Adaptive Data Hiding Based Telephony Speech Enhancement

N Prasad[a], P Sitaramanjaneyulu[b], E.Praveen Kumar[c], G. R. L.V. N. S. Raju [d]

[a,b,c,d] Dept. of ECE, Shri Vishnu Engineering College for Women (A), Bhimavaram, Andhra Pradesh, India
[a]prasadnece@svecw.edu.in

Abstract: Public telephone systems transmit speech across a limited frequency range, about 300–3400 Hz, called narrowband (NB) which results in a significant reduction of quality and intelligibility of speech. This paper proposes a fully backward compatible novel method for bandwidth extension of NB speech. The method uses adaptive data hiding technique to provide a perceptually better wideband speech signal. The spectral envelope parameters are extracted from the high frequency components of speech signal existing above NB, which are then spread by using spreading sequences, and are embedded in the NB speech signal using adaptive data hiding technique. The embedded information is extracted at the receiving end to reconstruct the wideband speech signal. Theoretical and simulation analyses show that the proposed method is robust to quantization and channel noises. The log spectral distortion test clearly show that the reconstructed wideband signal gives a much better performance in terms of speech quality when compared to the conventional speech bandwidth extension methods employing data hiding..

Keywords: Telephony speech enhancement, Telephonic networks, Speech quality, Spread spectrum, adaptive data hiding

_____

1. Introduction

Speech transmission through existing telephone networks face difficulties in respect of losing a portion of speeches as frequency of human speeches lies beyond its bandwidth ranging between 300 and 40000 Hz. As a result it generates problems regarding decreasing the quality of voices along with accessibility of the speech spectrum across the telephonic network. It is found that utilisation of wideband (WB) can overcome this problem as the frequency of this WB lies between 0 and 8000 Hz in comparison to NB. However establishing this bandwidth with larger speech frequency requires complete change in existing infrastructures that is both time consuming and requires more cost [1]. Thus it is suggested to implement the technique of speech bandwidth extension (BWE) at the receivers for amplifying speech frequency [2]. It would result in enhancing the quality of the existing system without changing the overall framework.

A significant amount of enhancement in the quality of speech is attained by inclusion of BWE techniques in the telephone network in existence. Artificial bandwidth extension (ABE) is considered as an effective BWE technique that works on the principle of mutual dependency of NB signal with out-of-band. It helps in estimating the information regarding out-of-band from NB signal and utilise those information to reconfigure the WB for enhancing the quality of speech. Therefore, it is considered as an effective speech production model that contributes in improvement of frequency bands of the existing network of telephonic communication.

Based on the source-filter model that is incorporated with the ABE framework it results in estimation of WB excitation signal. Along with that a system is utilised for filtering the vocal tract. Both of these operations contribute to modifying the bandwidth spectral. Several operations involved with this modification are modulating noise [3], harmonic and noise modelling [4] and sinusoidal synthesis [5]. Apart from that spectral folding and translation, modulation of pitch and non-linear processing [2] are also used to estimate the WB excitation. Suitable techniques involved with this estimation are codebooks [6], Gaussian mixture models [8], linear mapping [7], hidden Markov models [9] and neural networks [10]. On the other hand, a certain set of limitations are also involved with this technique to improve the quality of existing speech framework [11].

A suitable set of existing literature is present to suggest an alternative solution for improving the quality of telephonic speech in comparison to ABE techniques. It includes transmission of additional information regarding out-of-band that results in improving the quality effectively [1]. Data hiding methods are utilised for hiding the confirmation related to out-of-band that ensures a backward compatibility in the existing network. As proposed by Siyue Chen and Henry Leung there is an upper band(UB) consisting of a line spectrum pair in the speech BWE method. It belongs to a range of higher frequency of 4 to 8 KHz [12]. A composite NB speech signal is obtained by encoding the UB signals and adequate information is embedded in this band. A better quality of WB is

N Prasad[a], P Sitaramanjaneyulu, E.Praveen Kumar , G. R. L.V. N. S. Raju

obtained by extracting the embedded set of information and decoding it at the end of receivers. However this technique has resulted in development of poor quality of NB speech signals. Hence the technique of phonetic classification has been adopted by Siyue Chen and Henry Leung [13] for improving the quality of composite NB signals and the WB speech that is reconfigured [12]. This technique results in encoding of the UB signal in a more effective manner hence it increases the quality of speech signals. However it can also produce poor quality speeches on getting corrupted by noises included in the channel and those results in giving poor performance of BWE [12, 13]. As proposed by Siyue Chen et. al. undetectable elements involved with the NB signals are removed from the hidden channel to improve the quality of BWE in NB speech [14]. Regeneration of hidden audible elements results in configuration of a WB speech of good quality. However limitation of this process involves some missing frequency in hidden components embedded in the audible channel. As proposed by Zhe Chen et. al. [15] there is another technique of perception based least significant bit watermark method used for reconfiguring a high quality WB signal by embedding UB elements in the NB speech framework. These components are extracted at the receiver's end for reconstructing the WB signal. For developing a backward compatible WB codec the technique of joint coding along with data hiding is utilised by Peter Vary and Bernd Geiser [16] that results in embedding of additional information at a rate of 600 bit/s. Bernd Geiser and Peter Vary [17] proposed a NB coder method to develop a backward compatible WB telephony by acquiring additional information regarding NB speeches at a rate of 400 bit/s. However it gave a poor quality WB while corrupted by noises in the speech channel.

An adaptive data hiding technique is proposed in [18] that include embedding the components of secret speech signal in the host speech signal. It is termed as audio steganography that results in keeping the quality of host signal intact. A stego speech signal is found to be produced from the host speech in an unidentifiable form while it is useful in extracting the secret speech signal maintaining the quality of host signal.

Transmission of speech at low frequency ranging between 0 to 300 Hz does not create any problem in the telephonic network hence BWE towards UB is considered in this network [19-21]. A novel NB speech of BWE technique is proposed in this paper that includes an adaptive data hiding technique [18]. A linear predictive coding (LPC) is used to analyse the UB signal and for embedding spectral envelope parameters in NB speech. This proposed scheme is utilised to use real UB information rather based on estimation. It is also found to be compatible with the conventional equipment associated with NB terminals such as plain ordinary telephone sets (POTS). These NB receivers are capable of assessing the NB speech appropriately that even do not need additional hardware. On the other hand a customised receiver is utilised to extract embedded information that results in delivery of better quality WB signals.

Quantization noise is considered in the proposed techniques in [12, 13, 22] for implementing BWE of the NB speech. Apart from channel noise QN is also treated in this work by incorporating spread spectrum (SPSP) technique. However different kinds of QN techniques involved in this paper are such as pulse code modulation (PCM), μ-law, ADPCM (adaptive delta pulse code modulation) and EFR (Enhanced full rate).

For extracting the embedded information successfully the SPSP technique is included in this work. This method is well known for its potentiality in respect of interference. A specific SS (spreading sequence) is multiplied with an individual parameter that is considered to embed for spreading. Embedded information is generated by summing all the spread signals. However the embedded information can be recovered successfully as spread sequences have a low cross correlation with each other [23]. A correlator is used to recover spread sequences that are orthogonal over each other.

Low correlations in spreading sequences are preferred as it also results in minimising the chances of interference. In this regard Hadamard codes are found with an orthogonal structure and an optimum cross correlation. On the other hand a varied cross correlation attributes are found in several other codes such as m-sequences, Gold and Kasami codes [24, 25]. Therefore Hadamard codes are preferred in this study to minimise interference while extracting embedded information.

An adaptive data hiding technique is discussed in section 2, the proposed NB speech BWE method is described in section 3, both subjective and objective test results are discussed in section 4 and section 5 includes the conclusion.

2. Technique of adaptive data hiding for BWE

Estimation of embedded NB signal in the temporal domain is done for embedding the extended band signal ($Y_{eb}(n)$) to NB signal ($Y_{nb}(n)$). Initially samples $\{Y_{nbi}\}$ in this regard are classified as given below:

$$Et_i = \begin{cases} 3, if\ 0 < |Y_{nbi}| \leq 100 \\ 4, if\ 100 < |Y_{nbi}| \leq 2000 \\ 5, if\ 2000 < |Y_{nbi}| \leq 10000 \\ 6, if\ 10000 < |Y_{nbi}| \end{cases} \qquad (1) \qquad \text{where } Et_i \text{ signifies the term ETH}$$

that represents high embedding capacity of samples having higher magnitude as comparative to the lower one. Maximum number of parameters is restricted for embedding in each NB frame by inclusion of ETH.

Assume that *D(h)* is the representation vector of $Y_{eb}(n)$

Parameters selected to be embedded are assumed to be spread after multiplying with pseudo-noise (PN) code

$$D_l(h) \bullet \vec{s}^{\,l}, 1 \leq l \leq M \qquad (2)$$

Where $\vec{s}^{\,l}$ PN code is whose length is *M* and *l* denotes index of a specific parameter of *(D$_l$(h))*. For spreading the vectors corresponding parameters are utilised. Summation of spreading vectors for embedding information is given as below:

$$H(j) = \sum_{l=1}^{M} D_l(h)s^l(j) \qquad (3)$$

$s^l(j)$ is the jth element of vector $\vec{s}^{\,l}$. Hidden data is denoted by *H(j)* and the encoded data is denoted by *T(j)* whereas data packets are denoted by $\{dp_i\}$

$$dp_i = \sum_{k=0}^{Et_i-1} 2^k \bullet T(j) \qquad (4)$$

$\{Y_{nbi}\}$ are quantized as per ETH $\{Et_i\}$ for hiding data as given below:

$$Qua_i = Y_{nbi} - \text{Reminder}\,(Y_{nbi}, 2^{Et_i}) \qquad (5)$$

Where Y_{nbi} by 2^{Et_i} is obtained as remainder of division

$\{dp_i\}$ are hidden in $\{Qua_i\}$ [18] as given below:

$$Y^1_{nbi} = Q_{uai} + dp_i \bullet sign(Qua_i) \qquad (6)$$

Where where $sign(Qua_i) = \begin{cases} 1, if\, Quai \geq 0 \\ -1, if\, Qua_i < 0 \end{cases} \qquad (7)$

As a result a composite NB signal is obtained and it is then transmitted to receivers over the telephonic network channel. Both the channel and quantization noises are introduced in this telephonic network. Assume received signal is denoted by $\hat{Y}^1_{nb}(n)$ or $\hat{Y}^1_{nb}(n) = Y^1_{nb}(n) + er$. $\hat{Y}^1_{nb}(n)$ denotes the combined version of noises involved with channel and quantization. A conventional telephonic terminal is used to treat this combined signal. However differences between $Y_{nb}(n)$ and $Y^1_{nb}(n)$ are negligible hence the quality of $Y_{nb}(n)$ is not degraded significantly in this paper.

Estimation of ETH in equation (1) is dome to retrieve $\hat{Y}_{eb}(n)$. The data is extracted as follows:

$$dp_i = |Y^1_{nbi}|\&(2^k) \quad k = \{0,1,2,\ldots,Et_i-1\} \qquad (8)$$

The correlation can be expressed as follows:

$$\hat{D}_{lo}(h) = \frac{1}{M} \sum_{j=1}^{M} \hat{H}(j)s^{lo}(j) \tag{9}$$

$\hat{H}(j)$ is considered as corrupt version and expressed as:

$$\hat{H}(j) = H(j) + e(j) \tag{10}$$

$e(j)$ denotes combination of channel and quantization noise

On substituting (10) in (9) we get:

$$\hat{D}_{lo}(h) = \frac{1}{M} \sum_{j=1}^{M} \hat{H}(j)s^{lo}(j)$$

$$= \frac{1}{M} \sum_{j=1}^{M} s^{lo}(j) \left(\sum_{l=1}^{M} D_l(h)s^l(j) + e(j) \right)$$

$$= \frac{1}{M} \sum_{j=1}^{M} s^{lo}(j) \times \left(D_{lo}(h)s^{lo}(j) + \sum_{l \neq lo} D_l(h)s^l(j) + e(j) \right)$$

$$= D_{lo}(h) + \frac{1}{M} \sum_{j=1}^{M} \sum_{l \neq lo} D_l(h)s^l(j)s^{lo}(j) + \frac{1}{M} \sum_{j=1}^{M} s^{lo}(j)e(j) \tag{11}$$

Mutually orthogonal PN codes are expressed as:

$$\sum_{j=1}^{M} s^l(j)s^{lo}(j) = 0 \tag{12}$$

Where $l \neq l_o$

$$\sum_{j=1}^{M} \sum_{l \neq lo} D_l(h)s^l(j)s^{lo}(j) = \sum_{l \neq lo} D_l(h) \sum_{j=1}^{M} s^l(j)s^{lo}(j) = 0 \tag{13}$$

As $s^{lo}(j)$ is not correlated with $e(j)$

$$\frac{1}{M} \sum_{j=1}^{M} s(j)e(j) \to 0 \tag{14}$$

By getting $M \to \infty$

On substituting (13) and (14) in (11) we get

$$\hat{D}_{lo}(h) = D_{lo}(h) \tag{15}$$

When $M \to \infty$. It signifies that parameters of extended band signals can be extracted as a result of including SPSP technique and suppressing both channel and quantization noises.

3. Bandwidth extension of NB speech using adaptive data hiding

3.1 Transmitter

Bandwidth is the factor that is mainly based on the limitation of the signals in the particular system. Separation of the bandwidth in the higher and lower stage will also deliver a better knowledge in the system. High and low pass filters are another important part of the system that enhances the performance better in the runtime. The frequency range of low band signals in the system is 0-4Khz whereas a high bank transmits the signal level 4-8 KHz. It can also be said that speech information is another crucial part that is engaged in this particular area.

Moreover, evaluation of the structure will also transmit the compatible frequency with the speech information in the system area [26]. $Y_{eb}(n)$ is the extended band signal that might be crucial for implementing the system better? Bandwidth extension will also increase the performance of the particular system better in the segment.
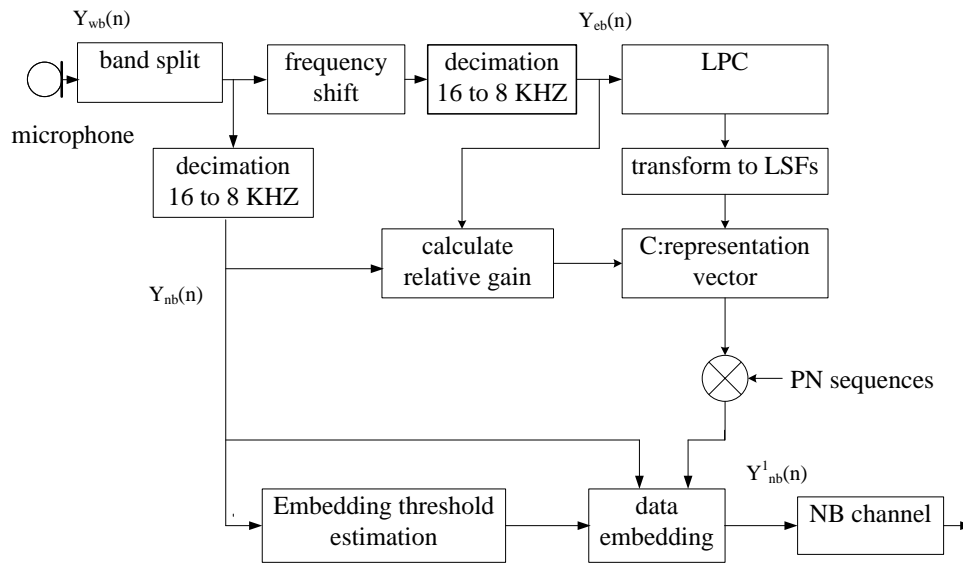

Figure 1: Proposed transmitter

Managing the parameters in the transmission will also generate a better value in the estimation process of the system. $Y_{eb}(n)$ is the extended bandwidth that will implement the performance better in the segment. $Y_{nb}(n)$ is the NB value which plays a crucial role in the transmission of the frequency in the application. Enhancement of the number of parameters in the system will also implement the structure better for effective performance in the area. Spectral frequency is another important part that also needs to be managed for implementing the process of

estimation better. However, it can also be stated that $g_r = \dfrac{g_{eb}}{g_{nb}}$ is the value that is mainly based on the estimation

of the gain in the system. Moreover, Levinson-Durbin algorithm is one of the most crucial parts that can be helpful for increasing the performance better in the area [13]. Representation of the vector is another important part that also needs to be managed for implementing the process better in the area. $Y_{eb}(n)$ is the relational vector that is present in the case which will deliver a better result in the case. Telephone network channel receiver mainly works on data hiding techniques that will deliver a better result at the end of the transmission.

3.2 Reciever

In the receiver side, spectral receiver and gain value is another factor that is also engaged in the structure. Linear predictive coefficient is another crucial part that will also generate a better value in the receiver for proper transmission. Filtering of the extended band will also generate a better value in the segment which will implement the performance of the system better in the area. Furthermore, evaluation of $\hat{g}_{eb} = \hat{g}_r \bullet \hat{g}_{nb}$ will also enhance the generation of the result better in the area. Furthermore, determination of the $\hat{Y}^1_{nb}(n)$ might also play a crucial role in the segment that will emphasize the signalling better in the segment. Interpolation of the value will also deliver a better value that will increase the performance of the system [13]. Transmitting the value from 8000Hz to 16000Hz will maintain a better sample speech which will increase the success level of the performance better in the segment. Estimation of $Y^{11}_{nb}(n)$ will also restore the functionality in the area that will generate a better signal in the receiver side and extended band signal.
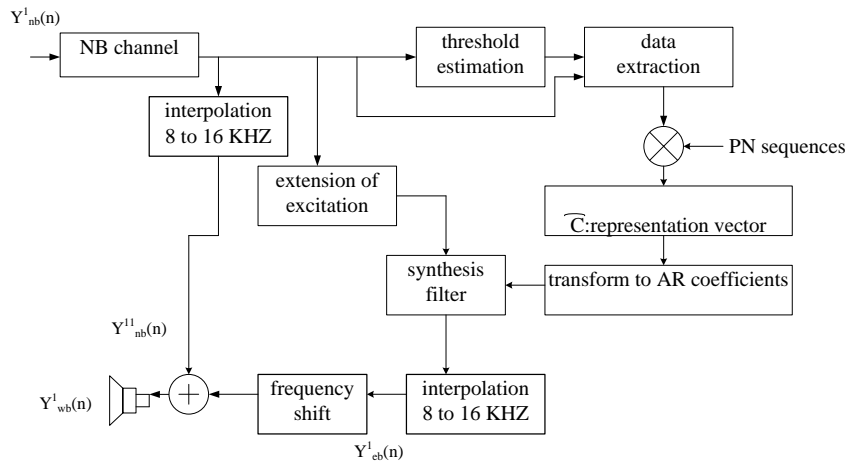
Figure 2: Proposed receiver

4. Experimental Results

To evaluate the proposed methods speech splames are used that were collected from the TIMIT database[27]. Ten different speakers including males and females spoke different sentences. According to the proposed methods, these sentences were generally 2 to 2.5 sec long for evaluating the performance. The NB samples were segregated into 20-ms frames which were non overlapped and processed one by one.

Objective and subjective both measurements have been taken to evaluate the performance of the selected methods. These proposed methods have been compared with various methods and these are the ABE of telephony speech [12] which is proposed by data hiding. Phonetic classification and data hiding introduced BWE speech[13]. Data hiding enhances telephony speech[14]. BSE speech is based on an audio watermark [22]. Narrowband speech codes are used in stenographic WB telephony and ABE of speech is used to spread the information which is supported by watermark. Data hiding with phonetic classification, conventional bit stream data hiding, conventional joint coding and data hiding, conventional signal domain data hiding, conventional data hiding and conventional WTSI are present in the analysis. The vectorial form of quantization modulation index (QIM) is used by Conventional WTSI for speech BWE. Two channel models are used by the experiments which are regulated in this study also provided below:

(i) μ -law channel model.

(ii) AWGN channel model with a signal to noise ratio (SNR) of 35 dB.

4.1. Subjective quality evaluation

Through using mean opinion scores, perceptual clarity had been assessed in this paper and the score of the test was [12, 13, 36]. This listening test has been performed to compare the original WB signal, reconstructed WB signal and CNB signal. These speech samples were given personally so they could not listen to the others. Those samples were provided in a noiseless room to individual listeners. After that the opinion of every listener had been taken to evaluate the speech sounds through using pre-set scale. In this test, 10 females and 10 males had taken part and they belonged to the age of 22 to 32 years.

4.1.1. Perceptual Transparency

Information should be implanted distinctly by the proposed methods. CNB signal cannot be differentiated from the NB signal in this proposed method. In this study proposed methods used the MOS test[12,13,26] of which result was average MOS. Subjects have been taken part in the MOS test for comparison of the pairs of samples with CNB signal and NB signal. In Table 1 their opinion has been recorded in terms of MOS. The resultant average MOS is included in Table 2. All samples and all subjects of standard speech BWE methods and proposed methods. A transparent perceptual clarity advantage of the proposed methods over standard speech BWE methods was found from the average MOS which is shown in Table 2. In addition, an MOS of 3.90 is presented by the proposed methods. As MOS 3.90 is near to the MOS, these two signals sound similar to each other. This quality of sound presents that NB signal is almost similar to CNB signal. The data implanted that speech BWE methods has a little consequences on perception.

Table 1 Mean opinion scores (MOS)

Score	Instruction
1	Two signals are different
2	Two signals are similar, but the difference is easy to see
3	Two signals sound very similar, only little difference exists
4	Two signals sound identical

Table 2 Comparative performance in terms of average MOS

Method	MOS
Conventional data hiding[12]	3.02
Data hiding with Phonetic classification [13]	3.75
Conventional signal domain data hiding [14]	3.58
Conventional bit stream data hiding [15]	3.02
Conventional joint coding and data hiding [16]	3.53
Conventional WTSI [22]	3.88
Proposed method	3.90

4.1.2 Subjective comparison of original WB speech, CNB speech and reconstructed WB speech

This subjective listening test had been performed for evaluation of performance. It had also been done for the comparison between speech BWE methods[12-15,22] and the proposed methods. Actual WB speech has been collected from TIMIT database. It was characterized by I; where CNB speech was signified by II and III was denoted to the reconstructed WB speech. Listeners were interested to compare the speech samples which were included in I and III. It was taken from the listeners to receive an opinion whether the first speech sample is worse, equal or better in comparison to the second speech sample. The results of comparing I and II with III are classified in Table 3 (a) and (b). In the table the numbers of the listeners are presented in Arabic numbers and their preferences are also included in the table. It is found by the researchers that WB speech is better in comparison to the other signals which are CNB speech methods and the proposed methods. Better WB reconstruction performances are seen in the proposed methods.

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

	I	II	III
Conventional data hiding[12]	▷	20	13
	◁	0	0
	≈	0	7
Data hiding with Phonetic classification [13]	▷	20	11
	◁	0	0
	≈	0	9
Conventional bit stream data hiding [15]	▷	20	12
	◁	0	0
	≈	0	8
Conventional WTSI [22]	▷	20	9
	◁	0	0
	≈	0	11
Proposed method	▷	20	8
	◁	0	0
	≈	0	12

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

		II	III
Conventional data hiding[12]	▷		4
	◁		9
	≈		7
Data hiding with Phonetic classification [13]	▷		2
	◁		11
	≈		7
Conventional bit stream data hiding [15]	▷		3
	◁		9
	≈		8
Conventional WTSI [22]	▷		2
	◁		12
	≈		6
Proposed method	▷		0
	◁		14
	≈		6

4.2. Objective quality evaluations

Evaluation on the proposed methods is being done for further studies and similar data is used in the analysis. Objective measurements are taken to evaluate the performance. LSD measurements[12] are used in assessing the quality of WB speech. Through using ITU-T PESQ tool[28] perceptual clarity is measured. ITU-T recommended the WB-PESQ measurement[29] to check the quality of WB speech.

4.2.1. Comparison of original and reconstructed UB speech

LSD measurement is presented in the study to identify the similarity between true signal and UB signal.

$$LSD = \frac{1}{2\Pi} \int_{-\Pi}^{\Pi} (20\log_{10}\frac{G}{|A(e^{jw})|} - 20\log_{10}\frac{\hat{G}}{|\hat{A}(e^{jw})|})^2 dw$$

(16)

Linear prediction is calculated by the spectral envelopes which are used for short frames like 20ms long. Superior quality is present in the LSD measurement. Average result of the LSD measurement is presented in Table no. 4.

Table 4: Comparative performance through average LSD

Method	LSD
Conventional data hiding[12]	12.17
Data hiding with Phonetic classification [13]	10.46
Conventional signal domain data hiding [14]	6.3
Conventional bit stream data hiding [15]	5.89
Conventional joint coding and data hiding [16]	5.05
Conventional WTSI [22]	2.9
Proposed method	2.61

From Table no. 4, it is found that proposed methods exceeded the conventional speech BWE methods[12-16,22]. As numbers of parameters are errors, performance of the LSD methods decreases. In addition it is found that LSD methods are conventional and small errors implanted in the study.

4.2.2. Perceptual transparency

Through providing NB signal and CNB signal, NB-PESQ measurement has been performed to evaluate the perceptual clarity. It gives a higher score and superior quality than the other speech methods.

Table 5: Comparative performance through average NB-PESQ

Method	NB-PESQ
Conventional data hiding[12]	3.21
Data hiding with Phonetic classification [13]	3.96
Conventional signal domain data hiding [14]	3.75
Conventional bit stream data hiding [15]	3.17
Conventional joint coding and data hiding [16]	3.68
Conventional WTSI [22]	3.98
Proposed method	4.01

4.2.3. Robustness of embedded information

The effects of noise corruption is noted next, where AWGN is included in composite NB signals that include 35 dB SNR. MSE is used for measuring performance while the SS length is considered as 16. As the smaller value of MSE shows superior quality, the SNR of 35 dB occurred through MSE is considered as successful while using SPSS technique. Even though the law causes issues, after the application of MSE the law derived effective results.

4.2.4 . WB Speech quality

In order to measure the quality of the WB speech, WB-PESQ is used that includes information from the TIMIT database, WB data as speech. The results of all the results regarding the WB-PESQ speech is provided in Table 6 through BWE method along with developed methods [12-16,22]. The PESQ score of the proposed method is 4.10, which displays the effective result of the WB speech quality; that is further derived through subjective listening examination.

Table 6: Comparative Performance through average WB-PESQ

Method	WB-PESQ
Conventional data hiding[12]	2.62
Data hiding with Phonetic classification [13]	2.85
Conventional signal domain data hiding [14]	3.79
Conventional bit stream data hiding [15]	3.83
Conventional joint coding and data hiding [16]	3.19
Conventional WTSI [22]	4.05
Proposed method	4.10

5. Conclusion

In consideration of the research, an existing NB telephone network is recommended. The transmitting consisting of the temporal domain of NB signal carries the spreaded spectral envelope parameters of elongated band signals. The information is used to catch the network or the speech signal at the other end.

SPSS technique has been deployed to strengthen the band signal using quantization and channel noises that are found in the spectral envelope parameters. The LSD tests display the improvement of speech quality through the proposed method. On the other hand, the MOS results show that UB information is transparent compared to conventional ideas. The proposed method is more suitable and can be used for improving the bandwidth of prevailing telephone networks without considering the changes within it.

Acknowledgments

References

P. Jax, P. Vary, Bandwidth extension of speech signals: A catalyst for the introduction of wideband speech coding?. IEEE Communications Magazine. 44(5), 106–111 (2006)

P. Jax, Enhancement of bandlimited speech signals: Algorithms and theoretical bounds. Ph.D. thesis, (RWTH Aachen University, 2002)

Y. Qian, P. Kabal, Dual-mode wideband speech recovery from narrowband speech, in Proc. EUROSPEECH 2003, Geneva, September 2003, pp. 1433–1436

S. Vaseghi, E. Zavarehei, Q. Yan, Speech bandwidth extension: Extrapolations of spectral envelop and harmonicity quality of excitation, in Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing (ICASSP), Toulouse, May 2006, pp. 844–847

J. Epps, W.H. Holmes, A new technique for wideband enhancement of coded narrowband speech, in Proc. IEEE Workshop on speech coding, Porvoo, June 1999, pp. 174–176

H.u. Rongqiang, V. Krishnan, D.V. Anderson, Speech bandwidth extension by improved codebook mapping towards increased phonetic classification, in Proc. INTERSPEECH 2005, Lisbon, Portugal, September 2005, pp. 1501–1504

Y. Nakatoh, M. Tsushima, T. Norimatsu, Generation of broadband speech from narrowband speech using piecewise linear mapping, in Proc. EUROSPEECH, Rhodes, Greece, September 1997, pp. 1643-1646

H. Pulakka, U. Remes, K. Palomaki, M. Kurimo, P. Alku, Speech bandwidth extension using gaussian mixture model-based estimation of the highband Mel spectrum, in Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing (ICASSP), Prague, May 2011, pp. 5100–5103

P. Bauer, T. Fingscheidt, An HMM based artificial bandwidth extension evaluated by cross-language training and test, in Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing (ICASSP), Las Vegas, NV, April 2008, pp. 4589–4592

H. Pulakka, P. Alku, Bandwidth extension of telephone speech using a neural network and a filter bank implementation for highband Melspectrum. IEEE Transactions on Audio, Speech, and Language Processing. 19(7), 2170–2183 (2011)

P. Jax, P. Vary, An upper bound on the quality of artificial bandwidth extension of narrowband speech signals, in Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing (ICASSP), Orlando, FL, USA, May 2002, vol. 1(2002), pp. 237–240

S. Chen, H. Leung, Artificial bandwidth extension of telephony speech by data hiding, in Proc. IEEE International Symposium on Circuits and Systems (ISCAS 2005), Kobe, Japan, May 2005, pp. 3151–3154

S. Chen, H. Leung, Speech bandwidth extension by data hiding and phonetic classification, in Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing (ICASSP), Honolulu, HI, April 2007, vol. 4(2007), pp. 593–596

S. Chen, H. Leung, H. Ding, Telephony speech enhancement by data hiding. IEEE Transactions on Instrumentation and Measurement. 56(1), 63-74 (2007)

Z. Chen, C. Zhao, G. Geng, F. Yin, An audio watermark based speech bandwidth extension method. EURASIP Journal on Audio, Speech and Music Processing. 2013(10), 1-8 (2013)

P. Vary, B. Geiser, Steganographic wideband telephony using narrowband speech codecs, in Proc. Asilomar Conference on Signals, Systems, and Computers (ACSSC 2007), Pacific Grove, CA, November 2007, pp.1475-1479

B. Geiser, P. Vary, Backwards Compatible Wideband Telephony in Mobile Networks: CELP Watermarking and Bandwidth Extension, in Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing (ICASSP), Honolulu, HI, April 2007, vol. 4(2007), pp. 533-536

A. Delforouzi, M. Pooyan, Adaptive and Efficient Audio Data Hiding Method in Temporal Domain. In Proceedings of the IEEE International Conference on Information, Communications and Signal Processing, Macau, China, Dec 2009.

H. Ding, Wideband audio over narrowband low-resolution media, in Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing (ICASSP), Montreal, Quebec, Canada, March 2004, pp. 489–492

E. Hansler, G. Schmidt, Speech and Audio Processing in Adverse Environments (Springer, 2008)

B. Iser, W. Minker, G. Schmidt, Bandwidth extension of speech signals (Springer, New York, 2008)

B. Geiser, P. Jax, P. Vary, Artificial bandwidth extension of speech supported by watermark-transmitted side information, in Proc. INTERSPEECH 2005, Lisbon, Portugal, September 2005, pp. 1497–1500

J.G. Proakis, Digital Communications, 2nd edn. (McGraw-Hill, New York, 1989)

E.H. Dinan, E.H. Jabbari, Spreading codes for direct sequence CDMA and wideband CDMA cellular networks. IEEE Communications Magazine. 36(9), 48–54 (1998)

A. Goldsmith, Wireless communications (Cambridge University Press, New York, 2005)

S. Chen, H. Leung, Concurrent data transmission through analog speech channel using data hiding. IEEE Signal Processing Letters. 12(8), 581-584 (2005)

J.S. Garofolo, Getting started with the DARPA TIMIT CD-ROM: An acoustic phonetic continuous speech database, National Institute of Standards and Technology (NIST), Gaithersburg, MD, USA,1988

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

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