first frame of a whale sound.

## 3.3Peak Tracking:

Once the spectral peaks of the current frame have been detected, the peak continuation algorithm adds them to the incoming peak trajectories. Figure 5 shows the peak

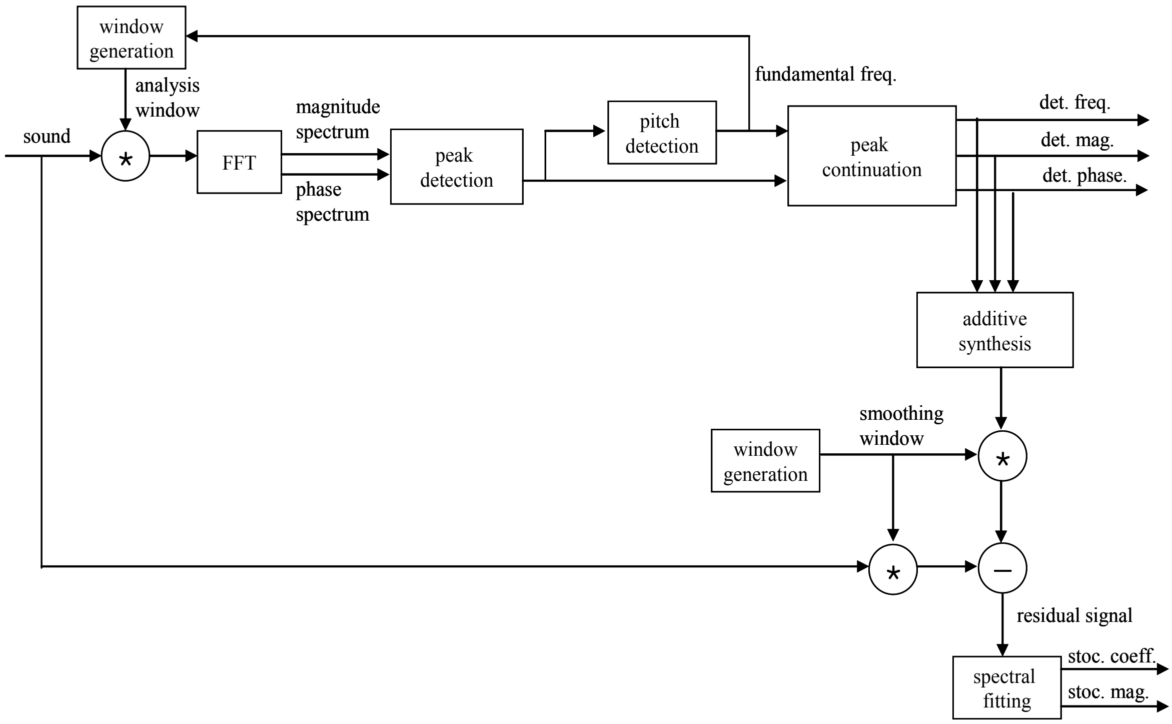


Fig.1. Block diagram of the SMS analysis process

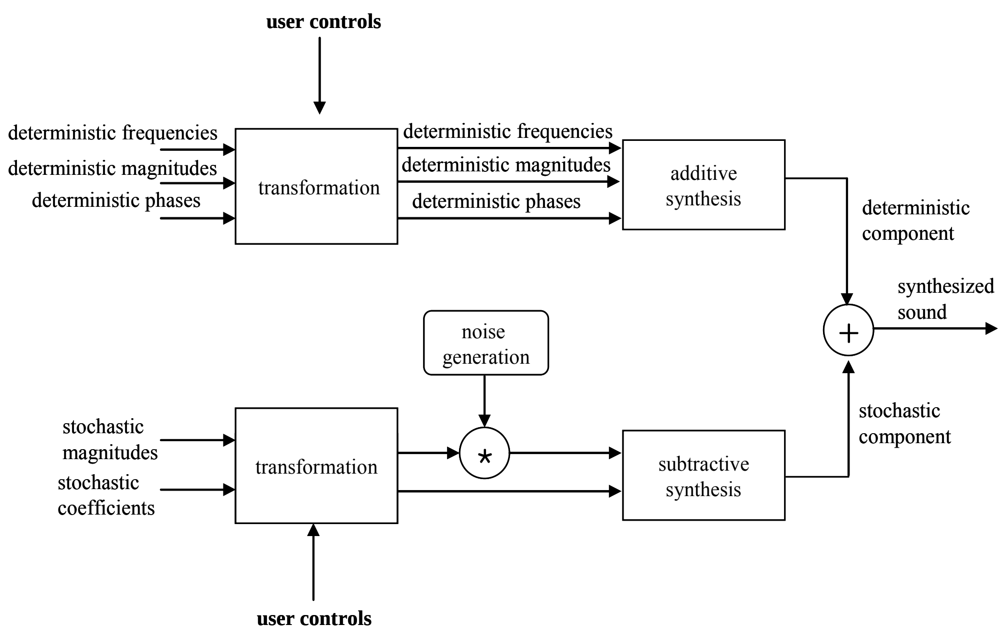


Fig.2. Block diagram of the SMS synthesis process

tracking of a whale sound.

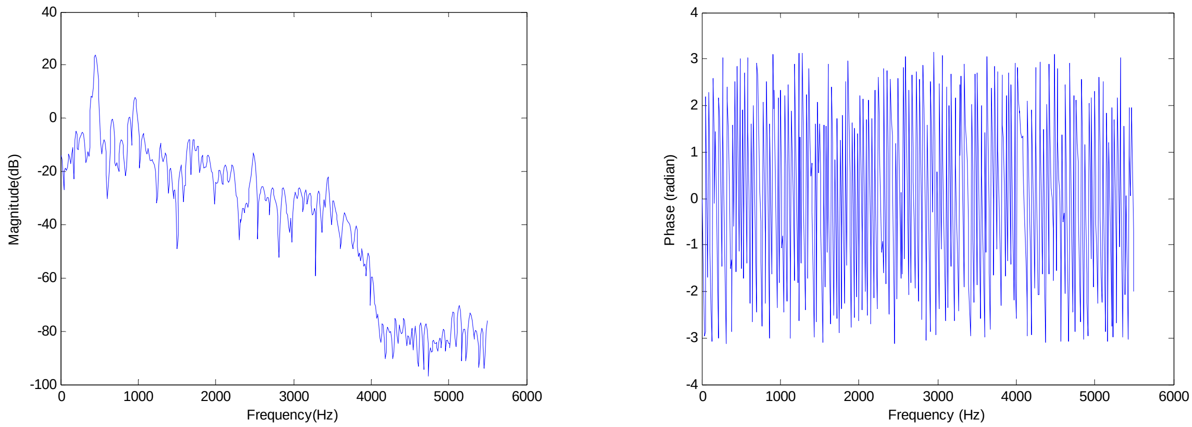


Fig.3. Magnitude and phase spectrum for the first frame of whale sound

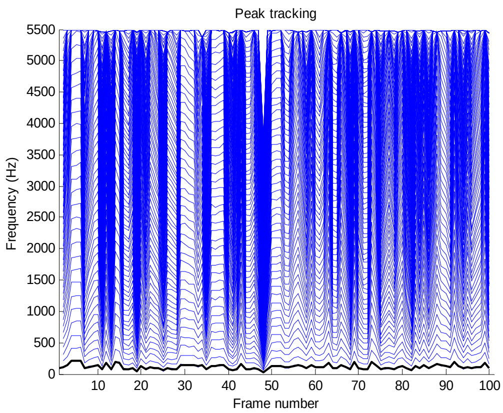


Fig.5. Peak tracking of whale sound

## 4. Proposed Method

In this section, we propose a modified SMS which includes a new transformation technique to synthesize the deterministic component of the target sound. The block diagram of the proposed transformation process is shown in figure 6. In this process, we assume that we have one base sound and we want to synthesize N whale sounds by utilizing the parameter set of the base sound. We can select any whale sound as a base sound. The magnitude, frequency and phase of each sinusoid of the sound are denoted as Mag, Freq, and Phase respectively. For example, the magnitude, frequency and phase of each sinusoid of the base sound are represented as $Mag_B$, $Freq_B$, and $Phase_B$ respectively.

The transformation process is implemented in the following three steps:

Step 1: Calculate the difference parameter set (magnitude, frequency, and phase of the partial) by performing subtraction between the parameter set of base sound and

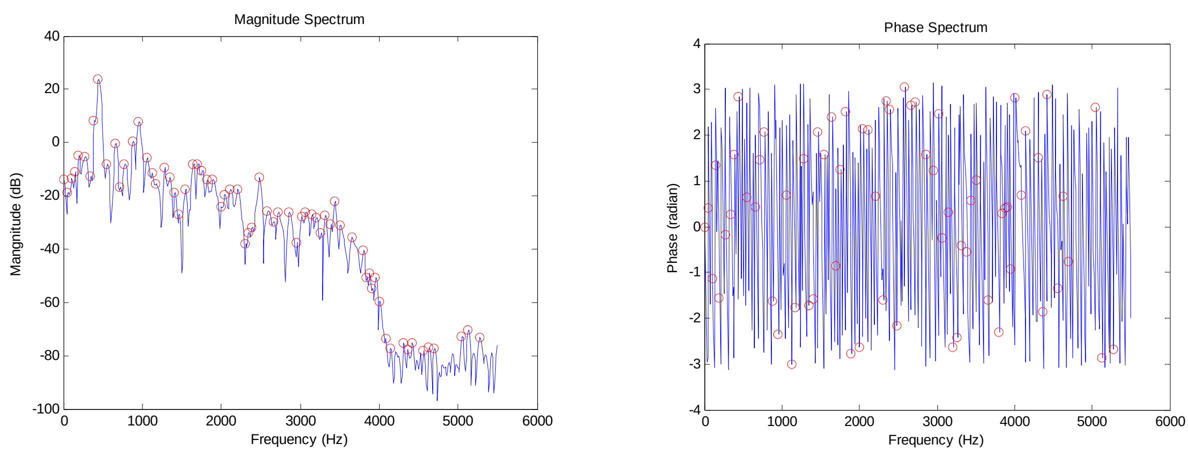


Fig.4. Peak detection in magnitude and phase spectrum

the parameter set of the target sound.
Step 2: Store the difference parameter set (magnitude, frequency and phase) as well as the parameter set of the base sound in database.
Step 3: Add the difference parameters with the parameters of the base sound to generate the parameter set for the

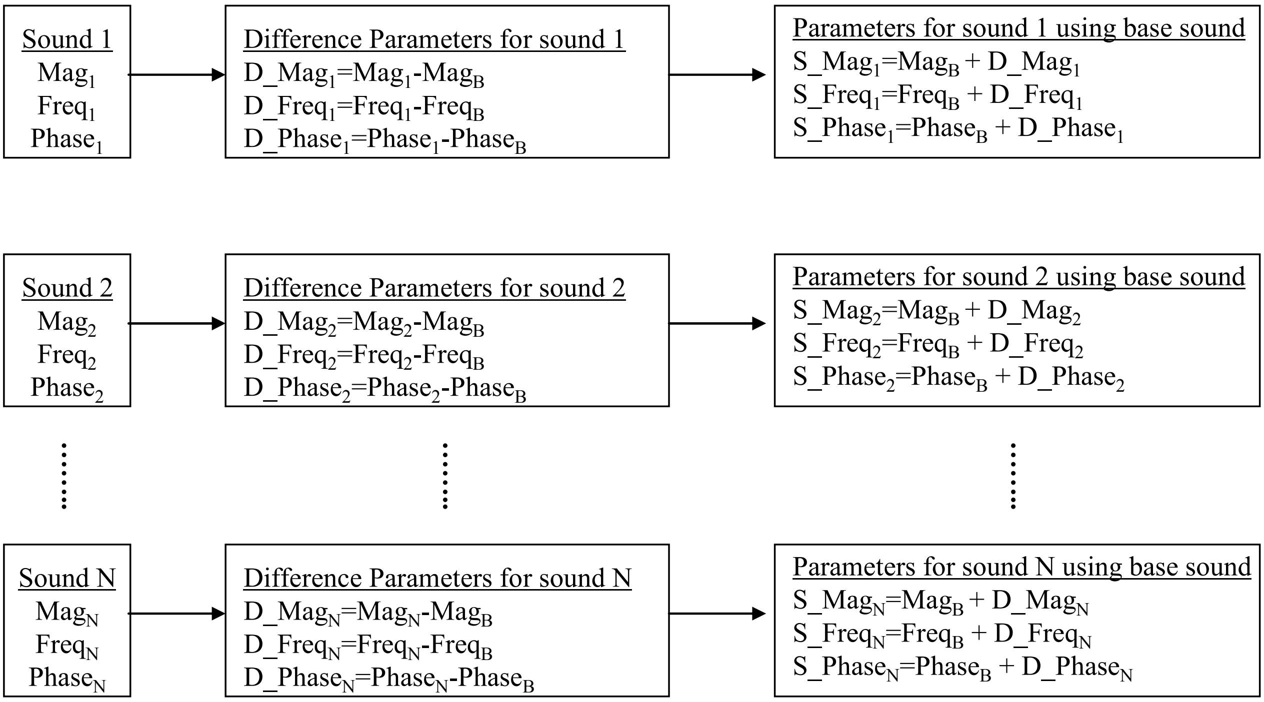


Fig.6. Block diagram of the proposed transformation process

desired sound. This parameter set is used as a component to synthesize the deterministic part of the sound.
The main advantage of this transformation technique is that we only need to store the parameter set of the base sound and the difference parameter set to synthesize the desired sound. As a result, this process reduces the data size of the parameters.
The success of the analysis process depends on the selection of the program parameters such as STFT window, window size, and hop size. One of successful parameter sets is the hanning window, window size of 512, hop size of 256. These selected parameters provide a better result for the whale sound analysis. The sampling frequency of the whale sound used in our simulation is 11 KHz. The duration of the each recorded whale sound is 1.6 second (17,600 samples). By considering frame size of 512 samples and hope size of 256 samples, we have 69 overlapping frames for each sound. From each frame we detect 65 peaks. Thus, in each frame we extract 65 number of magnitude, frequency and phase for each sinusoid of the sound.
To synthesize the stochastic part of the target sound, we assume that the whale sounds have similar stochastic (noise) signal. Thus, we can use the residual signal of base sound as the stochastic component of the target sound.

## 5. Simulation Results and Discussion

In this section, we evaluate the performance of our proposed SMS to synthesize whale sound. The metrics of time domain representation, frequency domain representation, spectrum

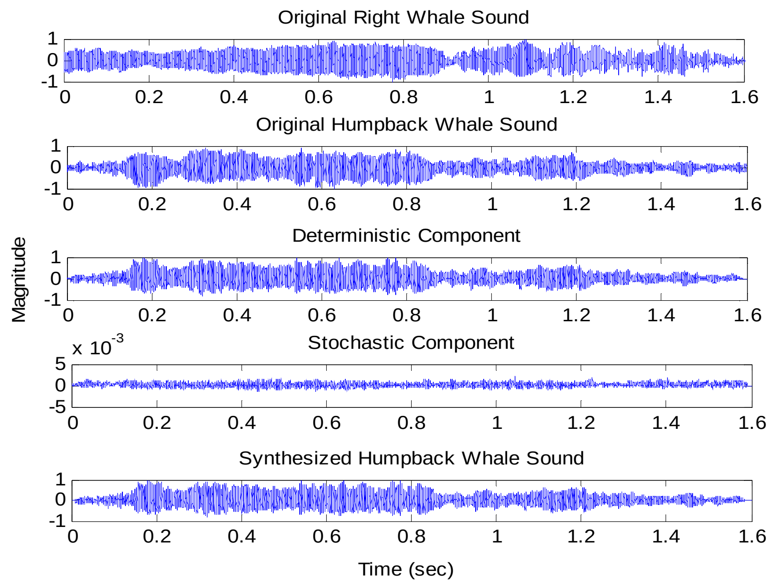


Fig.7. Time domain representation of original and synthesized whale sound

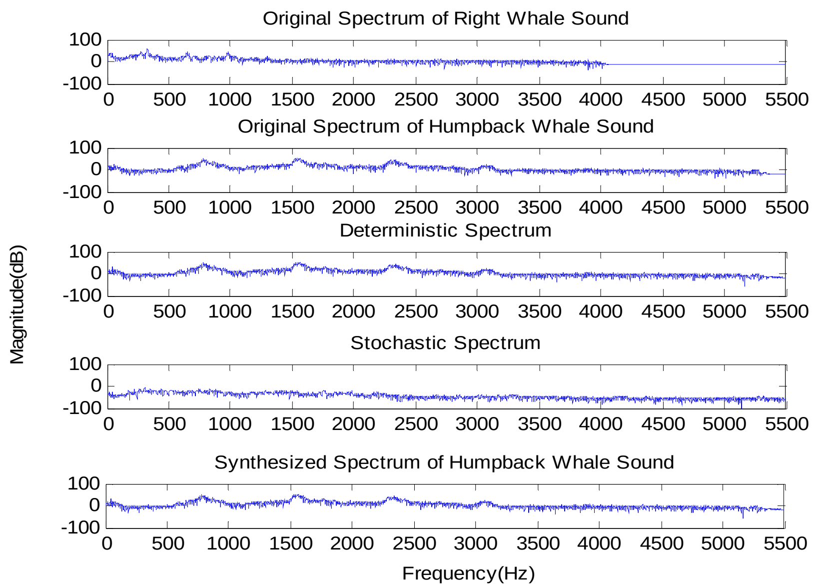


Fig.8. Frequency domain representation of original and synthesized whale sound

matching, and listening of each case form the basis of the study comparison.

In our simulation, we select the right whale sound as the base sound and humpback whale sound as the target sound. Figure 7 and 8 shows the time domain and frequency domain representation of original and synthesized whale sound.

Figure 9 shows the spectrum matching between original and synthesized humpback whale sound.

From these results, we observed that the proposed SMS generate good synthesized sound from base sound which resembles much more closely the original sound.

## 6. Conclusion

In this paper, we have proposed a modified SMS that includes a new transformation technique to synthesize whale sounds. The main advantage of using this transformation technique is to reduce the data size for storing the synthesis parameters to generate the desired sound. Experimental results demonstrate that our proposed SMS can efficiently synthesize the target whale sound from the base sound. This evaluation technique can provide solutions for synthesizing different sounds from one base sound.

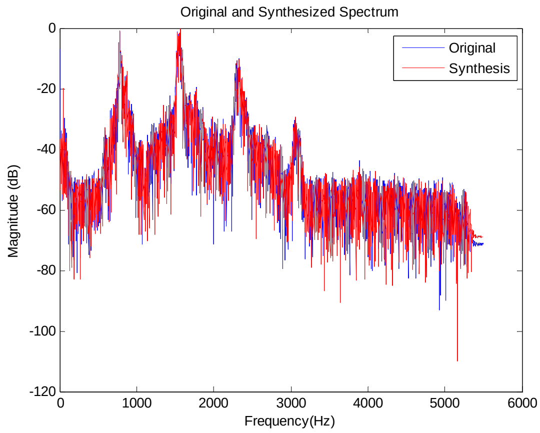


Fig.9. Spectrum matching of original and synthesized humpback whale sound using the proposed method

## References


[1] D. M. Green, A. Ferrarri, D. McFadden, J. S. Pearse, A. N. Popper, W. J. Richardson,S.H. Ridgway, and P. L. Tyack, "Low-frequency sound and marine mammals: Current knowledge and research needs," National Academy Press ,1994.

[2] Ph. Depalle, G. Garcia and X. Rodet, "Tracking of Partials for Additive Sound Synthesis Using Hidden Markov Models," in Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Vol. 1, pp. 225-228, 1993.

[3] J. B Allen, "Short term spectral analysis, synthesis and modification by discrete Fourier transform," IEEE transaction on Acoustics, Speech and Signal Processing, Vol. ASSP-25, pp. 235-238, 1977.

[4] J. B. Allen and R. Lawrenc, "A Unified Approach to Short-Time Fourier analysis and Synthesis," in Proceedings of IEEE, Vol. 65, pp. 1556-1564, 1977.

[5] X. Serra, "Musical Sound Modeling with Sinusoid plus Noise," Musical Sound Processing, published in C. Roads, S. Pope, A. Picialli, G.De Poli editors by Sweets and Zeitlinger Publishers, pp. 91-122, 1997.

[6] 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.

[7] X. Serra, "A System for Sound Analysis/Transformation/Synthesis based on a Deterministic plus Stochastic Decomposition," Ph.D Thesis, Stanford University, 1989.

[8] E. B. George and M. J. T. Smith, "Analysis-by-synthesis/Overlap-add sinusoidal modeling applied to the analysis and synthesis of musical tones," Journal of Audio Engineering Society, Vol. 40, no. 6, pp. 497-516, 1992.

[9] J. B. Allen and R. Lawrenc, "A Unified Approach to Short-Time Fourier analysis and Synthesis," in Proceedings of IEEE, Vol. 65, pp. 1556-1564, 1977.


**Pranab Kumar Dhar** received the B.S. degree in Computer Science and Engineering from Chittagong University of Engineering and Technology (CUET), Chittagong, Bangladesh in 2004. He received M.S. degree from the School of Computer Engineering and Information Technology at University of Ulsan, South Korea in 2010. Since 2005, he has been serving as a faculty member in the Department of Computer Science and Engineering at Chittagong University of Engineering and Technology (CUET), Chittagong, Bangladesh. His research interest includes Sound Synthesis, Digital Watermarking, Digital Image Processing, Multimedia Systems, and Digital Signal Processing.

**Mohammad Ibrahim Khan** received the B.S. degree in Electrical and Electronic Engineering from Bangladesh University of Engineering and Technology (BUET), Dhaka, Bangladesh in 1999. He received M.S. degree in Computer Science and Engineering from the same University in 2002. He received his Ph.D degree in Computer Science and Engineering from Jahangirnagar University in 2010. Since 1999, he has been serving as a faculty member in the Department of Computer Science and Engineering at Chittagong University of Engineering and Technology (CUET), Chittagong, Bangladesh. His research interest includes Digital Image Processing, Graph Theory, Sound Synthesis, Cryptography, Digital Watermarking, Multimedia Systems, and Digital Signal Processing.

**Kaushik, Deb** received the B. Tech. and M. Tech. Degrees from Department of Computer Science and Eng. of Tula State University, Tula, Russia, in 1999 and 2000 respectively. Since 2001, he has been serving as a faculty member in Department of Computer Science and Eng., Chittagong University of Engineering and Technology, Chittaong, Bangladesh. He is currently a Ph.D. candidate at the School of Electrical Engineering and Information Systems, University of Ulsan, Korea. His research interests include Computer Vision, Pattern Recognition, Intelligent Transportation System (ITS) and Human-Computer Interaction.

**Jong-Myon Kim** received the B.S. degree in Electrical Engineering from Myongji University, South Korea in 1995. He received M.S. degree from the School of Electrical and Computer Engineering at University of Florida, Gainesville, FL, USA in 2000. He received his Ph.D. degree from the School of Electrical and Computer Engineering at Georgia Tech, Atlanta, GA in 2005. Currently he is a faculty member in the School of Computer Engineering and Information Technology at University of Ulsan, South Korea. His research interest includes high performance, energy efficient computer architectures as well as application specific SOC design for embedded color image and video processing. Prior to joining academia, he was a processor architect in both multi-format CODEC chip design and prototyping team and reconfigurable processor team at Samsung Advanced Institute of Technology (SAIT), South Korea. He is actively participating as a member of variety of professional research societies, like IEEE, KIPS, KIISE, SPIE, KAS and EASCO.