

Research Article

Superwideband Bandwidth Extension Using Normalized MDCT Coefficients for Scalable Speech and Audio Coding

Young Han Lee¹ and Seung Ho Choi²

¹ *Speech Team, Future IT Convergence Laboratory, LG Advanced Research Institute, Seoul 137-724, Republic of Korea*

² *Department of Electronic and IT Media Engineering, Seoul National University of Science and Technology, Seoul 139-743, Republic of Korea*

Correspondence should be addressed to Seung Ho Choi; shchoi@seoultech.ac.kr

Received 30 April 2013; Accepted 12 June 2013

Academic Editor: Jinan Fiaidhi

Copyright © 2013 Y. H. Lee and S. H. Choi. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

A bandwidth extension (BWE) algorithm from wideband to superwideband (SWB) is proposed for a scalable speech/audio codec that uses modified discrete cosine transform (MDCT) coefficients as spectral parameters. The superwideband is first split into several subbands that are represented as gain parameters and normalized MDCT coefficients in the proposed BWE algorithm. We then estimate normalized MDCT coefficients of the wideband to be fetched for the superwideband and quantize the fetch indices. After that, we quantize gain parameters by using relative ratios between adjacent subbands. The proposed BWE algorithm is embedded into a standard superwideband codec, the SWB extension of G.729.1 Annex E, and its bitrate and quality are compared with those of the BWE algorithm already employed in the standard superwideband codec. It is shown from the comparison that the proposed BWE algorithm relatively reduces the bitrate by around 19% with better quality, compared to the BWE algorithm in the SWB extension of G.729.1 Annex E.

1. Introduction

In early speech communication services, narrowband codecs having a bandwidth of around 3.4 kHz were commonly used since the available network bandwidth was quite limited. These services could provide sufficient quality for comprehension, but it was generally agreed that they did not satisfy users' increasing expectations for higher sound quality. Due to the advances in network technologies, however, this transmission bandwidth has recently been increased [1–3]. Thus, a great deal of research has been focused on further extending the bandwidth of speech and/or audio signals from narrowband to wideband, superwideband, and audio band [4–6].

There are two different kinds of approaches for extending the bandwidth according to whether or not the side information is available, as shown in Figure 1. As depicted in Figure 1(a), it is usual to realize bandwidth extension by using the side information that is transmitted from the encoder. On the other hand, it is also possible to extend bandwidth

only at the decoder without any side information [7], which is shown in Figure 1(b). In other words, instead of using the side information, artificial bandwidth extension can estimate the higher band signal from the lower band signal by using a pattern recognition algorithm such as hidden Markov models (HMMs) [8] Gaussian mixture models (GMMs) [9] and [10–15]. While artificial bandwidth extension algorithms do not require any additional bits for sending the side information, their performance is somewhat restricted depending on the performance of the pattern recognition algorithm used in the extension, compared with the bandwidth extension using the side information [16–20]. Thus, an issue that arose in this bandwidth extension approach is how to improve the performance of the bandwidth extension algorithm with lower increase of the bitrate for the side information.

In this paper, we propose a superwideband bandwidth extension (BWE) algorithm using normalized spectral coefficients for scalable speech and audio coding, where speech and audio signals are transformed into the spectral domain via a modified discrete cosine transform (MDCT). To this end,

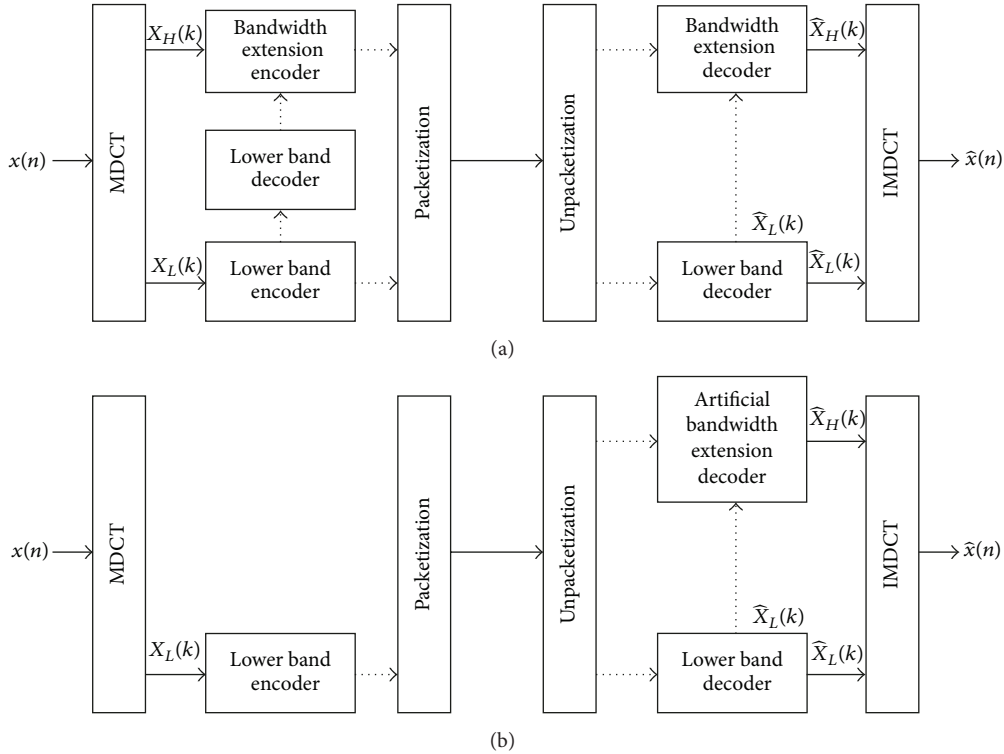


FIGURE 1: Classification of bandwidth extension algorithms applied in the MDCT domain: (a) bandwidth extension with the side information and (b) without any side information (artificial bandwidth extension).

the superwideband is split into several subbands, and MDCT coefficients belonging to each subband are normalized by the subband gain that is obtained by squared sum of each MDCT coefficient. We then estimate normalized MDCT coefficients appropriate for the superwideband in order to fetch them from the wideband. Moreover, to improve the coding efficiency, we incorporate a quantization scheme that deals with the relative ratio of gains between adjacent subbands, resulting in a reduced bitrate for the proposed BWE algorithm.

The remainder of this paper is organized as follows. A brief review on superwideband extension algorithm is given in Section 2. The proposed superwideband extension algorithm is described in Section 3. Performance evaluation is illustrated in Section 4. Finally, conclusions are given in Section 5.

2. Superwideband Extension Algorithm

In this section, we briefly review an existing superwideband (SWB) extension algorithm employed in G.729.1 Annex E [17]. The conventional SWB extension algorithm operates in the MDCT domain performed with a length of 640 samples, and it is comprised of a generic mode and a sinusoidal mode. In the generic mode, the higher band signals are reconstructed by transposing the lower band signals with a gain adjustment. However, since tonal signals, which are generated by an instrument such as a guitar, are characterized as the magnitude and the position of several tones, it is not

proper to model them by using the generic mode. Thus, the sinusoidal mode tries to model directly tonal components. In fact, the mode selection is done by estimating the tonality of the input audio signals. Therefore, we are interested in how to improve the BWE algorithm realized as the generic mode in G.729.1 Annex E with the same mode selection and sinusoidal mode.

Figure 2 shows the structure of the conventional BWE algorithm employed as the generic mode in G.729.1 Annex E. As shown in the figure, the conventional BWE algorithm is composed of two blocks: one is a fetch index search and the other is gain compensation and scaling factor quantization block. As a first step, in order to obtain appropriate MDCT coefficients of the higher band, the maximum cross-correlation lag, d_j^* , for the j th subband is determined as

$$d_j^* = \arg \max_{B_j^{\text{lo}} \leq d_j \leq B_j^{\text{hi}}} C(d_j), \quad 0 \leq j \leq M-1, \quad (1)$$

where

$$C(d_j) = \frac{\left| \sum_{k=0}^{N_j-1} X_{H,j}(k) \widehat{X}_L(d_j+k) \right|}{\sqrt{\sum_{k=0}^{N_j-1} \widehat{X}_L^2(d_j+k)}} \quad (2)$$

and M is the total number of subbands. In addition, N_j is the number of MDCT coefficients belonging to the j th subband. In (2), $X_{H,j}(k)$ is the k th higher band MDCT coefficient of the j th subband, and $\widehat{X}_L(k)$ is the k th lower band

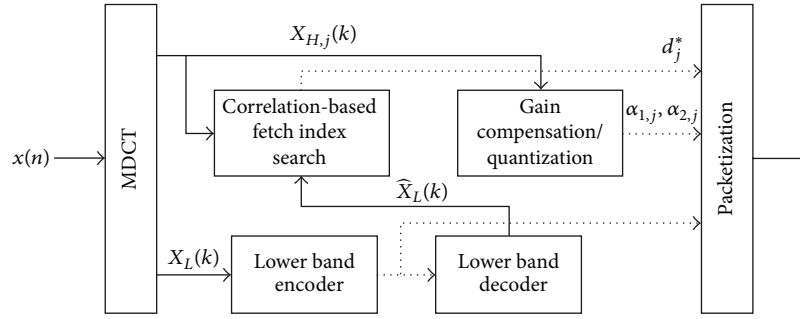


FIGURE 2: Structure of a conventional superwideband extension algorithm employed as the generic mode in G.729.1 Annex E [17].

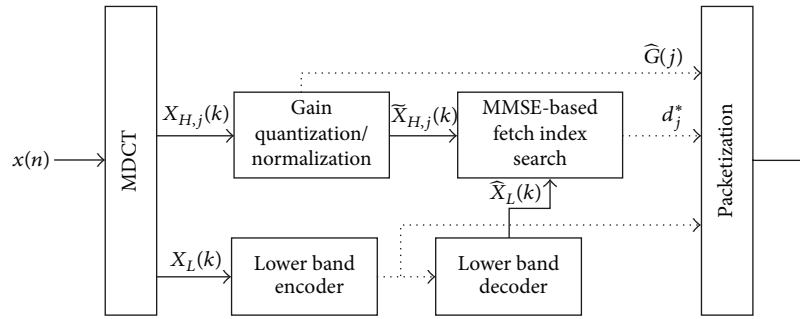


FIGURE 3: Structure of the proposed SWB extension algorithm.

MDCT coefficient reconstructed by the lower band decoder. In addition, B_j^{lo} and B_j^{hi} are the low and high boundaries for the fetch index search in the j th subband, respectively. The parameters are set as $M = 4$, $N_0 = 40$, $N_1 = N_2 = 70$, and $N_3 = 100$ [17].

Next, in order to compensate for the mismatch between $X_{H,j}(k)$ and $\hat{X}_L(d_j^* + k)$ for the j th subband, we compute two different scaling factors in the linear spectral domain as well as in the log spectral domain, $\alpha_{1,j}$ and $\alpha_{2,j}$, which are defined as

$$\alpha_{1,j} = \frac{\sum_{k=0}^{N_j-1} X_{H,j}(k) \hat{X}_L(d_j^* + k)}{\sum_{k=0}^{N_j-1} \hat{X}_L^2(d_j^* + k)}, \quad (3)$$

$$\alpha_{2,j} = \frac{\sum_{k=0}^{N_j-1} (M_j(k) - M_j) D_j(k)}{\sum_{k=0}^{N_j-1} (M_j(k) - M_j)^2},$$

where $M_j(k) = \log_{10}|\alpha_{1,j} \hat{X}_L(d_j^* + k)|$, $M_j = \arg \max_k M_j(k)$, and $D_j(k) = \log_{10}|X_{H,j}(k) - M_j|$.

3. Proposed SWB Extension Algorithm

The proposed BWE algorithm has three major differences from the SWB extension algorithm described in Section 2. First, instead of quantizing the scale factor for each subband, the proposed algorithm quantizes the gain of each subband. Also, the replication of the higher band is done by using the MDCT coefficients normalized by the subband gains. Second, in order to improve the coding efficiency, the numbers of

subbands are differently assigned for subband gains and fetch indices, where subband gains and fetch indices are related to the spectral shape and the spectral fine structure, respectively. However, since the threshold in quiet for the higher band is much greater than that for the lower band [21], a detailed representation of the spectral shape is more important than that of the spectral fine structure for the higher band. Thus, it would better to assign more bits to the subband gain quantization than the fetch index quantization. Therefore, we decide to provide more subbands for gain, which is referred to as *gain subband*, but less for the fetch index, which is referred to as *fetch subband*. Third, instead of using a maximum correlation criterion, we apply a minimum mean square error (MMSE) criterion for the fetch index search, which can provide better performance to minimize the spectral errors.

Figure 3 shows the structure of the proposed BWE algorithm. Compared to Figure 2, the proposed BWE algorithm consists of a gain normalization/quantization block and a fetch index search block. In the proposed BWE algorithm, we first estimate the subband gain of the j th gain subband, $G(j)$, using the equation of

$$G(j) = \frac{1}{N_{g,j}} \sqrt{\sum_{k=0}^{N_{g,j}-1} X_{H,j}^2(k)}, \quad 0 \leq j \leq M_g - 1, \quad (4)$$

where M_g is the total number of gain subbands and $N_{g,j}$ is the number of MDCT coefficients in the j th gain subband. In addition, $X_{H,j}(k)$ is the k th higher band MDCT coefficient in the j th gain subband.

In this proposed BWE algorithm, we incorporate the concept of differential quantization to reduce the number

TABLE 1: Comparison of bit allocation between the conventional BWE algorithm and the proposed one with a typical set of the design parameters, where the algorithms are realized in the SWB extension of G.729.1 Annex E.

Parameter	Conventional BWE	Proposed BWE	
		Bit allocation	Typical: $M_g = 14, M_f = 2$
Initial gain		m	5
Subband gains	28	$n \cdot (M_g - 1)$	26
Fetch indices	32	$7M_f$	14
Single sinusoidal/mode indicator/envelop shaping	19	19	19
Total	79		64

of bits for quantizing the subband gains. In other words, we quantize the relative ratios of the subband gains, which corresponds to differential quantization applied in the log spectral domain. This is because the dynamic range of the relative ratio is smaller than that of the subband gain. That is, each subband gain can be quantized as

$$\widehat{G}(j) = \begin{cases} Q_m(G(j)), & j = 0, \\ \widehat{G}(j-1) \cdot Q_n\left(\frac{G(j)}{\widehat{G}(j-1)}\right), & 1 \leq j \leq M_g - 1, \end{cases} \quad (5)$$

where $Q_n(x)$ is an n -bit scalar quantizer for x . Specifically, the gain of the first gain subband should be quantized with a higher bit quantizer than those of the rest of subband gains; that is, $m > n$. Next, the MDCT coefficients are normalized by using the corresponding quantized gain, such that

$$\widetilde{X}_{H,j}(k) = \frac{X_{H,j}(k)}{\widehat{G}(j)}, \quad (6)$$

where $\widetilde{X}_{H,j}(k)$ is the k th higher band normalized MDCT coefficient belonging to the j th gain subband.

The fetch indices in the proposed BWE algorithm are then obtained based on the MMSE criterion. In other words, a fetch index, d_l^* , for the l th fetch subband is found by using the equation of

$$d_l^* = \arg \min_{B_l^{\text{lo}} \leq d_l \leq B_l^{\text{hi}}} E(d_l), \quad 0 \leq l \leq M_f - 1, \quad (7)$$

where $E(d_l) = \sum_{k=0}^{N_{f,l}-1} (\widetilde{X}_H(k) - \widetilde{X}_L(d_l + k))^2$ and $\widetilde{X}_L(k) = \widehat{X}_L(k) / \sqrt{\sum_{k=0}^{N_{f,l}-1} \widehat{X}_L^2(k)}$. In addition, M_f is the total number of fetch subbands, $N_{f,l}$ is the number of MDCT coefficients belonging to the l th fetch subband, and B_l^{lo} and B_l^{hi} are the low and high boundaries of the l th fetch subband, respectively.

In the proposed BWE algorithm, several combinations are available by varying the settings for M_g , M_f , m , and n , whereas parameters such as the sinusoidal coding and envelope shaping parameters are used for the same operation as in [17].

4. Performance Evaluation

In this section, we compared the performance of the proposed BWE algorithm with that of the conventional BWE algorithm

employed in G.729.1 Annex E [17] in terms of bitrate, spectrogram, and quality. Note here that we only replaced the generic mode of the BWE algorithm in the codec.

4.1. Bitrate. The number of bits/frame for the conventional BWE algorithm was 79 bits/frame, and 60 bits/frame out of them were for d_j , $\alpha_{1,j}$, and $\alpha_{2,j}$ (for $j = 0, \dots, 3$) in (1) and (3). Even if the BWE algorithm operated as the generic mode, a single sinusoidal component was always modeled, and its location, sign, and amplitude were quantized with 10 bits/frame. The remaining 9 bits/frame out of the 79 bits/frame were assigned for the mode indicator and the envelope shaping that was applied to reduce the pre- and postecho artifacts. For a more detailed explanation on the bit assignment for the conventional BWE algorithm, refer to the literature [17].

For the proposed BWE algorithm, there were a number of combinations for the bit assignment. To this end, we had performed exhaustive experiments by changing the design parameters of the proposed BWE algorithm. Eventually, it was found that by setting $M_g = 14$, $M_f = 2$, $m = 5$, and $n = 2$, the proposed BWE algorithm could provide better quality with a lower bitrate than the conventional BWE algorithm. It was also shown from the Table 1 that the proposed BWE algorithm achieved a relative bitrate reduction of around 19%, compared to the conventional BWE algorithm.

4.2. Spectrogram Comparison. Figures 4(b) and 4(c) show the spectrograms for the original audio signal (Figure 4(a)) when the conventional and proposed BWE algorithms were applied, respectively. Note here that the original signal taken from the sound quality assessment material (SQAM) [22] was sampled at a rate of 32 kHz, but the bandwidth of the decoded signal by G.729.1 Annex E was limited up to 14 kHz. In this paper, the spectrograms were obtained by applying a Blackman window whose length was 1,024 samples with an overlap of 512 samples to the decoded signals.

As shown in Figure 4(b), the spectrogram of decoded signal by the conventional BWE algorithm had horizontal lines at round 9.6 kHz and 11.6 kHz, as denoted by a dotted box, which were caused by the gain mismatch. This mismatch, however, was mitigated by using the proposed BWE algorithm; thus, there were no horizontal lines in Figure 4(b). These results implied that the proposed BWE algorithm could provide better quality than the conventional BWE algorithm.

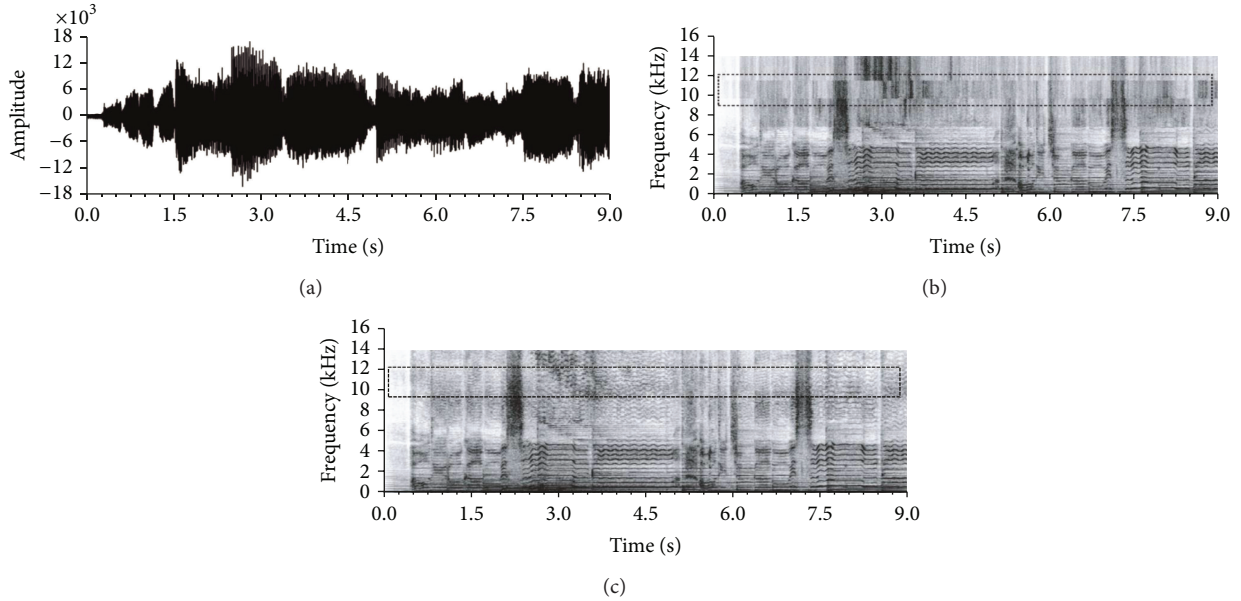


FIGURE 4: Comparison of spectrograms: (a) audio signal, (b) spectrogram of decoded signal by the conventional BWE algorithm, and (c) that by the proposed BWE algorithm.

TABLE 2: Preference test results.

Signal type	Preference score (%)		
	Conventional BWE	No difference	Proposed BWE
Speech	8.89	48.89	42.22
Music	11.11	31.11	57.78
Noisy speech	11.11	48.89	40.00
Average	10.37	42.96	46.67

4.3. Quality Preference Test. As a subjective measure, we conducted an AB preference test between the conventional BWE algorithm and the proposed one. Here, we chose three different signal types for the preference test, including speech, music, and noisy speech signals. Note here that speech signals were taken from the database [6] and others from SQAM [22]. We prepared five files for each signal type, where each file was preprocessed according to the G.729.1/G.718 SWB processing plan [23]. All files were then presented to nine listeners having no auditory disorders. Table 2 shows the preference test results. The proposed BWE showed higher scores for all the signal types than the conventional BWE. In particular, the strong periodicity of music signal resulted in the best preference score. It could be concluded from the table that the listeners preferred the signals decoded by the proposed BWE algorithm for all signals types, rather than those decoded by the conventional BWE algorithm.

5. Conclusion

In this paper, we proposed a superwideband bandwidth extension algorithm using normalized MDCT coefficients for scalable speech and audio coding. In the proposed algorithm, the gain of each subband is quantized, and the replication

of the higher band is done by using the MDCT coefficients normalized by the subband gains. Also, the numbers of subbands are differently assigned according to subband gains and fetch indices. Moreover, we apply a minimum mean square error (MMSE) criterion for the fetch index search, which can provide better performance to minimize the spectral errors. As a result, it was shown from the spectrogram comparison and the AB preference test that the proposed BWE algorithm provided better quality, especially for the music signal than the BWE algorithm currently employed in G.729.1 Annex E, while it achieved a relative bitrate reduction of 18.99%.

Acknowledgment

This research was supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (no. 2013R1A1A2007971).

References

- [1] C. Lamblin, "Recent audio/speech coding developments in ITU-T and future trends," in *Proceedings of the European Signal Processing Conference (EUSIPCO '08)*, Plenary Lecture, Lausanne, Switzerland, 2008.
- [2] 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, pp. 6773–6783, 2011.
- [3] 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, pp. 8469–8484, 2011.

- [4] "ITU-T Temporal Document 298 R1," Report of Q23/16 Rapporteur's Meeting, 2008.
- [5] N. I. Park and H. K. Kim, "Artificial bandwidth extension of narrowband speech applied to CELP-type speech coding, Information," *International Interdisciplinary Journal*, vol. 16, no. 3(B), pp. 3153–3164, 2013.
- [6] Y. R. Oh, Y. G. Kim, M. 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, pp. 1459–1466, 2011.
- [7] P. Vary and R. Martin, *Digital Speech Transmission: Enhancement, Coding and Error Concealment*, Wiley, Chichester, UK, 2006.
- [8] P. Jax and P. Vary, "On artificial bandwidth extension of telephone speech," *Signal Processing*, vol. 83, no. 8, pp. 1707–1719, 2003.
- [9] G.-B. Song and P. Martynovich, "A study of HMM-based bandwidth extension of speech signals," *Signal Processing*, vol. 89, no. 10, pp. 2036–2044, 2009.
- [10] U. Kornagel, "Techniques for artificial bandwidth extension of telephone speech," *Signal Processing*, vol. 86, no. 6, pp. 1296–1306, 2006.
- [11] H. Pulakka, L. Laaksonen, M. Vainio, J. Pohjalainen, and P. Alku, "Evaluation of an artificial speech bandwidth extension method in three languages," *IEEE Transactions on Audio, Speech and Language Processing*, vol. 16, no. 6, pp. 1124–1137, 2008.
- [12] K.-T. Kim, M.-K. Lee, and H.-G. Kang, "Speech bandwidth extension using temporal envelope modeling," *IEEE Signal Processing Letters*, vol. 15, pp. 429–432, 2008.
- [13] J. H. Park, H. K. Kim, M. B. Kim, and S. R. Kim, "A user voice reduction algorithm based on binaural signal separation for portable digital imaging devices," *IEEE Transactions on Consumer Electronics*, vol. 58, no. 2, pp. 679–684, 2012.
- [14] J. A. Kang, C. J. Chun, H. K. Kim, M. B. Kim, and S. R. Kim, "A smart background music mixing algorithm for portable digital imaging devices," *IEEE Transactions on Consumer Electronics*, vol. 57, no. 3, pp. 1258–1263, 2011.
- [15] Y. R. Oh, J. S. Yoon, H. K. Kim, M. B. Kim, and S. R. Kim, "A voice-driven scene-mode recommendation service for portable digital imaging devices," *IEEE Transactions on Consumer Electronics*, vol. 55, no. 4, pp. 1739–1747, 2009.
- [16] J. Herre and M. Dietz, "MPEG-4 high-efficiency AAC coding," *IEEE Signal Processing Magazine*, vol. 25, no. 3, pp. 137–142, 2008.
- [17] M. Tammi, L. Laaksonen, A. Rämö, and H. Toukoma, "Scalable superwideband extension for wideband coding," in *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP '09)*, pp. 161–164, Taiwan, April 2009.
- [18] B. Geiser, P. Jax, P. Vary et al., "Bandwidth extension for hierarchical speech and audio coding in ITU-T Rec. G.729.1," *IEEE Transactions on Audio, Speech and Language Processing*, vol. 15, no. 8, pp. 2496–2509, 2007.
- [19] H. Ehara, T. Morii, and K. Yoshida, "Predictive vector quantization of wideband LSF using narrowband LSF for bandwidth scalable coders," *Speech Communication*, vol. 49, no. 6, pp. 490–500, 2007.
- [20] Y. H. Lee, H. K. Kim, M. S. Lee, and D. Y. Kim, "Bandwidth extension of a narrowband speech coder for music delivery over IP," *Lecture Notes in Artificial Intelligence*, vol. 4413, pp. 198–208, 2007.
- [21] T. Painter and A. Spanias, "Perceptual coding of digital audio," *Proceedings of the IEEE*, vol. 88, no. 4, pp. 451–515, 2000.
- [22] EBU Tech Document 3253, Sound Quality Assessment Material (SQAM), 1988.
- [23] ITU-T WP3/16, Processing Test Plan for the ITU-T Joint (G.718/G.729.1) SWB/Stereo Extension Optimisation/Characterization Phase, 2008.



Hindawi

Submit your manuscripts at
<http://www.hindawi.com>

