THE JOURNAL OF THE ACOUSTICAL SOCIETY OF KOREA Vol.29, No.3E 2010. 9. pp. 131~139

Enhanced Spectral Hole Substitution for Improving Speech Quality in Low Bit-Rate Audio Coding

Chang-Heon Lee* and Hong-Goo Kang*
*Department of Electrical and Electronic Engineering, Yonsei University
(Received June 24, 2010; accepted August 17, 2010)

Abstract

This paper proposes a novel spectral hole substitution technique for low bit-rate audio coding. The spectral holes frequently occurring in relatively weak energy bands due to zero bit quantization result in severe quality degradation, especially for harmonic signals such as speech vowels. The enhanced aacPlus (EAAC) audio codec artificially adjusts the minimum signal-to-mask ratio (SMR) to reduce the number of spectral holes, but it still produces noisy sound. The proposed method selectively predicts the spectral shapes of hole bands using either intra-band correlation, i.e. harmonically related coefficients nearby or inter-band correlation, i.e. previous frames. For the bands that have low prediction gain, only the energy term is quantized and spectral shapes are replaced by pseudo random values in the decoding stage. To minimize perceptual distortion caused by spectral mismatching, the criterion of the just noticeable level difference (JNLD) and spectral similarity between original and predicted shapes are adopted for quantizing the energy term. Simulation results show that the proposed method implemented into the EAAC baseline coder significantly improves speech quality at low bit-rates while keeping equivalent quality for mixed and music contents.

Keywords: *Enhanced aacPlus, spectral hole substitution, prediction, just noticeable level difference (JNLD), audio coding, Speech coding*

I. Introduction

The transform-based audio coding utilizes the psychoacoustic model (PAM) as a reference metric for bit-allocation [1]. Since low bit-rate audio codecs are not capable of fulfilling required bits by the PAM, the PAM model should be adjusted for meeting bit constraints while minimizing perceptual distortions. The 3GPP enhanced aacPlus audio coding standard (EAAC) takes an approach of adding offsets to masking thresholds which corresponds to providing equal loudness quantization noises for all frequency bands [2]. Even with the PAM adjustment, however, no bits are assigned to some frequency bands in low bit-rate encoding option. The bands that no bits are assigned are called spectral hole bands, which is the main cause of severe quality degradation.

To solve the problem the EAAC allocates small number of bits to the selected hole bands to satisfy the minimum signal-to-mask (SMR) requirement. Since the requirement is simply set by equally distributing a certain portion of total perceptual entropy to each bark band, however, the strategy still has the problem of producing noisy sounds at low bit-rates especially for speech signals. The perceptual noise substitution (PNS) technique used in MPEG-4 AAC can be an alternative way to remove spectral holes [3]. It only transmits a flag bit for defining a type of encoding and an energy

Corresponding author: Chang-Heon Lee (leech@dsp.yonsei.ac.kr)
B601 Department of Electrical and Electronic Engineering, Yonsei University134 shinchondong seodaemoon-gu, 120-749, Seoul, Korea

parameter for the noisy frequency band. Then, spectral coefficients of the noisy band are artificially synthesized by random values in the decoding stage. However, the effective region for applying the PNS method is limited to high frequencies, e.g. above 5 kHz because human can easily perceive small frequency differences in low frequency region [4, 5]. It should be noted that spectral holes frequently occur at low frequency regions in recently developed low bit-rate audio codecs. Therefore, the PNS technique cannot be directly applicable to the problem of spectral holes in low frequency band. Recently, the noise filling tool was newly introduced to improve the performance of the PNS in a low bit-rate unified speech and audio coding - MPEG RM0 version [6]. Since this tool mainly addresses the coarse quantization problem occurring in wide scalefactor bands, e.g. related to high frequencies, it still has a limitation to enhance the quality in low frequency bands.

This paper proposes a novel spectral hole substitution algorithm using a predictive scheme and a criterion of the just noticeable level difference (JNLD) thresholds [4] to improve the quality of speech signals in low bit-rate audio coding. Even tonal or harmonically related spectral components are often defined as spectral hole bands if they have relatively weak energy compared to others. It usually happens in spectral null or valley regions of speech vowel sounds. In such cases, the proposed algorithm predicts the spectral shapes of hole bands either from harmonically related coefficients nearby, i.e. intra-prediction, or from the ones in the previous frame, i.e. inter-prediction. To further save bits, the shapes of bands that have low prediction gain are replaced with random values at the decoding stage as the PNS does. Finally, the energy-related parameters for the hole bands are determined by using the criterion of JNLD threshold and the maximum correlation between predicted and original shapes. The JNLD threshold provides a guideline for controlling the energy-related gains according to the correlation values to prevent unexpected perceptual distortion from being newly produced. The spectral band replication (SBR) technique only estimates high frequency components by copying or predicting from low frequency spectral coefficients [7], however the proposed approach extends the predictive technique to a larger frequency range including low frequency regions with an additional inter-frame prediction. Besides, the most important difference is the control mechanism of energy-related parameters. To improve perceptual quality, the energy of spectral hole bands is controlled by prediction performance instead of conserving the original band energy. Experimental results with subjective listening tests confirm the effectiveness of the proposed algorithm.

II. Conventional Methods for Reducing Spectral Holes

Low bit-rate audio codecs often cannot fulfill bit-demands set by a typical psychoacoustic model, which results in serious quality degradation of decoded sounds. Though the psychoacoustic model can be artificially adjusted to satisfy bit-rate constraints, it is still unavoidable to have spectral hole bands. Since the perceptual noise substitution (PNS) used in the MPEG-4 AAC leads to a great saving of bits by coding noisy frequency bands with a few bits, it can reduce spectral holes somehow. The hole avoidance technique was also employed into the EAAC codec to lessen quality degradation caused by spectral holes.

1.1. Perceptual Noise Substitution

The PNS technique utilizes the fact that the actual fine structure of noise-like signal band is of minor importance for subjective perception. If the band is classified as a noise band, the PNS saves bits by transmitting only a flag bit to define the type of encoding and the band energy information instead of actual spectral components that are more bit-demanding. At first, the PNS detects noise-like frequency bands from the input audio signal in the encoding process. Only a noise encoding flag and the

total power of spectral coefficients are used for modeling the noise-like bands. In the decoder, spectral coefficients for noise bands are substituted by pseudo random vectors and its total power is adjusted by the transmitted band energy. However, Schulz claimed that the noise substitution technique is only effective for high frequency regions, i.e. above 5 kHz because the human ear can detect frequency differences of as little as 1 Hz at low frequencies [5]. Actually, the MPEG-4 AAC codec adopts the PNS technique only to high frequency regions, i.e. above 4 kHz for 48 kHz sampling frequency. In the low bit-rate encoding option, i.e. 12 and 16 kbps mono, of recently developed audio codecs such as the high efficiency AAC (HE-AAC) and the EAAC, a typical core encoding is applied only for low frequency coefficients whose highest frequency is around 3.5 kHz, and remaining high frequency components are synthesized by the spectral band replication (SBR) technique to improve coding efficiency [8]. If the PNS needs to be applied to the core coding layer, therefore, its efficiency cannot be guaranteed.

1.2. Avoidance of Spectral Holes

In case available bits for the audio codec are not large enough to satisfy the bit requirement of the typical psychoacoustic model, one option would be adjusting the masking thresholds to reducing the number of allocated bits to each band. The EAAC audio codec continuously increases masking thresholds by allowing equal loudness quantization noises to each frequency band until the thresholds are high enough to be covered by given number of bits [2] such as:

$$T_r(n) = \left(T(n)^{0.25} + r\right)^4, \quad (1)$$

where $T(n)$ means the initial masking threshold of n-th frequency band derived from the psychoacoustic model, and the power of 0.25 is used to approximately represent loudness scale. The additional term, r, denotes the constant loudness value that adjusts the masking threshold by linearly adding to each scalefactor band.

However, the adjustment of masking thresholds described above cannot guarantee high audio quality because there remain many bands where spectral values are still set to zero after the quantization process. To solve the problem, more efficient algorithm has been proposed [2]. In the enhanced method, the bands that must not be set to zero are selected based on the energy distribution. Then, the masking threshold for the selected band is modified as:

$$\widetilde{T}_r(n) = \min\left(\left(T(n)^{0.25} + r\right)^4, \frac{E(n)}{R_m(n)}\right), \quad (2)$$

where $E(n)$ denotes the energy of n-th frequency band, and $R_m(n)$ means the minimum signal-to-mask ratio (SMR) required for n-th band. The minimum SMR values are calculated for each frequency band at the given bit-rate by equally distributing a certain percentage of total perceptual entropy to active bark bands.

Therefore, the avoidance of spectral hole (ASH) technique can be an alternative approach to compensate spectral hole effect in the core bandwidth. While increasing masking thresholds to meet the bit-rate constraint, bands having relatively lower energy than neighbor bands are led to be quantized by the ASH method. However, the output sound is still noisy because of inaccurate quantization at low energy bands. The effect becomes more severe for speech signals because spectral components located at between harmonics or at valley regions between formants are perceptually more sensitive [9]. The proposed algorithm improves the quality by employing a novel encoding scheme that also utilizes perceptual criterion.

III. Spectral Hole Substitution Algorithm

Although the bands contain tonal or harmonically related spectral components, they can be quantized as hole bands because of relatively weak energy compared to other bands, which usually occurs for

signals having spectral tilt or valley regions such as speech vowel. Assuming that spectral peak coefficients are well-encoded by the core coding because they have higher energy than others, spectral coefficients of those hole regions can be predicted from harmonically related coefficients nearby or previous frames. By utilizing the spectral similarity of inter- and intra-frames in harmonically related or tonal frequency bands, we propose a method to substitute spectral holes with the predicted values. Fig. 1 describes a block diagram of the proposed prediction-based spectral hole substitution technique. The procedures of PAM adjustment and spectral hole detection are designed based on the EAAC audio codec.

To predict the effectiveness of the proposed algorithm depending on encoding bit-rates, we first observed how often the algorithm could be applied into an actual coding process. Fig. 2 shows the histogram of hole occurrence for each scalefactor band at 12 and 16 kbps mono encoding modes. In the figure, the highest frequency of the 32nd scalefactor band is about 3.5 kHz. for 48 kHz sampling frequency. From the histogram given in Fig. 2, we can expect that the quality improvement by the proposed approach will increase as the bit-rate decreases if the proposed algorithm has better performance. Once the bands are classified as the ones for substitution, the substitution types are selected based on the criterion of inter- and intra-frame correlation values in unquantized modified discrete cosine transform (MDCT) domain. If the maximum correlation value is larger than the threshold, δ, the spectral shape of the band is predicted from nearby frequencies or previous frame in quantized MDCT domain. Otherwise, corresponding shape vectors are not predicted but replaced with random sequences in the decoder without spending additional bits. In Fig. 1, 'P' and 'NP' routines mean predictive and non-predictive, i.e. PNS-like approaches for generating shapes, respectively.

Finally, gain values for each band are calculated and controlled perceptually based on the maximum correlation value and the just noticeable level difference (JNLD) threshold. If the correlation between predicted and original shapes is high, the gain is determined for the band to have the energy of original signal as close as possible. Otherwise, the gain is attenuated to reduce perceptual distortion caused by boosting of undesirable coefficients, which is based on the psychoacoustic background that the decrement of spectral level by the quantization is less perceptible than the increment [11]. This gain control mechanism also reflects the concept of the adaptive post-filtering process widely used in speech coding standards [12]. The JNLD threshold is introduced for defining a lower bound of gain decrement. For 'NP'-related bands, only the JNLD threshold is reflected to control the gain because the correlation is pretty low in this case.

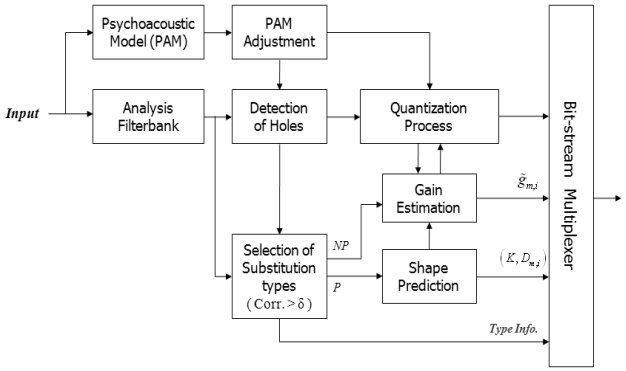


Fig. 1. Block diagram of the proposed substitution technique.

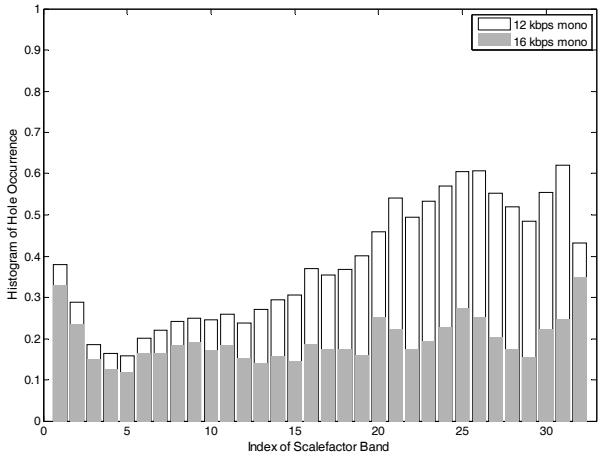


Fig. 2. Histogram of spectral hole occurrence.

3.1. Predictive Spectral Shape Estimation

To estimate the spectral shape of the i-th frequency band at the m-th frame, spectral similarities of both intra-frame and inter-frame basis are considered. For stationary signals, spectral components of tonal or harmonic bands in a current frame may be highly related to those of previous frames, which has been utilized to estimate the tonality and to detect noise bands in audio coding standards [5]. However, for onset segments frequently occurring in speech signals, the intra-frame based prediction becomes more efficient because the correlation with signals of previous frames is generally low. Based on these characteristics, the spectral shape can be estimated by taking a switched prediction method. The spectral shape vector of the i-th frequency band in the m-th frame, $\tilde{\mathbf{X}}_{m,i}$, is determined to have a unit energy from the predicted vector, $\hat{\mathbf{X}}_{m,i}$, such as:

$$\begin{aligned}\tilde{\mathbf{X}}_{m,i} &= \frac{\hat{\mathbf{X}}_{m,i}}{\sqrt{\hat{\mathbf{X}}_{m,i}\hat{\mathbf{X}}_{m,i}^{T}}},\\ \hat{\mathbf{X}}_{m,i} &= \left[X_{q,m-K}(T_i - D_{m,i}),\cdots,X_{q,m-K}(T_i + N_i - 1 - D_{m,i})\right],\end{aligned} \tag{3}$$

where $X_{q,m}(n)$ means quantized MDCT coefficients of the m-th frame, N_i and T_i denote the number of frequency bins and the first bin index of the i-th band, respectively. The type of predictor and the optimal lag of frequency bin, K and $D_{m,i}$, are determined by maximizing the normalized cross-correlation in MDCT domain between the original target vector, $X_m(n)$, and candidate vectors as follows:

$$\begin{aligned}[K, D_{m,i}] &= \arg\max_{[k\in\{0,1\},d_k]}\left(R_{m,i}^{k}(d_k)\right),\\ R_{m,i}^{k}(d_k) &= \frac{\sum_{n=0}^{N_i-1} X_m(n+T_i)X_{q,m-k}(n+T_i-d_k)}{\sqrt{\sum_{n=0}^{N_i-1} X_m^2(n+T_i)\cdot\sum_{n=0}^{N_i-1} X_{q,m-k}^2(n+T_i-d_k)}},\end{aligned} \tag{4}$$

where k = 0 and 1 are related to intra- and inter-frame cross-correlations, respectively. Considering the pitch range of speech signals is about 60 Hz to 400 Hz [10], the range of d_k has been set to cover the pitch frequencies. For intra-prediction cases, d_k has the range of $[N_i, N_i + \triangle - 1]$. At 48 kHz sampling frequency, for example, since a single frequency bin corresponds to around 11.7 Hz in 2:1 downsampled domain actually operated in core coding layer, $\triangle$ needs to be set to satisfy the constraint, $11.7 \cdot \triangle > 400$. In case of applying inter-prediction, the range of d_k is set to $[-\triangle/2, \triangle/2 - 1]$.

Fig. 3 shows a histogram of maximum normalized cross-correlation values determined from eq. (4) for speech signals at 12 kbps mono mode where $\triangle$ was set to 64 (6 bits). This figure verifies that the predictive spectral shape estimation can be very effective especially at low frequencies, which can be explained from the fact that harmonically related spectral coefficients are highly correlated in low frequency regions.

3.2. Perceptual Gain Control

To match the energy of substituted components with that of original spectral coefficients, the gain value can be calculated such as:

$$g_{m,i} = \sqrt{\sum_{n=0}^{N_i-1} X_m^2(n+T_i)}. \tag{5}$$

However, the gain value needs to be properly adjusted if predicted spectral shapes and original

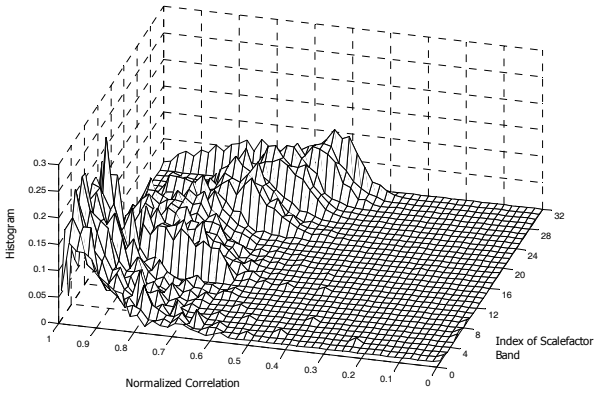


Fig. 3. Histogram of normalized correlation values for speech signals at 12kbps mono.

coefficients are different. If the shape has been predicted well, the gain calculated by eq. (5) would provide less quantization noises. However, the predicted values may not be accurate in general especially for non-tonal or non-harmonic spectral coefficients. To minimize perceptual distortions it is more appropriate to further reduce the gain value because it prevents boosting undesirable coefficients. The rationale behind of the idea can be found from the psychoacoustic background that the decrement of spectral level during the quantization process is less perceptible than the increment of the level [11]. As well, especially for speech signals, since quantization noises located at between harmonics or at valley regions between formants are very sensitive, the decrement of gain can be more effective to reduce perceptual distortions. A lower limit of decreasing gain values needs to be set because a large decrease can also generate unexpected perceptual distortion, which confirms by the theory of the just noticeable level difference (JNLD) concept.

The JNLD introduced by Zwicker is the detection threshold for magnitude differences, which explains that the human ear cannot sensitively perceive differences in spectral magnitude within the JNLD thresholds [4]. The JNLD depends on only a level of excitation pattern, and it can be approximated as [13]:

$$\begin{aligned} J_{m,i} = {} & 5.95072 \cdot \left(\frac{6.39468}{E_{m,i}} \right)^{1.713} + 9.01033 \cdot 10^{-11} \cdot E_{m,i}^4 \\ & + 5.05622 \cdot 10^{-6} \cdot E_{m,i}^3 - 0.00102438 \cdot E_{m,i}^2 \\ & + 0.0550197 \cdot E_{m,i} - 0.198719, \end{aligned} \tag{6}$$

where $E_{m,i}$ denotes the excitation pattern in dB scale of the i-th frequency band at the m-th frame. The excitation pattern can be obtained by smoothing the energy pattern of each frequency band with a spreading function. The JNLD value, $J_{m,i}$, is defined only if $E_{m,i} > 0$, otherwise $J_{m,i}$ is set to 10^{30} as in the ITU-R Rec. BS.1387-1 [13]. The JNLD has the characteristic that large level differences are required for detecting the level change of weak signals whereas the sensitivity to small differences increases for loud signals.

The gain value is controlled based on the psychoacoustic theory explained above, and the similarity between predicted shape vectors and originals is also reflected in the gain control. The mechanism for gain control is proposed as follows:

$$\tilde{g}_{m,i} = \alpha \cdot g_{m,i} + (1-\alpha) \cdot \sqrt{g_{m,i}^2 \cdot 10^{-J_{m,i}/10}}, \tag{7}$$

where $\alpha\,(0 < \alpha \leq 1)$ denotes the normalized correlation value obtained from eq. (4) between the predicted and original shapes, and $\sqrt{g_{m,i}^2 \cdot 10^{-J_{m,i}/10}}$ was determined from the assumption that the band has the JNLD threshold energy of $\left(\sum_{n=0}^{N_i-1} X_m^2(n+T_i)\right) \cdot 10^{-J_{m,i}/10}$. As shown in eq. (7), the gain value is adaptively controlled according to the similarity of predicted spectral shapes.

Fig. 4 shows the modified gain values, $\widetilde{g_{m,i}}$, in dB scale according to α values where α=1.0 means $\widetilde{g_{m,i}} = g_{m,i}$ and α=0.0 corresponds to the lower bound determined by the JNLD threshold energy. As shown in this figure, when the shape is predicted as close as the original one, α value becomes closer to one, thus the adjusted gain will be similar to $g_{m,i}$, i.e. keeping the energy of substituted band with that of original spectral band. Whereas, as the difference between the predicted and original shapes increases,

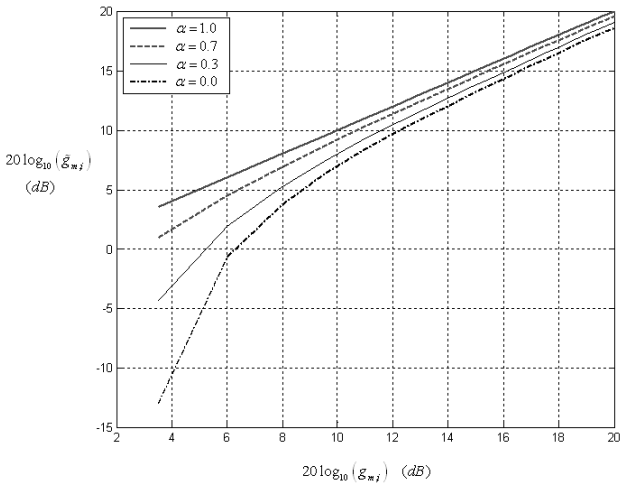


Fig. 4. The modified gain, $\tilde{g}_{m,i}$, according to α values.

the gain decreases toward the lower bound determined by the JNLD threshold. For the bands of which shape vectors are replaced with random sequences because of too low correlation, α in eq. (7) was set to 0.3 empirically.

IV. Experimental Results and Discussions

The δ value for determining the substitution type in Fig. 1 was set to 0.4 based on experimental results. Table 1 summarizes the bit allocation of the proposed method for spectral hole substitution. In case of P-type, a flag bit indicating the inter-frame or the intra-frame prediction is needed, and the prediction lag is encoded with 6 bits to cover the maximum pitch frequency of speech signals. The proposed method was implemented into the Enhanced aacPlus audio codec. It encodes low-frequency spectral coefficients in the core layer, and remaining high-frequency components are encoded by the SBR technique for improving the coding efficiency [8]. Since the spectral hole substitution technique works only in the core layer, signals for performance evaluation were encoded and synthesized without using the SBR technique.

To evaluate the subjective quality of the proposed algorithm, we performed the multiple stimuli with hidden.

reference and anchor (MUSHRA) test [14]. In the MUSHRA test, an anchor was generated by using a low-pass filter with a cut-off frequency of 3.5 kHz, and two coding modes of 12 and 16 kbps mono were utilized for comparing the proposed method to the conventional one. Test materials were taken from the MPEG exploration data set sampled at 48 kHz [15]. Test samples consist of five speech, five mixed and five music contents. Twelve experienced listeners were asked to make a quality judgement by using a headphone (Sennheiser HD600) in a quiet environment. Fig. 5 shows the results of subjective quality measures, the mean values of the MUSHRA test scores and the 95% confidence intervals for the tolerance. As shown in Fig. 5, the proposed algorithm has much higher speech quality than the conventional method especially for 12 kbps mono coding mode. Since the core layer of 12 and 16 kbps mono modes has the bandwidth of about 3.5 kHz [2], the scores of anchor signals are always higher than those of test samples. To more clearly observe the quality difference, as shown in Fig. 6 the differential values of MUSHRA scores were measured with 95% confidence intervals. The differential value was calculated by subtracting the score of the conventional method from that of the proposed algorithm. Fig. 6 shows that the proposed algorithm significantly improves the speech quality for both 12 and 16 kbps coding modes while keeping the quality of mixed and music contents. Also, the

Table 1. Bit-allocation for spectral hole substitution.

	Subs. Type	Pred. Type	Pred. Lag	Gain	Total
P	1	1	6	7	15
NP	1	·	·	7	8

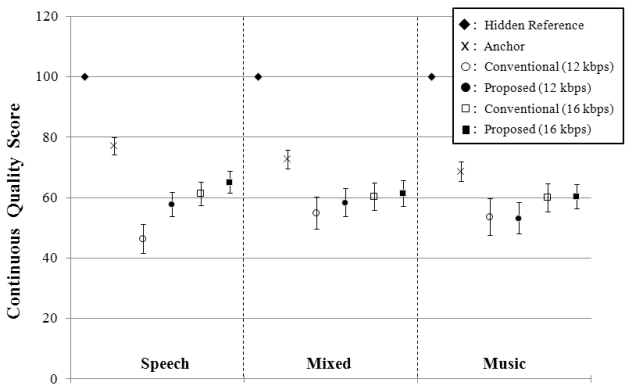


Fig. 5. MUSHRA test results (Absolute scores).

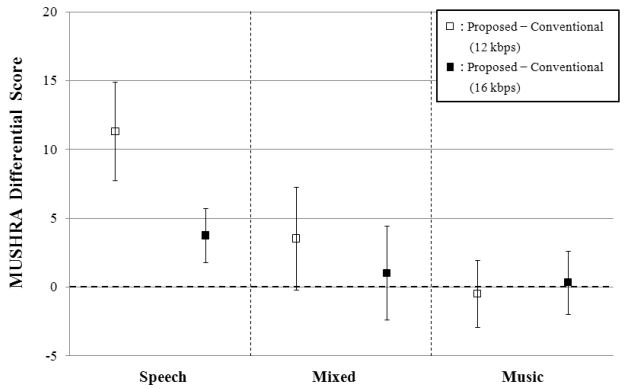


Fig. 6. MUSHRA test results (Differential scores).

Table 2. Comparison of PESQ scores.

Coding mode	12 kbps mono		16 kbps mono	
	Convent.	Proposed	Convent.	Proposed
Female	3.410	3.632	3.724	3.830
Male	3.454	3.661	3.758	3.871
Total	3.432	3.646	3.741	3.850

subjective listening test results verify that the effectiveness of the proposed algorithm for speech signals significantly increases as the bit-rate decreases.

Since the bandwidth of core layer signals is about 3.5 kHz at 12 and 16 kbps mono modes, the perceptual evaluation of speech quality (PESQ) can be utilized to evaluate the objective sound quality for speech signals [16]. Thus, we additionally measured the PESQ scores with 96 speech samples of NTT-AT Korean database. Table II compares PESQ scores of the proposed algorithm to those of the conventional method. As shown in the table, the proposed method improves quality of synthesized speech signals at both coding modes. Also, it confirms the effectiveness of the proposed spectral hole substitution algorithm that the quality improvement for 12 kbps mode is higher than 16 kbps mode.

V. Conclusion

This paper proposed a novel spectral hole substitution algorithm to improve speech quality in low bit-rate audio coding. The spectral shapes of detected hole bands were modeled with predicted ones from harmonically related coefficients nearby or previous frames based on the assumption that tonal or harmonic bands would have high spectral similarity. For the bands that have low prediction gain, random sequences were utilized for artificially synthesizing spectral shapes at a decoding side. The gain values were controlled by using the criterion of the JNLD threshold and the maximum correlation between predicted and original shapes to minimize perceptible distortion. The MUSHRA test results verified that the proposed method had superior speech quality to the conventional method at low bitrates while showing comparable performance for mixed and music signals.

References

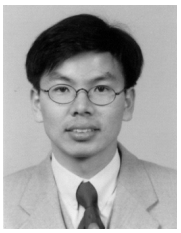
1. J. D. Johnston, "Transform coding of audio signals using perceptual noise criteria," *IEEE J. Select. Areas Commun.*, vol. 6, pp. 314-323, 1988.
2. 3GPP TS 26.403 v7.0.0, *Enhanced aacPlus general audio codec; Encoder specification; Advanced audio coding (AAC) part*, June, 2006.
3. J. Herre and D. Schulz, "Extending the MPEG-4 AAC codec by perceptual noise substitution," *AES 104th Convention*, Amsterdam, May 1998.
4. E. Zwicker and H. Fastl, *Psychoacoustics, Facts and Models, Second Updated Edition*. New York: Springer, 1999.
5. D. Schulz, "Improving audio codecs by noise substitution," *J. Audio Eng. Soc.*, vol. 44, no. 7/8, pp. 593-598, July/August, 1996.
6. M. Neuendorf, P. Gournay, M. Multrus, J. Lecomte, B. Bessette, R. Geiger, S. Bayer, G. Fuchs, J. Hilpert, N. Rettelbach, F. Nagel, J. Robilliard, R. Salami, G. Schuller, R. Lefebvre, and B. Grill, "A novel scheme for low bitrate unified speech and audio coding - MPEG RM0," *AES 126th Convention*, Munich, Germany, May 2009.
7. 3GPP TS 26.404 v6.0.0, *Enhanced aacPlus general audio codec; Encoder specification; Spectral Band Replication (SBR) part*, Sep., 2004.
8. 3GPP TS 26.401 v6.2.0, *Enhanced aacPlus general audio codec; General description*, Mar., 2005.
9. M. R. Schroeder, B. S. Atal, and J. L. Hall, "Optimizing digital speech coders by exploiting masking properties of the human ear," *J. Acoust. Soc. Amer.*, vol. 66, pp. 1647-1652, 1979.
10. Thomas F. Quatieri, *Discrete-Time Speeh Signal Processing, Principles and Practice*. Prentice Hall PTR, 2002.
11. T. Sporer, "Objective audio signal evaluation-applied psychoacoustics for modeling the perceived quality of digital audio," *AES 103rd Convention*, preprint 4280, 1997.
12. J. H. Chen and A. Gersho, "Adaptive postfiltering for quality enhancement of coded speech," *IEEE Trans. Speech Audio Processing*, vol. 3, no. 1, January, 1995.
13. ITU-R Rec. BS.1387-1, *Method for objective measurements of perceived audio quality*, 1999.
14. ITU-R BS.1534-1, *Method for the Subjective Assessment of Intermediate Quality Level of Coding Systems*, 2003.
15. ISO/IEC JTC1/SC29/WG11 MPEG2007/N9095, *Framework for Exploration of Speech and Audio Coding*, San Jose, USA, Apr., 2007.
16. ITU-R Rec. P.862, *Perceptual evaluation of speech quality (PESQ), an objective method for end-to-end speech quality assessment of narrowband telephone networks and speech coders*, Feb. 2001.

【Profile】

• Chang-Heon Lee

Received the B.S., M.S., and Ph.D. degrees in electrical and electronic engineering from Yonsei University, Seoul, Korea, in 2003, 2005, and 2010, respectively. He is currently a post-doctoral researcher at the France Telecom Orange Labs. His research interests include speech signal processing in the field of speech compression and synthesis.

• Hong-Goo Kang



Received the B.S., M.S., and Ph.D. degrees in electronic engineering from Yonsei University, Seoul, Korea, in 1989, 1991, and 1995, respectively. He was a Senior Member of Technical Staff of AT&T, Labs-Research, from 1996 to 2002. In2002, he joined the Department of Electrical and Electronic Engineering, Yonsei University, where he is currently an Associate Professor. His research interests include speech signal processing, array signal processing, and communication signal processing.