Hybrid Transform Based Speech Band Width Enhancement Using Data Hiding

Sunil Kumar Koduri*, Kishore Kumar T

Department of ECE, NIT Warangal, Telangana 506004, India

Corresponding Author Email: sunilk21@student.nitw.ac.in

ABSTRACT

The limited narrow band (LNB) speech signal spread in the range of 300 to 3400Hz used in public switched telephone networks results in poor-quality telephony speech. Bandwidth extension techniques are performed to expand the frequency range from LNB speech to a clear wideband (CWB) speech signal range of 50Hz-7000Hz over existing public telephone networks. In this paper, a novel robust speech bandwidth extension algorithm by Discrete Wavelet Transform- Discrete Cosine Transform- Based Data Hiding (DWT-DCT-DH) Hybrid transform model was used to spread the out-of-band (3400Hz to 7000Hz) speech frequencies over the LNB speech. In this proposed technique the out-of-band speech frequencies are embedded in LNB speech and imperceptibly spread over the network. These Embedded out-of-band speech frequencies are recovered steadily at the receiver end to generate a restored CWB telephony speech of considerably better quality. The proposed technique simulation results show more intelligible and better-quality telephony speech generated compared to the other bandwidth extension techniques.

1. INTRODUCTION

Human speech may have frequencies more than conventional telephone networks operating at 300-3400Hz. When a human speech signal is transmitted through the telephone network leads to losing information due to the limited narrow bandwidth of the telephone network. This results in significantly low quality and lucidity of speech transmission. This problem can be solved by using a comprehensive wideband transmission whose spectrum ranges from 50-7000Hz. As a traditional telephone network installed to operate at 300-3400 Hz it is not feasible to work at a wideband spectrum. Hence use of a wideband spectrum needs to install a new network which turns out to be very expensive and time-consuming [1]. Therefore, other techniques are to be adopted for the improvement of speech quality. To use the existing infrastructure and to improve the quality, bandwidth extension (BE) techniques [2] can be implemented.

In the artificial bandwidth extension techniques (ABETs), a clear wideband (CWB) signal is generated by predicting the lost portion of the signal from the limited narrowband (LNB) speech alone. Most of the ABETs proposed in the literature are based on the source-filter (SF) model of speech production. The SF model divides the BE technique into excitation signal extension and CWB speech signal spectral envelope estimation. Many methods for excitation enhancement are found [3]. Many methods for CWB spectral envelope approximation are illustrated [3-7]. Even though ABETs have many advantages there are a few limitations like their performance is limited. Thus, it will not able to reconstruct high-quality CWB [8].

The quality of CWB can be further improved when some supplementary information from out-of-band is communicated by hiding with the LNB signal [1]. At the receiver, when the embedded information is extracted, a CWB signal with a much better speech quality can be reconstructed by combining the out-of-band signal that was transmitted by hiding within the LNB signal and the LNB signal. The speech bandwidth extension using data hiding approaches uses the real out-of-band information instead of its estimation which makes the reconstruction of the CWB speech more accurate compared to the conventional ABETs. In order to ensure the desired backward compatibility with respect to the existing telephone networks, data hiding methods would be used to hide the out-of-band information in the LNB. Several methods have been developed for this problem as a result of research efforts. A BE technique has been stated [9], accordingly that, the encoded spectral envelop parameters (SEPs) of the missing spectral frequencies (MSFs) in the range of 4 to 8 kHz and known as missing band (MB) signal, is concealed into the LNB to generate a composite limited narrowband (CLNB) speech. A Technique for producing high-quality CWB over the above method was reported [10], in which the MB signal was encoded with high efficiency through phonetic classification. A BE approach was reported [11], accordingly that, SEPs of MB signal was concealed into the least significant bits of LNB. BE based on the quantization-based data hiding technique has been stated [12]. The noticeable components of the MB signal are implanted within the hidden channel [13]. The concealed data can be consistently reproduced at the destination. The audio signal of better quality is regenerated [14] using pitch-scaling. Enhancing the bandwidth using combined coding and data hiding (CCDH) is introduced [15]. A High-quality CWB signal is reproduced [16, 17] based on CCDH Method.

The existing methodologies [8-17] failed to deliver high-quality CLNB and reconstructed CWB (RCWB) signals along with vigor towards quantization and channel (QACN) noises. Speech BE using data hiding techniques could deliver high-

quality CLNB and RCWB signals and also be able to offer vigor towards QACN. Therefore, innovative speech BE algorithms with data hiding methods development is vital for the enhancement of the quality of CLNB and RCWB signals and also able to effectively manage QACN.

A Discrete Wavelet Transform-Fast Fourier Transform - based data hiding (DWT-FFT-DH) method is reported in the study of Rekik et al. [18] for embedding the secrete signal in detailed parameters of the host speech signal in DWT coefficients of the cover signal without lowering the cover signal quality. It is observed that the DWT-FFT-DH method could produce a stego signal which is indistinguishable from the cover signal and also be able to restore the secrete signal without lowering the quality. FFT is replaced with discrete Cosine transform in the DWT-FFT-DH technique. A novel robust BE algorithm using DWT-DCT-DH is proposed to embed the out-of-band spectral frequencies within LNB signal. These embedded spectral frequencies are recovered steadily at the receiver side to produce a better-quality CWB signal.

The effect of noises like QACN is discussed in this work. The effect of quantization noise is reported [9, 10]. Though, the influence of the channel noise was not assessed [9, 10]. The current development uses a code division multiple access (CDMA) approach for reproducing the concealed information which is appealed as robust towards noises like QACN. Especially, every information bit entrenched within the LNB signal is spread out as the product by a definite spreading sequence (SPSE). Then, the spread signals were summed to create the concealed information. The concealed information could be consistently retrieved since the correlation among the SPSEs is low.

The rest of the paper is segmented into four sections. Section 2 introduced the DWT-DCT-DH method. An innovative BE by DWT-DCT-DH technique is presented in Section 3. The experimental analyses are deliberated in section 4. Lastly, section 5 presents conclusions.

2. DWT-DCT BASED DATA HIDING

Consider an MB signal $A_{mb}(n)$ which is to be hidden in the LNB signal $A_{lnb}(n)$. At first, DWT is performed on $A_{lnb}(n)$ to calculate the detailed coefficients and then DCT is applied to detailed coefficients to compute DCT coefficients. Consider that $A_{mb}(n)$ is encoded into a sequence of data bits, i.e., $D_b \varepsilon \{-1,1\}, b = 0,1, \ldots, B-1$, where B represents the total number of bits.

Every information bit entrenched within the LNB signal is spread out as the product with a definite SPSE, i.e., $D_b p^b$. The SPSE p^b length is K. Then, the spread signals were added up to create the concealed data and are given by:

$$V(m) = \sum_{b=1}^{B-1} E_b p^b(m) \qquad (1)$$

The concealed data $V(m)$ is embedded into the last 8 DCT coefficients [18] which results in a CLNB signal spectrum. The time-domain CLNB signal is obtained by applying an inverse DCT and then inverse DWT on the CLNB signal spectrum. The obtained CLNB signal $A_{lnb}^1(n)$ is transmitted through the telephone network channel (TNC) to the destination. The channel injects noises like QACN. Consider $\hat{A}_{lnb}^1(n)$ represent the received signal, i.e., $\hat{A}_{lnb}^1(n) =$

$A_{lnb}^1(n) + ë$. The mixture of QACN is represented by ë. The traditional phone terminal treats $\hat{A}_{lnb}^1(n)$ as a normal signal. $A_{nb}(n)$ quality is not significantly tarnished as the observed changes among $A_{lnb}(n)$ and $A_{lnb}^1(n)$ are very low.

Extraction of the concealed data $\hat{A}_{mb}(n)$ needs a receiver which can calculate the DCT coefficients by performing DCT on $\hat{A}_{lnb}^1(n)$. The concealed data is then extracted from the last 8 DCT coefficients [18]. The data bits are decoded by employing a multiuser detector [19]. i.e.,

$$\check{D}_b = sign\left(\sum_{m=1}^{M-1} \mathbb{V}_m p_m{}^b\right) \qquad (2)$$

In a noise-free environment, $\mathbb{V}_m = V_m$ substitute it into (2), we have

$$\check{D}_b = sign\left(\sum_{m=1}^{M-1} V_m p_m{}^b\right)$$
$$= sign\left(\sum_{m=1}^{M-1}\left(D_m p_m{}^b p_m{}^b\right.\right.$$
$$\left.\left. + \sum_{g=0,g\neq b}^{B-1} D_g p_m{}^g p_m{}^b\right)\right) \qquad (3)$$
$$= sign\left(MD_m\right.$$
$$\left. + \sum_{g=0,g\neq b}^{B-1} D_g \sum_{m=0}^{M-1} p_m{}^g p_m{}^b\right)$$

The SPSEs are orthogonal, i.e., $\sum_{m=0}^{M-1} p_m{}^g p_m{}^b = 0$, where $g \neq b$.

Therefore

$$\sum_{g=0,g\neq b}^{B-1} D_g \sum_{m=0}^{M-1} p_m{}^g p_m{}^b = 0 \qquad (4)$$

This concludes that the parameters of $A_{mb}(n)$ could be efficiently retrieved by using the CDMA approach.

3. BE UTILIZING DWT-DCT-DH TECHNIQUE

3.1 Transmitter

The transmitter is depicted in Figure 1. Primarily, CWB speech $A_{cwb}(n)$ is sampled at 16000 Hz and is passed through a low pass filter (LPF) and high pass filter (HPF) to generate an LNB signal and an MB signal. LNB signals have speech information ranging from 0 to 4000 Hz and MB have speech information in the range of 4000 to 8000 Hz. Then LNB signal $A_{lnb}(n)$ is generated by decimating the output of LPF. The HPF output is then decimated to generate an upper band (UB) signal $A_{mb}(n)$.

Reduce the parameters which characterize $A_{mb}(n)$ to insert UB signal imperceptibly in LNB signal and linear predictive (LP) analysis [20] is used here to fulfill. The LP coefficients are evaluated by applying the Levinson-Durbin method [20]

on $A_{mb}(n)$ and later these are transformed into line spectral frequencies (LSFs) as there is a slight change in coefficients leading to the distortions while reproducing. Furthermore, the gain of $A_{mb}(n)$ needs to be hidden in order to evade over-approximation [21]. Hence, the gain is assessed as $g_r = \frac{g_{mb}}{g_{lb}}$ and pooled with LSFs to generate a representation vector, that is, $D = [lsf_1 lsf_2 \dots lsf_r, g_r]$. Quantize D to the nearby entry of a vector quantization (VQ) codebook that is produced by the fuzzy c-means (FCM) algorithm [22]. The binary equivalent of entry index, i.e., $D_0, D_{1,\dots} D_{B-1}$ is concealed into the LNB signal based on the DWT-DCT-DH technique which results in a CLB signal and it is communicated through TNC to the destination.

The excitation signal (ES) has many parameters which are not implanted to lessen parameters that could be implanted since exceeding 3.4 kHz the human ear remains insensitive to alterations of ES [23]. Thus, the prediction of the UB excitation signal from $A_{lnb}(n)$ at the destination assurances the reproduction performance.

A synchronization sequence (SYSE) like 1111….111 is introduced subsequently every frame of $A_{lnb}^1(n)$ to achieve frame synchronization [24] among the sender and receiver. The appearance of a novel frame is designated by the receiving of an SYSE.

3.2 Receiver

The receiver is shown in Figure 2. Recuperate the entry index appropriately using the DWT-DCT-DH method and then the VQ codebook is used for regaining the corresponding quantized LSFs accurately. The recovered LSFs are then used for generating LP coefficients. Meanwhile, inverse filtering is performed on $A_{lnb}^1(n)$ which produces LNB residual signal. Then the residual signal is extended which results in UB excitation. The synthesis filter defined by the recovered LP coefficients is excited through UB excitation in order to reconstruct $A_{mb}^1(n)$. At this instant, the sampling rate of $A_{lnb}^1(n)$ and $A_{mb}^1(n)$ is 8000 Hz and then these signals are interpolated. $A_{mb}^{11}(n)$ represents the interpolated $A_{mb}^1(n)$. The interpolated CLNB $A_{lnb}^{11}(n)$ and recovered $A_{mb}^1(n)$ signals are added in order to regenerate the high-quality CWB signal $A_{cwb}^1(n)$.
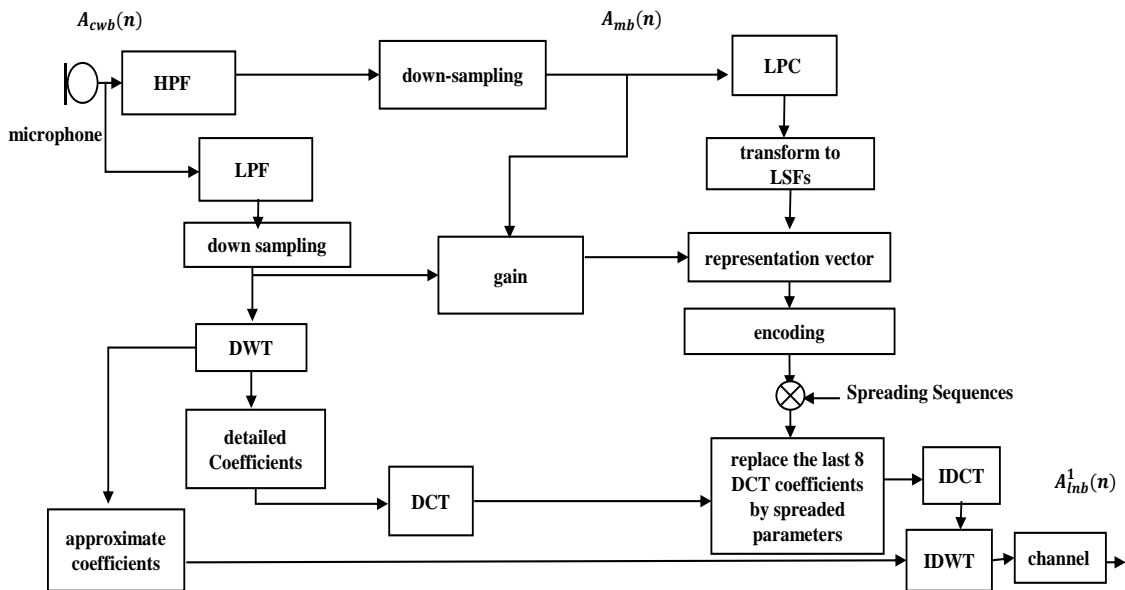

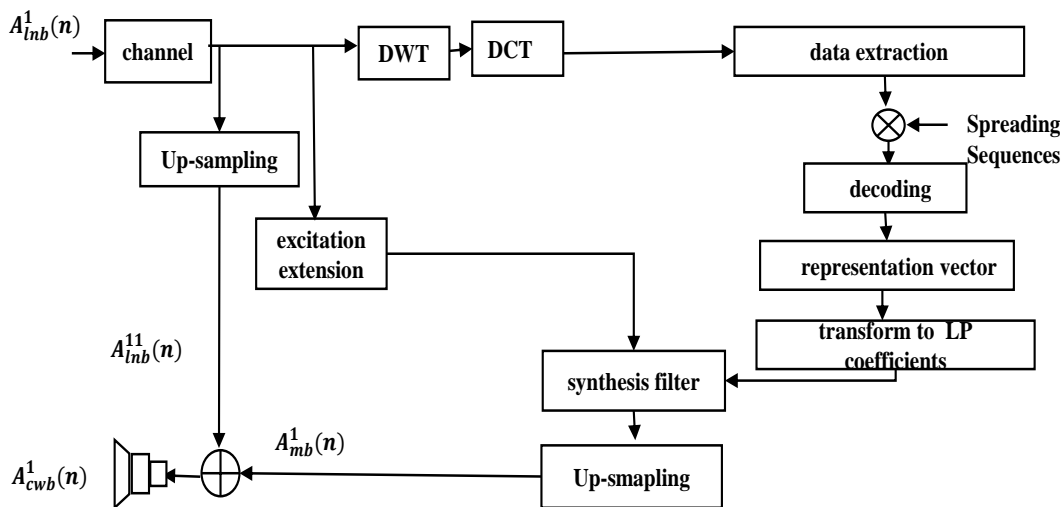**Figure 1.** Proposed transmitter

Figure 2. Proposed receiver

4. EXPERIMENTAL RESULTS

To assess the performance of the proposed method, fifty sentences spoken by 40 talkers were collected from the TIMIT database [25]. The LNB signal is decomposed into frames of 20-ms duration with an overlap of 10-ms among frames. The frames are then processed one by one. Subjective and objective tests are used to assess performance. Speech quality can be better evaluated with subjective tests. The proposed methodology competency is explored by comparing with the traditional techniques, such as traditional telephony speech bandwidth enhancement (TTSBE) by data hiding (TTSBEDH) [9], TTSBE by phonetic classification (TTSBEPC) [10], TTSBE using watermark (TTSBEUW) [11], and TTSBE using speech codecs (TTSBEUSC) [16]. The channel models considered here are μ-law and additive white Gaussian noise (AWGN).

4.1 Subjective assessments

The perceptual clearness is assessed based on the mean opinion score (MOS) test [9, 10]. The listening test for comparison of various speech signals like CWB, LNB, CLNB, and RCWB is performed [26]. These tests were performed in a silent room using headsets. During each test, thirty participants are considered.

4.1.1 Perceptual Clearness (PCL)

In the proposed technique the information must be transparently concealed, i.e., LNB and CLNB are subjectively indistinguishable. High PCL means low perceptible degradation in the CLNB signal. PCL is assessed based on the MOS test. Listeners those are participated in the test compare LNB and CLNB to provide a decision in terms of MOS which are tabulated in Table 1. Table 2 showcases the results of the averaged MOS for conventional [9-11, 16] and proposed approaches. The remarkable perceptual clearness of the proposed approach over the traditional approaches is shown by MOS values in Table 2.

Table 1. MOS

Score	Instruction
1	LNB and CLNB signals are dissimilar
2	Noticeable dissimilarity among LNB and CLNB signals
3	Small dissimilarity among LNB and CLNB signals
4	LNB and CLNB signals are similar

Table 2. MOS assessment outcomes

Approach	MOS
TTSBEDH [9]	2.81
TTSBEPC [10]	3.01
TTSBEUW [11]	3.12
TTSBEUSC [16]	3.46
Proposed method	3.98

4.1.2 Subjective contrasts among CWB, LNB, CLNB, and RCWB signals

I, II, III, and IV in Table 3 represent the CWB signal, LNB signal, CLNB signal, and RCWB signal respectively. The subjects are asked to do a pairwise analysis of signals among I to IV and must tell whether the first signal is paramount (>), deprived (<), or alike (≈) to the second signal. Table 3 provides the responses of pairwise comparison of I, II, and III to the other signal, Table 4 provides the responses of pairwise comparison of II and III to the other signal, and Table 5 provides the responses of comparison among III and IV. The number of subjects with an exact preference (> or < or ≈) is mentioned with Arabic digits in the table. The CWB signal outperforms the CLNB signal for conventional [9-11, 16] and proposed methods which are endorsed by Table 3. Table 3 also endorsed that a clearly enhanced RCWB signal quality of the proposed technique over the conventional methods.

Table 3. Subjective contrast outcomes among I, II, III, and IV

Technique	I	II	III	IV
TTSBEDH [9]	>	30	30	15
	<	0	0	0
	≈	0	0	15
TTSBEPC [10]	>	30	30	14
	<	0	0	0
	≈	0	0	16
TTSBEUW [11]	>	30	30	13
	<	0	0	0
	≈	0	0	17
TTSBEUSC [16]	>	30	30	10
	<	0	0	0
	≈	0	0	20
Proposed method	>	30	30	2
	<	0	0	0
	≈	0	0	28

Table 4. Subjective contrast outcomes among II, III, and IV

Technique	II	III	IV
TTSBEDH [9]	>	8	3
	<	4	18
	≈	18	9
TTSBEPC [10]	>	8	1
	<	2	19
	≈	20	10
TTSBEUW [11]	>	5	2
	<	3	20
	≈	22	8
TTSBEUSC [16]	>	5	2
	<	2	22
	≈	23	6
Proposed method	>	1	0
	<	0	28
	≈	29	2

Table 5. Subjective contrast outcomes among III and IV

Technique	III	IV
TTSBEDH [9]	>	6
	<	15
	≈	9
TTSBEPC [10]	>	5
	<	16
	≈	9
TTSBEUW [11]	>	3
	<	17
	≈	10
TTSBEUSC [16]	>	4
	<	19
	≈	7
Proposed method	>	0
	<	28
	≈	2

The remarkable perceptual clearness of the proposed approach over the traditional approaches is endorsed by Table 4. Compared to conventional methods, the RCWB signal is better than the LNB signal for the proposed approach which is endorsed in Table 4. Compared to conventional methods, the RCWB signal is better than the CLNB signal for the proposed approach which is endorsed in Table 5.

4.2 Objective quality evaluation

The RCWB signal quality is assessed with log spectral distortion (LSD) [9, 10] and CWB-perceptual evaluation of speech quality (CWB-PESQ) [27] tests. The perceptual clearness is assessed with the LNB-PESQ test [28]. The robustness of concealed data to QACN is assessed with a bit error rate (BER) measure [13].

4.2.1 Perceptual clearness (PCL)

LNB-PESQ test is utilized to assess PCL by comparing LNB signal with CLNB signal. LNB-PESQ ranges from 0.5 to 4.5, lower values such as 0.5 represents the worsened PCL and higher values like 4.5 represent the best PCL. Table 6 lists the responses of mean scores for the traditional [9-11, 16] and proposed techniques. A clear PCL enhancement of the proposed approach over the traditional techniques is witnessed from the scores as listed in Table 6.

Table 6. LNB-PESQ test outcomes

Approach	LNB-PESQ
TTSBEDH [9]	2.81
TTSBEPC [10]	3.02
TTSBEUW [11]	3.33
TTSBEUSC [16]	3.40
Proposed method	4.43

4.2.2 RCWB signal quality

The quality of RCWB speech is evaluated by comparing CWB and RCWB signals in the CWB-PESQ test. Table 7 presents the mean CWB-PESQ scores of the conventional [9-11, 16] and proposed methods. The proposed technique produces a score of 4.38 which specifies that the RCWB signal quality attained is remarkable. Thus, the proposed technique improved the speech quality when compared to the traditional methods.

Table 7. CWB-PESQ test outcomes

Approach	CWB-PESQ
TTSBEDH [9]	2.31
TTSBEPC [10]	2.63
TTSBEUW [11]	3.54
TTSBEUSC [16]	3.62
Proposed method	4.38

4.2.3 Comparison of original and reconstructed MB speech

LSD is a very reliable measure for assessing the resemblance among true and restored MB signals and is given by

$$LSD = \frac{1}{2\pi} \int_{-\pi}^{\pi} \left(20\log_{10} \frac{g_p}{a_s(e^{jw})} - 20\log_{10} \frac{\hat{g}_p}{|\hat{a}_s(e^{jw})|} \right)^2 dw \tag{5}$$

where, g_p and $\hat{g}_p$ are gains of true and restored MB signals respectively, $\frac{1}{a_s(e^{jw})}$ and $\frac{1}{\hat{a}_s(e^{jw})}$ are the spectral envelopes of true and restored UB signals respectively. In general, the best quality of reproduced MB signal has a low value of LSD. Table 8 lists the mean LSD scores for the existing [9-11, 16] and proposed schemes under μ-coding. There is a clear enhancement in the quality of the proposed scheme over the traditional schemes [9-11, 16] is witnessed from the values as listed in Table 8.

Table 8. LSD test outcomes

Approach	Log Spectral Distortion
TTSBEDH [9]	13.56
TTSBEPC [10]	11.56
TTSBEUW [11]	7.12
TTSBEUSC [16]	6.67
Proposed method	2.31

4.2.4 Vigour of concealed data

AWGN with SNR ranges from 15 to 35 dB is summed up with CLNB signal [29]. The proposed method's robustness is assessed based on BER. The PN code size is 8. The lower value of BER designates the RCWB signal of high quality. The BER values which were attained with SNR in the range of 15 to 35 dB are beneath $7.7036 * 10^{-4}$ which endorses the RCWB signal of high quality. The BER value which is attained with μ-law coding is $4.61 * 10^{-4}$ which endorses the RCWB signal of high quality.

5. CONCLUSION

A novel speech BE based on the DWT-DCT-DH (Hybrid transform model) technique is presented in this paper for embedding spread SEPs of MB signal within LNB signal DCT coefficients. The concealed information is extracted for generating an extraordinary quality CWB signal at the receiver.

The concealed information is vulnerable to QACNs. CDMA approach is, thus, employed for reproducing the concealed information which is appealed as robust towards QACNs. Especially, every information bit entrenched within the LNB signal is spread out as the product by definite SPSE. Then, the spread signals were summed to create the concealed information. The concealed information could be consistently retrieved since the correlation among the SPSEs is low. It is evident that the proposed approach was a vigorous solution for speech BE. Subjective, CWB-PESQ, and LSD test results confirmed that there is excellent and improved performance in RCWB signal using the proposed method over the conventional techniques. A clear PCL enhancement of the proposed approach over the traditional techniques is witnessed from MOS and LNB-PESQ tests. Studying the performance of the proposed approach on noisy input speech signals will be the future work.

REFERENCES

[1] Jax, P., Vary, P. (2006). Bandwidth extension of speech signals: A catalyst for the introduction of wideband speech coding. IEEE Communications Magazine, 44(5): 106-111. https://doi.org/10.1109/MCOM.2006.1637954

[2] Jax, P. (2002). Enhancement of bandlimited speech

signals: algorithms and theoretical bounds. Doctoral dissertation, Ph. D Thesis, RWTH Aachen.

[3] Prasad, N., Kumar, T.K. (2016). Bandwidth extension of speech signals: A comprehensive review. International Journal of Intelligent Systems and Applications, 8(2): 45-52. https://doi.org/ 10.5815/ijisa.2016.02.06

[4] Ling, Z.H., Ai, Y., Gu, Y., Dai, L.R. (2018). Waveform modeling and generation using hierarchical recurrent neural networks for speech bandwidth extension. IEEE/ACM Transactions on Audio, Speech, and Language Processing, 26(5): 883-894. https://doi.org/10.1109/TASLP.2018.2798811

[5] Lee, B.K., Noh, K., Chang, J.H., Choo, K., Oh, E. (2018). Sequential deep neural networks ensemble for speech bandwidth extension. IEEE Access, 6: 27039-27047. https://doi.org/10.1109/ACCESS.2018.2833890

[6] Abel, J., Fingscheidt, T. (2017). A DNN regression approach to speech enhancement by artificial bandwidth extension. In 2017 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA), pp. 219-223. https://doi.org/10.1109/WASPAA.2017.8170027

[7] Wang, Y., Zhao, S., Qu, D., Kuang, J. (2016). Using conditional restricted Boltzmann machines for spectral envelope modeling in speech bandwidth extension. 2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pp. 5930-5934. https://doi.org/10.1109/ICASSP.2016.7472815

[8] Jax, P., Vary, P. (2002). An upper bound on the quality of artificial bandwidth extension of narrowband speech signals. 2002 IEEE International Conference on Acoustics, Speech, and Signal Processing, 1: I-237. https://doi.org/10.1109/ICASSP.2002.5743698

[9] Chen, S., Leung, H. (2005). Artificial bandwidth extension of telephony speech by data hiding. In 2005 IEEE International Symposium on Circuits and Systems. pp. 3151-3154. https://doi.org/10.1109/ISCAS.2005.1465296

[10] Chen, S., Leung, H. (2007). Speech bandwidth extension by data hiding and phonetic classification. In 2007 IEEE International Conference on Acoustics, Speech and Signal Processing-ICASSP'07, 4: IV-593. https://doi.org/ 10.1109/ICASSP.2007.366982

[11] Chen, Z., Zhao, C., Geng, G., Fuliang, Y. (2013). An audio watermark based speech bandwidth extension method. EURASIP Journal on Audio, Speech and Music Processing, 10: 1-8. https://doi.org/10.1186/1687-4722-2013-10

[12] Sagi, A., Malah, D. (2007). Bandwidth extension of telephone speech aided by data embedding. EURASIP Journal on Advances in Signal Processing, 1: 37-52. https://doi.org/10.1155/2007/64921

[13] Chen, S., Leung, H., Ding, H. (2007). Telephony speech enhancement by data hiding. IEEE Transactions on Instrumentation and Measurement, 56(1): 63-74. https://doi.org/ 10.1109/TIM.2006.887409

[14] Geiser, B., Vary, P. (2013). Speech bandwidth extension based on in-band transmission of higher frequencies. 2013 IEEE International Conference on Acoustics, Speech and Signal Processing, pp. 7507-7511. https://doi.org/10.1109/ICASSP.2013.6639122

[15] Geiser, B., Vary, P. (2007). Backwards compatible wideband telephony in mobile networks: CELP watermarking and bandwidth extension. 2007 IEEE International Conference on Acoustics, Speech and Signal Processing-ICASSP'07, 4: IV-533. https://doi.org/10.1109/ICASSP.2007.366967

[16] Bhatt, N., Kosta, Y. (2015). 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, 18(1): 57-64. https://doi.org/10.1007/s10772-014-9249-1

[17] Bhatt, N. (2016). 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, 19(4): 881-893. https://doi.org/10.1007/s10772-016-9378-9

[18] Rekik, S., Guerchi, D., Selouani, S.A., Hamam, H. (2012). Speech steganography using wavelet and Fourier transforms. EURASIP Journal on Audio, Speech, and Music Processing, 2012(1): 1-14. https://doi.org/10.1186/1687-4722-2012-20

[19] Proakis, J.G. (1989). Digital Communications, 2nd edition. McGraw-Hill, New York.

[20] Somerville, F.C.A., Woodard, J.P. (2001). Voice Compression and Communications: Principles and Applications for Fixed and Wireless Channels. New York: John Wiley & Sons.

[21] Nilsson, M., Kleijn, W.B. (2001). Avoiding over-estimation in bandwidth extension of telephony speech. In 2001 IEEE International Conference on Acoustics, Speech, and Signal Processing. Proceedings (Cat. No. 01CH37221), 2: 869-872. https://doi.org/10.1109/ICASSP.2001.941053

[22] Bezdek, J.C. (2013). Pattern recognition with fuzzy objective function algorithms. Springer Science & Business Media. https://doi.org/10.1007/978-1-4757-0450-1

[23] Jax, P., Vary, P. (2003). On artificial bandwidth extension of telephonespeech. Signal Processing, 83(8): 1707-1719. https://doi.org/10.1016/S0165-1684(03)00082-3

[24] Europea Telecommunications Standards Institute (ETSI) Standard. (2000). Speech Processing, Transmission and Quality Aspects (STQ); Distributed speech recognition; Front end feature extraction algorithm; Compression algorithms, ETSI ES 201 108 V1.1.2.

[25] Garofolo, J.S., Lamel, L.F., Fisher, W.M., Fiscus, J.G., Pallett, D.S. (1988). Getting started with the DARPA TIMIT CD-ROM: An acoustic phonetic continuous speech database. National Institute of Standards and Technology (NIST), Gaithersburgh, MD, 107: 16.

[26] Prasad, N., Kishore Kumar, T. (2017). Speech bandwidth extension aided by spectral magnitude data hiding. Circuits, Systems, and Signal Processing, 36(11): 4512-4540. https://doi.org/10.1007/s00034-017-0526-5

[27] International Telecommunications Union. (2011). 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, pp. 863-863.

[28] International Telecommunications Union. (2001).

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, pp. 862-862.

[29] Keiser, B.E., Strange, E. (1995). Digital Telephony and Network Integration. New York: Van Nostrand Reinhold.