DOI: 10.3745/KIPSTB.2010.17B.1.69

수정된 스펙트럴 모델링을 이용한 수염고래 소리 합성

전희성[†] · 파르나브 다르[††] · 김철홍[†††] · 김종면[††††]

요 약

스펙트럴 모델링 합성 (Spectral Modeling Synthesis, SMS)은 뮤지컬 사운드 모델링을 위한 강력한 툴로써 사용되어 왔다. 이 기술은 사운드를 결정적 (deterministic) 성분과 통계적 (stochastic) 성분의 조합으로 간주한다. Deterministic 성분은 크기 (amplitude), 주파수 (frequency), 위상 (phase) 함수에 따른 사인파의 연속으로 표현되는 반면, stochastic 성분은 백색 잡음 (white noise)으로 자극된 시간 변화 필터로서 동작하는 크기 스펙트럼 엔블로프 (spectrum envelop)의 연속으로 표현된다. 이러한 표현들은 원음의 모든 지각적인 특징들을 활용해 합성된 사운드를 구현 가능케 한다. 하지만, 고래 소리와 같은 복잡한 사운드에 대해 기존의 SMS를 사용할 때 연속적인 프로임에 있는 부분 주파수가 다른 경우 결정적 성분에서 상당한 위상 변화가 발생한다. 왜냐하면 기존의 SMS는 사운드의 결정적 성분을 합성하기 위해서 계산된 위상을 이용하기 때문이다. 그 결과 기존의 SMS는 높은 주파수 영역에서 원래 스펙트럼과 합성된 스펙트럼 사이에서 좋은 스펙트럼 매칭을 제공하지 못한다. 이러한 문제를 해결하기 위해 본 논문은 수정된 SMS를 제안한다. 제안하는 SMS는 결정적 성분을 합성하기 위해 원래 주파수 정보를 이용할 뿐만 아니라 주파수 영역에서 복잡한 잔재 (residual) 스펙트럼을 계산함으로써 원음과 합성음 사이에서 좋은 스펙트럼 매칭을 제공한다. 다양한 고래 소리 합성을 모의 실험한 결과, 제안된 방법은 시간 및 주파수 영역에서 기존의 SMS와 유사한 성능을 보였다. 하지만, 제안된 방법은 기존의 SMS보다 스펙트럼 매칭에서 더 좋은 성능을 보였다.

키워드 : 고래소리 합성, 스펙트럴 모델링, 짧은 시간 푸리에 변환, 가산 합성, 위상 변화

Baleen Whale Sound Synthesis using a Modified Spectral Modeling

Hee-Sung Jun[†] · Pranab K. Dhar[††] · Cheol Hong Kim[†††] · Jong-Myon Kim[††††]

ABSTRACT

Spectral modeling synthesis (SMS) has been used as a powerful tool for musical sound modeling. This technique considers a sound as a combination of a deterministic plus a stochastic component. The deterministic component is represented by the series of sinusoids that are described by amplitude, frequency, and phase functions and the stochastic component is represented by a series of magnitude spectrum envelopes that functions as a time varying filter excited by white noise. These representations make it possible for a synthesized sound to attain all the perceptual characteristics of the original sound. However, sometimes considerable phase variations occur in the deterministic component by using the conventional SMS for the complex sound such as whale sounds when the partial frequencies in successive frames differ. This is because it utilizes the calculated phase to synthesize deterministic component of the sound. As a result, it does not provide a good spectrum matching between original and synthesized spectrum in higher frequency region. To overcome this problem, we propose a modified SMS that provides good spectrum matching of original and synthesized sound by calculating complex residual spectrum in frequency domain and utilizing original phase information to synthesize the deterministic component of the sound. Analysis and simulation results for synthesizing whale sounds suggest that the proposed method is comparable to the conventional SMS in both time and frequency domain. However, the proposed method outperforms the SMS in better spectrum matching.

Keywords : Whale Sound Synthesis, Spectral Modeling, Short Time, Fourier Transform(STFT), Additive Synthesis, Phase Variation

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 [9]. Different species of whales produce distinct

※ This work was supported by 2004 Research Fund of University of Ulsan.

[†] 정 회 원 : 울산대학교 컴퓨터정보통신공학부 교수
[††] 준 회 원 : 울산대학교 컴퓨터정보통신공학부 석사과정
[†††] 정 회 원 : 전남대학교 전자컴퓨터공학부 교수
[††††] 종신회원 : 울산대학교 컴퓨터정보통신공학부 교수(교신저자)

논문접수 : 2009년 9월 24일
수 정 일 : 1차 2009년 12월 7일
심사완료 : 2009년 12월 7일

sounds, such as songs, moans, clicks, roars, and sighs. Some of the sounds produced are not only particular to a species but are also unique in certain areas. Some baleen whales such as bowhead, minke, right, fin whales can produce complex sound. In this study we synthesized different kinds of baleen whale sound using spectral modeling synthesis.

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 [5-6, 8]. 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 [1-2]. However, sometimes phase variations occur in the deterministic component using the SMS for complex sound when the partial frequencies in successive frames differ. This is because it utilizes the calculated phase to synthesize deterministic component of the sound. As a result, it does not provide a good spectrum matching between original and synthesized spectrum in the higher frequency region.

To overcome this problem, we propose a modified SMS that utilizes original phase information to synthesize the deterministic component of the sound. The stochastic spectrum is calculated by subtracting the deterministic spectrum from the original spectrum and then using spectral fitting. The stochastic signal is generated by using an inverse short time Fourier transform (STFT) on a series of magnitude spectrum envelopes that function as a time varying filter excited by white noise. We then add the deterministic and stochastic signal in time domain for each frame. In this paper, we synthesize whale sounds using the proposed method and compare the proposed method to the SMS technique. The analysis and simulation results illustrate that the proposed method are comparable to the SMS in both time and frequency domain. However, the proposed method outperforms the SMS in better spectrum matching with original spectrum because of the use of original phase to synthesize the deterministic component of the sound. Thus, the proposed SMS technique can efficiently synthesize the complex whale sound which resembles much more closely the original sound.

The rest of this paper is organized as follows. Section 2 presents background information regarding the deterministic plus stochastic model, an overview of the SMS analysis and synthesis process, magnitude and phase spectra computation, peak detection, pitch detection and peak continuation process. Section 3 presents our proposed method for the higher quality of whale sound synthesis. Section 4 summarizes and discusses experimental results of the different whale sounds for both the SMS and the proposed method, and Section 5 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 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

$$\mathrm{s(t)} = \sum_{r=1}^{R} \mathrm{Ar(t)cos[\Theta r(t)]+e(t)} \qquad (1)$$

where R is the number of sinusoids, *Ar(t)* and *Θr(t)* is the instantaneous amplitude and phase of the rth 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 *ω r(t)* and therefore satisfies

$$\Theta\mathrm{r(t)} = \int_0^t \omega\mathrm{r}(\tau)d\tau \qquad (2)$$

where *ωr(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,

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

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 [1, 3].

2.2 An Overview of SMS Analysis and Synthesis Process

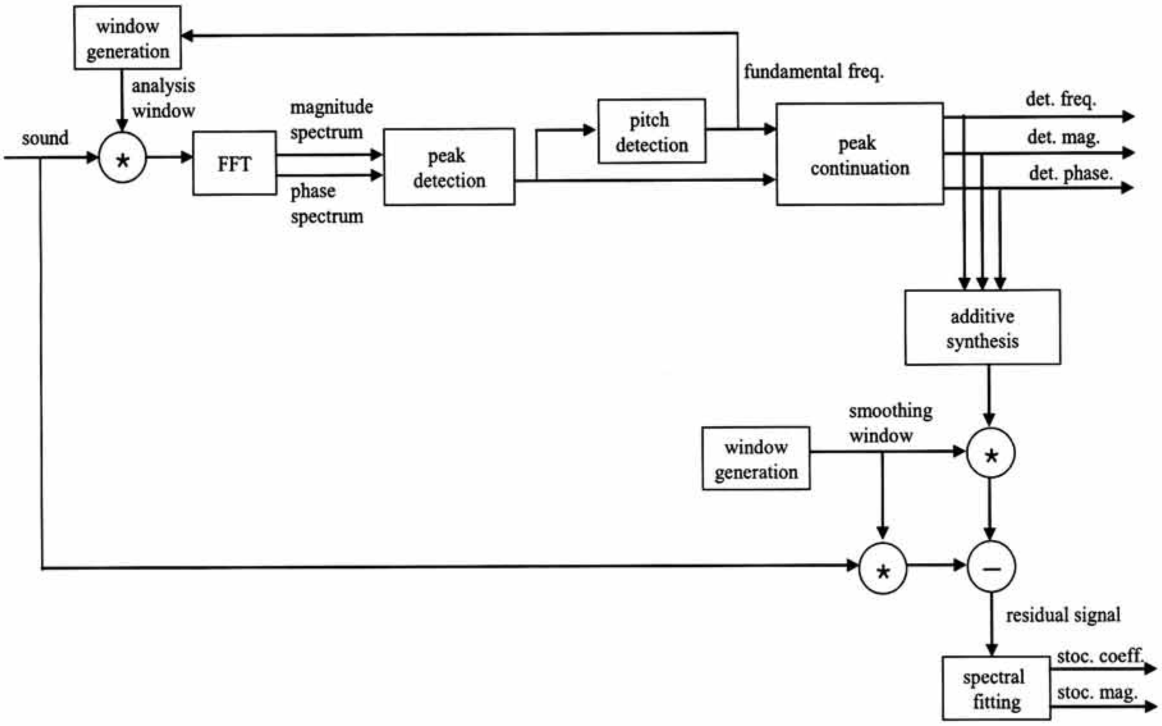
The deterministic plus stochastic model supports many possible implementations. Both analysis and synthesis models are the frame-based process with the computation done one frame at a time. (Fig. 1) shows a block diagram for the SMS analysis process. We have analyzed the sound by multiplying it with 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 [1-3].

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.

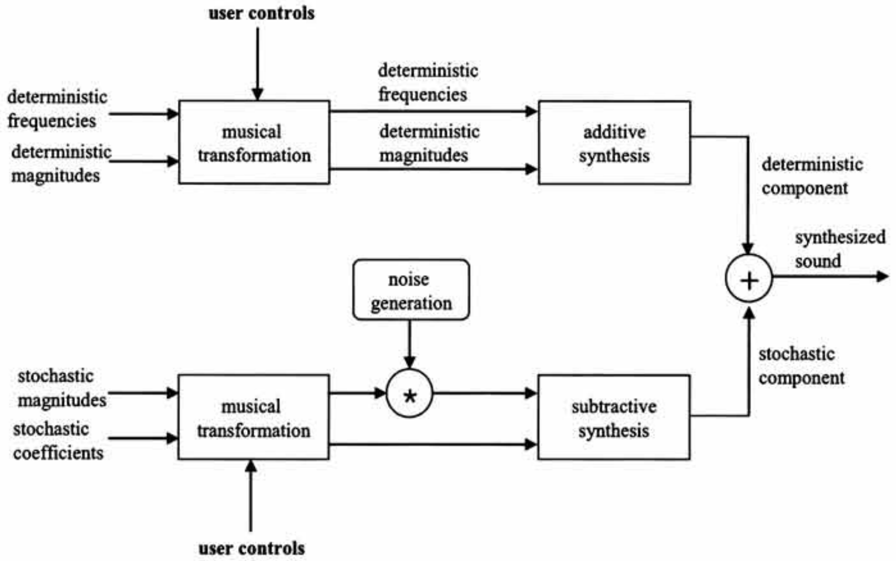
(Fig. 2) shows a block diagram of the SMS synthesis process. The deterministic component (sinusoidal component) is calculated from the frequency and magnitude trajectories. The result of the synthesized stochastic signal is a noise signal by time varying spectral shape obtained in the analysis (i.e., 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.

2.3 Magnitude and phase spectra computation

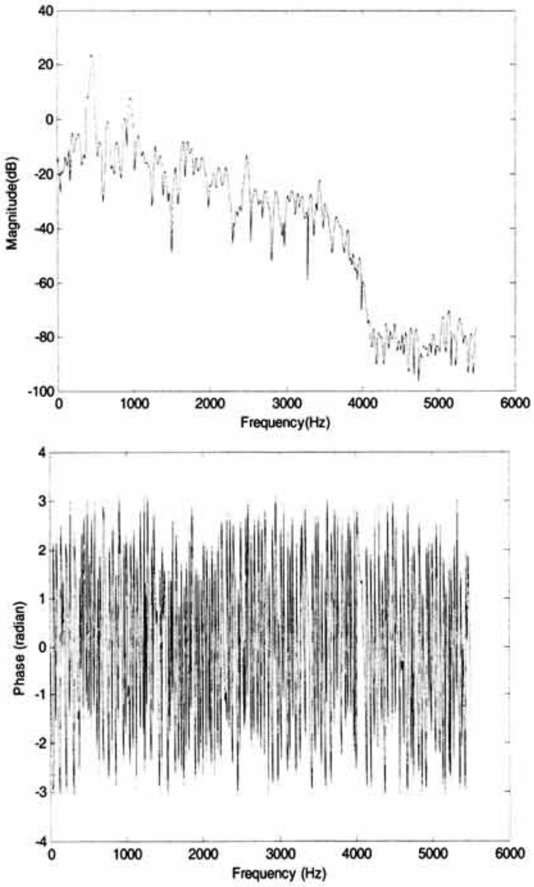
The computation of the magnitude and phase spectra of the current frame is the first step in the analysis. By analyzing the spectra the sinusoid are tracked and decided whether a part of the signal is considered as deterministic or noise. The computation of the spectra is carried out by the short time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



(Fig. 1) Block diagram of the SMS analysis process



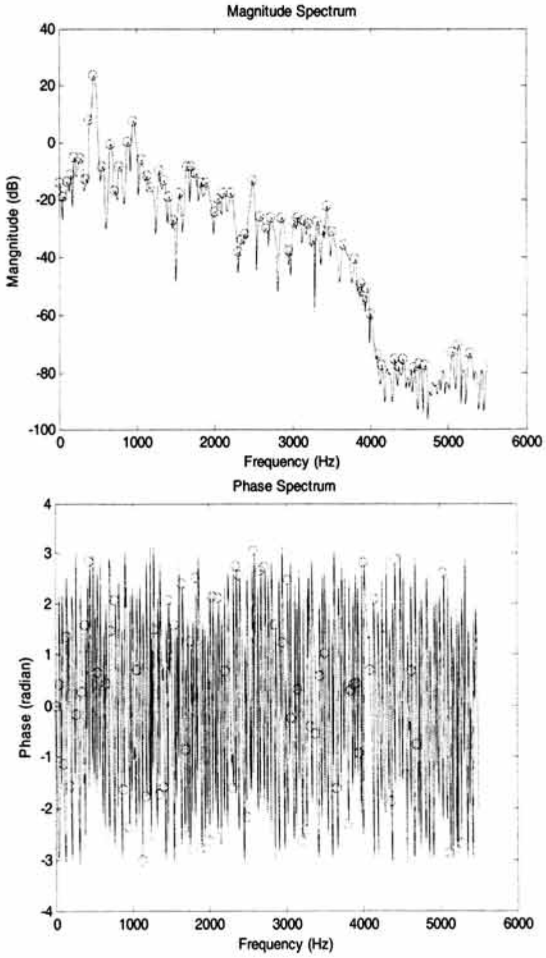
(Fig. 2) Block diagram of the SMS synthesis process



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

2.4 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. (Fig. 4) shows the peak detection of the first frame of bowhead whale sound.



(Fig. 4) Peak detection in magnitude and phase spectrum

2.5 Pitch detection

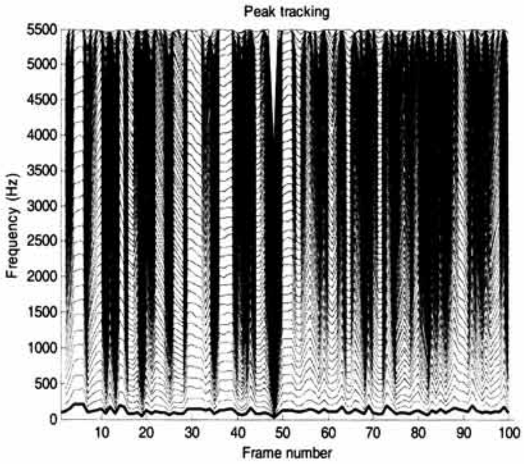
Before continuing a set of peak trajectories from the current frame it is useful to search for a possible fundamental frequency for periodicity. If it exists, we have more information to simplify and improve the tracking of partials. This fundamental frequency can also be used to set the size of the analysis window, in order to maintain the constant number of periods to be analyzed at each frame and to get the best time-frequency trade-off possible. This is called as a pitch-synchronous analysis.

2.6 Peak Continuation

Once the spectral peaks of the current frame have been detected, the peak continuation algorithm adds them to the incoming peak trajectories. (Fig. 5) shows the peak tracking of the bowhead whale sound.

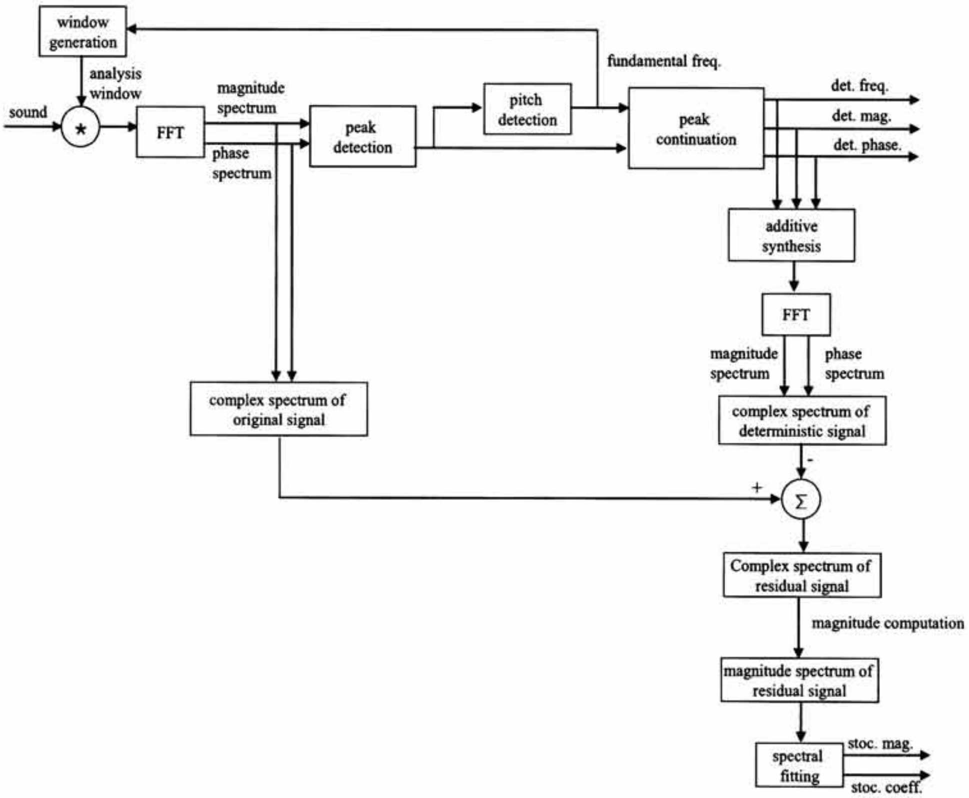
3. Proposed Method

To provide better spectrum matching, we propose a modified SMS that calculates the complex residual spectrum in frequency domain and utilizes original phase information to synthesize the deterministic component of sound. We can obtain the stochastic representation of the residual signal by subtracting the deterministic spectrum from the original spectrum and then utilizing spectral fitting (line segment approximation) of the magnitude spectrum.

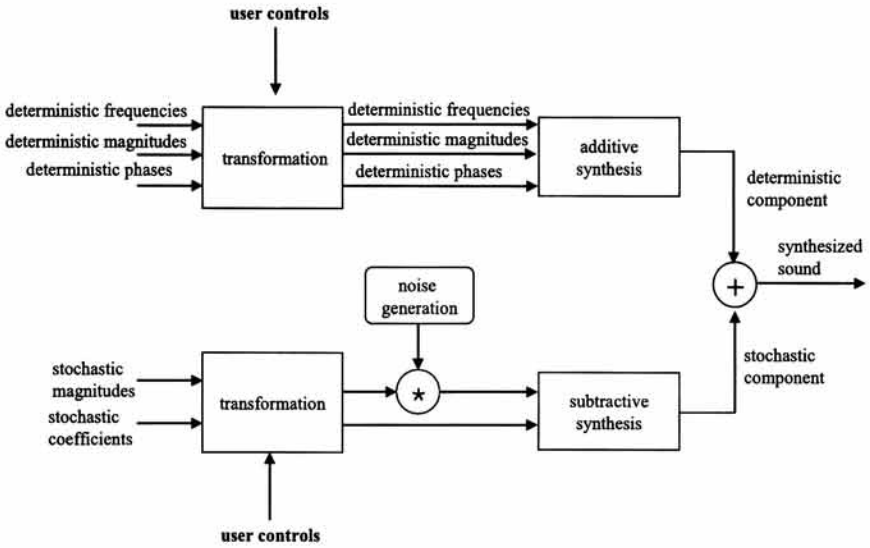


(Fig. 5) Peak tracking of the bowhead whale sound

In the synthesis process, the deterministic signal is calculated by a sine wave for each magnitude, frequency, and phase trajectory. The stochastic signal is calculated by a complex spectrum envelope of the residual and an inverse STFT. We then add the deterministic component with stochastic one using an overlap add method [4, 7] in time domain for each frame to obtain the synthesized whale sound. (Fig. 6) and (Fig. 7) show the analysis and synthesis processes of the proposed method, respectively.



(Fig. 6) Block diagram of the analysis process in the proposed SMS



(Fig. 7) Block diagram of the synthesis process in the proposed SMS

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, and hop size of 256. These selected parameters provide a better result for the whale sound analysis. The sampling frequency of the whale sounds used in our simulation is 11 KHz. The duration of the recorded minke, right, bowhead and fin whale sounds is 3.072, 2.365, 2.406, and 1.226 second respectively.

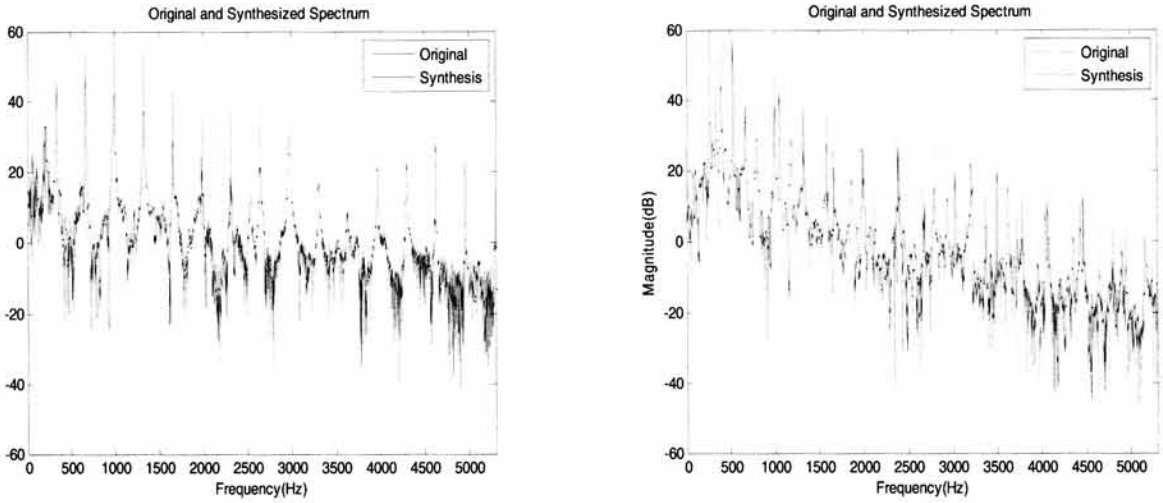
4. Results and Discussion

In this section, we evaluate the performance of our proposed method to synthesize different whale sound, and compare the proposed method to the SMS. The metrics of time domain representation, frequency domain representation, spectrum matching, and listening of each case form the basis of the study comparison.

We observe that both the proposed method and the SMS generate a good synthesized sound which resembles much more closely the original sound. However, phase variations were occurred in the deterministic component using the SMS for complex whale sounds when the partial frequencies in successive frames differ. This is because it utilizes the calculated phase to synthesize deterministic component of the sound. This results in not providing a good spectrum matching between original and synthesized spectrum in higher frequency range. Sometimes we can ignore phase variation to synthesize deterministic component for simple harmonic sound when the partial frequencies in successive frames are similar. However, some whale sounds are more complex sound than the harmonic sound such as musical instrument sounds. Thus, to synthesize such complex whale sounds we should consider the phase variation for good spectrum matching between original and synthesized spectrum. The proposed SMS can overcome this phase variation problem by utilizing original phase information to synthesize the deterministic component of the sound.

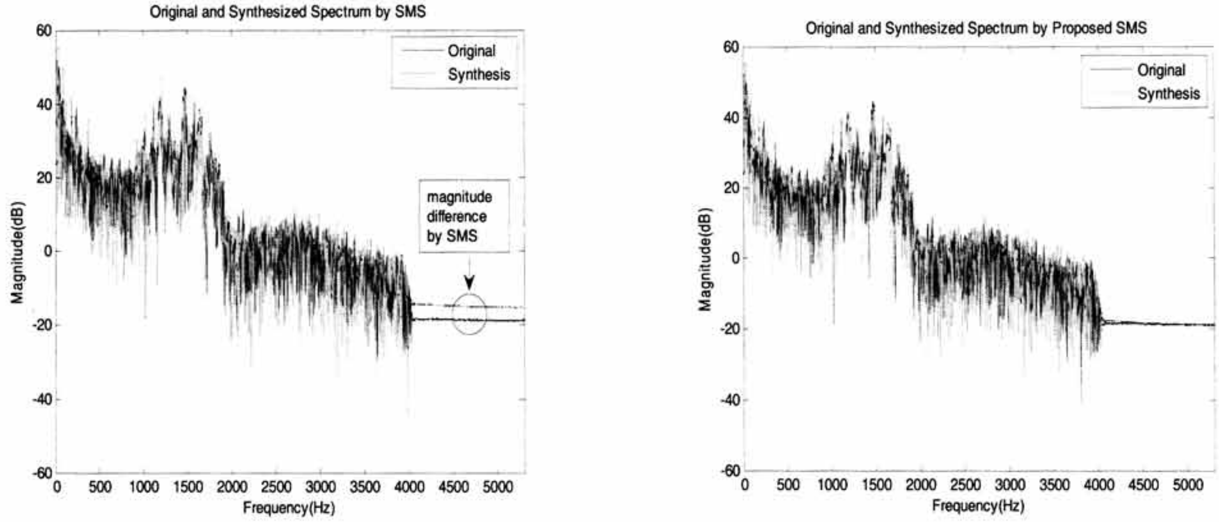
In the synthesis process, conventional SMS utilizes the calculated phase to generate the synthesized deterministic component of the sound. This method provides good spectrum matching of original and synthesized sound when the partial frequencies are only stationary in successive frames. (Fig. 8) (a) and (Fig. 8) (b), for example, show the original and synthesized sounds of the guitar and piano generated by using the SMS which provides good spectrum matching.

However, for the complex whale sounds including bowhead, minke, right, fin whales this method does not provide good spectrum matching in higher frequency region when partial frequencies in successive frames differ, resulting in phase variation in the deterministic component of the sound. This is because it uses the calculated phase to synthesize the deterministic signal. (Fig. 9) (a), (Fig. 10) (a), (Fig. 11) (a), (Fig. 12) (a) show the spectrum matching of original and synthesized minke, right, bowhead, and fin whale sound generated using the SMS, respectively. We observed that in lower frequency region (between 0 to 3.5 KHz approximately) the conventional SMS provides good spectrum matching but in higher frequency region (between 3.5 KHz to 5 KHz approximately) it does not provide good spectrum



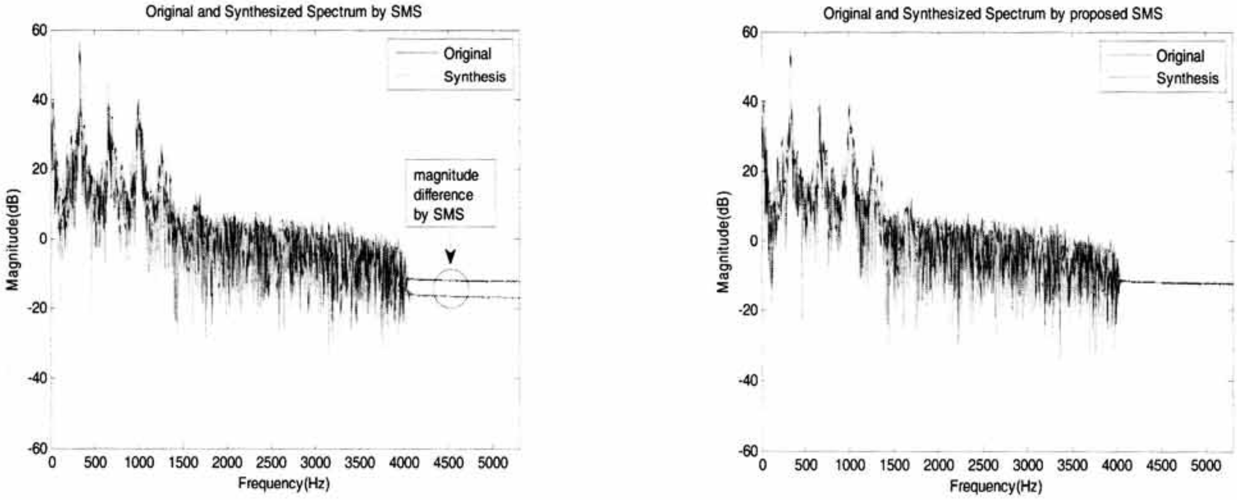
(a) (b)

(Fig. 8) Spectrum matching of original and synthesized sounds: (a) guitar and (b) piano



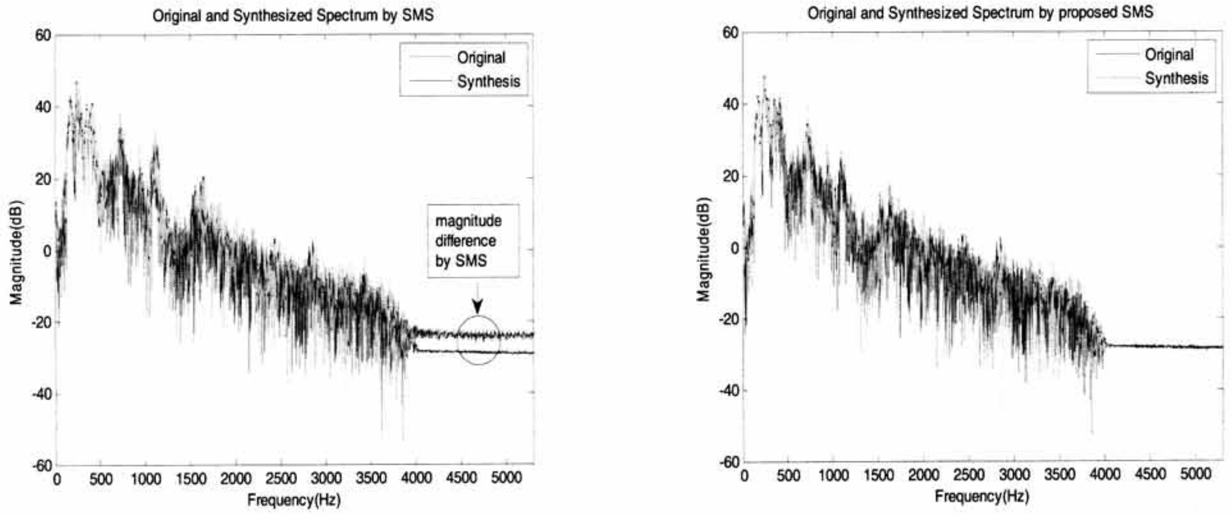
(a) (b)

(Fig. 9) Spectrum matching of original and synthesized sound for minke whale: (a) using SMS and (b) using proposed SMS



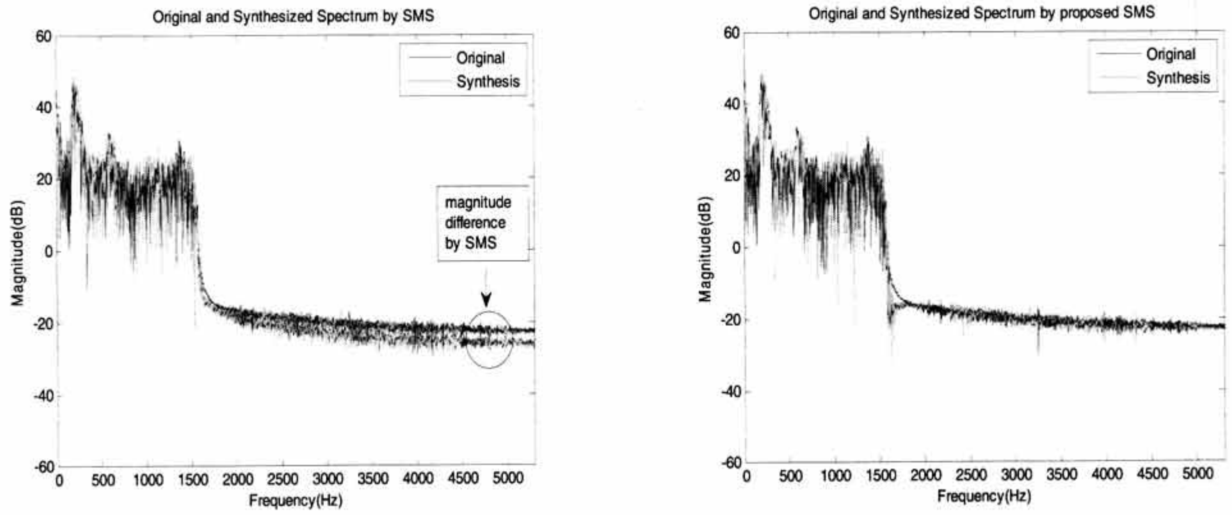
(a) (b)

(Fig. 10) Spectrum matching of original and synthesized sound for right whale: (a) using SMS and (b) using proposed SMS



(a) (b)

(Fig. 11) Spectrum matching of original and synthesized sound for bowhead whale: (a) using SMS and (b) using proposed SMS



(a) (b)

(Fig. 12) Spectrum matching of original and synthesized sound for fin whale: (a) using SMS and (b) using proposed SMS

matching of the original and synthesized sound. The proposed SMS overcome this problem by utilizing original phase information to synthesize the deterministic component of the sound. (Fig. 9) (b), (Fig. 10) (b), (Fig. 11) (b), (Fig. 12) (b) show the spectrum matching of original and synthesized minke, right, bowhead, and fin whale sounds using the proposed SMS. We observed that the proposed SMS provides better spectrum matching of the original and synthesized sound than the conventional SMS in the higher frequency region (between 3.5 KHz to 5 KHz approximately) as well as in the lower frequency region (between 0 to 3.5 KHz approximately). Overall, the proposed method outperforms the SMS in better spectrum matching of the original and synthesized whale sounds.

5. Conclusion

In this paper, we have proposed a modified spectral modeling synthesis (SMS) to synthesize whale sounds. Since whale sounds are more complex than musical instrument sounds, the conventional SMS cannot be used directly to the whale sound. We observed that the conventional SMS has occurred considerable phase variations in the deterministic component of whale sounds when the partial frequencies in successive frames differ each other. This was because it utilizes the calculated phase to synthesize the deterministic component of the sound. As a result, it could not provide a good spectrum matching between the original and synthesized spectrum in the higher frequency region. To overcome this problem, we have presented our modified SMS which utilizes

original phase information to synthesize the deterministic component of the sound and calculates the complex residual spectrum in frequency domain. This provides good spectrum matching of the original and synthesized sound. Analysis and simulation results for different baleen whale sound synthesis indicate that the proposed method outperforms the conventional SMS in spectrum matching between original and synthesized spectrums.

References

[1] 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.

[2] 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.

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

[4] 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.

[5] 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.

[6] 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.

[7] R. J. McAulay and T. F. Quatieri, "Speech Analysis/Synthesis based on a Sinusoidal Representation," IEEE transaction on Acoustics, Speech and Signal Processing, Vol.34, No.4, pp. 744-754, 1986.

[8] 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.

[9] D. M. Green et. Al., 'Low-frequency Sound and Marine Mammals: Current Knowledge and Research needs,' National Academy Press, 1994.

전 희 성

e-mail : hsjun@ulsan.ac.kr
1981년 서울대학교 전기공학과(학사)
1983년 서울대학교 전기공학과(공학석사)
1983년~1986년 금성반도체(주) 연구소 연구원
1992년 Electrical & Computer Engineering, Rutgers - The State University of New Jersey, USA(공학박사)
1992년~1993년 삼성전자 통신연구소 수석연구원
1993년~현 재 울산대학교 컴퓨터정보통신공학부 교수
관심분야 : 컴퓨터비전, 디지털 신호처리 등

파르나브 다르

e-mail : pranab_cse@yahoo.com
2004년 Computer Science and Engineering, Chittagong University of Engineering and Technology (CUET), Chittagong, Bangladesh (BS)
2005년~2008년 Computer Science and Engineering, Chittagong University of Engineering and Technology (CUET), Chittagong, Bangladesh, Lecturer
2008년~현 재 울산대학교 컴퓨터정보통신공학부 석사과정
관심분야 : 사운드 합성, 디지털 워터마킹, 멀티미디어 압축, 임베디드시스템 등

김 철 홍

e-mail : cheolhong@gmail.com
1998년 서울대학교 컴퓨터공학과(학사)
2000년 서울대학교 컴퓨터공학부(공학석사)
2006년 서울대학교 전기컴퓨터공학부(공학박사)
2005년~2007년 삼성전자 반도체총괄 SYS.LSI사업부 책임연구원
2007년~현 재 전남대학교 전자컴퓨터공학부 교수
관심분야 : 임베디드시스템, 컴퓨터구조, SoC 설계, 저전력 설계 등

김 종 면

e-mail : jongmyon.kim@gmail.com

1995년 명지대학교 전기공학과(학사)

2000년 Electrical & Computer Engineering, University of Florida, USA (공학석사)

2005년 Electrical & Computer Engineering, Georgia Institute of Technology, USA(공학박사)

2005년~2007년 삼성종합기술원 전문연구원

2007년~현 재 울산대학교 컴퓨터정보통신공학부 교수

관심분야 : 임베디드시스템, 시스템-온-칩, 컴퓨터구조, 병렬처리, 신호처리 등