International Journal of Multimedia and Ubiquitous Engineering
Vol.8, No.5 (2013), pp.51-60
http://dx.doi.org/10.14257/ijmue.2013.8.5.06

Sinusoidal Coding and Spectral Band Replication for Low Bit-Rate Super-Wideband Speech and Audio Coding

Kosangrok Oh, Dong Hoon Sung and Seung Ho Choi*

Dept. of Electronic and IT Media Engineering
Seoul National University of Science and Technology
Seoul 139-743, Korea

E-mail: shchoi@seoultech.ac.kr
** Corresponding author*

Abstract

In this paper, we present a new sinusoidal coding and spectral band replication (SBR) method for a low bit-rate super-wideband speech and audio coding. The sinusoidal coding algorithm utilizes the local rms energy of partial frequency bands in modified discrete cosine transform (MDCT) domain to select peak MDCTs. The SBR is based on the pseudo-spectral correlation between lower band and higher band. The proposed techniques are shown experimentally to give improved performance compared to ITU-T G.718 Annex B from the objective and subjective evaluation.

Keywords: *Sinusoidal coding, Spectral band replication, Super-wideband, MDCT, Pseudo-spectrum*

1. Introduction

While wideband (50-7000 Hz) speech codecs can give the basic requirement of intelligibility of speech signals [1], the naturalness and quality of speech can be further enhanced by rendering super-wideband (SWB) speech signals that has bandwidth up to 14000 Hz. Moreover, the wideband speech codec has the limitation for representing other signals such as music and mixed (speech and music) signals. Conventionally speech codec and audio codec are individually developed by ITU-T and MPEG, respectively. However, recently speech-based audio codecs are being developed, which merge speech and audio codecs into single codec. SWB extension is an efficient technique for the low bit-rate SWB codecs by using a small amount of side information [2-5]. This paper deals with the SWB extension techniques that extend the frequency band from wideband to SWB.

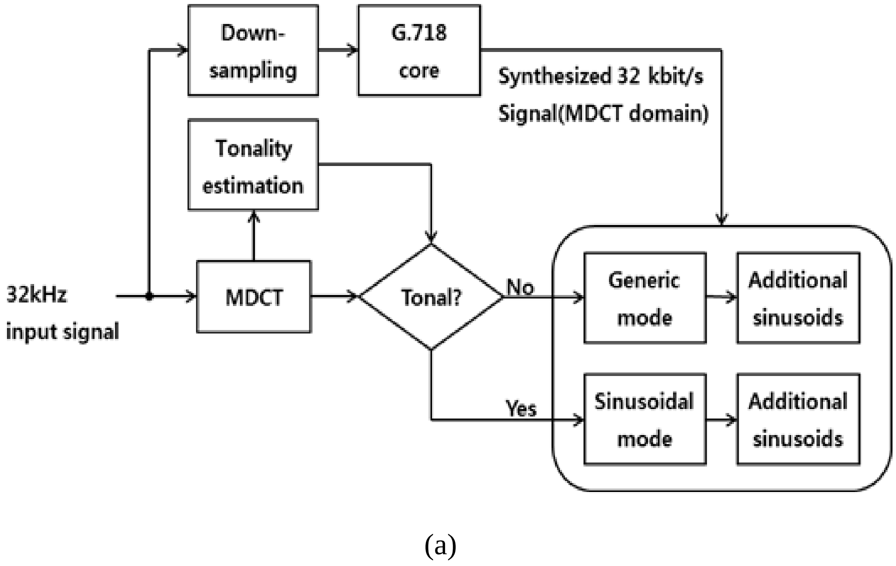
The proposed SWB extension techniques utilize sinusoidal coding [6] and spectral band replication (SBR) [7]. The sinusoidal coding technique utilizes the local rms energy of modified discrete cosine transform (MDCT) coefficients in selecting peak MDCTs and the SBR is based on the spectral correlation between low-band and high-band, namely wideband (50-7000 Hz) and higher band (7-14 kHz). The proposed techniques have been implemented based on ITU-T G.718 Annex B [4] that extends the frequency band from the wideband to SWB and has bit-rate of 4 kbit/s (80 bits in 20 msec frame). The SWB extension algorithms developed in this work can be utilized for the development of Enhanced Voice Service (EVS) codec [8] and be applied to high-quality service terminal of mobile Voice over IP (VoIP) and convergence multimedia [9-11].

ISSN: 1975-0080 IJMUE

The remainder of this paper is organized as follows. First, we briefly describe the SWB extension method of ITU-T G.718 Annex B in Section 2. The proposed low bit-rate sinusoidal coding techniques and the correlation-based spectral band replication methods are described in Sections 3 and 4, respectively. Performance evaluation is provided in Section 5. Finally, conclusions are summarized in Section 6.

2. G.718 Super-Wideband Extension

In this section, the super-wideband extension algorithm of ITU-T G.718 codec is briefly introduced [4].



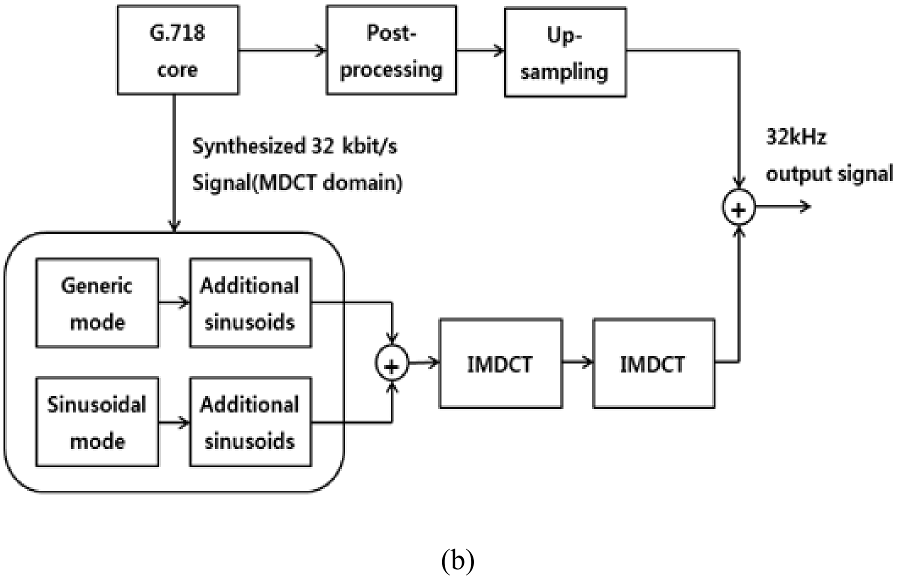


Figure 1. Structural block diagram of the decoder in G.718 SWB [4]

The structure of the G.718 SWB encoder is presented in Figure 1(a). The SWB extension uses a 32 kHz signal, while the ITU-T G.718 core codec operates on a 16 kHz signal. The SWB encoding is performed in the MDCT domain. The generic mode and the sinusoidal mode are set by decision based on the estimated tonality of the input signal. Higher SWB layers are coded using additional sinusoids, which improve the quality of the high frequency content. The structure of the ITU-T G.718 SWB extension decoder is presented in Figure 1(b). The ITU-T G.718 core codec decodes a 16 kHz signal, and the SWB extension decodes high frequency to provide 32 kHz output. Higher SWB layers comprise of additional sinusoidal mode coding, which improves the quality of the high frequency content.

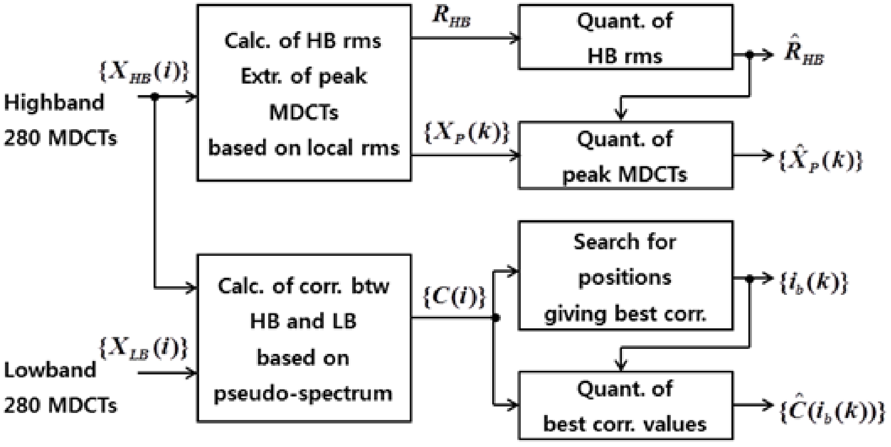
3. Low bit-rate sinusoidal coding

The overall block diagrams of the proposed encoder and decoder for SWB coding are illustrated in Figure 2. The high-band is divided into 7 subbands and each subband has one MDCT. For more efficient bit stream, we determined each length of subband as 32, 32, 32, 32, 32, 56, and 64. G.718 SWB can occasionally have the case that the peak MDCT does not exist in a subband as shown in Figure 3(a). However, in the proposed method, the distribution of the peak MDCTs is balanced as shown in Figure 3(b).

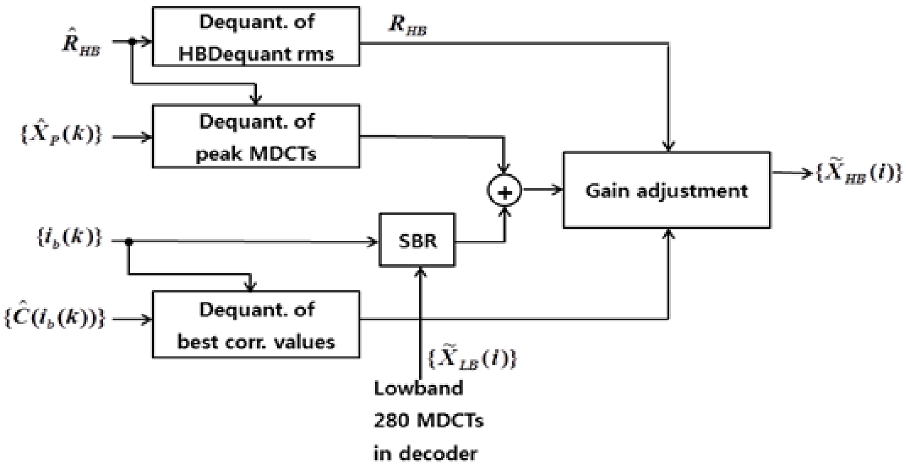
First, the *rms* of high-band MDCTs can be represented as

$$R_{HB} = \sqrt{\frac{\sum_{n=0}^{279} (X_{HB}(i))^2}{280}} \ , \tag{1}$$

where $X_{HB}(i)$ indicates high-band MDCT. The R_{HB} is quantized as $\hat{R}_{HB}$. In decoder, $\hat{R}_{HB}$ is used to adjust the gain in SBR. The gain is represented as $G(\hat{R}_{HB}) = \alpha \times \hat{R}_{HB} + \beta$, where α and β can be obtained experimentally.

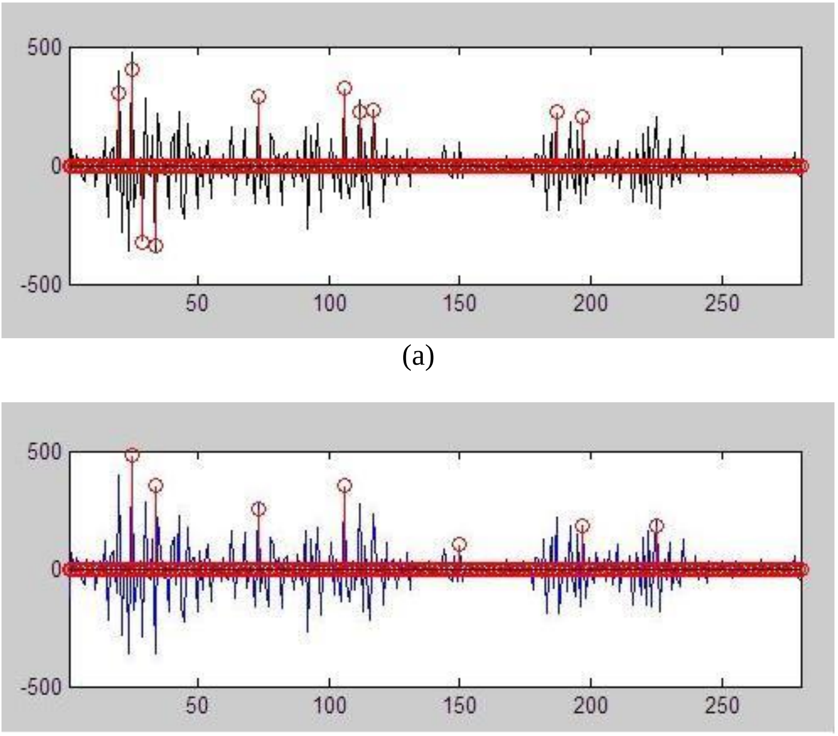


(a)


$\hat{R}_{HB}$
Dequant. of HBDequant rms
R_{HB}
$\{\hat{X}_P(k)\}$
Dequant. of peak MDCTs
+
Gain adjustment
$\{\widetilde{X}_{HB}(i)\}$
$\{i_b(k)\}$
SBR
$\{\hat{C}(i_b(k))\}$
Dequant. of best corr. values
$\{\widetilde{X}_{LB}(i)\}$
Lowband 280 MDCTs in decoder

(b)

Figure 2. Proposed structural block diagram of (a) the SWB encoder and (b) the SWB decoder



(a)

(b)

Figure 3. Examples of peak MDCT selection: (a) G.718, (b) the proposed method

Next, in each subband, the peak MDCT is selected at either the lower half or the higher half of the subband based on the local *rms* values. Figure 4 shows the example that the peak

MDCT is selected in the higher half with higher *rms* value, while G.718 SWB selects the peak MDCT in the left half. According to the position of peak MDCT, the subband gains are adjusted in the decoder. The peak MDCT $X_P(k)$ is transformed into the ratio of peak MDCT to $\hat{R}_{HB}$ as $\tilde{X}_p(k) = X_p(k)/\hat{R}_{HB}$, and quantized as $\hat{X}_P(k)$.

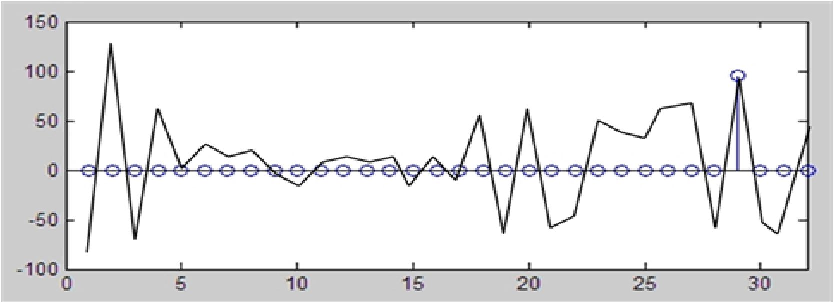


Figure 4. Example of peak MDCT selection based on local *rms*

4. Correlation-based Spectral Band Replication

In G.718 SWB, if current frame is similar to previous frame, generic mode is set. In this case, low-band MDCTs are replicated to high-band without any consideration whether MDCTs between low-band and high-band are similar or not. However, the proposed method enhances the SBR by using the cross-correlation of spectral parameters between low-band and high-band.

First, we utilize pseudo-spectrum [12, 13] in order to approximate the DFT magnitude spectrum. When $X(k)$ is the MDCT, the pseudo-spectrum is $S(k) = \sqrt{X^2(k) + (X(k-1) - X(k+1))^2}$. The normalized cross-correlation $C(i)$ of pseudo-spectra between low-band and high-band is obtained by

$$C(i) = \frac{\sum_j (S_{HB}(j) S_{LB}(i+j))}{\sqrt{\sum_j (S_{HB}(j))^2} \sqrt{\sum_i (S_{LB}(i+j))^2}} \quad , \tag{2}$$

where $S_{HB}(\cdot)$ and $S_{LB}(\cdot)$ are the pseudo-spectra of low-band and high-band, respectively.

As illustrated in Figure 5, the starting positions for SBR are not fixed but varied by searching the best starting positions that give the best cross-correlation values between low-band and high-band. These searches are conducted in the encoder around initial offsets that have been obtained statistically in advance. In the decoder, by using the information of the transmitted starting positions, SBR is conducted except the transmitted peak MDCTs. Furthermore, the cross-correlation values at the best starting positions are quantized and transmitted. In the decoder, gain adjustment is conducted based on the transmitted cross-correlation values.

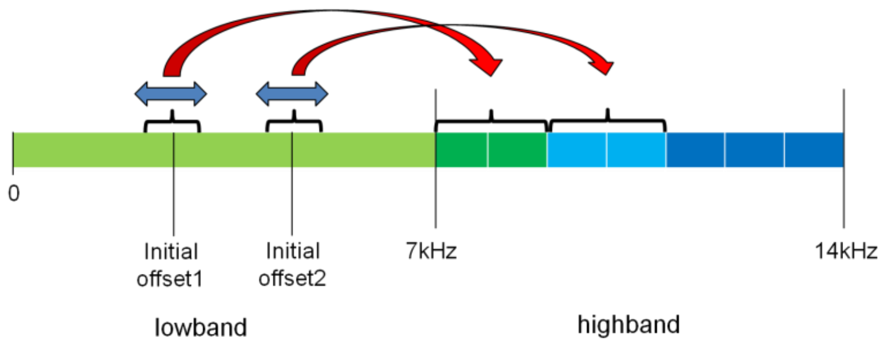
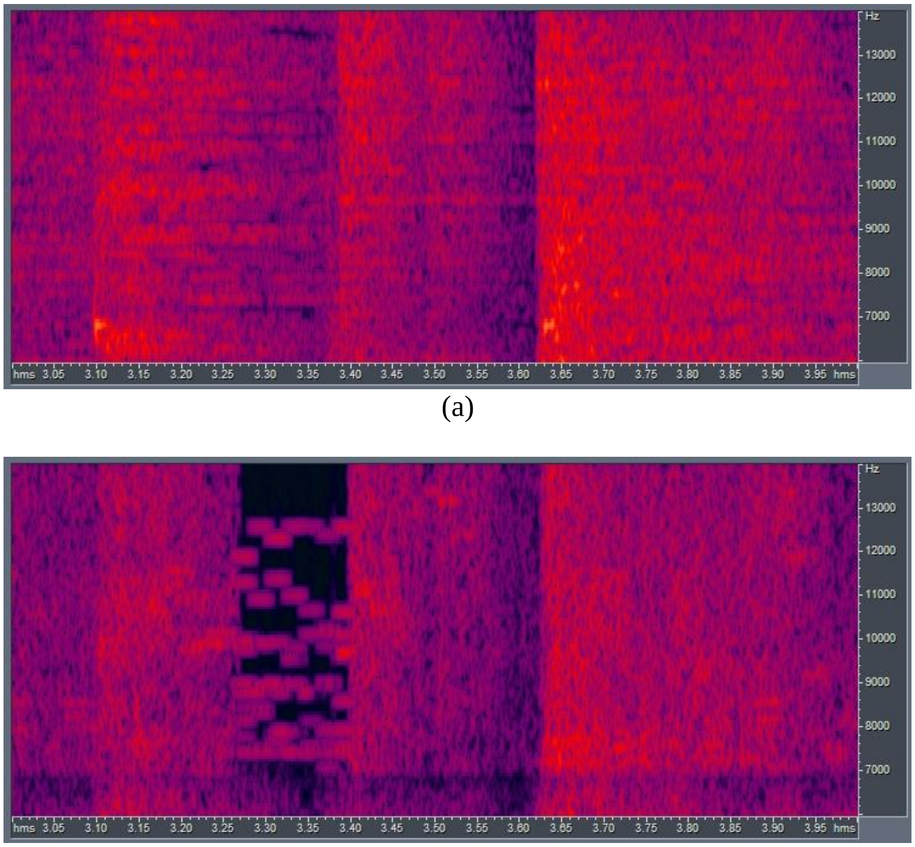


Figure 5. Illustration of proposed SBR

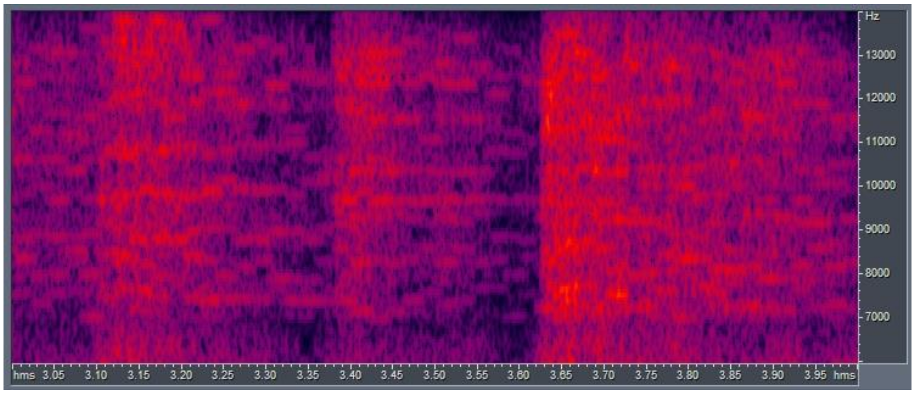
5. Performance Evaluation

As mentioned previously, in G.718 SWB, only one mode is selected at each frame, namely sinusoidal or generic. However the proposed method utilizes both sinusoidal coding and SBR techniques. In Figure 6, we show the examples for the difference of spectrograms by G.718 SWB and the proposed method. As shown in the figure, the spectral similarity is enhanced.



(a)

(b)



(c)

Figure 6. Spectogram Examples: (a) Original, (b) G.718 SWB, and (c) the proposed method

Table 1 shows the comparison of bit allocations for G.718 SWB and the proposed SWB. In the proposed method, the bits for sinusoidal signs are not allocated since the change of sign results in just phase-shifting of sine wave signal and performance differences are slight in the preliminary experiments.

We evaluated the performance of the proposed method by objective and subjective evaluations. First, an objective test was conducted; that is PEAQ [14, 15]. The sound sources of 5-10 seconds are two kinds of male and female voices, three kinds of music, and mixed (music and voice) sounds. As shown in the Table 2, our proposed method has better performance than G.718 for all kind of signal types.

Subjective test was also conducted; that is MUSHRA (Multiple Stimuli with Hidden Reference and Anchor) [16]. 10 subjects with no auditory diseases participated in the test with same sound sources. As shown in the Table 3, the proposed method has better performance than G.718 SWB.

Table 1. Comparison of Bit Allocations

	G.718		Proposed
	Generic	Sinusoidal	
SWB/Stereo	1	1	1
Generic/Sinusoidal	1	1	0
Sinusoidal positions	10	51	37
Sinusoidal signs	1	6	0
Sinusoidal magnitudes	8	21	28
Subband lags	30	0	8
Subband gain magnitudes	24	0	5
Subband gain sign	4	0	0
Reserved	1	0	1
total	80	80	80

Table 2. PEAQ results

	G.718	Proposed
Speech	-3.767	-3.743
Music	-3.754	-3.700
Mixed	-3.722	-3.716
Average	-3.751	-3.721

Table 3. MUSHRA test results

	G.718	Proposed
Speech	84.7	84.9
Music	76.8	77.9
Mixed	74.8	80.0
Average	79.4	80.9

6. Conclusion

In this paper, we presented novel low bit-rate super-wideband extension techniques for both speech and audio coding, which are based on low bit-rate sinusoidal coding and correlation-based spectral band replication. We showed that the proposed techniques can efficiently encode the super-wideband by utilizing the local *rms* energy of MDCTs and the pseudo-spectral correlation between low-band and high-band MDCTs. From the objective and subjective evaluation, namely PEAQ and MUSHIRA, the proposed techniques were shown experimentally to give the improved performance compared to ITU-T G.718 Annex B.

Acknowledgements

This work was supported in part by Electronics and Telecommunications Research Institute (ETRI) and in part by Seoul National University of Science and Technology.

References

[1] J. Schnitzler and P. Vary, "Trends and perspectives in wideband speech coding", Signal Processing, vol. 80, no. 11, **(2000)**, pp. 2267-2281.

[2] ITU-T Recommendation, G.722.1 Annex C, Low-complexity coding at 24 and 32 kbit/s for hands-free operation in systems with low frame loss, **(2005)**.

[3] ITU-T Recommendation, G.729.1 Annex E, Superwideband Scalable Extension for G.729.1, **(2010)**.

[4] ITU-T Recommendation, G.718 Annex B, Superwideband Scalable Extension for G.718, **(2010)**.

[5] Y. R. Oh, Y. G. Kim, M. A. Kim, H. K. Kim, M. S. Lee and H. J. Bae, "Phonetically balanced text corpus design using a similarity measure for a stereo super-wideband speech database", IEICE Transactions on Information and Systems, vol. E94-D, no. 7, **(2011)** July, pp. 1459-1466.

[6] H. Purnhagen, N. Meine and B. Edler, "Sinusoidal coding using loudness-based component selection", Proc. ICASSP, Orlando, Florida, **(2002)** May, pp. 1817-1820.

[7] M. Dietz, L. Liljeryd and K. Kjorling, "Spectral band replication, a novel approach in audio coding", Preprint 5553, 112th AES Convention, Munich, Germany, **(2002)** May.

[8] TDOC s4-100712, EVS Permanent Document (EVS-2): EVS Project Plan, **(2010)**.

[9] J. A. Kang and H. K. Kim, "Adaptive redundant speech transmission over wireless multimedia sensor networks based on estimation of perceived speech quality", Sensors, vol. 11, no. 9, **(2011)** September, pp. 8469-8484.

[10] J. A. Kang and H. K. Kim, "An adaptive packet loss recovery method based on real-time speech quality assessment and redundant speech transmission", International Journal of Innovative Computing, Information and Control, vol. 7, no. 12, **(2011)** December, pp. 6773-6783.

[11] N. I. Park and H. K. Kim, "Artificial bandwidth extension of narrowband speech applied to CELP-type speech coding", Information: an International Interdisciplinary Journal, vol. 16, no. 3(B), **(2013)** March, pp. 3153-3164.

[12] L. Daudet and M. Sandler, "MDCT analysis of sinusoids: exact results and applications to coding artifacts reduction", IEEE Trans. Speech and Audio Processing, vol. 12, no. 3, **(2004)**, pp. 302-312.

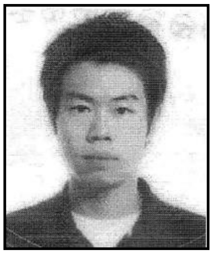
[13] B. Geiser, H. Kruger and P. Vary, "Super-wideband bandwidth extension for wideband audio codecs using switched spectral replication and pitch synthesis", Pro. DAGA, **(2010)**, pp. 663-664.

[14] Method for objective measurements of perceived audio quality, ITU-R BS.1387 **(2001)**.

[15] F. Baumgarte and A. Lerch, "Implementation of Recommendation", ITU-R BS.1387, Delayed Contribution Document 6Q/18-E, **(2001)** February.

[16] Method for the subjective assessment of intermediate quality level of coding systems, ITU-R BS.1534, **(2001)**.

Authors



Kosangrok Oh

Kosangrok Oh is an undergraduate student in Electronic and IT Media Engineering at Seoul National University of Science and Technology. His current research interests include speech and audio coding, speech recognition.

Dong Hoon Sung

Dong Hoon Sung is an undergraduate student in Electronic and IT Media Engineering at Seoul National University of Science and Technology. His current research interests include speech and audio coding, speech signal processing.

Seung Ho Choi

Seung Ho Choi received a B.S. degree in Electronic Engineering from Hanyang University, Korea in 1991. He then received both M.S. and Ph.D. degrees in Electrical Engineering from Korea Advanced Institute of Science and Technology (KAIST), Korea in 1993 and 1999, respectively. He was a senior researcher at Samsung Advanced Institute of Technology, Korea, from 1996 to 2002. He was a visiting professor at University of Florida, USA, from 2008 to 2009. Since August 2002, he has been with the Department of Electronic and IT Media Engineering at Seoul National University of Science and Technology as a professor. His current research interests include speech recognition, speech and audio coding, human-computer interaction. E-mail: shchoi@seoultech.ac.kr.