DWT-DCT-Based Data Hiding for Speech Bandwidth Extension

Sunil Kumar KODURI, T. Kishore KUMAR

Department of Electronics and Communication Engineering, National Institute of Technology Warangal, India

{sunil.veena10, kishorefr}@gmail.com

Submitted July 30, 2020 / Accepted March 20, 2021

Abstract. *The limited narrowband frequency range, about 300–3400Hz, used in telephone network channels results in less intelligible and poor-quality telephony speech. To address this drawback, a novel robust speech bandwidth extension using Discrete Wavelet Transform-Discrete Cosine Transform Based Data Hiding is proposed. In this technique, the missing speech information is embedded in the narrowband speech signal. The embedded missing speech information is recovered steadily at the receiver end to generate a wideband speech of considerably better quality. The robustness of the proposed method to quantization and channel noises is confirmed by the mean square error test. The enhancement in the quality of reconstructed wideband speech of the proposed method over conventional methods is reasserted by subjective listening and objective tests.*

Keywords

Telephone networks, speech bandwidth extension, telephony speech enhancement, speech quality, DWT-DCT-based data hiding

1. Introduction

Most of the traditional telephone networks allow only a narrow band (NAB) signal which is band-limited to 300 Hz–3400 Hz. Usually, human speech contains frequencies far beyond the NAB frequency range. Thus, the transmission of human speech through the networks leads to the muffled sound and poor-quality telephony speech. Wideband (WIB) speech transmission in the range of 50 Hz–7000 Hz would be desirable for better speech quality. To allow WIB speech services, the essential changes required within the network infrastructure are quite expensive and time-taking [1]. This happens to be a major hurdle for the transmission of high-quality speech in the telephone networks. Therefore, it is very important to enable WIB speech transmission using speech bandwidth extension (SBWEX) techniques to enhance the quality of speech [2].

Artificial bandwidth extension is one among various methods of SBWEX which improves the quality and intelligibility of telephony speech [2]. In this approach, the out-of-band information i.e. the frequencies below 300 Hz and above 3400 Hz are estimated from the NAB signal. The excitation signal and spectral envelope are estimated by most artificial bandwidth extension techniques which are used to regenerate the out-of-band signal. Different approaches for extension of excitation signal are presented in [2], [3]. Different techniques for estimating WIB spectral envelop are presented in [3–7]. However, traditional artificial bandwidth extension methods are suffering from reconstructing WIB speech with high quality under all conditions [8].

Compared to artificial bandwidth extension techniques, a WIB speech with high quality is reconstructed when the out-of-band information is transmitted by hiding it in the NAB signal using data hiding methods [1]. Several techniques for SBWEX using data hiding are proposed in the state-of-the-art literature. An SBWEX technique is proposed in [9] to embed the encoded spectral envelope parameters of the lost speech frequency components within the NAB signal. A better-quality WIB signal is reconstructed at the receiver end using the embedded information. A much better-quality WIB signal over [9] has been reconstructed in [10], where the spectral envelope parameters are efficiently encoded using phonetic classification. The pitch-scaled frequencies of the out-of-band signal are embedded into the unused frequencies of traditional telephony speech to enhance the quality of reconstructed WIB speech in [11]. The WIB signal of better quality is regenerated in [12–14] using joint source coding and data hiding technique. A high-quality WIB signal is reconstructed in [15], [16] using various frequency-domain data hiding techniques. The enhancement in the quality of reconstructed WIB speech is achieved by restoring the hidden audible components of the out-of-band signal [17]. The spectral envelope parameters of an out-of-band signal are embedded into the NAB signal bitstream to improve the quality of reconstructed WIB speech in [18]. The WIB signal of better quality is regenerated in [19] using a quantization based watermarking technique.

SBWEX techniques with data hiding are expected to deliver high quality composite narrowband (CNAB) alongside reconstructed WIB (RWIB) signals. Also, these techniques must be able to handle issues pertaining to quanti-

DOI: 10.13164/re.2021.0435

zation and channel noises. Nevertheless, most of the traditional techniques fail to provide high-quality CNAB and RWIB signals [9–19]. Also, they are less robust to channel and quantization noises. Thus, developing a novel SBWEX technique using data hiding is essential to improve the quality of CNAB and RWIB signals and more robust to channel and quantization noises.

An audio steganography technique presented in [20] used discrete Wavelet transform-fast Fourier transform-based data hiding technique to insert the secret message signal in detailed coefficients of a host speech signal without degrading the perceptual quality of the host signal. It was shown that this approach is producing a stego signal that is indistinguishable from the host signal, while being able to reliably recover the secret message signal at the receiver end without any degradation in quality. DCT is used here instead of FFT in the discrete Wavelet transform-fast Fourier transform-based data hiding technique.

A novel SBWEX algorithm using the discrete Wavelet transform-discrete Cosine transform-based data hiding technique [20] is proposed to embed the parameters of the lost speech frequency components within the detailed coefficients of the NAB signal. These hidden parameters are retrieved at the receiver side to produce a better-quality WIB signal by combining the missing speech signal that was transmitted through the detailed coefficients and the NAB signal. The proposed scheme uses the real missing speech information instead of its estimation which makes the reconstruction of the WIB speech more accurate compared to the conventional artificial bandwidth extension methods. Furthermore, the proposed method is compatible with conventional NAB terminal equipment, e.g., a plain ordinary telephone set. In other words, conventional NAB receivers can still access the NAB speech properly without additional hardware, while a customized receiver can extract the embedded information and provide WIB signal with much better quality.

The telephone network channel introduces channel and quantization noises. Techniques proposed in [9, 10, 33] for SBWEX are considering only the quantization noise and ignoring the channel noise. The quantization noise and channel noise effects are considered in this paper. The spread spectrum technique [21] is used in this work for retrieving the embedded information as it is claimed to be more robust against quantization and channel noises. In particular, each parameter to be inserted is spread by multiplying with a particular spreading sequence. The embedded information is then formed by adding the spread signals. Due to orthogonality among spreading sequences, the embedded information is retrieved reliably by using a correlator.

To minimize the interference caused by the other embedded components, spreading sequences with low cross-correlations are preferred. Hadamard codes have an optimal cross-correlation performance, i.e., orthogonal to each other, whereas the m-sequences, Gold-codes, and Kasami-codes are with varying cross-correlation properties [36], [37]. Because the Hadamard codes are well recognized by their optimal cross-correlation performance, they are employed in this work for minimizing the interference caused by the other embedded components.

The paper is ordered as follows. In Sec. 2, the DWT-DCT-based data hiding method for SBWEX is introduced. Section 3 deals with a novel SBWEX algorithm using the DWT-DCT-based data hiding technique. The subjective and objective analyses are discussed in Sec. 4. Finally, in Sec. 5, conclusions are summarized.

2. DWT-DCT Based Data Hiding Technique for SBWEX

To embed upper band signal $S_{\mathrm{eb}}(n)$ into narrowband signal $S_{\mathrm{nb}}(n)$, firstly, detailed and approximation coefficients are computed by applying DWT on $S_{\mathrm{nb}}(n)$ and then DCT coefficients are computed by applying DCT on detailed coefficients. Assume that the representation vector which represents $S_{\mathrm{eb}}(n)$ is $\mathbf{R} = [LSF_1, LSF_2, \ldots, LSF_{10}, \bar{G}_{\mathrm{r}}]$, where line spectral frequencies are denoted by LSF and gain is denoted by $\bar{G}_{\mathrm{r}}$.

Every parameter of $\mathbf{R}$ is spread by multiplying it with a particular pseudo-random noise sequence, i.e., $\check{\mathrm{D}}_i \cdot p^{\rightarrow i}$, $1 \le i \le K$ where K is the pseudo-random noise sequence $p^{\rightarrow i}$ length. The hidden data is then produced by adding all of these spreading vectors and is given by

$$V(j) = \sum_{i=1}^{K} \check{\mathrm{D}}_i p^i(j) \tag{1}$$

where the jth element of $p^{\rightarrow i}$ is represented by $p^i(j)$. The last 16 coefficients of DCT coefficients are replaced by $V(j)$ resulting in a CNAB signal spectrum [20]. To convert back the CNAB signal spectrum to time-domain representation, inverse DCT and then inverse DWT is applied on the CNAB signal spectrum. Thus, a CNAB signal $S^1_{\mathrm{nb}}(n)$ is produced so that it can be communicated to the receiver on a telephone network channel and the quantization and channel noises are introduced by telephone network channel. Assume that the received signal is represented by $\hat{S}^1_{\mathrm{nb}}(n)$, i.e., $\hat{S}^1_{\mathrm{nb}}(n) = S^1_{\mathrm{nb}}(n) + \bar{e}$, where $\bar{e}$ represents the combination of channel and quantization noises. The conventional phone terminal treats $\hat{S}^1_{\mathrm{nb}}(n)$ as an ordinary signal. The NAB signal quality is not noticeably degraded since there is a very small perceived difference between $S_{\mathrm{nb}}(n)$ and $S^1_{\mathrm{nb}}(n)$.

At the receiver, retrieval of the embedded data requires applying DWT on the CNAB signal and then applying DCT on detailed coefficients to obtain the DCT coefficients. The spread parameters are then obtained from the last 16 DCT coefficients and a correlator is used to despread these parameters. Assuming a particular $\underset{\wedge}{\mathrm{D}}_i$ to be retrieved is denoted by $\underset{\wedge}{\mathrm{D}}_{io}$ and then the correlation is given by

$$\underset{\frown}{D}_{io} = \frac{1}{K} \sum_{j=1}^{K} \underset{\cdot}{V}(j) p^{io}(j) \tag{2}$$

where $\underset{.}{V}(j)$ represent noisy $V(j)$ and is given by

$$\underset{.}{V}(j)=V(j)+\bar{e}(j). \tag{3}$$

Equation (3) is substituted into (2), we have

$$\begin{aligned}\underset{\frown}{D}_{io}&=\frac{1}{K}\sum_{j=1}^{K}\underset{.}{V}(j)\,p^{io}(j)\\&=\frac{1}{K}\sum_{j=1}^{K}p^{io}(j)\left(\sum_{j=1}^{K}\check{D}_i p^i(j)+\bar{e}(j)\right)\\&=\frac{1}{K}\sum_{j=1}^{K}p^{io}(j)\times\left(\check{D}_{io}p^{io}(j)+\sum_{i\neq io}\check{D}_i p^i(j)+\bar{e}(j)\right)\\&=\check{D}_{io}+\frac{1}{K}\sum_{j=1}^{K}\sum_{i\neq io}\check{D}_i p^i(j)\,p^{io}(j)+\frac{1}{K}\sum_{j=1}^{K}p^{io}(j)\,\bar{e}(j).\end{aligned}\tag{4}$$

The pseudo-random noise sequences are orthogonal. i.e., $\sum_{j=1}^{K}p^i(j)\,p^{io}(j)=0$, where $i \neq io$. Therefore,

$$\sum_{j=1}^{K}\sum_{i\neq io}\check{D}_{io}p^i(j)\,p^{io}(j)=\sum_{i\neq io}\check{D}_{io}\sum_{j=1}^{K}p^i(j)\,p^{io}(j)=0 . \tag{5}$$

Also, since there was no correlation between $p^{io}(j)$ and $\bar{e}(j)$, i.e.

$$\frac{1}{K}\sum_{j=1}^{K}p^{io}(j)\,\bar{e}(j)=0 \tag{6}$$

when $K\to\infty$. Equations (5) and (6) are substituted into (4), thus we have

$$\underset{\frown}{D}_{io} = \check{D}_{io}. \tag{7}$$

This illustrates that the parameters which represent $\hat{S}_{eb}(n)$ can be effectively recovered due to the use of the spread spectrum technique.

3. SBWEX Utilizing DWT-DCT Based Data Hiding Technique

3.1 Transmitter

The proposed transmitter is shown in Fig. 1. A speech signal designated as WIB $S_{wb}(n)$ is sampled at a frequency of 16 kHz. This signal is further fragmented to form a low-band signal using a low-pass filter and a high-band signal using a high-pass filter respectively. The low-pass filter extracts speech signal information that is present between 0 and 4 kHz and is designated as a low-band signal while the high-pass filter extracts speech information that is present between 4 kHz and 8 kHz designated as a high-band signal. The low-pass filter output is decimated by a factor of two in order to produce NAB signal $S_{nb}(n)$. The high-band signal is decimated to produce an upper-band (UPB) signal $S_{eb}(n)$. Therefore, 8 kHz is the sampling frequency of $S_{nb}(n)$ and $S_{eb}(n)$.

To imperceptibly embed $S_{eb}(n)$ into $S_{nb}(n)$, minimize the number of the parameters which represents $S_{eb}(n)$. To produce linear predictive coefficients, linear predictive analysis is carried out on $S_{eb}(n)$ [22]. A small variation in linear predictive coefficients results in substantial distortions when reconstructing $S_{eb}(n)$; hence linear predictive coefficients are modified into line spectral frequencies (LSFs) [22]. Also, the gain of $S_{eb}(n)$ has to be embedded to evade over-estimation [23]. Thus, the representation vector which represents $S_{eb}(n)$ is formed by combining LSFs and gain, $\mathbf{R} = [LSF_1, LSF_2, \ldots, LSF_{10}, \bar{G}_r]$. The parameters which represent $S_{eb}(n)$ are hidden using the DWT-DCT-based data hiding technique in the NAB signal. Thus, a CNAB signal $S^1{}_{nb}(n)$ is produced so that it can be communicated to the receiver on a telephone network channel.

The excitation parameters are not embedded to reduce the number of parameters of $S_{eb}(n)$ to be hidden. This is because the ear is not sensitive to the distortions of the excitation signal at above the NAB frequency range [24]. Thus, estimating the excitation of $S_{eb}(n)$ at the receiver from $S_{nb}(n)$ is well-suited for the reconstruction performance.

A synchronization sequence like 111…..11 is added after every frame of the CNAB signal to achieve frame synchronization [25] between the transmitter and receiver. The arrival of a new frame of the CNAB signal at the receiver is indicated by the reception of a synchronization sequence.

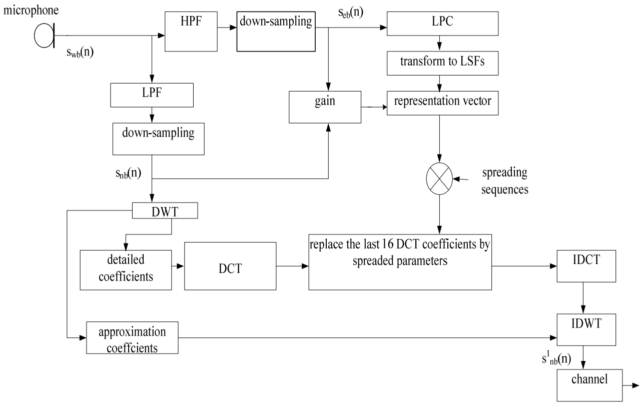

Fig. 1. Proposed transmitter.

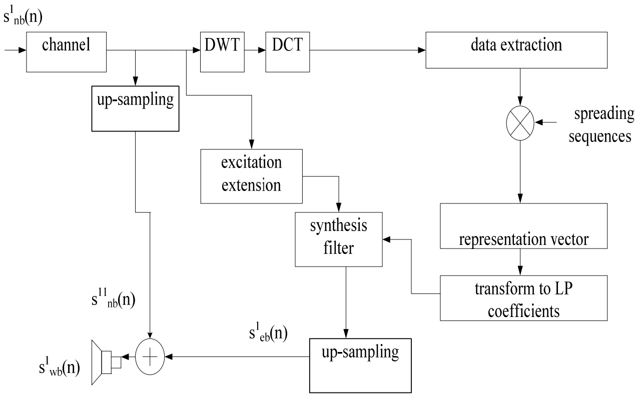

Fig. 2. Proposed receiver.

3.2 Receiver

The proposed receiver is shown in Fig. 2. The DWT-DCT-based data hiding technique properly recovers the representation vector and then linear predictive coefficients are obtained from LSFs. Meanwhile, NAB residual signal is obtained by inverse filtering $\hat{S}^{1}{}_{\mathrm{nb}}(n)$ using linear predictive coefficients of $\hat{S}^{1}{}_{\mathrm{nb}}(n)$ and then obtain the UPB excitation signal by extending the NAB residual signal. The UPB signal $\hat{S}_{\mathrm{eb}}(n)$ that was embedded is synthesized by exciting the synthesis filter described by the recovered linear predictive coefficients by a UPB excitation signal. The received CNAB and reconstructed UPB signals are sampled at an 8 kHz sampling rate. These signals are then interpolated by a factor of two. The interpolated CNAB $S^{11}{}_{\mathrm{nb}}(n)$ and UPB $S^{1}{}_{\mathrm{eb}}(n)$ signals are added up for reproducing a WIB signal $S^{1}{}_{\mathrm{wb}}(n)$ of good quality.

4. Experimental Results

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

The speech utterances used for the performance evaluations of traditional and proposed SBWEX techniques were obtained from the TIMIT database [26]. The evaluations were done by taking thirty different speech utterances which were spoken by thirty female and male speakers. The performance assessment of the methods was done by considering the subjective as well as objective measures. Each speech signal was split to form frames of 20 ms long and between frames, an overlap of 10 ms was maintained. Each frame was processed individually. Existing SBWEX algorithms like data hiding [9], phonetic classification [10], audio watermark [18], and steganographic WIB telephony [13] were compared with the proposed method. These are represented, respectively, by traditional SBWEX using data hiding (TSBWEDH), traditional SBWEX using phonetic classification (TSBWEPC), traditional SBWEX using bit-stream data hiding (TSBWEBD), and traditional SBWEX using watermark transmitted side information (TSBWEWS) respectively, in the analysis. Additive white Gaussian noise (AWGN) channel model is used in this paper.

4.1 Subjective Listening Test Results

The speech utterances used for the subjective listening tests were a subset of the eighteen hundred speech utterances. The subjective listening tests were made on two hundred speech utterances drawn randomly from the eighteen hundred speech utterances. The obtained speech quality of the proposed method and conventional SBWEX methods [9, 10, 13, 18] is assessed using an absolute category rating (ACR) listening test [34] recommended by the International Telecommunications Union (ITU-T). The perceptual transparency is assessed with the mean opinion score (MOS) test [9], [10]. The subjective comparison between WIB, CNAB, NAB, and RWIB signals is also employed [15]. The obtained speech quality of the proposed method in comparison with the outputs of the conventional methods [9, 10, 13, 18] was assessed using comparison category rating (CCR) listening test [35] recommended by the International Telecommunications Union (ITU-T). Each person is made to hear the speech utterances through headphones in a silent chamber. An evaluation was done using a predefined scale by examining participant's views on speech sounds. Thirty normal listeners (15 females and 15 males) between the age of 20 and 35 years have participated in these tests.

ITU-T Test Results

The speech samples used in the listening test were taken from the TIMIT database. Two hundred sentences were taken for evaluating the performance of conventional methods [9, 10, 13, 18] and the proposed method. Since the main application of the SBWEX 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 NAB signal for conventional SBWEX methods [9, 10, 13, 18] and the proposed method.

The ACR test was conducted to evaluate the quality of the bandwidth extended speech signal generated by the conventional SBWEX methods [9, 10, 13, 18] 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. Thirty subjects participated in the test. MOS values for the conventional SBWEX methods [9, 10, 13, 18] and the proposed method are presented in Tab. 1. A clearly improved reconstructed WIB signal quality of the proposed method over the traditional methods is observed from Tab. 1.

Pairwise Comparisons

The speech samples used in the CCR listening test were taken from the TIMIT database. Two hundred sentences were taken for evaluating the performance of conventional methods [9, 10, 13, 18] and the proposed method. CCR listening test compares two speech samples and provides information on which sample is better in terms of quality based on the comparison mean opinion score (CMOS) given in Tab. 2. Subjects participating in a CCR listening test compared pairs of speech samples from the outputs of the conventional methods [9, 10, 13, 18] and proposed method. The second signal of the pair is rated compared to the first signal. The results of the listening test are presented in Tab. 3 in terms of CMOS and respective 95% confidence interval (CI95) for each of the CCR conditions. A clearly improved reconstructed wideband signal quality of the proposed method over the traditional methods [9, 10, 13, 18] is observed from Tab. 3.

Technique	MOS
TSBWEDH [9]	2.45
TSBWEPC [10]	2.76
TSBWEBD [18]	3.54
TSBWEWS [13]	3.61
Proposed technique	4.57

Tab. 1. ACR listening test results.

Score	Rating of second signal compared to that of the first signal
3	Much better
2	Better
1	Slightly better
0	About the same
–1	Slightly worse
–2	Worse
–3	Much worse

Tab. 2. Comparison mean opinion score (CMOS).

CCR Condition	CMOS	CI95
TSBWED VS. TSBWEPC	0.63	[0.39; 0.87]
TSBWEDH VS.TSBWEBD	1.31	[1.18; 1.44]
TSBWEDH VS.TSBWEWS	1.37	[1.22; 1.51]
TSBWEDH VS. Proposed method	2.45	[2.33; 2.56]
TSBWEPC VS. TSBWEBD	0.80	[0.55; 1.05]
TSBWEPC VS. TSBWEWS	1.03	[0.86; 1.20]
TSBWEPC VS. Proposed method	1.93	[1.78; 2.18]
TSBWEBD VS. TSBWEWS	0.81	[0.60; 1.03]
TSBWEBD VS. Proposed method	1.78	[1.65; 1.91]
TSBWEWS VS. Proposed method	1.67	[1.52; 1.81]

Tab. 3. CCR listening test results.

Score	Instruction
1	NAB and CNAB signals sounds different
2	Observable difference between NAB and CNAB signals
3	Minute difference between NAB and CNAB signals
4	NAB and CNAB signals sounds alike

Tab. 4. MOS.

Technique	Mean opinion score
TSBWEDH [9]	2.89
TSBWEPC [10]	3.07
TSBWEBD [18]	3.18
TSBWEWS [13]	3.54
Proposed method	3.97

Tab. 5. Results of the MOS.

Perceptual Transparency

The information should be transparently hidden by the proposed method. That is the CNAB and NAB signals should be subjectively indistinguishable. High perceptual transparency means low noticeable NAB signal degradation. There should be high perceptual transparency even after embedding the spectral envelope parameters of the UPB signal into detailed Wavelet coefficients of the NAB signal. The perceptual transparency was assessed with the MOS test. Listener's on comparing CNAB and NAB signals come out with a decision in terms of MOS as given in Tab. 4. The average MOS values of traditional [9, 10, 13, 18] and the proposed techniques are given in Tab. 5. The proposed technique gives a MOS value of 3.97 which indicates that the proposed technique has excellent perceptual transparency over the traditional techniques [9, 10, 13, 18]. The proposed technique gives a MOS value of 3.97 which was almost near the standard MOS value of 4 which indicates that CNAB and NAB signals were more or less identical.

Subjective Comparisons between WIB, NAB, CNAB and RWIB Speech Samples

A listening test was done for comparing performances between the proposed and conventional methods [9, 10, 13, 18]. Here, the WIB signal, NAB signal, CNAB signal, and RWIB signal were labeled I, II, III, and IV respectively. Participants are asked to do a pairwise comparison between the samples to tell whether the first sample was superior to, inferior than or equal to the second. The responses after comparing I, II, and III with the other signals respectively are tabulated in Tab. 6 (a), (b), and (c).

The number of participants with a specific preference is indicated by Arabic numerals in the table. It is observed that the WIB signal is superior to NAB and CNAB signals of traditional [9, 10, 13, 18] and the proposed methods from Table 6(a). Also, we observe that RWIB signal quality is far superior using the proposed method over traditional methods [9, 10, 13, 18] from Tab. 6(a). Thus, the speech quality was enhanced by the proposed technique. Compared to traditional methods [9, 10, 13, 18], it is observed that the RWIB signal of the proposed method is superior to that of the NAB signal, as may be seen from Tab. 6(b). Compared to conventional methods [9, 10, 13, 18], it is observed that RWIB speech of the proposed technique is better than CNAB speech from Tab. 6(c).

4.2 Objective Quality Evaluations

The database which was used in subjective listening tests was also used in evaluating objective measures. The perceptual transparency was assessed with the narrowband-perceptual objective listening quality assessment (NAB-POLQA) measures [27]. RWIB speech quality was evaluated with the log spectral distortion (LSD) [9], [10] and wideband-perceptual objective listening quality assessment (WIB-POLQA) measures [27]. The robustness of hidden data against quantization and channel noises was evaluated with the help of a mean square error (MSE) measure [17, 31, 32].

Technique	I	II	III	IV
TSBWEDH [9]	>	30	30	14
	<	0	0	0
	≈	0	0	16
TSBWEPC [10]	>	30	30	12
	<	0	0	0
	≈	0	0	18
TSBWEBD [18]	>	30	30	11
	<	0	0	0
	≈	0	0	19
TSBWEWS [13]	>	30	30	9
	<	0	0	0
	≈	0	0	21
Proposed method	>	30	30	1
	<	0	0	0
	≈	0	0	29

Tab. 6. (a) Subjective comparison test results between I, II, III, and IV.

Technique	II	III	IV
TSBWEDH [9]	>	8	3
	<	4	18
	≈	18	9
TSBWEPC [10]	>	8	1
	<	2	19
	≈	20	10
TSBWEBD [18]	>	5	2
	<	3	20
	≈	22	8
TSBWEWS [13]	>	5	2
	<	2	22
	≈	23	6
Proposed method	>	1	0
	<	0	28
	≈	29	2

Tab. 6. (b) Subjective comparison test results between II, III, and IV.

Technique	III	IV
TSBWEDH [9]	>	6
	<	18
	≈	6
TSBWEPC [10]	>	5
	<	17
	≈	8
TSBWEBD [18]	>	3
	<	18
	≈	9
TSBWEWS [13]	>	4
	<	20
	≈	6
Proposed method	>	0
	<	29
	≈	1

Tab. 6. (c) Subjective comparison results between III and IV.

Perceptual Transparency

The evaluation of perceptual transparency is done by providing NAB and CNAB signals as inputs and comparing them to rate speech quality. The NAB-POLQA value will range between 1 and 5, where the higher the value, the more superior the quality. The average NAB- POLQA values of conventional [9, 10, 13, 18] and proposed methods are tabulated in Tab. 7. The proposed technique gives a NAB-POLQA value of 4.12 which indicates that the proposed technique has excellent perceptual transparency over traditional techniques [9, 10, 13, 18], which was already confirmed by subjective listening tests.

WIB Speech Quality

The evaluation of the quality of RWIB speech is done by giving WIBA and RWIB signals as inputs and comparing them in order to rate speech quality. The average WIB-POLQA values of the conventional [9, 10, 13, 18] and proposed methods are shown in Tab. 8. A WIB-POLQA value of 4.24 confirms that the RWIB signal quality that was obtained by the proposed technique is excellent compared to traditional techniques [9, 10, 13, 18], which was already confirmed by subjective listening tests on a set of participants. Thus, the speech quality was improved from using the proposed technique.

RWIB Speech Quality

The quality of RWIB speech is evaluated using LSD measure and is calculated using the formula

$$LSD = \frac{1}{2\pi}\int_{-\pi}^{\pi}\left(20\log_{10}\frac{g_{\mathrm{p}}}{a_{\mathrm{s}}\left(\mathrm{e}^{\mathrm{j}w}\right)} - 20\log_{10}\frac{\hat{g}_{\mathrm{p}}}{\left|\hat{a}_{\mathrm{s}}\left(\mathrm{e}^{\mathrm{j}w}\right)\right|}\right)^2 \mathrm{d}w \tag{8}$$

where g_{p} is the gain of UPB signal, $(a_{\mathrm{s}}(\mathrm{e}^{\mathrm{j}w}))^{-1}$ is the spectral envelope of UPB signal, $\hat{g}_{\mathrm{p}}$ is the gain of the reconstructed UPB signal and $(\hat{a}_{\mathrm{s}}(\mathrm{e}^{\mathrm{j}w}))^{-1}$ is the spectral envelope of the reconstructed UPB respectively. An RWIB signal with the least value of LSD is said to be of good quality. The resultant LSD for conventional [9, 10, 13, 18] and proposed techniques with a μ-law coding are presented in Tab. 9 and

Technique	NAB- POLQA
TSBWEDH [9]	2.54
TSBWEPC [10]	2.99
TSBWEBD [18]	3.21
TSBWEWS [13]	3.34
Proposed method	4.12

Tab. 7. Results of the NAB-POLQA.

Technique	WIB- POLQA
TSBWEDH [9]	2.08
TSBWEPC [10]	2.45
TSBWEBD [18]	3.13
TSBWEWS [13]	3.34
Proposed method	4.24

Tab. 8. Results of the WIB-POLQA.

Technique	Log Spectral Distortion
TSBWEDH [9]	12.83
TSBWEPC [10]	10.69
TSBWEBD [18]	6.07
TSBWEWS [13]	5.94
Proposed method	2.23

Tab. 9. Results of the LSD.

it was evident that the RWIB signal quality of the proposed technique was far superior to the signal quality generated using conventional techniques [9, 10, 13, 18]. In addition, the proposed technique offers LSD of 2.23 indicating that RWIB speech of the proposed technique and original WIB speech qualities are almost equal. Good RWIB signal performance of the proposed technique which was already found in the subjective tests is now supported by these LSD values also. The proposed technique offers LSD of 2.41 with the AWGN channel model.

Robustness of Hidden Information

AWGN with a signal to noise ratio ranges between 15 and 35 dB [28] is added to the CNAB signal. The evaluation of the robustness of the proposed technique is done by utilizing MSE and is calculated using the formula

$$MSE = \frac{1}{N}\sum_{n=0}^{N-1}\left(S_{\text{wb}}^{1}\left(n\right) - S_{\text{wb}}\left(n\right)\right)^{2} \tag{9}$$

where the RWIB signal is represented by $S^1{}_{\text{wb}}(n)$ and the original WIB signal is represented by $S_{\text{wb}}(n)$. The spreading sequence length is 16. An RWIB signal with a small value of MSE is said to be of good quality. The proposed technique gives MSE values as a function of the signal to noise ratio ranges between 15 and 35 dB, which are below 7.88×10^{-4} indicating that the RWIB signal quality obtained by the proposed technique is excellent. The proposed technique gives an MSE value after adding quantization noise (μ-law) to $S^1{}_{\text{nb}}(n)$ is 6.07×10^{-4} which indicates RWIB signal quality that was obtained by the proposed technique is excellent.

5. Conclusion

In this paper, SBWEX utilizing the DWT-DCT-based data hiding technique has been proposed. The parameters of the UPB signal are embedded within the NAB signal. The embedded information is used to reconstruct the WIB signal of good quality at the receiver end. The robustness of the proposed method is confirmed by the mean square error test. The RWIB signal quality was enhanced by the proposed technique over conventional techniques and it was evident through subjective listening and objective tests.

Acknowledgments

The authors sincerely thank the Department of Science and Technology-Science and Engineering Research Board (DST-SERB) New Delhi, India for providing funds for a research grant (SERB Project File No: EMR/2016/006970).

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] LING, Z.-H., AI, Y., GU, Y., et al. Waveform modelling and generation using hierarchical recurrent neural networks for speech bandwidth extension. *IEEE/ACM Transaction on Audio, Speech, and Language Processing*, 2018, vol. 26, no. 5, p. 883–894. DOI: 10.1109/TASLP.2018.2798811

[5] LEE, B.-K., NOH, K., CHANG, J.-H., et al. Sequential deep neural networks ensemble for speech bandwidth extension. *IEEE Access*, 2018, vol. 6, p. 27039–27047. DOI: 10.1109/ACCESS.2018.2833890

[6] ABEL, J., FINGSCHEIDT, T. A DNN regression approach to speech enhancement by artificial bandwidth extension. In *Proceedings of the IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA)*. New Paltz (NY, USA), 2017, p. 219–223. DOI: 10.1109/WASPAA.2017.8170027

[7] WANG, Y., ZHAO, S., QU, D., et al. Using conditional restricted Boltzmann machines for spectral envelope modelling in speech bandwidth extension. In *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing*. Shanghai (China), 2016, p. 5930–5934. DOI: 10.1109/ICASSP.2016.7472815

[8] 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 (FL, USA), 2002, p. 237–240. DOI: 10.1109/ICASSP.2002.5743698

[9] 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

[10] 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

[11] 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

[12] 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

[13] 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

[14] 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

[15] 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

[16] KODURI, S. K., KUMAR, T. K. Speech bandwidth extension aided by hybrid model transform domain data hiding. In *Proceedings of the IEEE International Symposium on Circuits and Systems (ISCAS)*. Sapporo (Japan), 2019, p. 1–5. DOI: 10.1109/ISCAS.2019.8702323

[17] 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

[18] 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

[19] 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

[20] 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

[21] HASSAN, A. A., HERSHEY, J. E., SAULNIER, G. J. *Perspectives in Spread Spectrum.* Boston/Dordrecht/London: Kluwer Academic Publishers, 1998. ISBN: 978-0-792-38265-2

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

[23] 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

[24] 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

[25] 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.

[26] 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), 1993. ISBN: 1-58563-019-5

[27] INTERNATIONAL TELECOMMUNICATIONS UNION. Perceptual objective listening quality assessment: An advanced objective perceptual method for end-to-end listening speech quality evaluation of fixed, mobile, and IP-based networks and speech codecs covering narrowband, wideband, and super-wideband signals. *ITU-T Recommendation P.863*, January 2011.

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

[29] PRASAD, N., KUMAR, T. K. 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. 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

[32] 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

[33] 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.

[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 Recommendation G.191*, September 2005.

[36] DINAN, E. H. JABBARI, B. Spreading codes for direct sequence CDMA and wideband CDMA cellular networks. *IEEE Communications Magazine*, 1998, vol. 36, no. 9, p. 48–54. DOI: 10.1109/35.714616

[37] GOLDSMITH, A. *Wireless Communications*. New York (USA): Cambridge University Press, 2005**.** ISBN: 978-0521837163

About the Authors...

Sunil Kumar KODURI received his M. Tech in Digital Communication from Kakatiya University, Warangal, Telangana, India. Presently he is a JRF (Junior Research Fellow) Full Time Ph.D. Scholar at NIT Warangal, India. His research interests include speech bandwidth extension and speech enhancement.

T. Kishore KUMAR received his Ph.D. degree in the area of Signal Processing in 2004. He is presently a Professor in the Department of E.C.E, NIT Warangal, India. He published 25 research papers in various international journals. He was the former Head of the Dept. of ECE, N.I.T Warangal, India, and also Visiting Professor to AIT Bangkok sponsored by MHRD, Govt. of India. He has R&D projects worth about 1 Crore sponsored by SERB, DRDO, and MHRD, Govt. of India.