Fusion Based Method for Wide Band Speech Reconstruction Over Legacy Telephone Networks

R. Praveen Kumar EMANI [1], Pitchaiah TELAGATHOTI [1], Prasad NIZAMPATANAM [2]

[1] Dept. of Electronics and Communication Engineering, Vignan's Foundation for Science, Technology and Research, Guntur, India
[2] Dept. of Electronics and Communication Engineering, Faculty of Science and Technology (icfaiTech), ICFAI Foundation for Higher Education, Hyderabad, Telengana, India

rpraveenkumar.emani@gmail.com, tp1977_ece@yahoo.com, prasadnit@ifheindia.org

Submitted April 19, 2025 / Accepted September 27, 2025 / Online first October 24, 2025

Abstract. *Public Switched Telephone Networks (PSTNs) are limited to Restricted Band (RB) speech in the 0–4 kHz range, which reduces speech quality compared to Wideband (WB) speech (0–8 kHz). This study proposes a transform-based hybrid steganography technique combining Curvelet Transform (CT) and Fast Fourier Transform (FFT) to enhance RB speech quality while maintaining full PSTNs compatibility. The Curvelet Transform, capable of representing directional and edge-like features across multiple scales and angles, allows efficient capture of speech components such as formant transitions and unvoiced consonants, which are critical for improving quality and intelligibility. In the proposed approach, WB speech is decomposed into RB and Extended Band (4–8 kHz) components. Detailed coefficients of the RB are extracted using CT, and spread parameters of the Extended Band are embedded within the RB signal. Inverse CT and FFT are performed to form a Composite Restricted Band (CRB) signal for transmission. At the receiver, the embedded parameters are recovered to reconstruct the Extended Band, which is merged with the CRB signal to yield a high-quality Reconstructed Wideband (RWB) signal. Tests with participants aged 60–75 years demonstrated superior performance over conventional methods, with strong robustness to channel and quantization noise.*

Keywords

Curvelet transform, fast Fourier transform, hybrid steganography, speech quality, public switched telephone networks, linear predictive coding

1. Introduction

The limited acoustic bandwidth of typical PSTNs, which ranges from 0 to 4 kHz, results in substandard quality and incomprehensible speech. While the voiced consonants of speech sounds can be adequately represented by the RB of conventional PSTNs, the unvoiced consonant sounds that contribute significantly to speech intelligibility, with a frequency range of approximately 0 to 8 kHz, are poorly represented. This inadequate representation results in a degradation in speech quality and clarity, making listening more challenging, particularly for aged individuals.To enhance the listener's audio perception, the WB speech, spanning 0 to 8 kHz, must be transmitted through PSTNs.

The Adaptive Multirate Wide Band (AMR-WB) codecs [1] facilitate WB signal transmission. However, achieving this would require constructing entirely new communication infrastructure capable of handling greater bandwidths. This process is both prohibitively expensive and time-consuming [1]. The major challenge to implement AMR-WB codecs for telephony speech enhancement is the requirement for both transmitting and receiving devices to be WB-compatible, which many existing devices are not. Additionally, network constraints and the coexistence of RB and WB devices often result in suboptimal call quality or call failures. It is, therefore, desirable to enhance the bandwidth without modifying the infrastructure of the existing PSTNs [2]. One approach is Artificial Bandwidth Extension (ABE), where WB signals are regenerated by approximating the EB signal from the RB signal. ABE techniques are inspired by the interdependencies between RB and the EB signal, as indicated by the speech production model. The correlation between RB and EB signals supports the predicted values of the EB determined from the RB input.

The ABE techniques derived from the widely used conventional source-filter approximation involve estimating the vocal tract filter that remodels the speech source's spectral structure and its original excitation signal. Multiple approaches of EB excitation techniques namely non-linear processing techniques like modulation of fundamental frequency of speech segment [2], Gaussian noise modulation [3], excitation harmonic modeling [4], and sinusoidal synthesis [5], were reviewed. Historically, ABE methods employ techniques such as, codeword and matrix mapping [6], [7], Gaussian Mixture and Hidden Markov Models [8], [9], to en-

DOI: 10.13164/re.2025.0591

hance the limited frequency range of RB speech, improving intelligibility and quality. These methods, while effective, often struggled with robustness and adaptability to varied acoustic environments. In contrast, recent advancements in deep learning have revolutionized ABE by leveraging Neural Networks (NN) [10] to model complex speech patterns and features.

A sequential Deep Neural Network (DNN) proposed in [11] was used as a Bandwidth Extension (BE) tool to estimate and adjust the spectral energy of the EB signal by applying spectral folding and utilizing log-power spectral features. A method to enhance the RB speech quality using a deep learning model that utilizes the Wasserstein Generative Adversarial Network (WGAN) architecture was proposed in [12]. Unlike traditional models, this approach improves the reconstruction of the EB signal, resulting in more natural and realistic WB audio by combining adversarial learning with a perceptual loss function to achieve better generalization and output quality. A multi-scale GAN for BE proposed in [13] effectively improves the naturalness of the speech and its intelligibility, hence increasing its perceptual quality. The proposed sound enhancement network in [14] effectively reconstructs a WB speech signal using adversarial losses. A speech production model-based feed-forward NN with an H∞ optimization algorithm speech BE method was introduced in [15]. The proposed DNN architecture in [16] produces an explicit WB speech from a noisy RB speech by understanding the correlation between the log-power magnitude of RB and EB spectral components. A chronological codec network design proposed in [17] consists of an encoder-decoder network and Long Short-Term Memory (LSTM) for speech BE. A phase-sensible magnitude and complex branch NN architecture for BE of speech in the spectral domain is suggested in [18]. A low computational complexity blind BE method is proposed in [19] that uses a network architecture based on LPCnet. An analysis of a deep residual full-convolutional NN was presented in [20], and an LSTM network along with his previous work based on mixed-BE systems for BE extension of RB speech signal. Though these modern techniques provide superior quality in Reconstructed Wideband (RWB) signals, they might not always match the original signal. Thus, the reconstructed signal may sound unnatural. As discussed in [21], the best achievable estimate of the speech envelope by an ABE algorithm is limited due to the abundant dependencies between the RB and EB speech signals. ABE depends on a speaker-reliant trained database that not only is highly computationally demanding and costly for real-time processing but also is virtually impractical to construct, as indicated in [22]. In summary, it can be stated that the conventional ABE methods, which rely solely on estimation, often suffer from inaccurate spectral estimations and degraded audio quality. Therefore, it is necessary to explore alternative methods to address the limitations of the conventional ABE approach. A similar approach, except that the EB speech is embedded as side information within the RB signal, enhances the reconstruction of EB speech at the receiver. This approach improves audio fidelity, reduces artifacts, and ensures better performance, especially in PSTNs or lossy transmission environments, without requiring additional bandwidth.

Transmitting EB of the WB speech as side information allows accurate estimation of the speech spectral envelope [1]. Algorithms that use data-hiding techniques to conceal this side information in the RB signal are suggested in [23] to preserve the backward compatibility of current RB networks. A few such data-hiding schemes that offer better performance than conventional methods were discussed. The original WB is band-split, and the EB components are encoded and integrated inside the RB signal. The data-hiding approach described in [22] involves estimating Auto Regression (AR) coefficients, encoding them, and converting them into line spectrum pairs. These line spectrum pairs are integrated into the RB signal to create a CRB speech signal, and an RWB speech with better perceptual quality is reconstructed by extracting the hidden signal at the receiver side. However, poor CRB signal quality was observed with this method. To enhance the standard of the CRB signal and RWB speech, [24] endorsed acoustic phonetic grouping to more effectively encode the EB signal. Even so, the methods in [22], [24] produced poor BE performance when distorted by channel and spectral disturbances. A distinctive method of BE was suggested in [25] by hiding the perceptually key frequency components within the RB signal without compromising the speech quality. This is accomplished by separating the imperceptible frequency components in the RB signal and creating a hidden channel to hide audible frequency components within the hidden channel. By extracting the hidden components, a perceptually improved speech can be reconstructed at the reception side. However, the performance of the scheme is bounded by the channel noise and the number of significant audible components of missing frequencies of RB that can be integrated into the hidden channel. A modified LSB watermark approach-based BE technique is suggested in [26], where the parameters from the essential EB signal of the WB speech are extracted, compressed, and incorporated into the RB signal bit stream by an altered watermark scheme and transmitted to the receiver for reconstruction.

A backward-compatible transmission system to transmit the extra information embedded into the RB codec bit stream by joint coding and data-hiding is proposed in [27] for BE. Another backward compatibility with respect to PSTNs was proposed in [28] with a low-bit-rate BE algorithm based on the GSM Enhanced Full Rate (EFR) coder, and the extra information is incorporated into the EFR bit stream. A distinct BE approach was suggested in [29], where the key proposition is pitch-scaling of the EB signal of speech and placing them in the typical telephone speech unexploited frequencies.

A different concept on transmitting the side information is introduced in [30]. This technique uses the Linear Predictive Coding (LPC) technique to obtain attributes like

the envelope and gain of the EB signal from the WB signal, and these encoded components are embedded into the GSM FR 06.10 RB speech coder bit stream using joint source coding and data hiding methods. These attributes are extracted at the receiver and, along with the reconstructed lower frequencies, efficiently reproduce a WB signal. A similar approach, except that the high-frequency component extraction is based on both LPC and Mel-frequency cepstral coefficients techniques, is proposed in [31]. A BE algorithm built on an AMR RB coder offers backward-compatible transmission, which is asserted to ease the constantly changing channel circumstances, and is introduced in [32]. The RB signal is coded using Adaptive Differential Pulse Code Modulation in [33], and parameterization of the EB signal is done to transmit them together to reduce the bit rate while preserving the quality of the speech. A BE scheme associated with modeling sparse linear prediction is proposed in [34]. The LSB-based digital audio steganography approach was presented in [35], where the encoded classified data is inserted into the coefficients of the RB in the integer wavelet domain. A similar integer wavelet transform-based fully compatible BE technique was introduced in [36], where the retrieved EB signal by LPC is inserted into the wavelet coefficients of the RB speech signal obtained from the integer wavelet transform (IWT) [37]. This method generates lossless data hiding; in other words, accurate WB speech is recreated at the recipient's end. IWT enhances the perceptibility of the speech signals, making them closer to WB quality. Speech BE techniques utilizing the DWTFFTDH method and DWT-DCT-based data hiding technique have been proposed in [38], [39]. BWE techniques employing Discrete Hartley Transform (BWE-DHT) and Spectral Data Masking (BWE-SDM) are proposed in [40], [41] The encoded spectral envelope parameters of the EB signal are concealed within the intricate coefficients of the RB signal; these techniques also lack resilience towards channel and quantization distortions. The existing speech BE algorithms using data-hiding techniques struggle to deliver high-quality CRB and RWB signals. A distinctive ABE algorithm based on data hiding is needed to address these limitations by improving signal quality and making it resilient to noise.

A novel speech steganography technique [42] using a CT-FFT-based data hiding technique embeds secret speech specifications into detailed coefficients of original host speech. This method preserves the host signal's quality and ensures that the covert speech signal can be retrieved without degradation.

In digital telephone networks, speech bandwidth extension is typically applied only to the higher frequency range, specifically the Extended Band (4–8 kHz), since transmitting low-frequency components is generally not an issue in such networks [19–21]. A hybrid BE technique that integrates CT and FFT is proposed in this paper. At the transmitter side, this technique embeds the spread parameters of EB into the RB signal. The RB signal is first decomposed using a CT, and embedding occurs within the detailed coefficients to generate a CRB signal. On the receiver side, the hidden signal is extracted by performing the exact reverse operation, and a WB signal with considerably higher quality is reconstructed by extracting and adding the EB signal to the RB signal. The proposed approach utilizes actual EB information rather than estimated values, resulting in more accurate wideband (WB) speech reconstruction compared to traditional artificial bandwidth extension (ABE) techniques. Moreover, it remains fully compatible with standard PSTN systems. This means that conventional narrowband (NB) receivers can still decode the NB speech without requiring any hardware modifications, while specially designed receivers can extract the embedded data to reconstruct a significantly higher-quality WB signal.

The effects of the Telephone Network Channel (TNC) impairments, including quantization and channel noise, were discussed in [43]. Unlike prior methods proposed in [22] and [25] that address quantization noise but overlook channel noise, this approach employs the Spread Spectrum (SS) technique [44] for robust data recovery against channel and quantization noises. Additionally, adaptive channel equalization is applied to mitigate the effects of fading. Each parameter of the EB signal is spread using a specific sequence, and the resulting spreading vectors are combined to create hidden data. The minimal cross-correlation among spreading sequences ensures reliable recovery of the hidden data, even in noisy conditions. To minimize interference among the concealed components, the spreading sequences with the least cross-correlation are chosen. Hadamard codes are employed in this work as they are orthogonal to one another and are known for their performance in cross-correlation. Unlike m-sequences, Gold codes, and Kasami codes that exhibit varying cross-correlation properties [45–47], Hadamard codes effectively minimize interference, ensuring better data embedding and recovery. This paper also incorporates adaptive channel equalization [26] to counteract performance degradation caused by fading in TNCs. The Recursive Least Squares (RLS) algorithm is chosen for this purpose due to its fast convergence and strong tracking capability. Table 1 presents a comparison between the proposed and conventional methods considering key aspects such as robustness, quality of the RWB signal, and compatibility with PSTNs.

The article is organized as follows: Section 2 addresses the novel SBE technique employing CT and FFT. The subjective and objective assessments are covered in Sec. 3. Section 4 presents the conclusion.

Technique	Robustness	RWB signal quality	PSTN compatibility
AMR-WB CODECS [1]	High	Excellent	No
GMM [8]	Low	Low–Moderate	No
HMM [9]	Low	Low–Moderate	No
Neural networks [10]	Low	Low–Moderate	No
Proposed method	High	Good	Yes

Tab. 1. Evaluation of AMR-WB codecs, GMM, HMM, neural networks, and proposed method across key metrics.

2. Method

The proposed method enables the transmission of WB speech over legacy telephone networks by embedding high-frequency components into the low-frequency band using a CT-FFT based hybrid watermarking approach. The system is organized into three main stages: transmitter, channel modeling, and receiver.

2.1 Transmitter

The transmitter introduced is shown in Fig. 1. To ease the analysis, such as for time-frequency analysis of speech, the WB signal is divided into frames, each with a frame duration of 20 ms. Initiate band-splitting on each frame; this serves in the isolation of the different components of speech for analysis. The band splitter separates the low- and high-frequency components of the WB speech signal. The low-frequency 0 to 4 kHz components of the speech signal are down sampled by 2 to provide an RB signal, represented by $x_{\mathrm{RB}}(n)$, and the high-frequency components 4 to 8 kHz are also down sampled by 2 to provide an EB signal, represented by $x_{\mathrm{EB}}(n)$. This helps in efficient processing of RB and EB signals by reducing the computational load. The CT is well suited for non-stationary signals like speech since it allows good frequency and time-domain localization [48]. The CT represents directional and edge-like features across multiple scales and angles, enabling more efficient representation of speech components such as formant transitions and unvoiced consonants, which are essential for increasing speech quality and intelligibility. For this reason, CT is employed in this proposed method. The basic 2D representation of a speech signal in the curvelet domain, $S(t, f)$, is expressed as the sum of curvelets and is given by

$$S(t,f) = \sum_{r,d,p} C_{r,d,p} \cdot \phi_{r,d,p}(t,f) \tag{1}$$

where $\phi_{(r,d,p)}(t,f)$ is the curvelet basis function at scale (r), orientation (d), and spatial location (p). $C_{r,d,p}$ is the curvelet coefficient that evaluates the significance of each basis function.

The RB signal $x_{\mathrm{RB}}(n)$, is decomposed by applying CT into its constituent frequency components, i.e., Approximation and Detail Coefficients (DCs). Subsequently, FFT is applied to the DCs, which enables the analysis of the frequency spectrum, facilitating the computation of both magnitude $|x_{\mathrm{RB}}(k)|$ and phase $\phi_{\mathrm{RB}}(k)$ spectra, essential for accurately representing the signal's frequency characteristics. LPC [25] is applied to the EB signal to obtain Line Spectral Frequencies (LSFs) and relative gain factors [47], which compactly represent the EB information.The LSFs and the relative gain are used to form a representation vector. The $x_{\mathrm{EB}}(n)$ is characterized by 11 parameters, and to embed them, each of these parameters is multiplied by a unique Pseudo-random Noise (PN) code of length M, denoted as $\mathrm{CP}_i \cdot \vec{o}^{\,i}$, i for $1 \le i \le M$ to ensure that the embedded information is imperceptible while remaining robust to distortions. Summing these spread signal vectors produces the hidden data for embedding and is given by

$$D(p) = \sum_{i=1}^{M} \mathrm{CP}_i \, o^i(p). \tag{2}$$

The hidden data $D(p)$ are incorporated into the final 16 elements of the initial half of the $|x_{\mathrm{RB}}(k)|$ [38], resulting in a modified magnitude spectrum $|\hat{x}_{\mathrm{RB}}(k)|$, is given by

$$|\hat{x}_{\mathrm{RB}}(k)| = \begin{cases} |x_{\mathrm{RB}}(k)|, & k = 0, \ldots, \frac{L}{2} - 16 \\ D(p), & k = \frac{L}{2} - 15, \ldots, \frac{L}{2} - 1 \\ D(p), & k = \frac{L}{2}, \ldots, \frac{L}{2} + 16 \\ |x_{\mathrm{RB}}(k)|, & k = \frac{L}{2} + 17, \ldots, L - 1 \end{cases}. \tag{3}$$

The modified signal $\hat{x}_{\mathrm{RB}}(k)$ is the Composite Restricted Band (CRB) signal and represented in terms of its magnitude $|\hat{x}_{\mathrm{RB}}(k)|$ and phase spectrum $\hat{\phi}_{\mathrm{RB}}(k)$ as

$$\hat{x}_{\mathrm{RB}}(k) = |\hat{x}_{\mathrm{RB}}(k)| \, e^{j\hat{\phi}_{\mathrm{RB}}(k)}, \quad k = 0, \ldots, L-1. \tag{4}$$

To reconstruct the time-domain CRB signal $\hat{x}_{\mathrm{RB}}(n)$, an inverse FFT followed by an inverse CT to its spectrum is applied and transmitted over the TNC. During transmission, the channel introduces channel and quantization noises, which can degrade the signal quality, making the embedded hidden data more susceptible to corruption. The CRB signal, communicated through the TNC, is faded and noisy. If the proposed speech BE receiver operate on the faded signal, it would give a large Mean Square Error (MSE). To compensate for the channel fading, various channel equalization methods have already been proposed in the literature [37]. An equalizer that performs better, in terms of a lower MSE,

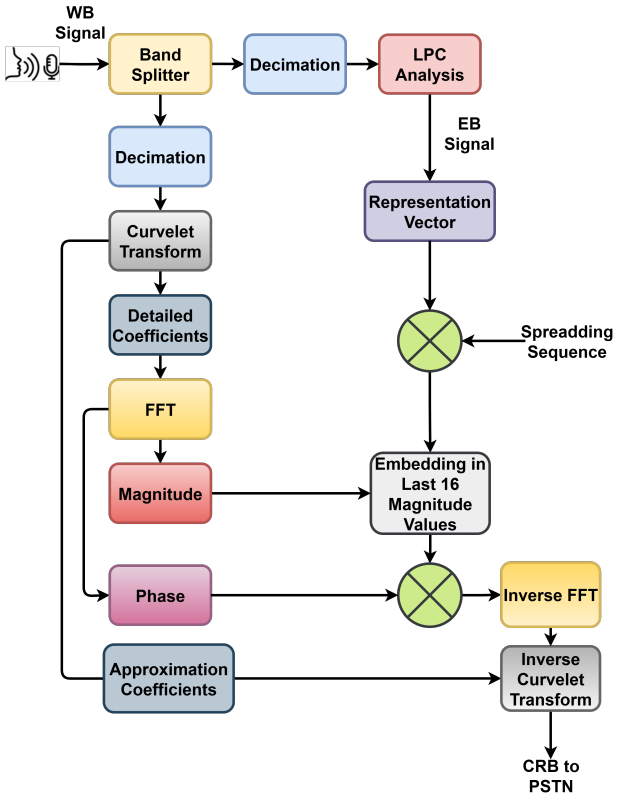


Fig. 1. Transmitter block diagram.

will usually result in accurate embedded information extraction. RLS algorithm is an adaptive filtering technique that estimates the parameters of a model by minimizing a weighted sum of squared errors and updating the estimate as new data arrives [49]. It is particularly effective when data is received sequentially and system parameters are time-varying. RLS iteratively refines its parameter estimates without recalculating the full solution, making it widely used for channel equalization in communication systems. Its principle is to adaptively adjust equalizer coefficients to minimize the squared error between the desired (transmitted) and equalized received signals, thereby compensating for channel distortions. As a recursive method, it converges much faster than the Least Mean Squares algorithm so RLS algorithm is used for channel equalization. Hence, in this paper, the RLS algorithm with 512 taps was employed for channel equalization. $\hat{x}^{\mathrm{r}}_{\mathrm{RB}}(n)$ represent the received signal at the receiver along with the noise; therefore $\hat{x}^{\mathrm{r}}_{\mathrm{RB}}(n) = \hat{x}_{\mathrm{RB}}(n) + E$, E is the error introduced by channel and quantization noises.

2.2 Receiver

The receiver introduced is shown in Fig. 2. To retrieve the hidden data $\hat{x}_{\mathrm{EB}}(n)$ from $\hat{x}^{\mathrm{r}}_{\mathrm{RB}}(n)$,the receiver first applies a CT to $\hat{x}^{\mathrm{r}}_{\mathrm{RB}}(n)$, to decompose the signal, followed by FFT on the DCs to obtain their magnitude and phase spectra, facilitating the extraction of the embedded data.

The spread parameters are extracted from the magnitude spectrum of $\hat{x}^{\mathrm{r}}_{\mathrm{RB}}(n)$ and then de-spread using a correlator [40]. For a given parameter, the corresponding retrieved correlation, denoted as $\widehat{CP}_{io}$, is computed as:

$$\widehat{(CP)}_{io} = \frac{1}{M}\sum_{p=1}^{M} \widehat{D}(p)\, o^{io}(p). \tag{5}$$

Then de-spreading is done using a correlator that uses the same PN sequence that is used at the transmitter during embedding. This isolates the desired parameter from the uncorrelated noise and the interference from the non-orthogonal components ensuring robust and faithful recovery of the embedded data. $\widehat{D}(p)$ is the noise-contaminated form of $D(p)$ which is given by

$$\widehat{D}(p) = D(p) + E(p). \tag{6}$$

By substituting (6) into (5), we get

$$\begin{aligned}
\widehat{(CP)}_{io} &= \frac{1}{M}\sum_{p=1}^{M} \widehat{D}(p)\, o^{io}(p) \\
&= \frac{1}{M}\sum_{p=1}^{M} (D(p) + E(p))\, o^{io}(p) \\
&= \frac{1}{M}\sum_{p=1}^{M} o^{io}(p) \left(\sum_{i=1}^{M} CP_i\, o^{i}(p) + E(p)\right) \\
&= \frac{1}{M}\sum_{p=1}^{M} o^{io}(p) \left(CP_{io}\, o^{io}(p) + \sum_{i\neq io} CP_i\, o^{i}(p) + E(p)\right).
\end{aligned} \tag{7}$$

Since the PN sequences are orthogonal,

$\sum_{p=1}^{M} o^{i}(p)\, o^{io}(p) = 0$ for all $i \neq io$.

Therefore, from (7), the term

$$\begin{aligned}
\sum_{p=1}^{M}\sum_{i\neq io} CP_i\, o^{i}(p)\, o^{io}(p) &= \sum_{i\neq io} CP_i \left(\sum_{p=1}^{M} o^{i}(p)\, o^{io}(p)\right) \\
&= 0.
\end{aligned} \tag{8}$$

Also, since PN sequence $o^{io}(p)$ and noise, $E(p)$ are uncorrelated, the term

$$\frac{1}{M}\sum_{p=1}^{M} o^{io}(p)\, E(p) = 0. \tag{9}$$

When the length of the PN sequence i.e., $M \to \infty$. Substitute (8) and (9) in (7), we get

$$\widehat{(CP)}_{io} = CP_{io}. \tag{10}$$

This highlights that the representation vector parameters $X_{\mathrm{EB}}(n)$ can be efficiently reconstructed using the SS technique. These LPCs are applied to the synthesis filter to obtain the Reconstructed EB (REB) signal. Both CRB and REB signals are interpolated and added to obtain the RWB signal with better quality so that it significantly improves the quality and intelligibility of the speech for the elderly.

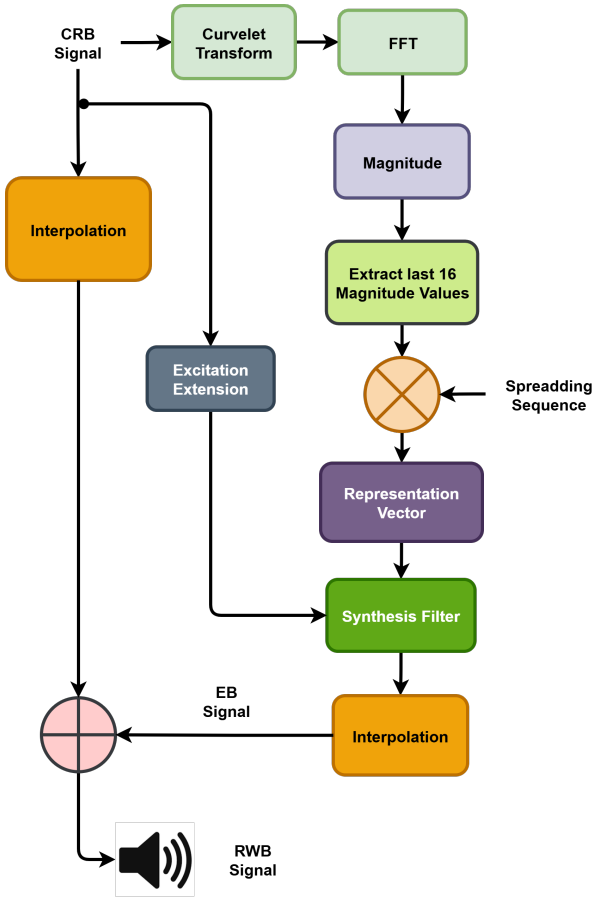


Fig. 2. Receiver block diagram.

3. Experimental Results

The quality evaluation techniques to assess the proposed method mainly considered two parameters. One is a good RWB quality at the receiver and maintains the RB speech quality even after embedding the spread parameters of EB into the RB signal compared to conventional methods. The evaluation of the proposed technique was carried out using speech transcripts from the Nippon Telegraph and Telephone (NTT) Corporation speech database [50]. The effectiveness of the recommended BE technique is assessed through both subjective and objective evaluation metrics. For comparative analysis, the proposed approach is compared with several existing conventional techniques [22, 24, 26, 30, 37, 40, 41, 45]. Furthermore, two telephony channel models μ-law and Additive White Gaussian Noise (AWGN) are considered in this study to assess the robustness of the proposed method under different transmission conditions.

3.1 Opinion-Based Listening Assessment Results

Perceptual transparency is determined using the Mean Opinion Score (MOS) test [26], [51]. A subjective evaluation of WB, RB, CRB, and RWB signals was also conducted [28]. The speech quality achieved by the suggested approach was checked against the outputs of conventional methods [22, 24, 26, 30, 37, 40, 41, 45] using a Pairwise Comparison Listening Test (PCLT) [52] endorsed by the ITU-T. Eight hundred distinct speech transcripts, each lasting two to five seconds and created by 40 male and female participants, were used to evaluate the efficacy of the suggested method. Subjects aged 60 to 75 assessed speech transcripts of RB, CRB, and RWB produced by both the suggested and traditional approaches [22, 24, 26, 30, 37, 40, 41, 45]. Elderly people were selected as subjects due to age-related auditory deterioration, which progressively affects hearing abilities, making older adults an optimal group to evaluate improvements in speech quality.

3.2 Perceptual Transparency

The EB signal parameters must be concealed cleanly by the suggested approach. That is, CRB and RB must be completely identical. The proposed CT method provides strong energy compaction and effectively captures directional features, helping to preserve important signal characteristics. This leads to high perceptual transparency, with minimal or no noticeable degradation of the RB speech signal. Perceptual transparency is determined by employing the MOS assessment [22, 24, 53]. The participants in the test analyze CRB and RB and articulate their views in terms of the MOS, as tabulated in Tab. 2. The study was performed in a soundproof environment with participants wearing headsets. Forty respondents aged between 60 and 75 years took part in the assessment. Figure 3 presents the Mean Opinion Score (MOS) results for the traditional methods [22, 24, 26, 30, 37, 40, 41, 45] and the proposed method,

Score	Instruction
1	Zero similarity between the RB signal and CRB signal
2	Very little similarity between the RB and CRB signals
3	RB and CRB signals are nearly equal
4	Both sound the same.

Tab. 2. MOS comparison.

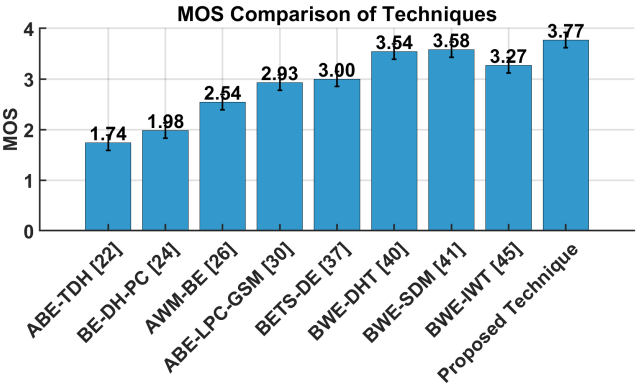


Fig. 3. MOS scores of speech quality with the 95 percent CI.

along with 95% Confidence Intervals (CI) for each technique. From the figure, it can be observed that the proposed technique demonstrates superior perceptual transparency compared to traditional methods, achieving a MOS value of 3.77 nearly equivalent to the conventional MOS value of 4 indicating that both the CRB and RB signals are perceptually indistinguishable. The confidence intervals further reflect the reliability of the subjective test results.

3.3 Subjective comparisons between WB, RB, CRB, and RWB Speech transcripts

A PCLT was carried out to assess the efficiency of the suggested approach against conventional approaches utilizing a μ-law channel model. Forty elderly participants took part in the experiment. In every instance, two speech transcripts selected from options WB, RB, CRB, and RWB are presented to the participants. They are requested to indicate whether the first transcript of the set was superior, inferior, or identical to the remaining 3 samples. Table 3 presents the outcomes of the comparison between WB and others, where $>$, $<$, and $\approx$ signify the participant's opinion that transcript WB is superior to, inferior to, or identical to RB, CRB, and RWB, respectively.

The pairwise comparisons between WB, RB, CRB, and RWB are presented in Tabs. 3–5. The tables display the total count of participants with a specific preference using Arabic numerals. From Tab. 3, it can be observed that the WB signal is superior when compared to the RB and CRB signals for traditional methods [22, 24, 26, 30, 37, 40, 41, 45] and the proposed method. Furthermore, we note that the quality of the RWB signal is significantly enhanced for the elderly people using the proposed method compared to traditional methods [22, 24, 26, 30, 37, 40, 41, 45].

Method	WB	RB	CRB	RWB
ABE-TDH [22]	>	40	40	28
	<	0	0	0
	≈	0	0	12
BE-DH-PC [24]	>	40	39	28
	<	0	0	0
	≈	0	1	12
AWM-BE [26]	>	40	40	24
	<	0	0	0
	≈	0	0	16
ABE-LPC-GSM [30]	>	39	40	16
	<	0	0	0
	≈	1	0	24
BETS-DE [37]	>	40	38	10
	<	0	0	0
	≈	0	2	30
BWE-DHT [40]	>	38	39	5
	<	0	0	0
	≈	2	1	35
BWE-SDM [41]	>	40	37	5
	<	0	0	0
	≈	0	3	35
BWE-IWT [45]	>	39	39	7
	<	0	0	0
	≈	1	1	33
Proposed method	>	40	40	3
	<	0	0	0
	≈	0	0	37

Tab. 3. Results of the subjective listening test comparisons of WB with RB, CRB, and RWB.

Method	RB	CRB	RWB
ABE-TDH [22]	>	18	28
	<	6	6
	≈	16	6
BE-DH-PC [24]	>	10	25
	<	6	9
	≈	24	6
AWM-BE [26]	>	18	24
	<	6	6
	≈	16	10
ABE-LPC-GSM [30]	>	12	13
	<	10	20
	≈	18	7
BETS-DE [37]	>	10	8
	<	5	22
	≈	25	10
BWE-DHT [40]	>	8	7
	<	4	23
	≈	28	10
BWE-SDM [41]	>	9	5
	<	6	27
	≈	25	8
BWE-IWT [45]	>	9	7
	<	6	25
	≈	25	8
Proposed method	>	6	2
	<	1	33
	≈	33	5

Tab. 4. Results of the subjective listening test comparisons of RB with CRB and RWB.

Method	CRB	RWB
ABE-TDH [22]	>	12
	<	20
	≈	8
BE-DH-PC [24]	>	8
	<	22
	≈	10
AWM-BE [26]	>	8
	<	21
	≈	11
ABE-LPC-GSM [30]	>	9
	<	22
	≈	9
BETS-DE [37]	>	8
	<	26
	≈	6
BWE-DHT [40]	>	4
	<	28
	≈	8
BWE-SDM [41]	>	3
	<	30
	≈	7
BWE-IWT [45]	>	4
	<	31
	≈	5
Proposed method	>	1
	<	36
	≈	3

Tab. 5. Results of the subjective listening test comparisons between CRB and RWB.

From Tab. 4, it is evident that when compared to conventional methods [22, 24, 26, 30, 37, 40, 41, 45], the RWB signal generated by the suggested technique demonstrates premier quality when compared to the RB signal. Compared to conventional approaches [22, 24, 26, 30, 37, 40, 41, 45], the proposed method generated a CRB signal that is more identical to the RB signal.The RWB speech produced by the suggested approach surpasses the CRB speech. This is indicated in Tab. 5. These subjective evaluations demonstrated that the speech quality and clarity for older adults were improved.

3.4 Objective Test Results

Objective quality evaluations are employed for a deeper examination of the suggested approach. Eight hundred speech transcripts articulated by forty male and forty female speakers are extracted from the NTT corporation database [50] for objective assessments. The achieved WB speech quality is evaluated by applying the Log Spectral Distortion (LSD) metric [54], [55] and WB-Perceptual Evaluation of Speech Quality (WB-PESQ). The perceptual transparency is assessed utilizing the ITU-T PESQ metric [56]. The MSE assesses the resilience of concealed data to quantization and noisy channels [25].

3.5 WB Speech Quality

The perceptual resemblance between WB and RWB signals is assessed by the LSD metric. The LSD value is calculated effectively from the FFT power spectra across the frequency band spanning 4 to 8 kHz as

$$\mathrm{LSD}(j) = \sqrt{\frac{1}{R_{\mathrm{high}} - R_{\mathrm{low}} + 1} \sum_{n=R_{\mathrm{low}}}^{R_{\mathrm{high}}} \left[10 \log_{10} \left(\frac{p_j(n)}{\widehat{p}_{jk}(n)} \right) \right]^2} \quad (11)$$

where $\mathrm{LSD}(j)$ is the LSD value of the j^{th} frame, and $\widehat{p}_{jk}(n)$ is the FFT power spectra of the original EB signal and, $(\widehat{p}_{jk})(n)$is FFT power spectra of the REB signal. R_{high} and R_{low} are the indices that correspond to the upper and lower bounds of the EB signal. The mean LSD, which is calculated by averaging the obtained LSD scores of all frames of every speech sample, is utilized as the metric. A low LSD score signifies superior quality of the REB signal. The μ-law channel model LSD assessment results for the traditional techniques [22, 24, 26, 30, 37, 40, 41, 45] and the proposed technique were tabulated in Tab. 6. The AWGN channel model-based LSD assessment was conducted and tabulated in Tab. 7.

Tables 6 and 7 show a clear improvement in the REB signal quality, which assures generating an RWB signal with superior quality with the suggested method in contrast to the classical techniques [22, 24, 26, 30, 37, 40, 41, 45].

Figure 4 illustrates the spectrograms of the WB, RB, and RWB speech generated by the proposed method. The RWB spectrogram closely matches the original WB spectrogram, particularly in the high-frequency regions between 4–8 kHz that are absent in conventional RB speech. This visual evidence supports the subjective and objective test results, confirming the effectiveness of the method in achieving near-wideband perceptual quality. Thus, the suggested approach markedly improved speech quality for elderly people.

3.6 WB-PESQ

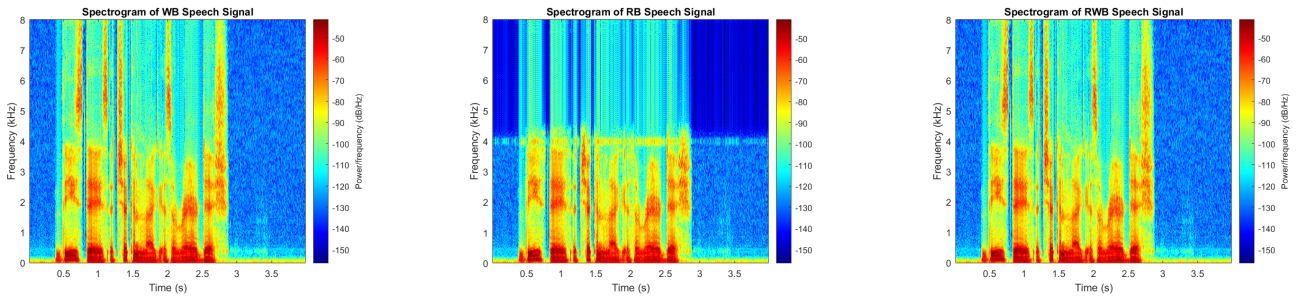
The quality of RWB speech is assessed by comparing WB and RWB signals. Table 8 presents the average WB-PESQ values for conventional [22, 24, 26, 30, 37, 40, 41, 45] and proposed methods. A WB-PESQ score of 3.97 indicates that the proposed technique achieves superior speech quality, which benefits elderly people, as confirmed by subjective listening tests. Considering these enhancements in clarity and comprehensibility for older adults, the methodology is more effective, for younger people as well. The proposed method yields MOS improvements in the range of +0.19 to +0.50 and PESQ gains of +0.12 to +0.38 compared to the best-performing baseline methods. These gains are not only statistically significant but also perceptually meaningful, especially for elderly listeners. Unlike transform-only or prediction-only (ABE) approaches in prior works, the hybrid proposed method, capable of representing directional and edge-like features across multiple scales and angles, allows efficient capture of speech components such as formant transitions and unvoiced consonants, which are critical for improving quality and intelligibility. Thus This framework better reconstructs the missing EB, over the conventional methods resulting in enhanced speech clarity and naturalness. Thus, the superiority of the proposed method is evident not only in objective scores but also in subjective evaluations, demonstrating its effectiveness over existing techniques [22, 24, 26, 30, 37, 40, 41, 45].

Method	LSD
ABE-TDH [22]	14.12
BE-DH-PC [24]	12.38
AWM-BE [26]	8.74
ABE-LPC-GSM [30]	7.77
BETS-DE [37]	4.79
BWE-DHT [40]	2.34
BWE-SDM [41]	2.08
BWE-IWT [45]	2.56
Proposed method	1.97

Tab. 6. μ-law channel model LSD assessment results.

Method	LSD
ABE-TDH [22]	13.72
BE-DH-PC [24]	13.11
AWM-BE [26]	11.93
ABE-LPC-GSM [30]	10.55
BETS-DE [37]	8.04
BWE-DHT [40]	2.94
BWE-SDM [41]	2.68
BWE-IWT [45]	3.14
Proposed method	2.02

Tab. 7. AWGN channel model LSD results.



(a) Spectrogram of WB signal (b) Spectrogram of RB signal (c) Spectrogram of RWB signal

Fig. 4. Spectrograms of WB, RB, and RWB signals.

Method	WB-PESQ Score
ABE-TDH [22]	1.83
BE-DH-PC [24]	2.53
AWM-BE [26]	2.63
ABE-LPC-GSM [30]	3.02
BETS-DE [37]	3.47
BWE-DHT [40]	3.84
BWE-SDM [41]	3.90
BWE-IWT [45]	3.82
Proposed method	3.97

Tab. 8. WB-PESQ scores.

Method	RB-PESQ score
ABE-TDH [22]	1.32
BE-DH-PC [24]	1.79
AWM-BE [26]	2.13
ABE-LPC-GSM [30]	2.78
BETS-DE [37]	3.06
BWE-DHT [40]	3.64
BWE-SDM [41]	3.78
BWE-IWT [45]	3.47
Proposed method	3.93

Tab. 9. RB-PESQ scores.

3.7 Perceptual Transparency

The perceptual transparency is evaluated by comparing RB and CRB input signals. In general, the RB-PESQ score will be between 0.5 and 4.5. The larger the value, the better its quality. Table 9 provides the RB-PESQ values for classical techniques [22, 24, 26, 30, 37, 40, 41, 45] and the suggested approaches. The suggested technique achieves an RB-PESQ score of 3.93, signifying superior perceptual transparency when compared to the traditional methods [22, 24, 26, 30, 37, 40, 41, 45], as validated by subjective listening assessments.

3.8 Resilience of Concealed Data

Subsequently, we examine the impact of noise distortion. To the CRB signal, AWGN having an SNR value ranging between 15 and 35 dB is added [57]. To evaluate the robustness of the suggested technique, evaluation is done by calculating the MSE value between the RWB signal $x_{\mathrm{RWB}}(n)$ and WB signal $x_{\mathrm{WB}}(n)$. MSE is calculated by

$$\mathrm{MSE} = \frac{1}{K}\sum_{k=0}^{K-1}\left(x_{\mathrm{RWB}}(k) - x_{\mathrm{WB}}(k)\right)^2 \tag{12}$$

where the value of K is the product of the sampling rate and the duration of the speech transcript. The received CRB signal is a combination of the transmitted signal, channel noise, quantization noise, and spectral distortions. In this proposed framework, the SS technique inherently suppresses the effects of channel distortion and quantization noise, while the RLS-based adaptive equalizer effectively compensates for spectral distortion. The lower MSE values observed under fading conditions are due to the application of the RLS-based channel equalizer. In the absence of equalization, the reconstructed WB signal deviates more from the transmitted WB due to spectral distortions, resulting in higher MSE. When equalization is employed, it effectively compensates for channel distortions, ensuring that the reconstructed WB signal closely resembles the transmitted WB. This compensation leads to reduced MSE values, even in the presence of fading. The suggested approach achieves MSE values below 2.95×10^{-5} in the SNR range of 15–35 dB without incorporating the fading channel model. When the fading channel model is introduced, the MSE further decreases to below 6.80×10^{-8} across the same SNR range, indicating that the proposed technique is more resilient to channel noise, quantization noise, and fading effects. The lower MSE value of 6.80×10^{-8} under fading conditions clearly demonstrates the improvement in the quality of the reconstructed signal.

3.9 Correlation Analysis

The Pearson correlation coefficient (r) was computed using the formula to evaluate the extent of linear concordance between subjective and objective quality judgments.

$$r = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2} \cdot \sqrt{\sum_{i=1}^{n}(y_i - \bar{y})^2}} \tag{13}$$

where x_i and y_i represent individual values of subjective and objective scores, respectively, $\bar{x}$ and $\bar{y}$ denote their means. If $r \approx 1$, there is a strong positive correlation; if $r \approx 0$, there is no correlation; and if $r \approx -1$, there is a strong negative correlation. The Pearson correlation coefficient of $r = 0.98$ for the proposed and conventional methods shows that there is strong linear concordance between subjective and objective quality judgments.

While the Pearson correlation measures the linear similarity between the original and reconstructed speech signals, Spearman rank correlation evaluates the structural fidelity between subjective and objective measures. The rank-based analysis ensures that the monotonic patterns of the speech signal like spectral envelopes, formant trajectories and temporal envelopes which are responsible for speech intelligibility and naturalness . In this paper, Spearman correlation was computed between the subjective and the objective scores to assess structural consistancy between subjective and objective measures. The process involves assigning ranks to each variable in the subjective and objective measures, calculating the rank differences for each pair. The Spearman rank coefficient, S_{r} is given by

$$S_{\mathrm{r}} = 1 - \frac{6\sum_{j=1}^{N} D_j^2}{N\left(N^2 - 1\right)} \tag{14}$$

where D_i is the difference in the ranks for the j^{th} value, N is the number of values. The Spearman rank coefficient S_{r} ranges from -1 to 1, where -1 represents reverse correlation, 0 indicates no correlation, and 1 represents a perfect

monotonic relationship between subjective perception and objective measurements. In our experiments, $S_r = 1$ was obtained for both the proposed and conventional methods, confirming structural consistency across perceptual and objective evaluations.

3.10 Complexity of Proposed Method

The traditional BE techniques utilize a segment of G.729 to execute LPC at the transmission end. The encoded AR coefficients are integrated into the RB speech using a straightforward quantization technique. The embedded information is retrieved at the receiving end with a minimum-distance decoder. The extraction of the embedded information is likewise conducted as a component of G.729. No intricate calculations are required. Consequently, traditional BE approaches are viable for real-time processing.

The proposed technique utilizes LPC at the transmitting end to extract spectral envelope parameters, which are disseminated via PN codes and subsequently incorporated in the DCs of the RB signal. Adaptive channel equalization is utilized at the receiving end to mitigate the impact of channel fading, after which the embedded information is retrieved using a standard correlator, commonly applied in telecommunications. The suggested system involves no complex computations, making it simpler to implement and suitable for real-time processing, akin to traditional SBE techniques.

In comparison to traditional speech BE techniques [12], [13], the proposed method demonstrates significantly superior speech signal quality, enhanced perceptual transparency, and resilience against quantization noise, channel noise, and channel spectral distortion, while also fulfilling real-time processing requirements. Consequently, the suggested system offers a pragmatic and efficient resolution to speech BE.

4. Conclusion

This paper introduces a novel approach for enhancing the quality of RB speech transmitted over PSTNs through a hybrid steganography technique based on CT and FFT. The method embeds the spread parameters of EB speech within RB speech, ensuring full compatibility with existing PSTNs infrastructure while improving overall speech quality. The reconstruction of the WB signal at the receiver side demonstrates notable improvements in intelligibility and perceptual quality, offering significant benefits for elderly individuals with age related hearing loss. Both subjective and objective evaluations indicate that the RWB speech obtained from the proposed method surpasses the quality of the RWB speech obtained by traditional methods in terms of clarity, naturalness, and listener satisfaction. The technique is also highly suited for real-time deployment on FPGAs or ASICs with minimal latency and can be integrated into existing networks or mobile devices without major modifications, enabling efficient bandwidth enhancement. Further research may investigate the potential of the suggested method to improve its robustness by implementing it on noisy WB speech signals.

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] GAJJAR, P., BHATT, N., KOSTA, Y. Artificial bandwidth extension of speech and its applications in wireless communication systems: A review. In *Proceedings of the 2012 International Conference on Communication Systems and Network Technologies (CSNT)*. Rajkot (India), 2012, p. 563–568. DOI: 10.1109/CSNT.2012.127

[3] QIAN, Y., KABAL, P. Dual-mode wideband speech recovery from narrowband speech. In *Proceedings of European Conference on Speech Communication and Technology*. Geneva (Switzerland), 2003, p. 1433–1436. DOI: 10.21437/EUROSPEECH.2003-250

[4] VASEGHI, S. S., ZAVAREHEI, E., YAN, Q., et al. Speech bandwidth extension: Extrapolations of spectral envelop and harmonicity quality of excitation. In *IEEE International Conference on Acoustics, Speech and Signal Processing Proceedings*. Toulouse (France), 2006, p. 1–4. DOI: 10.1109/ICASSP.2006.1660786

[5] EPPS, J., HOLMES, W. H. A new technique for wideband enhancement of coded narrowband speech. In *IEEE Workshop on Speech Coding Proceedings: Model, Coders, and Error Criteria (Cat. No.99EX351)*. Porvoo (Finland), 1999, p. 174–176. DOI: 10.1109/SCFT.1999.781522

[6] HU, R., KRISHNAN, V., ANDERSON, D. V. Speech bandwidth extension by improved codebook mapping towards increased phonetic classification. In *Proceedings of Interspeech*. Lisbon (Portugal), 2005, p. 1501–1504. DOI: 10.21437/INTERSPEECH.2005-527

[7] NAKATOH, Y., TSUSHIMA, M., NORIMATSU, T. Generation of broadband speech from narrowband speech using piecewise linear mapping. In *Proceedings of the 5th European Conference on Speech Communication and Technology (EUROSPEECH)*. Rhodes (Greece), 1997, p. 1643–1646. DOI: 10.21437/EUROSPEECH.1997-469

[8] PULAKKA, H., REMES, U., PALOMAKI, K., et al. Speech bandwidth extension using Gaussian mixture model-based estimation of the highband mel spectrum. In *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. Prague (Czech Republic), 2011, p. 5100–5103. DOI: 10.1109/ICASSP.2011.5947504

[9] BAUER, P., FINGSCHEIDT, T. An HMM-based artificial bandwidth extension evaluated by cross-language training and test. In *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. Las Vegas (NV, USA), 2008, p. 4589–4592. DOI: 10.1109/ICASSP.2008.4518678

[10] PULAKKA, H., ALKU, P. Bandwidth extension of telephone speech using a neural network and a filter bank implementation for highband mel spectrum. *IEEE Transactions on Audio, Speech, and Language Processing*, 2011, vol. 19, no. 7, p. 2170–2183. DOI: 10.1109/TASL.2011.2118206

[11] 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

[12] CHEN, X., YANG, J. Speech bandwidth extension based on Wasserstein generative adversarial network. In *Proceedings of the IEEE 21st International Conference on Communication Technology (ICCT)*. Tianjin (China), 2021, p. 1356–1362. DOI: 10.1109/ICCT52962.2021.9658055

[13] LIN, H., YANG, J., CHEN, X. Multi-scale generative adversarial network for speech bandwidth extension. In *IEEE 22nd International Conference on Communication Technology (ICCT)*. Nanjing (China), 2022, p. 1189–1193. DOI: 10.1109/ICCT56141.2022.10073331

[14] LI, Y., TAGLIASACCHI, M., RYBAKOV, O., et al. Real-time speech frequency bandwidth extension. In *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. Toronto (Canada), 2021, p. 691–695. DOI: 10.1109/ICASSP39728.2021.9413439

[15] GUPTA, D., SHEKHAWAT, H. S. Artificial bandwidth extension using H∞ optimization, deep neural network, and speech production model. In *Proceedings of the IEEE International Conference on Signal Processing and Communications (SPCOM)*. Bangalore (India), 2022, p. 1–5. DOI: 10.1109/SPCOM55316.2022.9840805

[16] TAHER, T., MAMUN, N., HOSSAIN, M. A. A joint bandwidth expansion and speech enhancement approach using deep neural network. In *International Conference on Electrical, Computer and Communication Engineering (ECCE)*. Chittagong (Bangladesh), 2023, p. 1–4. DOI: 10.1109/ECCE57851.2023.10101546

[17] XU, C.-D., LING, X.-P., YING, D.-W. Codec network for speech bandwidth extension. In *IEEE 2nd International Conference on Big Data, Artificial Intelligence and Internet of Things Engineering (ICBAIE)*. Nanchang (China), 2021, p. 387–391. DOI: 10.1109/ICBAIE52039.2021.9389968

[18] RU, J., JIA, M., ZHAO, Y., et al. A dual-branch neural network for phase-aware speech bandwidth extension. In *6th International Conference on Information Communication and Signal Processing (ICICSP)*. Xi'an (China), 2023, p. 634–638. DOI: 10.1109/ICICSP59554.2023.10390706

[19] SCHMIDT, K., EDLER, B. Blind bandwidth extension of speech based on LPCNet. In *28th European Signal Processing Conference (EUSIPCO)*. Amsterdam (Netherlands), 2021, p. 426–430. DOI: 10.23919/EUSIPCO47968.2020.9287465

[20] NIDADAVOLU, P. S., IGLESIAS, V., VILLALBA, J., et al. Investigation on neural bandwidth extension of telephone speech for improved speaker recognition. In *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. Brighton (UK), 2019, p. 6111–6115. DOI: 10.1109/ICASSP.2019.8682992

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

[22] CHEN, S., LEUNG, H. Artificial bandwidth extension of telephony speech by data hiding. In *IEEE International Symposium on Circuits and Systems (ISCAS)*. Kobe (Japan), 2005, vol. 4, p. 3151–3154. DOI: 10.1109/ISCAS.2005.1465296

[23] GAJJAR, P., BHATT, N., KOSTA, Y. Artificial bandwidth extension of speech and its applications in wireless communication systems: A review. In *Proceedings of the International Conference on Communication Systems and Network Technology (CSNT)*. Rajkot (India), 2012, p. 563–568. DOI: 10.1109/CSNT.2012.127

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

[25] CHEN, S. 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

[26] 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. 1–10. DOI: 10.1186/1687-4722-2013-10

[27] VARY, P., GEISER, B. Steganographic wideband telephony using narrowband speech codecs. In *Conference Record of the Forty-First Asilomar Conference on Signals, Systems and Computers (ACSSC)*. Pacific Grove (CA, USA), 2007, p. 1475–1479. DOI: 10.1109/ACSSC.2007.4487475

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

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

[30] 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, p. 57–64. DOI: 10.1007/s10772-014-9249-1

[31] BHATT, N. S. 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, p. 881–893. DOI: 10.1007/s10772-016-9378-9

[32] BHATT, N. S. Implementation and overall performance evaluation of CELP based GSM AMR NB coder over ABE. In *Fifth International Conference on Communication Systems and Network Technologies (CSNT)*. Gwalior (India), 2015, p. 402–406. DOI: 10.1109/CSNT.2015.46

[33] GHADERI, M., SAVOJI, M. H. Wideband speech coding using ADPCM and a new enhanced bandwidth extension method. In *IEEE 7th International Symposium on Intelligent Signal Processing (WISP)*. Floriana (Malta), 2011, p. 1–4. DOI: 10.1109/WISP.2011.6051691

[34] ALIPOOR, G., SAVOJI, M. H. Wide-band speech coding based on bandwidth extension and sparse linear prediction. In *35th International Conference on Telecommunications and Signal Processing (TSP)*. Prague (Czech Republic), 2012, p. 454–459. DOI: 10.1109/TSP.2012.6256335

[35] DELFOROUZI, A., POOYAN, M. Adaptive digital audio steganography based on integer wavelet transform. In *Third International Conference on Intelligent Information Hiding and Multimedia Signal Processing (IIH-MSP 2007)*. Kaohsiung (Taiwan), 2007, p. 283–286. DOI: 10.1109/IIH-MSP.2007.69

[36] NIZAMPATNAM, P., RAJU, G. R. L. V. N. S. Transform domain speech bandwidth extension. *Circuits, Systems, and Signal Processing*, 2019, vol. 38, no. 12, p. 5717–5733. DOI: 10.1007/s00034-019-01142-w

[37] SAGI, A., MALAH, D. Bandwidth extension of telephone speech aided by data embedding. *EURASIP Journal on Advances in Signal Processing*, 2007, p. 1–16. DOI: 10.1155/2007/64921

[38] KURADA, P., MARUVADA, S., SANAGAPALLEA, K. R. Speech bandwidth extension using DWT-FFT-based data hiding. *Radioengineering*, 2020, vol. 29, no. 1, p. 174–181. DOI: 10.13164/re.2020.0174

[39] KODURI, S. K., KUMAR, T. DWT-DCT based data hiding for speech bandwidth extension. *Radioengineering*, 2021, vol. 30, no. 2, p. 435–442. DOI: 10.13164/re.2021.0435

[40] HOSODA, Y., KAWAMURA, A., IIGUNI, Y. Speech bandwidth extension using data hiding based on discrete Hartley transform domain. *Circuits, Systems, and Signal Processing*, 2022, vol. 41, p. 2290–2307. DOI: 10.1007/s00034-021-01893-5

[41] PRASAD, N. Speech bandwidth extension using spectral data masking. *Cognitive Computing and Cyber Physical Systems*, 2025, vol. 598, p. 84–93. DOI: 10.1007/978-3-031-77078-4

[42] REKIK, S., GUERCHI, D., SELOUANI, S., et al. Speech steganography using wavelet and Fourier transforms. *Journal of Audio, Speech, and Music Processing*, 2012, vol. 20, p. 1–14. DOI: 10.1186/1687-4722-2012-20

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

[44] HASSAN, A. A., HERSHEY, J. E., SAULNIER, G. J. *Perspectives in Spread Spectrum*. Boston (MA, USA): Kluwer Academic Publishers, 1998. DOI: 10.1007/978-1-4615-5531-5

[45] GOLDSMITH, A. *Wireless Communications*. Cambridge (U.K.): Cambridge University Press, 2005. DOI: 10.1017/CBO9780511841224

[46] PRASAD, N., KISHORE KUMAR, T. Speech bandwidth extension aided by magnitude spectrum data hiding. *Circuits, Systems, and Signal Processing*, 2017, vol. 36, p. 4512–4540. DOI: 10.1007/s00034-017-0526-5

[47] NIZAMPATNAM, P., TAPPETA, K. 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

[48] FAROUK, M. H. *Application of Wavelets in Speech Processing*. Cham (Switzerland): Springer International Publishing, 2013. DOI: 10.1007/978-3-319-02732-6

[49] LINGHUI, W., WEI, H., KAIHONG, Z., et al. Adaptive channel equalization based on RLS algorithm. In *International Conference on System Science, Engineering Design and Manufacturing Informatization*. Guiyang (China), 2011, p. 105–108. DOI: 10.1109/ICSSEM.2011.6081250

[50] NTT ADVANCED TECHNOLOGY CORPORATION. *MultiLingual Speech Database for Telephonometry*. 2021, [Online]. Available at: https://www.rd.ntt/e/research/HI0002.html

[51] DING, H. Wideband audio over narrowband low-resolution media. In *IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*. Montreal (QC, Canada), 2004, p. 489–492. DOI: 10.1109/ICASSP.2004.1326029

[52] LAMBLIN, C., QUINQUIS, C., USAI, P. ITU-T G.722.1 Annex C: The first ITU-T superwideband audio coder. *IEEE Communications Magazine*, 2008, vol. 46, no. 10, p. 116–122. DOI: 10.1109/MCOM.2008.4644128

[53] 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

[54] LIU, X., BAO, C. Audio bandwidth extension based on ensemble echo state networks with temporal evolution. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, 2016, vol. 24, no. 3, p. 594–607. DOI: 10.1109/TASLP.2016.2519146

[55] PULAKKA, H., LAAKSONEN, L., VAINIO, M., et al. Evaluation of an artificial speech bandwidth extension method in three languages. *IEEE Transactions on Audio, Speech, and Language Processing*, 2008, vol. 16, no. 6, p. 1124–1137. DOI: 10.1109/TASL.2008.925149

[56] ITU-T RECOMMENDATION. *Perceptual Evaluation of Speech Quality (PESQ): An Objective Method for End-to-End Speech Quality Assessment of Narrow-Band Telephone Networks and Speech Codecs.* Rec. ITU-T P. 862, 2001.

[57] KEISER, B. E., STRANGE, E. *Digital Telephony and Network Integration*. Boston (MA, USA): Springer US, 2012. DOI: 10.1007/978-1-4615-1787-0

About the Authors ...

R. Praveen Kumar EMANI received his B.Tech degree in Electronics and Communication Engineering from Al-Ameer College of Engineering and Information Technology, Visakhapatnam, India, in 2006, followed by an M.Tech degree in Electronics and Communication Engineering from Gudlavalleru Engineering College, India, in 2009. He is currently pursuing research as a scholar in the Department of Electronics and Communication Engineering at Vignan's Foundation for Science, Technology, and Research, India, with a focus on speech processing.

Pitchaiah TELAGATHOTI received his Ph.D. degree from the Department of Electronics and Communication Engineering at Andhra University, Andhra Pradesh, India, in 2019. He is currently a Professor in the Department of Electronics and Communication Engineering at Vignan's Foundation for Science, Technology, and Research, Guntur, Andhra Pradesh, India. His research interests include signal processing and 5G communications.

Prasad NIZAMPATNAM earned his doctoral degree in the Department of Electronics and Communication Engineering at NIT Warangal, specializing in speech processing. He is currently serving as an Assistant Professor in the Department of Electronics and Communication Engineering at ICFAI Tech, ICFAI Foundation for Higher Education, Hyderabad, Telangana, India.