A Packet Loss Concealment Technique Improving Quality of Service for Wideband Speech Coding in Wireless Sensor Networks

Nam In Park,1 Jin Ah Kang,2 Seong Ro Lee,3 and Hong Kook Kim1

1 School of Information and Communications, Gwangju Institute of Science and Technology (GIST), 123 Cheomdangwagi-ro, Buk-gu, Gwangju 500-712, Republic of Korea
2 Smart-Work Research Section, Wired & Wireless Convergence Research Department, Electronics and Telecommunications Research Institute (ETRI), 218 Gajeong-ro, Yuseong-gu, Daejeon 305-700, Republic of Korea
3 Department of Information & Electronics, College of Engineering, Mokpo National University, 1666 Youngsan-ro, Cheonggye-myeon, Muan-gun, Jeonnam 534-729, Republic of Korea

Correspondence should be addressed to Hong Kook Kim; hongkook@gist.ac.kr

Received 2 January 2014; Revised 23 March 2014; Accepted 23 March 2014; Published 17 April 2014

A packet loss concealment (PLC) algorithm is proposed to improve the quality of decoded speech when packet losses occur in a wireless sensor network. The proposed algorithm is mainly based on artificial bandwidth extension (ABE) from narrowband to wideband. It consists of three main functions: packet loss concealment in the narrowband, ABE in the modified discrete cosine transform (MDCT) domain, and smoothing of wideband MDCT coefficients with those of the last good frame. The performance of the proposed PLC algorithm is implemented by replacing the PLC algorithm employed in the ITU-T Recommendation G.729.1. The experimental results show that the proposed PLC algorithm provides significantly better speech quality than the PLC in the ITU-T G.729.1.

1. Introduction

There have been rapid developments in wireless sensor networks (WSNs) owing to recent advances in devices such as ultralow-power microcontrollers and short-range transceivers [1]. WSN technology is used in a wide range of applications like environmental monitoring, human tracking, biomedical research, military surveillance, and multimedia transmission [2, 3]. This paper addresses the issues regarding sensors used for multimedia transmission, called wireless multimedia sensors (WMSs) [4, 5]. These sensors deal with multimedia data like image, video, speech, and audio. Multimedia sensor nodes have resource constraints such as low energy capacity of battery, low storage space, and limited computing power. Many multimedia sensor nodes focus on speech data transmission suitable for speech transmission over WSNs. In such cases, each sensor node is linked by wireless local area network (WLAN) links and real-time transport protocol/user datagram protocols (RTP/UDPs).

Packet loss rate increases in this type of transmission because of increased network congestion [6, 7]. In addition, depending on the network resources, the possibility of burst packet losses also increases, which potentially results in severe quality degradation of the reconstructed speech [8]. Most speech coders in use today are based on telephone-bandwidth narrowband speech, nominally limited to about 300–3,400 Hz at a sampling rate of 8 kHz. In order to improve speech quality in voice services, wideband speech coders have been developed for smoothly migrating from narrowband to wideband quality. They operate with a bandwidth 50–7,000 Hz at a sampling rate of 16 kHz. For example, ITU-T Recommendation G.729.1, a scalable wideband speech coder, improves the quality of speech by encoding the frequency bands ignored by the narrowband speech coder, ITU-T Recommendation G.729. Encoding wideband speech using ITU-T Recommendation G.729.1 is performed by two different operations on the low band and high band in the time and frequency domain, respectively. When a frame loss occurs,
the low-band and high-band packet loss concealment (PLC) algorithms work separately. The low-band PLC algorithm reconstructs the excitation and spectral parameters of the lost frame from the last good frame, and the high-band PLC algorithm reconstructs the spectral parameters such as modified discrete cosine transform (MDCT) coefficients of the lost frame from the last good frame [9].

Several packet loss concealment (PLC) methods have been proposed to reduce the speech quality degradation due to a packet loss [7, 10]. The PLC algorithm proposed in [7] was developed to improve the narrowband speech quality by estimating the excitation using comfort noise and multiple codebooks. A technique based on the resynchronization of the glottal pulses in the low band was also proposed in [10], which was subsequently embedded into ITU-T Recommendation G.729.1 as the low-band PLC algorithm [10]. However, the high-band PLC algorithm for ITU-T Recommendation G.729.1 replaced spectral parameters in the modified discrete cosine transform (MDCT) domain with those of the previous frame [10, 11]. In this case, the high-band signal was reconstructed without regard to the low-band signal for the lost frames. The speech quality would have been better if the PLC algorithm estimated the high-band signal by taking into account the reconstructed low-band signal for the lost frames.

Therefore, this paper proposes an artificial bandwidth extension-(ABE-) based PLC algorithm for high-band signal reconstruction in order to improve the quality of decoded speech under packet loss conditions in a WSN. The proposed PLC algorithm is mainly composed of three functions: PLC in the narrowband, ABE in the MDCT domain, and smoothing of the wideband MDCT coefficients using those of the last good frame. The ABE algorithm performs different operations for the 4–4.6 kHz and 4.6–7 kHz bands. It reconstructs the MDCT coefficients of the 4–4.6 kHz band from the harmonic spectral band replication and correlation-based replication approaches. On the other hand, the MDCT coefficients for the 4.6–7 kHz band are obtained by spectral folding [12]. The performance of the proposed PLC algorithm is evaluated by implementing it in the G.729.1 decoder, and it is compared with that of the PLC algorithm employed in the ITU-T Recommendation G.729.1 decoder.

The remainder of this paper is organized as follows. Following this introduction, Section 2 discusses the PLC algorithm currently employed in the G.729.1 decoder. Section 3 proposes an ABE-based PLC algorithm that can also be applied to the ITU-T Recommendation G.729.1 decoder. Section 4 evaluates the performance of the proposed ABE-based PLC algorithm. Finally, this paper is concluded in Section 5.

2. Conventional PLC Algorithm

PLC algorithms can be classified into a sender-based and a receiver-based algorithm, depending on the position where the PLC algorithm works [13, 14]. The sender-based algorithms try to prevent packet errors by using error-robust transmission methods or by including error correction data. The lost speech packets are retransmitted or the sequential speech packets are interleaved to avoid burst losses. Moreover, the speech packets are transmitted with forward error correction (FEC) code or redundant data, which are used to recover the lost speech signals at the receiver. In addition, robust header compression (ROHC) provides a robust speech streaming method at the transmission protocol layer by reducing the overhead due to protocol headers [15]. On the other hand, the receiver-based algorithms conceal lost speech signals by using the speech signal characteristics. The lost speech signals are replaced with silence, noise, or previously reconstructed speech signals. In other words, lost speech signals can be reconstructed by interpolating previous and next good speech signals [16]. In practice, the parameters of a lost frame should be estimated by extrapolating the parameters of a previous good frame.

Figure 1 shows a block diagram of the PLC algorithm employed in the ITU-T Recommendation G.729.1 decoder [17]. The PLC algorithm is composed of low-band and high-band PLC modules. The PLC algorithm reconstructs speech signals of a lost frame based on the speech parameters correctly received from the last good frame, where the speech parameters are excitations in the low band and the MDCT coefficients in the high band. In the low-band PLC module, the excitation of the lost frame is replaced with that obtained from the last good frame, and the energy of the reconstructed excitation is gradually decayed. In addition, a synthesis filter for the lost frame is reconstructed using the linear predictive coding (LPC) coefficients from the last good frame, and the pitch period of the lost frame is estimated as the integer part of the pitch period of the last good frame.

In the high-band PLC module, the high-band signal is reconstructed by time-domain bandwidth extension (TDBWE) that convolves the excitation generated from the low-band PLC module with a spectral envelope estimated from the high-band energy parameters of the last good frame. Then, an MDCT is applied to the high-band signal, and subsequently, the MDCT coefficients corresponding to 7.8 kHz are set to zero. Next, an inverse MDCT (IMDCT) is applied to the modified MDCT coefficients in order to obtain
the high-band signal. Finally, the reconstructed wideband signal of the lost frame is obtained by quadrature mirror filter (QMF) synthesis using both the low-band signal and the high-band signal that are reconstructed by the low-band PLC and high-band PLC modules, respectively.

3. Proposed ABE-Based PLC Algorithm

Figure 2 shows a block diagram of the proposed ABE-based PLC algorithm. When a frame loss occurs, the low-band PLC module reconstructs the low-band speech signal of the lost frame, $s_l(n)$. Simultaneously, the high-band signal is reconstructed by extending the glottal pulse using the high-band spectral envelope of the last good frame, which is denoted by $S_h(k)$ in Figure 2. Next, the high-band signal is obtained by applying an ABE algorithm to extend the low-band signal. Figure 3 shows a block diagram of the ABE-based PLC algorithm. When a frame loss occurs, the low-band PLC module reconstructs the low-band speech signal of the lost frame, $s_l(n)$. Simultaneously, the high-band signal is reconstructed by extending the glottal pulse using the high-band spectral envelope of the last good frame, which is denoted by $S_h(k)$ in Figure 2. Next, the high-band signal is obtained by applying an ABE algorithm to extend the low-band signal in the MDCT domain, resulting in $S_{abe}(k)$. Subsequently, $S_h(k)$ is obtained by smoothing $S_{abe}(k)$ with $S_h(k)$. By applying an IMDCT to $S_h(k)$, a time-domain high-band signal, $\hat{s}_h(n)$, is obtained. Finally, $s_l(n)$ and $\hat{s}_h(n)$ are constructed by using QMF synthesis. In the following subsections, the ABE algorithm and the spectral smoothing method are described in detail.

3.1. Artificial Bandwidth Extension (ABE). ABE is used to generate the high-band MDCT coefficients, $S_{abe}(k)$, as shown in Figure 3. In this paper, the frame size, $N$, is set to 160. For a given set of low-band MDCT coefficients, the high-band MDCT coefficients are obtained in different ways depending on the frequency bands. The high band is divided into two frequency bands, such as 4–4.6 kHz and 4.6–7 kHz.

First, for the frequency band of 4.6–7 kHz, the MDCT coefficients are initially generated by a spectral folding operation, which is defined as

$$S_f(k) = S_l(159 - k), \quad 24 \leq k < 120, \quad (1)$$

where $S_l(k)$ denotes the low-band MDCT coefficient in the $k$th frequency bin. The spectral folding in (1) generates $S_f(k)$ by mirroring $S_l(k)$, where the range of $k$ corresponds to the high band from 4.6 to 7 kHz. However, the spectral folding tends to create an unnaturally prominent harmonic structure at high frequencies, resulting in audible distortion. To mitigate this, the high band of 4.6–7 kHz is further split into subbands of 4.6–5.5 kHz and 5.5–7 kHz.

For the frequency band of 5.5–7 kHz, $S_f(k)$ is smoothened as

$$S_s(k) = \left(\frac{1}{4} \cdot |S_f(k)| + \frac{3}{4} \cdot |S_s(k - 1)|\right) \cdot \text{sgn}(S_f(k)), \quad 0 \leq k < 120, \quad (2)$$

where $S_s(59) = S_f(59)$. In addition, $\text{sgn}(x) = 1$ if $x \geq 0$, but $\text{sgn}(x) = -1$ otherwise. For the frequency band of 4–4.6 kHz, the low-band MDCT coefficients are grouped into 20 subbands, where each subband is composed of eight MDCT coefficients. The energy of the $b$th subband, $E(b)$, is defined as

$$E(b) = \left(\sum_{k = 8b}^{8(b+1)} S_f^2(k)\right)^{1/2}, \quad 0 \leq b < 20. \quad (3)$$

By using $E(b)$ in (3), each low-band MDCT coefficient is normalized as

$$\bar{S}_l(k) = \frac{S_l(k)}{E(b)}, \quad 8b \leq k < 8(b + 1), \quad 0 \leq b < 19. \quad (4)$$

where $\bar{S}_l(k)$ denotes the $k$th normalized low-band MDCT coefficient. Next, the MDCT coefficients for this frequency band are obtained differently depending on the voicing
characteristics of narrowband speech. Each frame is classified as either a voiced or an unvoiced frame by using a spectral tilt parameter, $\theta$, which is identical to the first reflection coefficient, $k_1$, from the ITU-T G.729.1 decoder. If $\theta$ of the current frame is greater than a predefined threshold, $\theta_{th}$, then the frame is declared as a voiced frame; otherwise, it is declared as an unvoiced frame.

For a voice frame, the harmonic characteristics of the low-band should be maintained in the high band [18]. The harmonic spectral band replication approach determines the harmonic period as $\Delta_v = 2N/T$ where $N$ is the frame size and $T$ indicates the pitch period obtained from the ITU-T Recommendation G.729.1 decoder. Then, by using $\delta_l(k)$ in (4), the $k$th MDCT coefficient, $\tilde{S}_l(k)$, is expressed as

$$\tilde{S}_l(k) = \tilde{S}_l\left(k + \frac{N}{2} \left[ \Delta_v - \text{mod} \left( N, \Delta_v \right) \right] \right), \quad 0 \leq k < 24, \quad \text{for} \, 0 \leq m < 3N/4$$

where $\text{mod}(x, y)$ is the modulus operation defined as $x \% y$, and $\lfloor x \rfloor$ denotes the largest integer less than or equal to $x$. In (5), $k$ varies from 0 to 23, which corresponds to the frequency band 4 to 4.6 kHz.

On the other hand, for an unvoiced frame, a correlation-based replication approach is used to patch the high-band MDCT coefficients. Thus, the optimal shift, $\Delta_{uv}$, which maximizes the autocorrelation [19, 20] between the normalized low-band MDCT coefficients, is determined as

$$\Delta_{uv} = \arg \max_{0 \leq m < 3N/4} \text{corr} \left( \delta_l(k), \delta_l(k + m) \right), \quad \text{for} \, 0 \leq m < 3N/4$$

where $3N/4$ is the maximum shift range. Note that the search range is limited between zero and $3N/4$ in order to find an optimal shift in the frequency band of 3-4 kHz. In (6), $\text{corr}(\delta_l(k), \delta_l(k + m))$ is defined as

$$\text{corr} \left( \delta_l(k), \delta_l(k + m) \right) = \sum_{k=0}^{N/4-1} \delta_l\left(k + \frac{3}{4}N\right) \delta_l\left(k + m\right).$$

Therefore, the $k$th MDCT coefficient that is most correlated to $\delta_l(k)$, $\delta_{uv}(k)$, is obtained as

$$\delta_{uv}(k) = \tilde{S}_l\left(k + \frac{N}{4} - \Delta_{uv}\right), \quad 0 \leq k < 24.$$ \hspace{1cm} (8)

It is important to avoid an abrupt change in the boundary between the low band and the high band. This is achieved by adjusting $\delta_{uv}(k)$ for $0 \leq k < 24$ so that the energy of the frequency band of 4–4.6 kHz changes smoothly when compared to the low-frequency band of 3.4–4 kHz [21]. The allowable energy for the $b$th high band, $E_b(b)$, is defined from $E(b)$ in (3) as

$$E_b(b) = \begin{cases} \alpha E(b + 16), & \text{if} \ E(b + 17) > \alpha E(b + 16) \\ E(b + 17), & \text{otherwise}, \end{cases} \quad 0 \leq b \leq 2,$$ \hspace{1cm} (9)

where $\sigma$ denotes a scale factor used to mitigate the abrupt energy change and it is set to 1.25 in this paper. Note also that the range of $b$ is associated with the frequency band of 4–4.6 kHz. Next, each high-band MDCT coefficient in the frequency band of 4–4.6 kHz is modified as

$$\tilde{S}_h(k) = \tilde{S}_l(k) E_b(2 - b), \quad b = \left\lfloor \frac{k}{8} \right\rfloor, \quad 0 \leq k < 2. \quad (10)$$

Finally, the extended MDCT coefficients, $\tilde{S}_{he}(k)$, are obtained by concatenating the MDCT coefficients obtained from (1), (2), and (10), such that

$$\tilde{S}_{he}(k) = \begin{cases} \tilde{S}_h(k), & 0 \leq k < 24, \\ \tilde{S}_l(k), & 24 \leq k \leq 60, \\ \tilde{S}_l(k), & 60 \leq k \leq 120. \end{cases}$$ \hspace{1cm} (11)

The extended MDCT coefficients in (11) provide an excessively fine structure at high frequencies, which results in musical noise. Therefore, it should be smoothened. This is done by applying a shaping function to $\tilde{S}_{he}(k)$, where a cubic spline interpolation is used for the shaping function that has a not-a-knot condition around four control points at 4, 5, 6, and 7 kHz with 0, −6, −12, and −18 dB, respectively [22]. Consequently, the extended MDCT coefficients are further modified as

$$S_{abe}(k) = 10^{0.05\sigma(k)} E_{abe}(k), \quad 0 \leq k < 120,$$ \hspace{1cm} (12)

where $\sigma(k)$ is a value obtained after applying the spline function.

3.2. Reconstruction of High-Band MDCT Coefficients for a Lost Frame. As mentioned earlier, the proposed ABE-based PLC algorithm reconstructs the high-band signal from the low-band signal, which is mainly composed of three modules: low-band PLC, ABE in the MDCT domain, and smoothing of the wideband MDCT coefficients using those of the last good frame. The high-band PLC module in the ITU-T Recommendation G.729.1 decoder utilizes the high-band energy of the last good frame regardless of the signal class characteristics such as voiced, unvoiced, and transition period. In contrast, in the proposed ABE-based PLC algorithm, the high-band MDCT coefficient, $S_{abe}(k)$, is smoothed with the high-band MDCT coefficient, $S_h(k)$, that is obtained from the high-band PLC module in the ITU-T G.729.1 decoder. In other words, the smoothed high-band MDCT coefficient, $\tilde{S}_h(k)$, is obtained as

$$\tilde{S}_h(k) = (|S_h(k)| + |S_{abe}(k)|) \cdot \text{sgn}(S_h(k)), \quad 0 \leq k < 120.$$ \hspace{1cm} (13)

Next, $\tilde{S}_h(k)$ is transformed into the time domain by applying an IMDCT, as shown in Figure 2. Finally, $\tilde{s}(n)$ and $\tilde{s}_{he}(n)$ are concatenated by QMF synthesis using a 64-tap filter to reconstruct the decoded wideband speech for the lost frame.

4. Performance Evaluation

The effectiveness of the proposed ABE-based PLC algorithm is demonstrated by comparing its performance with that of
the PLC algorithm employed in the ITU-T Recommendation G.729.1 decoder, which is referred to as G.729.1-PLC. For comparison, eight audio files (three male voice files, three female voice files, and two music files) were excerpted from the sound quality assessment material (SQAM) database [23]. Since the files were originally recorded in stereo at a sampling rate of 44.1kHz, the right channel signal of each file was downsampled to 16kHz. In addition, two different packet loss conditions such as random and burst packet losses were simulated. Packet loss rates of 10, 20, and 30% were generated by the Gilbert-Elliot model defined in ITU-T Recommendation G.191 [24]. To simulate burst packet loss conditions, the burstiness of the packet losses was set to 0.99, where the mean and maximum consecutive packet losses were measured as 1.9 and 5.6 frames, respectively.

First, the log spectral distortion (LSD) [25] was measured between the original and decoded signal. It is defined as

\[
\text{LSD} = \left( \frac{1}{N/4} \sum_{k=N/4}^{N/2-1} \left( 10 \log_{10} |A(k)|^2 - 10 \log_{10} |A'(k)|^2 \right) \right)^{1/2},
\]

where \(A(k)\) and \(A'(k)\) denote the \(k\)th spectral components of the original signal and the proposed signal, respectively. In order to obtain LSD, an \(N\)-point discrete Fourier transform was applied to both signals and then summed from \(N/4\) to \(N/2\). This was because only the spectral components of the high band were compared. Tables 1 and 2 compare the LSD between the proposed PLC and the G.729.1-PLC algorithms at packet loss rates of 10, 20, and 30% under random and burst packet loss conditions for the speech and music files, respectively. It was shown from the tables that the proposed PLC algorithm provides smaller LSDs than G.729.1-PLC for all packet loss conditions.

Second, the waveforms decoded by different PLC algorithms were compared, as shown in Figure 4. It was seen that the decoded signal obtained by the proposed PLC algorithm (Figure 4(e)) is closer in fidelity to the decoded signal without any loss (Figure 4(b)) than the decoded signals obtained by G.729.1-PLC (Figure 4(d)) for a given packet error pattern (Figure 4(c)). Additionally, Figure 5 compares the spectrograms of the signals decoded by different PLC algorithms. As shown in Figure 5, the spectrogram of decoded signal obtained by the proposed PLC algorithm (Figure 5(d)) was more similar to the decoded signal without any loss.

Figure 4: Waveform comparison of (a) original signal, (b) decoded signal without packet loss, (c) packet error pattern, (d) decoded signal by the G.729.1-PLC algorithm, and (e) decoded signal by the proposed PLC algorithm.

| Burstiness | PLR (%) | G.729.1-PLC (dB) | Proposed PLC (dB) |
|------------|---------|------------------|-------------------|
| \(y = 0\)  | 10      | 10.04            | 10.00             |
|            | 20      | 10.90            | 10.81             |
|            | 30      | 11.78            | 11.63             |
| \(y = 0.99\) | 10      | 10.28            | 10.20             |
|            | 20      | 11.02            | 10.85             |
|            | 30      | 11.92            | 11.75             |
| Average    |         | 10.99            | 10.87             |
Figure 5: Spectrogram comparison of (a) original signal, (b) decoded signal without packet loss, (c) decoded signal by the G.729.1-PLC algorithm, and (d) decoded signal by the proposed PLC algorithm in the packet loss.

Table 2: Comparison of log spectral distortions (LSDs) of the proposed PLC and G.729.1-PLC algorithms under random and burst packet loss conditions with different packet loss rates (PLRs) for music files.

| Burstiness PLR(%) | G.729.1-PLC (dB) | Proposed PLC (dB) |
|-------------------|------------------|-------------------|
| γ = 0             |                  |                   |
| 10                | 17.93            | 17.89             |
| 20                | 18.24            | 18.16             |
| 30                | 18.55            | 18.28             |
| γ = 0.99          |                  |                   |
| 10                | 18.35            | 18.30             |
| 20                | 18.62            | 18.50             |
| 30                | 18.68            | 18.34             |
| Average           | 18.40            | 18.25             |

Table 3: Comparison of A-B preference scores (%) for speech files between the proposed PLC and G.729.1-PLC under random (γ = 0) and burst packet loss (γ = 0.99) conditions with different packet loss rates (PLRs).

| Burstiness PLR(%) | G.729.1-PLC (%) | No. Diff. | Proposed PLC (%) |
|-------------------|-----------------|-----------|------------------|
| γ = 0             |                 |           |                  |
| 10                | 21.43           | 45.24     | 33.33            |
| 20                | 28.57           | 35.71     | 35.72            |
| 30                | 19.05           | 54.76     | 26.19            |
| γ = 0.99          |                 |           |                  |
| 10                | 14.29           | 52.38     | 33.33            |
| 20                | 26.19           | 40.48     | 33.33            |
| 30                | 16.67           | 47.62     | 35.71            |
| Average           | 21.03           | 46.03     | 32.94            |

Table 4: Comparison of A-B preference scores (%) for music files between the proposed PLC and G.729.1-PLC under random (γ = 0) and burst packet loss (γ = 0.99) conditions with different packet loss rates (PLRs).

| Burstiness PLR(%) | G.729.1-PLC (%) | No. Diff. | Proposed PLC (%) |
|-------------------|-----------------|-----------|------------------|
| γ = 0             |                 |           |                  |
| 10                | 21.43           | 50.00     | 28.57            |
| 20                | 14.29           | 57.14     | 28.57            |
| 30                | 28.57           | 42.86     | 28.57            |
| γ = 0.99          |                 |           |                  |
| 10                | 21.43           | 42.86     | 35.71            |
| 20                | 21.43           | 35.71     | 42.86            |
| 30                | 7.14            | 57.71     | 35.72            |
| Average           | 19.05           | 47.71     | 33.33            |

(Figure 5(b)) than the spectrogram of the decoded signals obtained by G.729.1-PLC (Figure 5(c)) in the high band.

Third, an A-B preference listening test was performed to evaluate the subjective quality. The audio data used for the test consisted of six speech files (three male and three female voices) and two music files. All the files were processed under random and burst packet loss conditions by G.729.1-PLC and the proposed PLC algorithm, respectively. In this paper, seven people with no auditory diseases participated. Audio files processed by the G.729.1-PLC and proposed PLC algorithm were presented to the participants, and they were asked to choose their preference. Tables 3 and 4 show the test results for the speech and music data, respectively. Note that if a participant could not distinguish the difference between the file processed by the proposed PLC and G.729.1-PLC, then “No. Diff.” was selected. It was shown from Tables 3 and 4 that the speech and music signals decoded by the proposed PLC algorithm were preferred to those by the G.729.1-PLC algorithm.
Next, in order to demonstrate the effectiveness of the proposed PLC algorithm, the stimuli with a hidden reference and an anchor (MUSHRA) test [26] were performed as a subjective listening test. For the MUSHRA test, two anchors with cut-off frequencies of 7 and 3.4 kHz were prepared. Seven people with no auditory diseases also participated in this test. Each participant was presented with the eight stimuli and was asked to rate the audio quality from 0 to 100. Figure 6 compares the MUSHRA scores, where each column corresponds to the opinion score averaged over seven listeners and eight audio files. Note that the vertical line on the top of each bar denotes the standard deviation of the opinion score. As shown in Figure 6, the proposed PLC algorithm achieved an average score of 39, which was higher than that by the G.729.1-PLC algorithm.

Finally, in order to show how much more effective the proposed PLC algorithm was in comparison to the G.729.1-PLC algorithm, a paired $t$-test [27] was performed using their MUSHRA scores. Assuming that the differences in MUSHRA scores followed a normal distribution, the test statistic had a $t$-distribution based on $(n - 1)$ degrees of freedom [28], where $n$ was the number of stimuli for the MUSHRA test; thus, $n = 56$. The test statistic was given by $t = \frac{\overline{d} - \overline{d}_0}{s/\sqrt{n}}$, where $\overline{d}$ and $s_2$ are the sample mean and the sample standard deviation of the $n$ differences of MUSHRA scores, respectively.

For the paired $t$-test, the test statistic must be greater than $t_{0.05}$ if two methods are significantly different. According to the mathematical table [28], $t_{0.05} = 1.674$ when $n = 56$ for a confidence of 95%. The test statistic was 3.20, which implied that the audio quality of the stereo signals processed by the proposed PLC was significantly better than that of G.729.1-PLC.

### 5. Conclusion

In this paper, a packet loss concealment (PLC) algorithm has been proposed to improve the performance of decoded signal quality when frame erasures or packet losses occurred in wireless sensor networks. The proposed PLC algorithm was based on artificial bandwidth extension (ABE) from the low band to the high band in the MDCT domain. The performance of the proposed PLC algorithm was evaluated by replacing the PLC algorithm currently employed in the ITU-T Recommendation G.729.1 decoder, G.729.1-PLC under random and burst packet loss rates of 10, 20, and 30%. The comparisons were made based on log spectral distortion (LSD), waveform/spectrogram comparison, an A-B preference test, MUSHRA test, and the paired $t$-test. It was shown from the comparisons that the proposed PLC algorithm provided better quality of decoded speech and music signals than G.729.1-PLC for all the simulated packet loss conditions.

### Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

### Acknowledgments

This work was supported in part by the National Research Foundation of Korea (NRF) Grant funded by the government of Korea (MSIP) (no. 2012-010636), by the MSIP (Ministry of Science, ICT and Future Planning), Korea, under the ITRC (Information Technology Research Center) Support Program (NIPA-2013-H0301-13-4005) supervised by the NIPA (National IT Industry Promotion Agency), and by the Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (NRF-2009-0093828).

### References

[1] J. Du, W. Chen, Y. Liu, Y. Gu, and H. Liu, “Catch you as i can: indoor localization via ambient sound signature and human behavior,” International Journal of Distributed Sensor Networks, vol. 2013, Article ID 434301, 16 pages, 2013.

[2] M. Petracca, G. Litovsky, A. Rinotti, M. Tacca, J. C. De Martin, and A. Fumagalli, “Perceptual based voice multi-hop transmission over wireless sensor networks,” in Proceedings of the IEEE Symposium on Computers and Communications (ISCC ’09), pp. 19–24, Sousse, Tunisia, July 2009.

[3] R. Manghalam, A. Rowe, R. Rajkumar, and R. Suzuki, “Voice over sensor networks,” in Proceedings of the IEEE Real-Time Systems Symposium, pp. 291–302, Rio de Janeiro, Brazil, December 2006.

[4] Y. Yang, X. Qin, G. Sun, Y. Xu, Z. Yang, and Z. Zuo, “Data deduplication in wireless multimedia monitoring network,” International Journal of Distributed Sensor Networks, vol. 2013, Article ID 153034, 7 pages, 2013.

[5] P. C. Newton and L. Arockiam, “A quality of service performance evaluation strategy for delay classes in general packet radio service,” International Journal of Advanced Science and Technology, vol. 50, pp. 91–98, 2013.

[6] M. M. A. Ghazala, M. F. Zaghloul, and M. Zahra, “Performance evaluation of multimedia streams over wireless com-puter networks (WLANs),” International Journal of Advanced Science and Technology, vol. 13, pp. 61–74, 2009.

[7] N. I. Park, H. K. Kim, M. A. Jung, S. R. Lee, and S. H. Choi, “Burst packet loss concealment using multiple codebooks and comfort noise for CELP-type speech coders in wireless sensor networks,” Sensors, vol. 11, no. 5, pp. 5323–5336, 2011.
[8] W. Jiang and H. Schulzrinne, “Comparison and optimization of packet loss repair methods on VoIP perceived quality under bursty loss,” in *Proceedings of the 12th International Workshop on Network and Operating Systems Support for Digital Audio and Video (NOSSDAV ’02)*, pp. 73–81, Miami, Fla, USA, May 2002.

[9] P. Gournay, F. Rousseau, and R. Lefebvre, "Improved packet loss recovery using late frames for prediction-based speech coders," in *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP ’03)*, pp. 108–111, Hong Kong, April 2003.

[10] T. Vaillancourt, M. Jelinek, R. Salami, and R. Lefebvre, “Efficient frame erasure concealment in predictive speech codecs using glottal pulse resynchronisation,” in *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP ’07)*, pp. 1113–1116, Honolulu, Hawaii, USA, April 2007.

[11] ITU-T Recommendation G. 729.1, G.729 Based Embedded Variable Bit-Rate Coder: An 8-32 kbit/s Scalable Wide-band Coder Bitstream Interoperable with G.729, Geneva, Switzerland, 2006.

[12] 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.

[13] O. J. Wasem, D. J. Goodman, C. A. Dvorak, and H. G. Page, “The effect of waveform substitution on the quality of PCM packet communications,” *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 36, no. 3, pp. 342–348, 1988.

[14] V. Hardman, M. A. Sasse, M. Hadnly, and A. Watson, “Reliable audio for use over the Internet,” in *Proceedings of the International Networking Conference*, pp. 171–178, Honolulu, Hawaii, USA, June 1995.

[15] S. Rein, F. H. P. Fitzek, and M. Reisslein, “Voice quality evaluation in wireless packet communication systems: a tutorial and performance results for ROHC,” *IEEE Wireless Communications*, vol. 12, no. 1, pp. 60–67, 2005.

[16] H. Sanneck, A. Stenger, K. B. Younes, and B. Girod, “New technique for audio packet loss concealment,” in *Proceedings of the IEEE Communications (GLOBECOM ’96)*, pp. 48–52, London, UK, November 1996.

[17] S. Ragot, B. Kövesi, R. Trilling et al., “ITU-T G.729.1: an 8–32 KBIT/S scalable coder interoperable with G.729 for wideband telephony and voice over IF,” in *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP ’07)*, pp. 529–532, Honolulu, Hawaii, USA, April 2007.

[18] D. Y. Shingchern and C. M. Tsai, “Determining start-band frequency for spectral band replication tool in MPEG-4 advanced audio coding,” *Information-An International Interdisciplinary Journal*, vol. 15, no. 5, pp. 1839–1850, 2012.

[19] Y. H. Lee, S. D. Jo, J. H. Park et al., “A time-scale modification-based voice changing method with seamless switching and its real-time implementation on digital imaging devices,” *Information-An International Interdisciplinary Journal*, vol. 15, no. 5, pp. 1303–1316, 2012.

[20] A. K. Shingh and R. Saxena, "Correlation theorem for fractional Fourier transform," *Information Journal of Signal Processing, Image Processing and Pattern Recognition*, vol. 4, no. 2, pp. 31–40, 2011.

[21] Y. H. Lee and S. H. Choi, "Superwideband bandwidth extension using normalized MDCT coefficients for scalable speech and audio coding," *Advances in Multimedia*, vol. 2013, Article ID 909124, 6 pages, 2013.
