Speech Bandwidth Extension Using DWT-FFT-Based Data Hiding

Phaneendra KURADA [1], Sailaja MARUVADA [2], Koteswara Rao SANAGAPALLEA [3]

[1] Dept. of ECE, Raghu Institute of Technology, Dakamarri, Visakhapatnam, Andhra Pradesh, India
[2] Dept. of ECE, JNTU Kakinada, Kakinada, Andhra Pradesh, India
[3] Dept. of ECE, KL University, Vaddeswaram, Vijayawada, Andhra Pradesh, India

kuradaphaneendra@gmail.com, maruvada.sailaja@gmail.com, rao.sk9@gmail.com

Submitted March 8, 2019 / Accepted October 16, 2019

Abstract. *A novel transform-domain speech bandwidth extension algorithm is proposed to transmit information about the missing speech frequencies over a hidden channel, i.e., the related encoded spectral envelope parameters are hidden within the narrowband speech signal using discrete wavelet transform-fast Fourier transform-based data hiding (DWTFFTBDH) technique. The hidden information is recovered reliably at the receiver to produce a wideband speech signal of much higher quality. Obtained results confirm the excellent reconstructed wideband speech quality of the proposed method over traditional methods.*

Keywords

Public Switched Telephone Network (PSTN), bandwidth extension of narrowband speech, DWT-FFT-based data hiding, speech quality, spread spectrum

1. Introduction

Majority of the telephone connections are restricted to speech frequencies below 4 kHz termed as narrowband (NB) which causes the characteristic sound of telephony speech. In order to improve the speech quality, speech frequencies up to 8 kHz, called wideband (WB), are desired. The required modifications, which are expensive and time-consuming, of today's telephone network infrastructure turned out to be the main complication for the introduction of high-quality speech transmission in existing networks [1].

An alternative approach to enhance the quality of the received NB signal is the artificial bandwidth extension (ABE) [2], where the bandwidth of the NB signal is artificially extended at the receiving end. The source-filter model divides bandwidth extension (BE) into excitation signal extension and WB spectral envelope estimation. Several approaches for excitation signal extension can be found in [2], [3]. Several approaches for WB spectral envelope estimation can be found in [3–6]. However, ABE techniques do not provide adequately stable WB speech quality in all circumstances [7].

A new solution to resolve this problem is to communicate the information about the missing speech frequencies (MSFs) over a hidden channel, i.e., the related information is hidden within NB signal using data hiding techniques [1]. Several state-of-the-art techniques for speech BE based on data hiding are available. A speech BE method has been proposed in [8], where the encoded spectral envelope parameters (SEPs) of the MSFs in 4 to 8 kHz range, called high-band (HIB) signal, are hidden into the NB signal to produce a composite NB (CNB) signal. A method for generating a high-quality WB signal over the above method has been proposed in [9], where HIB signal is encoded more efficiently by phonetic classification. A method for the enhancement of telephone speech quality has been proposed in [10], where SEPs of HIB signal are inserted into the least significant bits of the bit stream of NB signal. A method for speech BE has been proposed in [11], where quantization based data hiding technique is employed through which reliable transmission is attained over many typical telephony channels. In [12], the audible components of HIB signal are embedded into the hidden channel. It was shown that hidden data can be reliably recovered. A speech BE method has been proposed in [13] based on insertion of the pitch-scaled frequencies of HIB signal into 3.4 to 4 kHz. A method based on joint coding and data hiding (JCD) has been reported in [14] for extending the bandwidth of an NB signal. The WB signal of high quality is reconstructed in [15], [16] using JCD technique.

Speech BE algorithms using data hiding should provide a high-quality CNB signal and a reconstructed wideband (RWB) signal. These algorithms should also be robust enough to withstand channel and quantization noises. However, many of the methods discussed above [8–16] fail to provide a high-quality CNB signal and RWB signal. Also, they fail to provide robustness to withstand channel and quantization noises. So, the development of a novel speech BE algorithm using the data hiding technique is

required to enhance CNB signal quality and RWB signal quality and efficient handling of channel and quantization noises.

A speech steganography method has been proposed in [17]. This employs DWTFFTBDH technique to embed the parameters of the secret speech signal in the detailed wavelet coefficients of host speech signal without degrading the quality of the host signal. It was found that this approach is producing a stego speech signal that is indistinguishable from the host speech while being able to recover the secret speech signal without any degradation in quality.

A novel speech BE algorithm using DWTFFTBDH technique [17] is proposed to insert the encoded SEPs of HIB signal into detailed wavelet coefficients of NB signal. The hidden data is retrieved at the receiving end to produce a high-quality WB signal. Furthermore, the proposed method is compatible with conventional NB terminal equipment's, e.g., a plain ordinary telephone set (POTS). In other words, conventional NB receivers can still access the NB speech properly without additional hardware, while a customized receiver is able to extract the embedded information and provide WB signal with much better quality.

The telephone network channel effects, such as quantization and channel noises, are incorporated in this paper. More elaborate proposals, such as [8] and [9], aiming at speech BE, account for the aspect of quantization noise. However, the impact of channel noise has not been evaluated. The present invention considers the code division multiple access (CDMA) technique for recovering the hidden data as it is claimed to be robust against channel and quantization noises. In particular, each data bit to be embedded into the NB signal is spread out by multiplying it with a specific spreading sequence. The spread signals are then added up to form the hidden data. The hidden data can be reliably recovered because of low cross-correlation between spreading sequences (Hadamard codes are employed in this work).

The paper is organized as follows. In Sec. 2, DWTFFTBDH technique for BE is introduced. Section 3 deals with the novel speech BE algorithm using DWTFFTBDH technique. The subjective and objective analyses are discussed in Sec. 4. Section 5 gives a conclusion.

2. DWT-FFT-Based Data Hiding Technique for BE

To hide HIB signal $\mathbf{Y}_{\mathrm{eb}}(n)$ within NB signal $\mathbf{Y}_{\mathrm{nb}}(n)$, initially, discrete wavelet transform (DWT) is applied on $\mathbf{Y}_{\mathrm{nb}}(n)$ to decompose it into detailed and approximation coefficients. Fast Fourier transform (FFT) is then applied on detailed coefficients to compute the spectrum, followed by calculation of magnitude spectrum $|\mathbf{Y}_{\mathrm{nb}}(K)|$ and phase spectrum $\mathbf{\Phi}_{\mathrm{NB}}(k)$. Assume that $\mathbf{Y}_{\mathrm{eb}}(n)$ is encoded into a sequence of data bits, i.e., $C_s \in \{-1,1\}$, $s = 0,1,\ldots,S-1$, where S denotes the total number of bits.

Spread each data bit to be embedded by multiplying with a specific pseudo-noise (PN) code, i.e., $C_s \cdot q^s$. The length of the PN code q^s is S. Adding all of these spreading vectors produces hidden data. It is given by

$$E = \sum_{s=0}^{S-1} C_s \cdot q^s \,. \tag{1}$$

The hidden data αE are inserted into the last L elements of the first half of $|\mathbf{Y}_{\mathrm{nb}}(k)|$ [17], and this results in a modified magnitude spectrum $|\mathbf{Y}^1{}_{\mathrm{nb}}(k)|$ and it is given by

$$|\mathbf{Y}^1{}_{\mathrm{s\ nb}}(k)| = \begin{cases} |\mathbf{Y}_{\mathrm{nb}}(k)|, & k = 0,1,\ldots,\dfrac{M}{2}-L \\ \alpha E_l, & k = \dfrac{M}{2}-L-1,\ldots,\dfrac{M}{2}-1 \\ \alpha E_l, & k = \dfrac{M}{2},\ldots,\dfrac{M}{2}+L \\ |\mathbf{Y}_{\mathrm{nb}}(k)|, & k = \dfrac{M}{2}+L+1,..,M-1 \end{cases} \tag{2}$$

where E_l denotes the l^{th} component of $\mathbf{E}$ and α is a scalar that will enhance the quality of CNB signal, i.e.,

$$\alpha^2 E_l^{\,2} \le \frac{1}{G_{\mathrm{nb}}} \tag{3}$$

where G_{nb} denotes the energy of $|\mathbf{Y}_{\mathrm{nb}}(K)|$. Hence, an appropriate value of α is found by $\alpha = \sqrt{\dfrac{1}{G_{\mathrm{nb}} E_l^{\,2}}}$.

Considering that $|C_s \cdot q^s| = 1$,

$$\alpha = \sqrt{\frac{1}{SG_{nb}}} \,. \tag{4}$$

These changes result in a CNB signal, and its spectrum can be expressed as,

$$\mathbf{Y}^1{}_{\mathrm{s\ nb}}(K) = |\mathbf{Y}^1{}_{\mathrm{s\ nb}}(K)| \mathrm{e}^{\mathrm{j}\Phi_{\mathrm{NB}}(K)}, K = 0,\ldots,M-1 \,. \tag{5}$$

Inverse transform the CNB signal spectrum to convert back to the time representation of the CNB signal by applying an inverse FFT and then inverse DWT. The resulting CNB signal $\mathbf{Y}^1{}_{\mathrm{nb}}(n)$ is transmitted over telephone network channel to the receiver and the channel introduces channel and quantization noises. Let $\hat{\mathbf{Y}}^1{}_{\mathrm{nb}}(n)$ denote the received signal, i.e., $\hat{\mathbf{Y}}^1{}_{\mathrm{nb}}(n) = \mathbf{Y}^1{}_{\mathrm{nb}}(n) + er$. The combination of channel and quantization noises is denoted by er. $\hat{\mathbf{Y}}^1{}_{\mathrm{nb}}(n)$ is treated as an ordinary signal by a conventional phone terminal. The quality of $\mathbf{Y}_{\mathrm{nb}}(n)$ is not considerably degraded since the perceived differences between $\mathbf{Y}_{\mathrm{nb}}(n)$ and $\mathbf{Y}^1{}_{\mathrm{nb}}(n)$ are very small.

Recovery of the hidden data $\hat{\mathbf{Y}}_{\mathrm{eb}}(n)$ requires the receiver to compute the spectrum of the signal by applying DWT on $\hat{\mathbf{Y}}^1{}_{\mathrm{nb}}(n)$, and then, FFT is applied on detailed

coefficients, followed by calculation of magnitude spectrum and phase spectrum. The hidden data are then recovered from the magnitude spectrum of $\hat{\mathbf{Y}}^1_{\text{nb}}(n)$ [17] by

$$\hat{E}_l = \left|\hat{\mathbf{Y}}^1_{\text{nb}}(k)\right|, k = \frac{M}{2} - L - 1, ..., \frac{M}{2} - 1. \quad (6)$$

The data bits are decoded by employing a multiuser detector [18]. That is,

$$\hat{C}_s = \text{sign}\left(\sum_{l=0}^{L-1} \hat{E}_l q_l^{\,s}\right). \quad (7)$$

In a noise-free environment, $\hat{E}_l = \alpha E_l$. Substituting it into (7), we have

$$\begin{aligned}\hat{C}_s &= \text{sign}\left(\sum_{l=0}^{L-1} \alpha E_l q_l^{\,s}\right) \\ &= \text{sign}\left(\alpha \sum_{l=0}^{L-1}\left(c_l q_l^{\,s} q_l^{\,s} + \sum_{g=0, g\neq s}^{S-1} c_g q_l^{\,g} q_l^{\,s}\right)\right) \\ &= \text{sign}\left(\alpha L c_l + \alpha \sum_{g=0, g\neq s}^{S-1} c_g \sum_{l=0}^{L-1} q_l^{\,g} q_l^{\,s}\right).\end{aligned} \quad (8)$$

The PN sequences are orthogonal. That is

$$\sum_{l=0}^{L-1} q_l^{\,g} q_l^{\,s} = 0 \quad (9)$$

where $g \neq s$. Therefore,

$$\alpha \sum_{g=0, g\neq s}^{S-1} c_g \sum_{l=0}^{L-1} q_l^{\,g} q_l^{\,s} = 0. \quad (10)$$

This demonstrates that the parameters of $\mathbf{Y}_{\text{eb}}(n)$ can be effectively retrieved by using the CDMA technique.

3. Speech BE Using DWTFFTBDH Technique

3.1 Transmitter

Figure 1 shows the proposed transmitter. Initially, WB speech $\mathbf{Y}_{\text{wb}}(n)$ that was sampled at 16 kHz is split into a low-band signal and a HIB signal by the low-pass filter (LPF) and a high-pass filter (HPF), respectively, where low-band signal contains speech information between 0 and 4 kHz and HIB signal contains speech information between 4 kHz and 8 kHz. The NB signal $\mathbf{Y}_{\text{nb}}(n)$ is then produced by decimating LPF output by a factor of two. The output of HPF is shifted to the frequency range of NB spectrum and then decimated to produce an upper-band (UB) signal $\mathbf{Y}_{\text{eb}}(n)$.

Minimize the number of parameters that represent $\mathbf{Y}_{\text{eb}}(n)$ to imperceptibly embed HIB signal into the NB signal. Here, the linear predictive (LP) analysis [19] is utilized to fulfil this objective. Computing LP coefficients from $\mathbf{Y}_{\text{eb}}(n)$ using the Levinson-Durbin algorithm [19] and then these coefficients are converted to line spectral frequencies (*lsfs*) since the minor change in coefficients results in distortions while reconstructing $\mathbf{Y}_{\text{eb}}(n)$. Also, the gain of $\mathbf{Y}_{\text{eb}}(n)$ has to be embedded to avoid over-estimation [20]. Thus, calculate the relative gain as $g_{\text{r}} = g_{\text{eb}}/g_{\text{nb}}$ and combined with *lsfs* to produce a representation vector of $\mathbf{Y}_{\text{eb}}(n)$, i.e., $\mathbf{C} = [lsf_1, lsf_2, \ldots, lsf_{10}, g_{\text{r}}]$. Quantize $\mathbf{C}$ to the closest entry of a vector quantization (VQ) codebook that is generated by the fuzzy *c*-means (FCM) algorithm [21]. The binary representation of the entry index, i.e., $(c_0 c_1 c_2 \ldots c_{S-1})$ is then hidden within the NB signal using DWTFFTBDH technique to provide a composite NB signal $\mathbf{Y}^1_{\text{nb}}(n)$ that can be transmitted over telephone network channel to the receiver.

The parameters of the excitation signal are not embedded to minimize the parameters to be embedded since above 3400 Hz, the human ear is not sensitive to the excitation signal distortions [22]. Therefore, estimation of UB signal excitation from the NB signal at the receiver guarantees the reconstruction performance.

A synchronization sequence such as 1111….111 is inserted after every frame of $\mathbf{Y}^1_{\text{nb}}(n)$ to accomplish frame synchronization [23] between the transmitter and receiver. The arrival of a new frame of $\mathbf{Y}^1_{\text{nb}}(n)$ is indicated by the reception of a certain number of consecutive identical waveforms (synchronization sequence) at the receiver.

3.2 Receiver

Figure 2 shows the proposed receiver. Recover the entry index properly by the proposed DWTFFTBDH technique and then the corresponding quantized *lsfs* are properly retrieved from the VQ codebook. Then, construct LP coefficients from the retrieved *lsfs*. Meanwhile, $\hat{\mathbf{Y}}^1_{\text{nb}}(n)$ is inverse filtered using LP coefficients of $\hat{\mathbf{Y}}^1_{\text{nb}}(n)$ to obtain an NB residual signal and extending the residual signal. This results in a UB excitation signal. Synthesizing $\hat{\mathbf{Y}}_{\text{eb}}(n)$ is carried out by exciting the synthesis filter described by the retrieved LP coefficients by a UB excitation signal. At this point, the sampling rate for both $\hat{\mathbf{Y}}^1_{\text{nb}}(n)$ and $\hat{\mathbf{Y}}_{\text{eb}}(n)$ is 8 kHz. Interpolating these signals by a factor of two, the sampling rate of WB signal. $\mathbf{Y}^1_{\text{eb}}(n)$ denotes the interpolated $\hat{\mathbf{Y}}_{\text{eb}}(n)$, which lies in 0 to 4 kHz and is shifted to 4 to 8 kHz. The interpolated composite NB ($\mathbf{Y}^{11}_{\text{nb}}(n)$) and restored $\mathbf{Y}^1_{\text{eb}}(n)$ signals are summed to produce WB signal ($\mathbf{Y}^1_{\text{wb}}(n)$) of high quality.

4. Evaluation

Two aspects need to be considered for the quality evaluation of the proposed system. First, a good WB quality must be guaranteed for customized receiver. Second, the NB speech quality must not be degraded even after embedding the encoded spectral envelope parameters of HIB signal into detailed Wavelet coefficients of NB signal for conventional NB receivers.

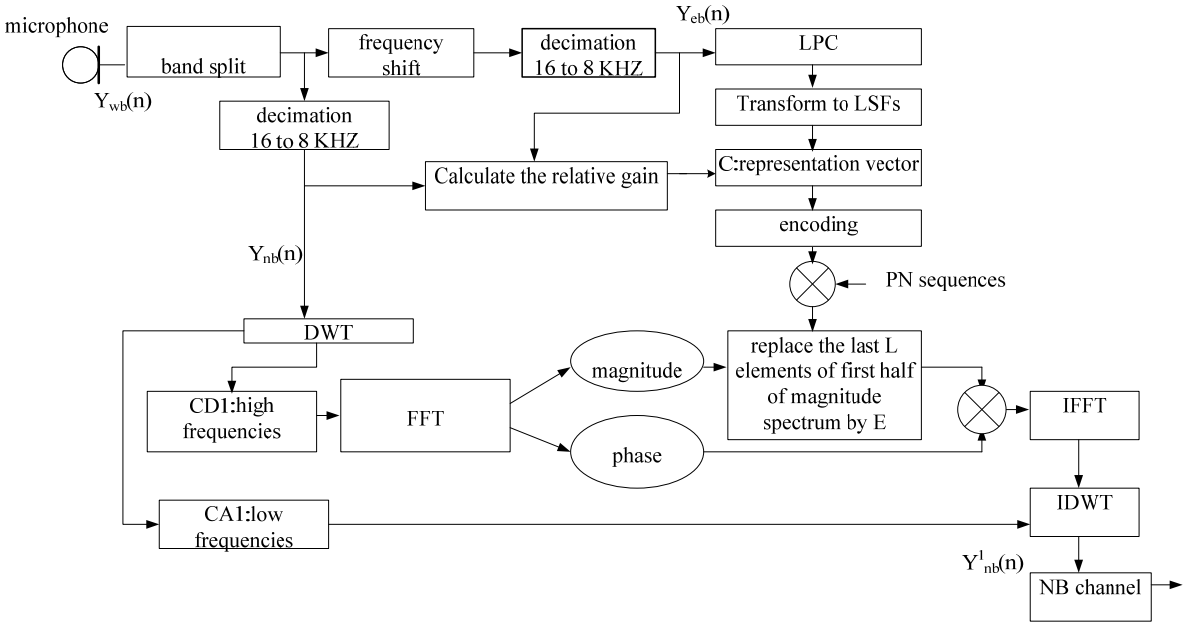

Fig. 1. The proposed transmitter.

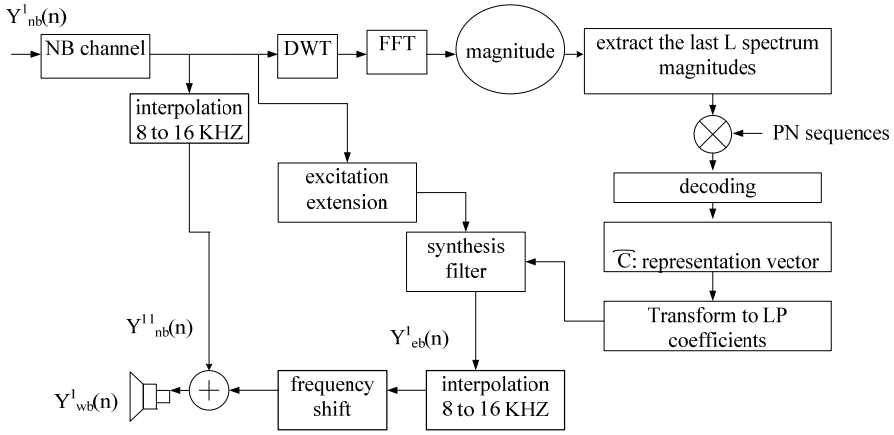

Fig. 2. The proposed receiver.

Twelve sentences spoken by 20 speakers including 10 men and 10 women (altogether two hundred forty speech utterances) are taken from the TIMIT corpus [24] for the performance evaluation. The NB signal is segmented into frames of length of 20 ms with 10-ms overlap between frames and is processed on a frame-by-frame basis. The performance of the proposed BE algorithm is assessed with subjective and objective tests. The different methods compared with the proposed method are: BE of telephony speech by data hiding [8], speech BE by data hiding and phonetic classification [9], an audio watermark-based speech BE [10] and steganographic WB telephony using NB speech codecs [15]. These are represented, respectively, by conventional speech BE using data hiding (CSBUDH), conventional speech BE using data hiding and phonetic classification (CSBUDHAPC), conventional speech BE using bit stream data hiding (CSBUBSDH) and conventional speech BE using watermark transmitted side information (CSBUWTSI) in the analysis. Telephony channel model used in this paper is additive white Gaussian noise (AWGN) channel model.

4.1 Subjective Listening Test Results

The speech utterances used for the subjective listening tests were a subset of the two hundred forty speech utterances. The subjective listening tests were made on one hundred speech utterances drawn randomly from the two hundred forty speech utterances. Here, the perceptual transparency (PET) is evaluated using a mean opinion score (MOS) test [8], [9]. The listening test for the comparison of the WB, CNB and RWB signals is conducted [12]. The obtained speech quality of the proposed method and conventional speech BE methods [8–10, 15] is also assessed using absolute category rating (ACR) listening test [34] recommended by international telecommunications union (ITU-T). Speech BE using different data hiding techniques [8, 9, 27–31] are used MOS test to evaluate the

perceptual transparency. These tests were conducted in a quiet environment using headphones. Twenty subjects participated in each test.

Perceptual Transparency

The information should be transparently hidden by the proposed method. That is $\mathbf{Y}^{1}_{nb}(n)$ and $\mathbf{Y}_{nb}(n)$ should be subjectively indistinguishable. High PET means low noticeable NB signal degradation. There should be high PET even after embedding the encoded spectral envelope parameters of HIB signal into detailed wavelet coefficients of NB signal. PET is evaluated using the MOS test [8], [9]. Subjects participating in the test compare $\mathbf{Y}_{nb}(n)$ and $\mathbf{Y}^{1}_{nb}(n)$ and provide their opinions in terms of MOS presented in Tab. 1. Table 2 illustrates the results of the averaged mean opinion scores for the traditional methods [8–10, 15] and the proposed method. As seen in Tab. 2, the proposed method reveals its distinct PET advantage over the traditional methods [8–10, 15].

Subjective Comparison of Original WB Speech, Composite NB speech and Reconstructed WB Speech

A subjective listening test has been conducted in order to compare the performance of the proposed technique with the existing techniques. WB speech is denoted as I, CNB speech and RWB speech are numbered as II and III respectively. The subjects have to compare speech samples pairs taken from I to III and rate the first sample of the pair sounded best ($\triangleright$), poor ($\triangleleft$) or similar ($\approx$) in relation to the second sample. Table 3(a) lists the results of comparing I with II and III and 3(b) lists the results of comparison of II with III. The number of subjects with a specific rating ($\triangleright$ or $\triangleleft$ or $\approx$) is presented in the table in Arabic numerals. Table 3(a) confirms that the consistent preference of original WB speech over CNB speech of traditional methods [8–10, 15] and the proposed method. A clearly improved RWB signal quality of the proposed method over the traditional methods is also observed from Tab. 3(a). Table 3(b) confirms that, compared to traditional methods, there is a clear RWB signal quality advantage of the proposed method over CNB speech.

Score	Instruction
1	NB and composite NB signals are dissimilar
2	NB and composite NB signals are alike, but easy to see the difference
3	NB and composite NB signals are very similar, the only minor difference exists
4	NB and composite NB signals are the same

Tab. 1. MOS.

Technique	Mean opinion score
CSBUDH [8]	2.97
CSBUDHAPC [9]	3.16
CSBUBSDH [10]	3.28
CSBUWTSI [15]	3.64
Proposed technique	3.94

Tab. 2. Results of the mean opinion score test.

	I	II	III
CSBUDH [8]	$\triangleright$	20	13
	$\triangleleft$	0	0
	$\approx$	0	7
CSBUDHAPC [9]	$\triangleright$	20	11
	$\triangleleft$	0	0
	$\approx$	0	9
CSBUBSDH [10]	$\triangleright$	20	12
	$\triangleleft$	0	0
	$\approx$	0	8
CSBUWTSI [15]	$\triangleright$	20	9
	$\triangleleft$	0	0
	$\approx$	0	11
Proposed method	$\triangleright$	20	2
	$\triangleleft$	0	0
	$\approx$	0	18

Tab. 3. (a) Subjective listening test results of the comparisons between I and the others.

	II	III
CSBUDH [8]	$\triangleright$	4
	$\triangleleft$	9
	$\approx$	7
CSBUDHAPC [9]	$\triangleright$	2
	$\triangleleft$	11
	$\approx$	7
CSBUBSDH [10]	$\triangleright$	3
	$\triangleleft$	9
	$\approx$	8
CSBUWTSI [15]	$\triangleright$	2
	$\triangleleft$	12
	$\approx$	6
Proposed method	$\triangleright$	0
	$\triangleleft$	17
	$\approx$	3

Tab. 3. (b) Subjective listening test results of the comparisons between II and III.

ITU-T Test Results:

The speech samples used in the listening test were taken from the TIMIT database. One hundred sentences were taken for evaluating the performance of conventional methods [8–10, 15] and the proposed method. Since the main application of the speech BE technique is in mobile communications, listening test samples are prepared so that they simulated speech transmitted over a cellular telephone network. The test samples were high-pass filtered with the mobile station input (MSIN) filter, which approximates the input response of a mobile station and the sound level of each test sample was normalized to 26 dB below overloading [35]. These pre-processed test samples were then down sampled to the 8-kHz sampling rate and used as NB

signal for conventional speech BWE methods [8–10, 15] and the proposed method.

The ACR test was conducted to evaluate the quality of the bandwidth extended speech signal generated by the conventional speech BE methods [8–10, 15] and the proposed method. The listeners were asked to evaluate the quality of the speech samples with the scale: 5 (excellent), 4 (good), 3 (fair), 2 (poor), 1 (bad). The test was conducted in a quiet environment using headphones. Twenty subjects participated in the test. MOS values for the conventional speech BE methods [8–10, 15] and the proposed method are presented in Tab. 4. A clearly improved reconstructed WB signal quality of the proposed method over the traditional methods is observed from Tab. 4.

4.2 Objective Test Results

The obtained WB speech quality is rated using the log spectral distortion (LSD) measure [8], [9] and the ITU-T WB perceptual evaluation of speech quality (WB-PESQ) measure [33]. The perceptual transparency is rated using the NB perceptual evaluation of speech quality (NB-PESQ) measure [25]. Bit error rate (BER) is used to evaluate the robustness of hidden data against quantization and channel noises. Speech BE using different data hiding techniques [11, 13, 27, 29, 30] are used NB-PESQ measure to evaluate the perceptual transparency.

Comparison of Original and Reconstructed HIB Speech

The perceptual similarity between original and reconstructed HIB signals is evaluated using the LSD measure and is given by

$$LSD = \frac{1}{2\pi}\int_{-\pi}^{\pi}\left(20\log_{10}\frac{g_{\mathrm{p}}}{\left|a_s\left(\mathrm{e}^{\mathrm{j}w}\right)\right|}-20\log_{10}\frac{\hat{g}_{\mathrm{p}}}{\left|\hat{a}_s\left(\mathrm{e}^{\mathrm{j}w}\right)\right|}\right)\mathrm{d}w \quad (11)$$

where g_{p} is the gain of the original HIB signal, $1/(a_s\mathrm{e}^{\mathrm{j}w})$ is the spectral envelope of the original HIB signal, $\hat{g}_{\mathrm{p}}$ is the gain of the reconstructed HIB signal and $1/(\hat{a}_s\mathrm{e}^{\mathrm{j}w})$ is the spectral envelope of the reconstructed HIB signal. Smaller LSD value indicates the best quality of the reconstructed HIB signal. Table 5 illustrates the results of the mean LSD values for the traditional techniques [8–10, 15] and the proposed technique under μ-law coding [8, 11, 12, 27–32]. A clearly improved reconstructed HIB signal quality and thus clearly improved reconstructed WB signal quality of the proposed technique over the traditional techniques is observed from Tab. 5. The mean LSD value for the proposed technique with AWGN channel model is 2.38.

Perceptual Transparency

NB-PESQ measure is used to rate PET by comparing $\mathbf{Y}_{\mathrm{nb}}(n)$ with $\mathbf{Y}^1{}_{\mathrm{nb}}(n)$. The NB-PESQ scale ranges from –0.5 for the worst PET up to 4.5 for the best PET. Table 6 illustrates the results of the averaged NB-PESQ scores for the traditional techniques [8–10, 15] and proposed technique.

Technique	MOS
CSBUDH [8]	2.63
CSBUDHAPC [9]	2.98
CSBUBSDH [10]	3.78
CSBUWTSI [15]	3.82
Proposed technique	4.63

Tab. 4. ACR listening test results.

Technique	Log spectral distortion
CSBUDH [8]	11.75
CSBUDHAPC [9]	9.57
CSBUBSDH [10]	4.98
CSBUWTSI [15]	4.87
Proposed technique	2.31

Tab. 5. Results of the log spectral distortion test.

Technique	NB-PESQ
CSBUDH [8]	2.99
CSBUDHAPC [9]	3.18
CSBUBSDH [10]	3.56
CSBUWTSI [15]	3.58
Proposed technique	4.28

Tab. 6. Comparative performance in terms of average NB-PESQ.

A clearly improved PET of the proposed technique over the traditional techniques is observed from Tab. 6.

Robustness of Hidden Information

The effect of noise corruption is considered now. AWGN is added to the CNB signal $\mathbf{Y}^1{}_{\mathrm{nb}}(n)$, with the signal to noise ratio (SNR) ranging from 15 to 35 dB [26]. The robustness of the proposed technique is evaluated using BER. The length of the PN code is 16. The smaller BER value indicates the better quality of the RWB signal. The obtained BER values as a function of SNR ranging from 15 to 35 dB are below 3.28×10^{-5} which confirms the better RWB signal quality.

The obtained BER value after applying μ-law coding to $\mathbf{Y}^1{}_{\mathrm{nb}}(n)$ is 1.32×10^{-5}, which confirms the better RWB signal quality.

WB Speech Quality

WB-PESQ measure [33] is used to evaluate the quality of reconstructed WB speech $\mathbf{Y}^1{}_{\mathrm{wb}}(n)$ by providing original WB speech $\mathbf{Y}_{\mathrm{wb}}(n)$ and reconstructed WB speech $\mathbf{Y}^1{}_{\mathrm{wb}}(n)$ as inputs. Here, the speech quality is rated using WB-PESQ measure by comparing $\mathbf{Y}_{\mathrm{wb}}(n)$ and $\mathbf{Y}^1{}_{\mathrm{wb}}(n)$. Table 7 illustrates the results of the averaged WB-PESQ scores for the traditional methods [8–10, 15] and the proposed method. Clear quality improvement of the proposed method over the traditional methods [8–10, 15] is observed from the average WB-PESQ scores as shown in Tab. 7.

Technique	WB-PESQ
CSBUDH [8]	2.60
CSBUDHAPC [9]	2.82
CSBUBSDH [10]	3.77
CSBUWTSI [15]	3.81
Proposed technique	4.17

Tab. 7. Comparative performance in terms of average WB-PESQ.

5. Conclusion

A novel speech BE algorithm using DWTFFTDH technique is proposed in this paper. The encoded spectral envelope parameters of HIB signal are hidden into detailed coefficients of NB signal. The hidden data is retrieved to produce a high-quality WB signal at the receiving end. The proposed technique proved to be a robust solution for BE of NB speech signals. Evaluation results confirm the excellent wideband performance of the proposed technique over the traditional speech BE techniques.

References

[1] JAX, P., VARY, P. Bandwidth extension of speech signals: A catalyst for the introduction of wideband speech coding? *IEEE Communications Magazine*, 2006, vol. 44, no. 5, p. 106–111. DOI: 10.1109/MCOM.2006.1637954

[2] JAX, P. Enhancement of bandlimited speech signals: Algorithms and theoretical bounds. *PhD Thesis*. RWTH Aachen University, Aachen, Germany, 2002.

[3] PRASAD, N., KISHORE KUMAR, T. Bandwidth extension of speech signals: A comprehensive review. *International Journal of Intelligent Systems and Applications*, 2016, vol. 8, no. 2, p. 45–52. DOI: 10.5815/ijisa.2016.02.06

[4] ABEL, J., FINGSCHEIDT, T. Artificial speech bandwidth extension using deep neural networks for wideband spectral envelope estimation. *IEEE Transactions on Audio, Speech, and Language Processing*, 2018, vol. 26, no. 1, p. 71–83. DOI: 10.1109/TASLP.2017.2761236

[5] LI, Y., KANG, S. Artificial bandwidth extension using deep neural network-based spectral envelope estimation and enhanced excitation estimation. *IET Signal Processing*, 2016, vol. 10, no. 4, p. 422–427, DOI: 10.1049/iet-spr.2015.0375

[6] WANG, Y., ZHAO, S., QU, D., et al., Speech bandwidth extension using recurrent temporal restricted Boltzmann machines. *IET Signal Processing Letters*, 2016, vol. 23, no. 12, p. 1877–1881. DOI: 10.1109/LSP.2016.2621053

[7] JAX, P., VARY, P. An upper bound on the quality of artificial bandwidth extension of narrowband speech signals. In *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing*. Orlando (USA), 2002, p. 237–240. DOI: 10.1109/ICASSP.2002.5743698

[8] CHEN, S., LEUNG, H. Artificial bandwidth extension of telephony speech by data hiding. In *Proceedings of the IEEE International. Symposium on Circuits and Systems*. Kobe (Japan), 2005, p. 3151–3154. DOI: 10.1109/ISCAS.2005.1465296

[9] CHEN, S., LEUNG, H. Speech bandwidth extension by data hiding and phonetic classification. In *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing*. Honolulu (Hawaii, USA), 2007, p. 593–596. DOI: 10.1109/ICASSP.2007.366982

[10] CHEN, Z., ZHAO, C., GENG, G., et al. An audio watermark based speech bandwidth extension method. *EURASIP Journal on Audio, Speech and Music Processing*, 2013, vol. 2013, no. 10, p. 1–8. DOI: 10.1186/1687-4722-2013-10

[11] SAGI, A., MALAH, D. Bandwidth extension of telephone speech aided by data embedding. *EURASIP Journal on Advances in Signal Processing*, 2007, vol. 2007, no. 1, p. 37–52. DOI: 10.1155/2007/64921

[12] CHEN, S., LEUNG, H., DING, H. Telephony speech enhancement by data hiding. *IEEE Transactions on Instrumentation and Measurement*, 2007, vol. 56, no. 1, p. 63–74. DOI: 10.1109/TIM.2006.887409

[13] GEISER, B., VARY, P. Speech bandwidth extension based on in-band transmission of higher frequencies. In *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing.* Vancouver (Canada), 2013, p. 7507–7511. DOI: 10.1109/ICASSP.2013.6639122

[14] GEISER, B., VARY, P. Backwards compatible wideband telephony in mobile networks: CELP watermarking and bandwidth extension. In *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing*. Honolulu (Hawaii, USA), 2007, p. 533–536. DOI: 10.1109/ICASSP.2007.366967

[15] BHATT, N., KOSTA, Y. A novel approach for artificial bandwidth extension of speech signals by LPC technique over proposed GSM FR NB coder using high band feature extraction and various extension of excitation methods. *International Journal of Speech Technology*, 2015, vol. 18, no. 1, p. 57–64. DOI: 10.1007/s10772-014-9249-1

[16] BHATT, N. Simulation and overall comparative evaluation of performance between different techniques for high band feature extraction based on artificial bandwidth extension of speech over proposed global system for mobile full rate narrow band coder. *International Journal of Speech Technology*, 2016, vol. 19, no. 4, p. 881–893. DOI: 10.1007/s10772-016-9378-9

[17] REKIK, S., GUERCHI, D., SELOUANI, S. A., et al. Speech steganography using wavelet and Fourier transforms. *EURASIP Journal on Audio, Speech, and Music Processing*, 2012, vol. 2012, no. 20, p. 1–14. DOI: 10.1186/1687-4722-2012-20

[18] PROAKIS, J. G. *Digital Communications*. New York: McGraw-Hill, 1989. ISBN: 978-0070509375

[19] HANZO, L. L., SOMERVILLE, F. C. A., WOODARD, J. P. *Voice Compression and Communications: Principles and Applications for Fixed and Wireless Channels*. New York: John Wiley & Sons, 2001. ISBN: 978-0-471-15039-8 (electronic)

[20] NILSSON, M., KLEIJN, W. B. Avoiding overestimation in bandwidth extension of telephony speech. In *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing*. Salt Lake City (UT, USA), 2001, vol. 2, p. 869–872. DOI: 10.1109/ICASSP.2001.941053

[21] BEZDEK, J. C. *Pattern Recognition with Fuzzy Objective Function Algorithms*. New York: Plenum, 1981. DOI: 10.1007/978-1-4757-0450-1

[22] JAX, P., VARY, P. On artificial bandwidth extension of telephone speech. *Signal Processing*, 2003, vol. 83, no. 8, p. 1707–1719. DOI: 10.1016/S0165-1684(03)00082-3

[23] EUROPEAN TELECOMMUNICATIONS STANDARDS INSTITUTE (ETSI) Standard. *Speech Processing, Transmission and Quality Aspects (STQ); Distributed speech recognition; Front-*

end feature extraction algorithm; Compression algorithms, ETSI ES 201 108 V1.1.2, April 2000.

[24] GAROFOLO, J. S., LAMEL, L. F., FISHER, W. M, et al. *Getting Started with the DARPA TIMIT CD-ROM: An Acoustic Phonetic Continuous Speech Database.* Gaithersburg (MD, USA): National Institute of Standards and Technology (NIST). ISBN: 1-58563-019-5

[25] INTERNATIONAL TELECOMMUNICATIONS UNION. *Perceptual Evaluation of Speech Quality (PESQ): An Objective Method for End to-end Speech Quality Assessment of Narrow-band Telephone Networks and Speech Codecs.* ITU-T Recommendation P.862, February 2001.

[26] KEISER, B. E., STRANGE, E. *Digital Telephony and Network Integration.* New York: Van Nostrand Reinhold, 1995. ISBN 978-1-4615-1787-0 (electronic)

[27] PRASAD, N., KISHORE KUMAR, T. Speech bandwidth extension aided by spectral magnitude data hiding. *Circuits, Systems, and Signal Processing*, 2017, vol. 36, no. 11, p. 4512–4540. DOI: 10.1007/s00034-017-0526-5

[28] CHEN, S., LEUNG, H. Concurrent data transmission through analog speech channel using data hiding. *IEEE Signal Processing Letters*, 2005, vol. 12, no. 8, p. 581–584. DOI: 10.1109/LSP.2005.851259

[29] PRASAD, N., KISHORE KUMAR, T. Bandwidth extension of narrowband speech using integer Wavelet transform. *IET Signal Processing*, 2017, vol. 11, no. 4, p. 437–445. DOI: 10.1049/iet-spr.2016.0453

[30] PRASAD, N., KISHORE KUMAR, T. Bandwidth extension of telephone speech using magnitude spectrum data hiding. *International Journal of Speech Technology*, 2017, vol. 20, no. 1, p. 151–162. DOI: 10.1007/s10772-016-9393-x

[31] CHEN, S., LEUNG, H. A bandwidth extension technique for signal transmission using chaotic data hiding. *Circuits, Systems, and Signal Processing*, 2008, vol. 27, no. 6, p. 893–913. DOI: 10.1007/s00034-008-9066-3

[32] GEISER, B., JAX, P., VARY, P. Artificial bandwidth extension of speech supported by watermark-transmitted side information. In *Proceedings of the 9th European Conference on Speech Communication and Technology*. Lisbon (Portugal), 2005, p. 1497–1500.

[33] INTERNATIONAL TELECOMMUNICATIONS UNION. *Wideband Extension to Recommendation P.862 for the Assessment of Wideband Telephone Networks and Speech Codecs.* ITU-T Recommendation P.862.2, November 2005

[34] INTERNATIONAL TELECOMMUNICATIONS UNION. *Methods for Subjective Determination of Transmission Quality*. ITU-T Recommendation P.800, August 1996.

[35] INTERNATIONAL TELECOMMUNICATIONS UNION. *Software Tools for Speech and Audio Coding Standardization.* ITU-T Rec. G.191, September 2005.

About the Authors...

Phaneendra KURADA was born on August 15, 1984 at Rajam in the state of Andhra Pradesh (AP). He obtained his M.Tech. in Communications and Radar Systems from Nagarjuna University, Guntur, India in 2008. He is currently a research scholar with the E.C.E department, JNTU Kakinada, India. His research interests include speech signal processing.

Sailaja MARUVADA was born on September 30, 1966 at Vizianagaram in the state of AP. She obtained her Ph.D. in ATM NETWORKS from JNTU Kakinada in 2009. She is currently serving as Professor, ECE Department, JNTU Kakinada. Her research interests include computer networks and adaptive signal processing

Koteswara Rao SANAGAPALLEA was born on March 5, 1952 at Nagula Varam, in the state of AP. He obtained his Ph.D. in Digital Statistical Signal Processing from AU Visakhapatnam, India in 1998. He is currently serving as Professor, ECE Department, KL University, India. His research interests include statistical signal processing and adaptive signal processing.