Development of Novel Robust Speech Bandwidth Extension Techniques

Submitted in partial fulfillment of the requirements
for the award of the degree of

DOCTOR OF PHILOSOPHY

by

K SUNIL KUMAR

(RollNo.718147)

Under the Supervision of
Prof. T.KISHOREKUMAR

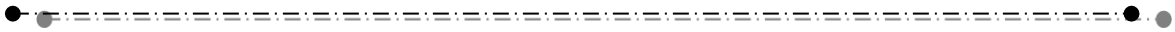
Professor

DEPARTMENTOFELECTRONICSANDCOMMUNICATIONENGINEERING
NATIONAL INSTITUTE OF TECHNOLOGY
WARANGAL–506004,T.S, INDIA

DECEMBER –2022

Dedicated to my beloved

Teachers, Parents, Children and wife



APPROVAL SHEET

This thesis entitled "**Development of Novel Robust Speech Bandwidth Extension Techniques"** by **Mr. K Sunil Kumar** is approved for the degree of **Doctor of Philosophy**.

Examiners

_____

_____

Supervisor

_____

Prof. T. Kishore Kumar

Professor, Electronics and Communication Engineering

Department, NIT WARANGAL

Chairman

_____

Prof. P. Sreehari Rao

Head, Electronics and Communication Engineering Department,

NIT WARANGAL

Date:

Place:

DECLARATION

I, hereby, declare that the matter embodied in this thesis entitled "**Development of Novel Robust Speech Bandwidth Extension Techniques"** is based entirely on the results of the investigations and research work carried out by me under the supervision of **Prof. T. Kishore Kumar**, Department of Electronics and Communication Engineering, National Institute of Technology Warangal. I declare that this work is original and has not been submitted in part or full, for any degree or diploma to this or any other University.

I declare that this written submission represents my ideas in my own words and where other ideas or words have been included. I have adequately cited and referenced the original sources. I also declare that I have adhered to all principles of academic honesty and integrity and have not misrepresented or fabricated or falsified any idea/date/fact/source in my submission. I understand that any violation of the above will cause for disciplinary action by the institute and can also evoke penal action from the sources which have thus not been properly cited or from whom proper permission has not been taken when needed.

K Sunil Kumar

Roll No: 718147

Date: 21/12/2022

Place: Warangal

Department of Electronics and Communication Engineering
National Institute of Technology
Warangal–506004,Telangana,India

CERTIFICATE

This is to certify that the dissertation work entitled **"Development of Novel Robust Speech Bandwidth Extension Techniques"**, which is being submitted by Mr. K Sunil Kumar (Roll No. 718147), a bonafide work submitted to National Institute of Technology Warangal in partial fulfilment of the requirement for the award of the degree of Doctor of Philosophy to the Department of Electronics and Communication Engineering of National Institute of Technology Warangal, is a record of bonafide research work carried out by her under my supervision and has not been submitted elsewhere for any degree.

Dr.T.KishoreKumar
(Supervisor)
Professor, Department of ECE
National Institute of Technology
Warangal, India – 506004

ACKNOWLEDGEMENTS

I would like to thank a number of people who have contributed to my PhD directly or indirectly and in different ways through their help, support and encouragement.

It gives me immense pleasure to express my deep sense of gratitude and thanks to my supervisor **Prof. T. Kishore Kumar** (National Institute of Technology Warangal, NITW),for his invaluable guidance, support and suggestions. His knowledge, suggestions, and discussions helped me to become a capable researcher. He has shown me the interesting side of this wonderful multidisciplinary area and guided me to get profound knowledge as well as publications in this area.

I am thankful to the current Head of the Dept. of ECE, **Prof. P. Sreehari Rao**, and the former Heads, **Prof. L. Anjaneyulu , Prof. N. Bheema Rao** and **Prof. T. Kishore Kumar** for giving me the opportunity and all the necessary support from the department to carry out my research work.

I thank **Prof. N.V. Ramana Rao**, Director, N.I.T. Warangal, for providing official support and financial assistance to carry out the research.

I take this privilege to thank all my Doctoral Scrutiny Committee members, **Prof. Vinod Kumar D M,** Department of EEE, **Prof.L. Anjaneyulu**, Department of Electronics and Communication Engineering and **Dr. G. Arun Kumar**, Assistant Professor, Department of Electronics and Communication Engineering for their detailed review, constructive suggestions and excellent advice during the progress of thisresearch work.

Special thanks to my seniors Sudeep Surendran and Sunnydayal V for their motivation, suggestions during publishing papers and being extremely supportive throughout my PhD period.

I take this opportunity to convey my regards to my speech lab-mates, NIT Warangal, Rakesh P, B surekha Reddy, Prasad Nizampatnam, Ravi Bolimera, S Siva Priyanka, M Sumeetha, M Manoj kumar and A Govind,for being always present next to me in time of need.

I thank my department co-scholars for being the very supportive friends. I also appreciate the help rendered from teaching, non-teaching members and fraternity of Dept. of ECE of NIT Warangal. They have always been encouraging and supportive.

I acknowledge my gratitude to all my teachers and colleagues at various aspects for supporting and co-operating to complete this work.

Finally, I appreciate and respect my family members (my father Mr K. Ramannna, my mother Mrs. K Rama Devi ,my son K. Vishnu Tej, my daughter K.Meenakshi and my beloved wife K Veena) for being very supportive while giving me mental support and inspiration that motivated me to complete the thesis work successfully. Especially, I thank my wife who has been a strong support throughout my PhD period.

SUNIL KUMAR KODURI

ABSTRACT

As of today, a major part of the world's telephone networks is still unusual to the frequency of human speech signals. When a human speech signal is transmitted through the telephone network leads to losing information due to the limited bandwidth of the telephone network. Due to historical and economic reasons, the frequency range of telephone voice is confined to around 300–3400 Hz, known as "limited narrowband" (LNB) speech. This leads to a thin and reduces the speech signal's naturalness. On the other hand, a high-frequency loss above the LNB reduces the clarity of fricative sounds such as 's' and 'f', resulting in muffled speech. Speech quality degrades due to the constrained frequency range, making it difficult for the listener to follow what is being said on the other end of the line. In addition, several characteristics that are exclusive to a speaker are removed.

For an obvious better speech quality and a new sense of presence, a Clear Wideband (CWB) speech is Significant and has a frequency range of 50-7000 Hz, which would improve the quality, intelligibility, and perceived naturalness of the spoken signal compared to the transmission of LNB speech. As a result, new telephone networks with higher bandwidths must be built, which will be costly and take some time to establish. By applying speech bandwidth extension (SBE) techniques, we may increase the receiving end's speech quality without having to make changes to the current telephone network's architecture.

The existing telephone networks can benefit greatly from an improvement in speech quality due to SBE technology. Artificial speech bandwidth extension (ASBE) is one approach that reconstructs the CWB signal only from the LNB signal by estimating out-of-band (the frequencies below and above the LNB) information. The speech production model reveals the dependency between LNB and out-of-band, which is the basis for ASBE approaches. There is sufficient evidence to support the use of LNB input to estimate out-of-band information. However, the intrinsic performance limitations of ASBE approaches prevent them from regenerating high-quality CWB speech [3]. Out-of-band information can be provided together with an LNB signal to improve the quality of

CWB speech significantly compared to ASBE approaches. Data concealing technique would be employed to hide out-of-band information in the LNB signal to guarantee required backward compatibility with current telephone networks. The SBE employing data concealing strategies is the focus of this thesis.

In this thesis, a significant focus is on creating and assessing innovative SBE algorithms that use data masking strategies. Composite LNB (CLNB) speech and reconstructed CWB (RCWB) speech quality can be improved using new techniques that are more robust to noises, such as channel and quantization noises (CAQNs). At frequencies above LNB, new SBE approaches are introduced. The SBE methods utilizing hybrid transform-based data hiding, frequency-domain data hiding, discrete wavelet transform-discrete cosine transform-based data hiding with encoding, and discrete cosine transform-based data hiding have all been developed and analyzed.

The performance of these methods has been evaluated using subjective and objective measures. The results from the subjective and objective evaluations of the speech bandwidth extension techniques presented in this thesis show a clear speech quality improvement of the proposed methods over the conventional speech bandwidth extension techniques. In a mean opinion score (MOS) subjective listening test, we verified that the proposed methods yield improved perceptual transparency compared to conventional methods. The log spectral distortion values obtained showed that the proposed techniques yield an improved speech signal quality compared to traditional methods.

CONTENTS

4 Speech Bandwidth Extension using DWT-DCT based Data Hiding

5 Speech Bandwidth aided by Frequency domain based Data Hiding

LIST OF FIGURES

Fig. No. Title…..……..……..... ….. Pg No.

LIST OF TABLES

Table No. Title…..……..……..……...…Pg No.

LIST OF ABREVATIONS

ASBE : Artificial speech bandwidth extension
ADPCM : Adaptive delta pulse code modulation
AMF : Adaptive modulation frequency
AMR : Adaptive multi rate
AMR-CWB : Adaptive multirate- clearwideband
AMT : Auditory masking threshold
AR : Auto regressive
AWGN : Additive white Gaussian noise
BER : Bit error rate
BSDH : Bit stream data hiding
BWE : Band width extension
CAQNs :Channel and Quantization Noises
CCR : Comparison category rating
CDH : Conventional data hiding
CEF : Complementary error function
CELP : Code excited linear prediction
CM : Codebook mapping
CMOS : Comparison mean opinion score
CPDF : Conditional probability density function
CWB : Clear Wideband
CWB- PESQ : Clear Wideband-Perceptual evaluation of speech quality
DFT : Discrete Fourier Transform
DS-CDMA : Direct-sequence code-division multiple access
DCT-BDH :Discrete cosine Transform Based Data Hiding
DWT-DCT-BDH : Discrete wavelet Transform -Discrete cosine Transform Based Data Hiding
DWT-DCT-BDHWE : Discrete wavelet Transform -Discrete cosine Transform Based Data Hiding with encoding
DWM: :Digital watermarking
DWT : Discrete wavelet transform
EFR : Enhanced Full-Rate
EM : Expectation maximization
ETSI : European telecommunications standards institute
EWBSC : Embedded wideband speech codec
FDF : Frequency-Domain Features

FDBDH : frequency domain-based data hiding
FE : Feature Extraction
FFT : fast Fourier transform
FIR : Finite impulse response
FR : Full-Rate
FT : Fourier transform
GMM :Gaussian mixture model
GSM : Global system for mobile communication
HMTDBDH : Hybrid model Transform-domain based data hiding
HMM : Hidden Markov model
HPF : High pass filter
IFFT : Inverse fast Fourier transform
ITU-T : International telecommunication union – telecommunication sector
IWCs : Integer wavelet coefficients
CCDH : Combined coding and data hiding
LB : Low band
LPC : Linear predictive coding
LPF : Low-pass filter
LPCs : Linear predictive coefficients
LSB : Least significant bit
LSBs : Least significant bits
LSD : Log-spectral distortion
LSFs : Line-spectral frequencies
MFCCs : Mel-frequency cepstral coefficients
MHB :Missing Highband
MOS : Mean opinion score
MSE : Mean square error
LNB : LimitedNarrowband
LNB- PESQ : Limited Narrowband-Perceptual evaluation of speech quality
NN : Neural network
NNs : Neural networks
PCM : Pulse-code modulation
PDF : Probability density function
PESQ : Perceptual evaluation of speech quality
PN : Pseudo-noise
PSD : Power spectral density
QN : Quantization noise

SBEDH : Speech bandwidth extension using data hiding
SC :Spectral centroid
SD : Spectral distortion
SF : Spectral folding
SFM :Source-filter model
SNR : Signal-to-noise ratio
SPEs : Spectral Envelop Parameters
SPSEs : Spreading sequences
SS : Spread spectrum
SISY : Sinsuiodal synthesis
TD : Time-domain
TDF : Time-Domain Features
TM : Tonal masker
UQ : Uniform quantization
VQ : Vector quantization

LIST OF SYMBOLS

f_0	Fundamental frequency
f_s	Sampling frequency
y_{spc}	Spectral centroid
y_{sf}	Spectral flatness
E_{fre}	Signal energy
y_{grix}	Gradient index
f_{gau}	Gaussian component densities
$S_{lnb}(n)$	Limited Narrowband signal
$S_{mhb}(n)$	Missing High band signal
$\|S_{lnb}(k)\|$	Magnitude spectrum of limited narrowband signal
$\Phi_{lnb}(k).$	Phase spectrum of limited narrowband signal
$S_{lnb}^{1}(n)$	Composite limited narrowband signal
$S_{cwb}(n)$	Original clear wideband signal
$S_{cwb}^{1}(n))$	Reconstructed clear wideband signal
g_p	Gain
p_{error}	Probability of detection error
$g'(n)$	Quantization noise

Chapter-1

1.1. Introduction

Most telephone networks are not usual to the frequency of human speech signals, and the frequency of the human speech signal is much beyond the telephone network's capacity. For economical and historical reasons, the telephony speech frequency range is confined to around 0.3–3.4kHz, known as limited narrowband (LNB) speech. Due to this, the transmission of the human voice over telephone networks reduces the high-frequency speech spectrum, which leads to a thin, unnatural-sounding voice and reduces the speech signal's naturalness. Conversely, a high-frequency loss above the LNB reduces the clarity of fricative sounds such as s, f, and k, which results in muffled speech. Furthermore, several distinct characteristics that are exclusive to a speaker are removed. Speech quality degrades due to the constrained frequency range, making it difficult for the listener to comprehend what is being said on the other end of the line.

For better speech quality and a new sense of presence, a Clear Wideband (CWB) speech is essential, which has a frequency range of 0.5-7kHz and could improve the intelligibility, quality, and perceived naturalness of the spoken speech signal compared to the LNB speech. As a result, a new telephone network channel (TNC) with higher bandwidth must be built, which will be costly and take significant time to establish [1]. By applying speech bandwidth extension (SBE) techniques [2], the quality of the received signal can be improved without making changes to the current telephone network's architecture.

Utilizing the pre-existing telephone infrastructure technology enhances the quality of telephone communication. Artificial speech bandwidth extension (ASBE) is one approach that reconstructs the CWB signal only from the LNB signal by predicting out-of-band (OOB) information (3.4 kHz above and 0.3 kHz below). The dependency between LNB and OOB revealed by the speech production mechanism is the basis for ASBE approaches. There is sufficient evidence to support the use of LNB input to estimate information that is not in the LNB band. However, the intrinsic performance limitations of the ASBE approach prevent them from regenerating high-quality CWB speech [3].

A portion of OOB information can be provided together with an LNB signal to significantly increase the quality of CWB speech compared to ASBE approaches [1, 4]. Data concealing approaches are employed to hide supplementary OOB information in the LNB signal in order to guarantee required backward compatibility with current telephone networks. The SBE employing data-concealing approaches is the main focus of this dissertation.

In this thesis, a significant focus is on creating and assessing innovative SBE algorithms that use a data-hiding strategy. At frequencies above the range of the LNB (traditional telephone band), new SBE approaches have been introduced. Composite LNB (CLNB) speech and reconstructed CWB (RCWB) speech quality is improved using new techniques that are more resilient to disturbances, such as channel and quantization noises (CAQNs). SBE methods utilizing hybrid model transform-based data hiding (HMTBDH), discrete wavelet transform-discrete Cosine transform-based data hiding (DWT-DCT-BDH),frequency-domain-based data hiding (FDBDH), and discrete cosine transform-based data hiding (DCTBDH) have all been designed and analyzed in this thesis and have been tested.

1.2. Human Speech production Scheme

Figure 1.1 depicts the human speech production scheme. The significant parts of the system are the oral cavity (mouth), trachea (air flow pipe), lungs, larynx (voice production organ), pharyngal cavity (throat), and nasal cavity (nose). Usually, the Pharyngeal- and oral cavities are referred to as the vocal tract, and the nasal cavity is the nasal tract. The Vocal cords, trachea, and vocal tract all get air from the lungs. The vocal tract and glottis are two different terms for the aperture between the vocal folds and the acoustic tube that extends above them.

The vocal folds vibrate, and a pulsating air flow occurs when voiced sounds are pronounced. The fundamental frequency f_0 is determined by the rate of vocal fold vibration. Unvoiced sounds are formed by making a limit at some point in the vocal tract. Varying the vocal tract shape results in different characteristics for distinct speech sounds. The vocal tract resonances, called the formants, alter by adjusting the vocal tract shape and subsequently change the sound spectrum.

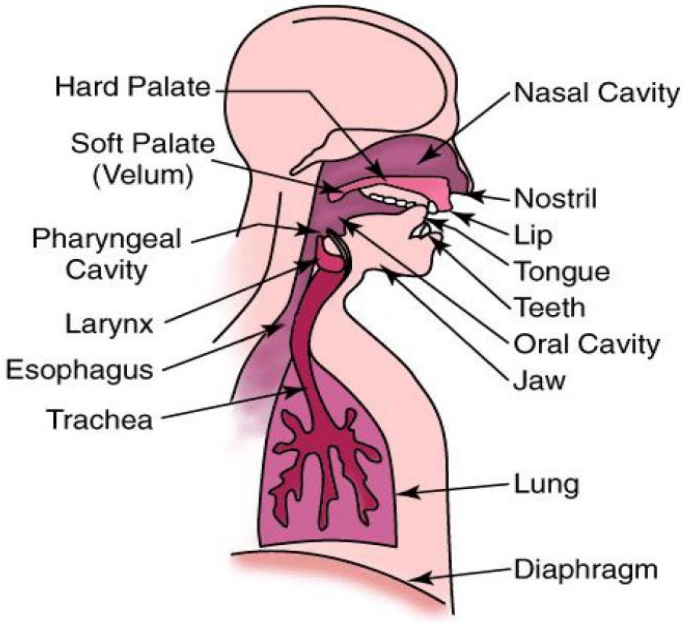


Figure 1.1: Schematic illustration of the human speech production

A certain speech sound is produced depending on the shape of the cavities and the position of the tongue, lips, jaw, etc. Speech sounds may be divided into numerous groups. Vowels and consonants are the two main categories of sounds. Vowels are voiced sounds formed when air flows freely through the vocal tract. Consonants can be unvoiced or voiced. Consonants are generated with complete or partial closure at a specific place in the vocal tract [5]. Vocalization, articulation location, and articulation style are used to categorize consonants [6] further. The following significant categories describe the consonants according to the manner of articulation.

- Block the airflow and suddenly release the air-producing plosives (/k/, /p/, /t/).
- By blocking the oral cavity and opening the passage to the nasal cavity would generate nasals (/m/, /n/).
- Fricatives (/s/, /f/) are generated by a turbulent, noise-like air flow through a constriction in the vocal tract.
- Approximants (/v/, /l/) are also produced by a relatively narrow constriction in the vocal tract.
- Closures of the vocal tract having a concise duration produce taps and flaps, and a series of rapid closures produce trills (/r/).

1.3. Nature Of Human speech signals

The most common approach to receiving human speech signals is face-to-face with human speech across the whole frequency range that is observable to the auditory system [7]. Fig. 1.2 (a) depicts the estimated PSD of a male voice sample recorded at a sampling rate (f_s) of 44.1 kHz. This signal is visible in the graphic, with a power of up to 22.05 kHz. Fig. 1.2 (b) depicts the estimated PSD of a female voice sample recorded at a sampling rate of 96 kHz. The LNB signal (ordinary telephone conversation) bandwidth of 0.3-3.4kHz is substantially smaller than what one would perceive in face-to-face interaction with a sound source.

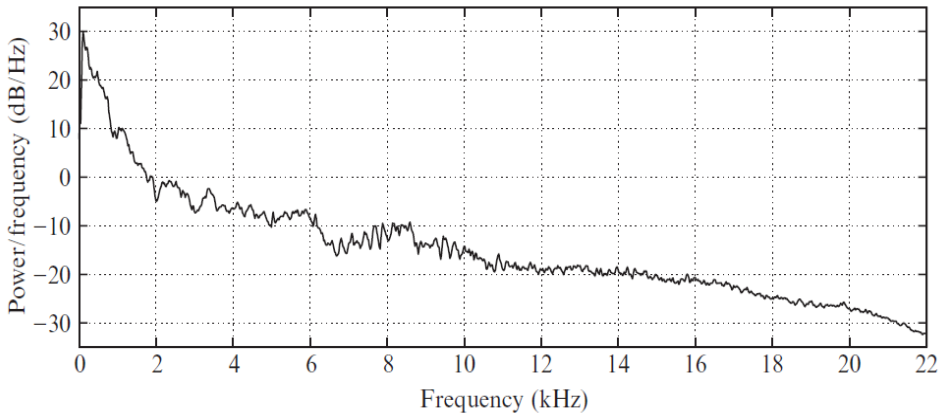


Figure 1.2 (a): Estimated PSD of a malespeech sample of 40s long

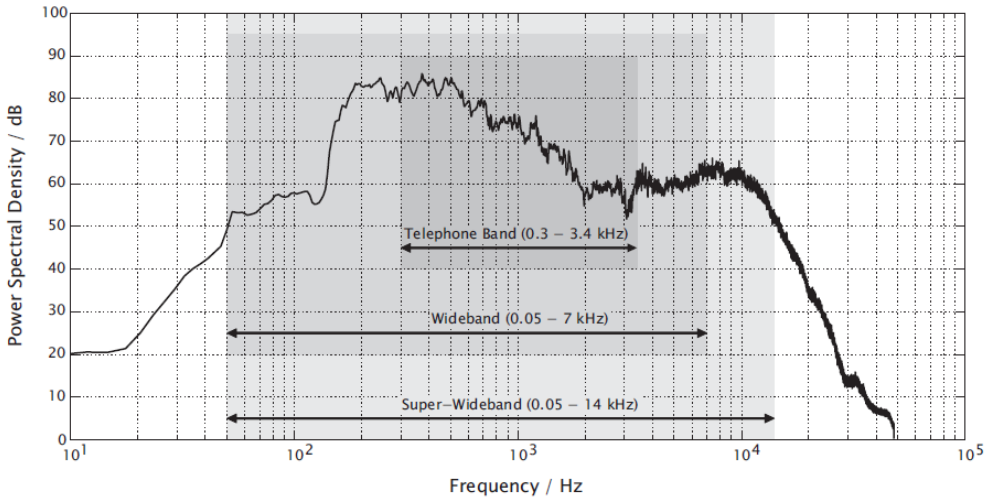


Figure 1.2 (b): Estimated PSD of a female speech sample of 50s long

1.4. Features of Speech Signal Sounds

1.4.1. Voiced sounds

The temporal and spectral features of a typical voiced speech sound are depicted in figure 1.3. It is clear from the figure that the voiced sound has a periodic structure and a significant variation in amplitude.

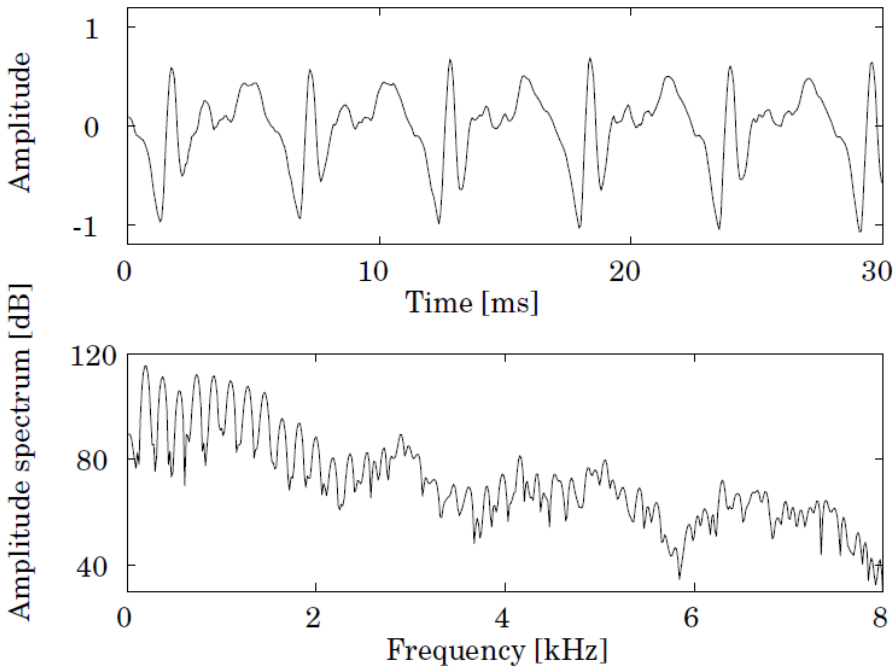


Figure 1.3: The temporal and spectral features of atypical voiced speech sound (/a/)

A clear harmonic structure at low frequencies is observed from the amplitude spectrum. The first harmonic corresponds to f_0. The maximum peaks in the spectral envelope (SE) are called formants, and the formant structure is distinctive for different vowels. The low-pass characteristics of the spectrum of voiced sounds originate from the excitation signal (ES).

In an LNB signal, most of the energy of voiced sounds is preserved because of the low-pass characteristics. Furthermore, an LNB signal also has the most important harmonics. The ear is still able to hear the pitch correctly although f_0 may be missing. The low-pass envelope extension at high frequencies is challenging for SBE techniques. On the other hand, at high frequencies, the precise reconstruction of the harmonic structure of speech is not perceptually significant [8].

1.4.2. Unvoiced sounds

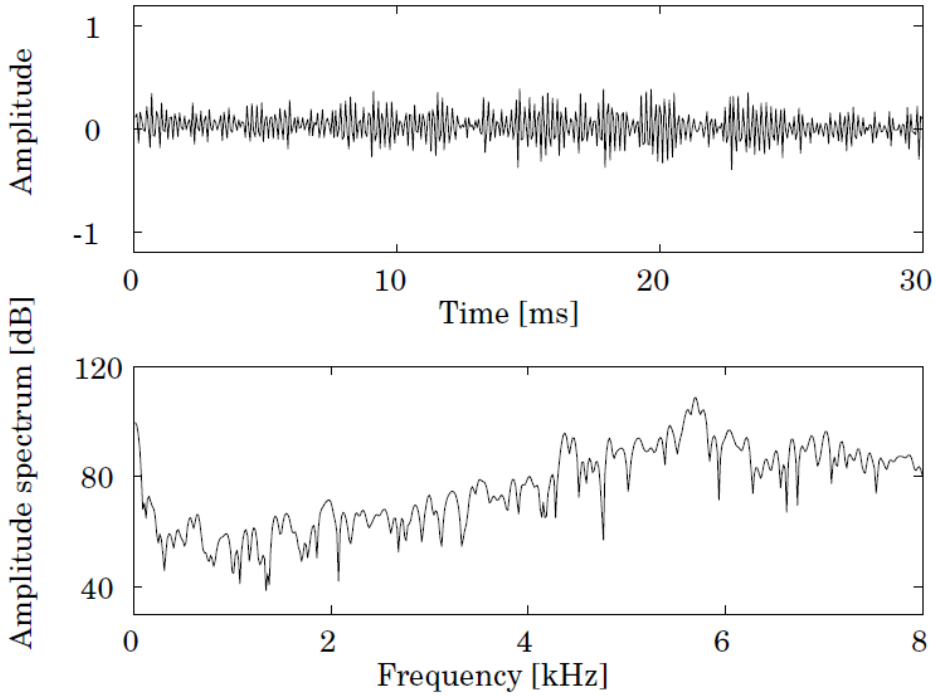


Figure 1.4: The temporal and spectral features of a typical unvoiced speech sound (/s/)

The temporal and spectral characteristics of a usual unvoiced speech sound are depicted in figure 1.4. The time-domain (TD) waveform of the unvoiced sound has little amplitude values and rapid direction changes because of the noisy excitation signal. The amplitude values in the amplitude spectrum increase with frequency, demonstrating that unvoiced sounds (fricatives) have a considerable portion of energy at high frequencies. It is apparent that this energy is lost from LNB signals. These sounds are particularly challenging for SBE techniques because natural sounds of fricatives are obtained only if enough amount of energy is added to the higher frequencies.

1.4.3. Plosives

Figure 1.5 depicts the typical plosive sound's temporal and spectral properties. In most cases, a plosive begins with a short pause, followed by a rush of friction. After the burst comes a voicing period that results in the subsequent vowel, these plosives are also complex

for SBE techniques. If the added higher frequencies amplitudes are too large, it is simply perceived as a tingle.

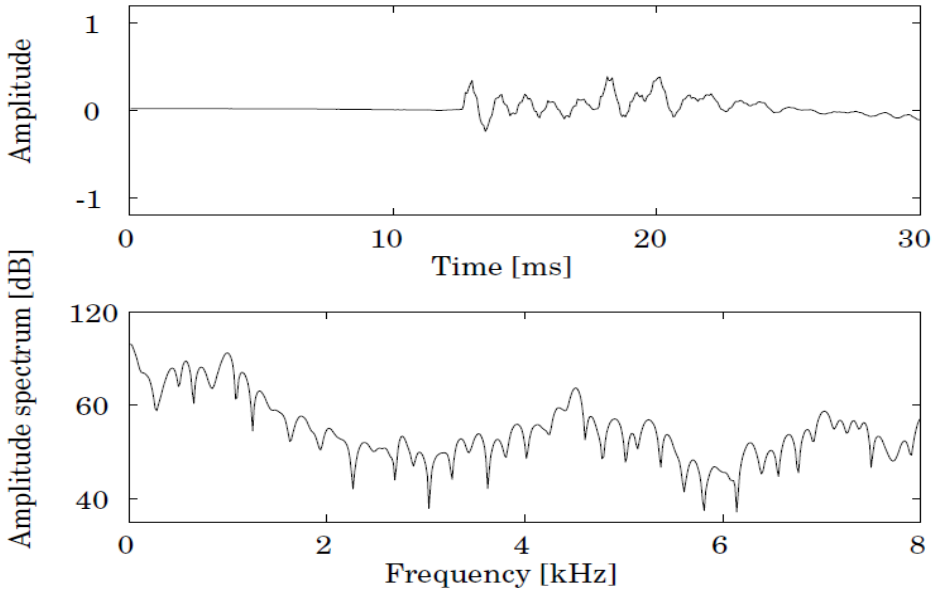


Figure 1.5:The temporal and spectral features of atypical plosive speech sound (/k/)

1.5. Limited Narrowband versus Clear Wideband Telephony

In the early days of telephony, the bandwidth of the conveyed speech was limited by technology. At that time, the limited bandwidth (0.3kHz to 3.4kHz) was due to the transducer's characteristics and other hardware and also the utilization of frequency division multiplex transmission. Due to this limited bandwidth, the syllables intelligibility is around 91%, and the intelligibility of cognizance of sentences is about 99% [9]. However, listening tests have demonstrated that the SBE techniques increase speech signals' quality, intelligibility, and naturalness [10].

Later, digital telephony started with the new era of pulse-code modulation (PCM) [11]. The limited bandwidth characteristic has been retained in PCM systems. The limited bandwidth was due to the sampling rate frequency of 8 kHz and the use of the low-pass filter (LPF) in the global system for mobile communication (GSM) system. For cellular telephone

systems, speech is transmitted in digital networks (GSM) with slightly higher quality. Since signal components underneath 0.3 kHz and the components from 3.4 to 4 kHz can be transmitted.

The intelligibility of telephony speech appears to be sufficient for a typical discussion, as we all know from our daily lives. If we are forced to grasp unknown words, we become aware of the limited intelligibility of syllables. In such situations, the user frequently needs to spell a word to distinguish between certain unique plosives (pot and cot) and fricatives (seed and feed) words. Another limitation is that several speaker-specific capabilities are not preserved on the phone. The CWB speech coding with the frequency range of 0.5–7kHz addresses these limitations. When Compared to LNB speech, the broader bandwidth improves the speech intelligibility, quality, and perceived naturalness of speech transmission. In addition, some plosive and fricative utterances are easier to differentiate from one another.

The Third-Generation Partnership Project started the standardization of the CWB speech codec for GSM in 1999 and published the first specifications in 2001 [12]. The adaptive multi-rate CWB (AMR-CWB) codec as the International Telecommunication Union – Telecommunication Sector (ITU-T) Recommendation G.722.2 was selected by ITU-T in 2002 [13]. Even though CWB standards were launched nearly two decades ago, CWB transmission adoption to end-users is still a continuing process. Today, mostly LNB telephony is offered to end-users. Using CWB speech coding involves the construction of new phone networks and terminals that support wideband spectrum, which turns out to be very expensive and will likely take much time to develop [1]. SBE methods may be used to boost receiver bandwidth without affecting the current telephone network infrastructure [2].

Existing telephone networks can benefit significantly from a notable improvement in speech intelligibility and quality due to SBE technology. The ASBE method only uses the LNB signal to estimate high band information (information above the LNB) and then uses that information to rebuild the CWB signal.

1.6. ASBE of LNB Signal

ASBE methods aim to increase the intelligibility and quality of the received LNB speech signals by artificially extending the limited frequency range of LNB speech. The Source-filter model (SFM) of speech generation provides the foundation for the majority of ASBE approaches discussed in the literature. Based on SFM [2], the block diagram in figure 1.6 depicts an implementation of the ASBE algorithm. The symbol $eb\ envel.$ in figure 1.6 is used to denote the missing high-band (MHB) SE signals. The input LNB signal, S_{lnb}, with a sampling rate of 8 kHz, is up-sampled to 16 kHz through interpolation, resulting in a CWB signal with LNB content, and then the signal is partitioned into 20 ms or 30 ms, and then frames will be processed in sequence.

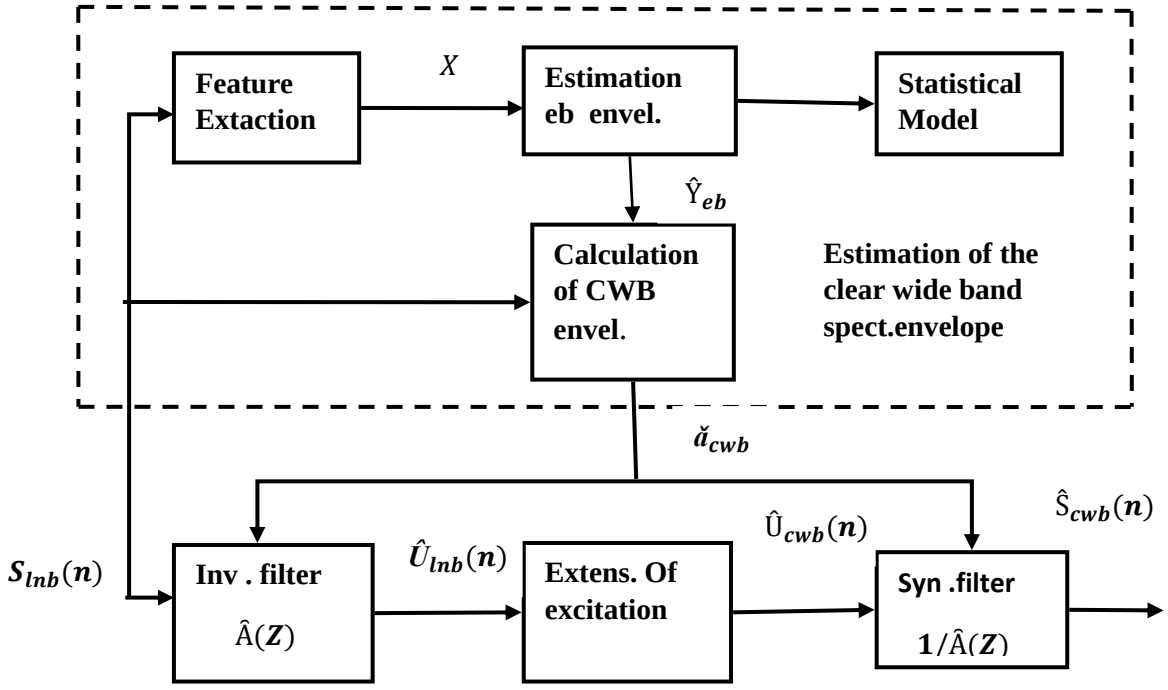


Figure 1.6: System for ASBE of LNB speech signal according to [2]

First, a feature vector X is derived from the LNB signal. This feature vector is supposed to describe the relevant characteristics of the LNB speech compactly. Second, with

the help of a pre-trained statistical model, the MHB spectral envelope $\hat{Y}_{eb}$ is estimated from the feature vector. The LNB frame is combined with the estimate $\hat{Y}_{eb}$ in a short-term power spectrum domain. The coefficients $ă_{cwb}$ are found through the auto-correlation function using the linear prediction analysis (LPA) technique.

The coefficients represent the estimated SE of the CWB signal $ă_{cwb}$ of the vocal tract filter $1/\hat{A}(Z)$. The LNB excitation signal $\hat{U}_{lnb}(n)$ is derived by applying the corresponding analysis filter $\hat{A}(Z)$ to the LNB signal $S_{lnb}(n)$, since the analysis filter is the inverse of the vocal tract (synthesis) filter.

The extension of the excitation signal changes over $\hat{U}_{lnb}(n)$ into a CWB excitation signal $\hat{U}_{cwb}(n)$. The estimate $\hat{U}_{cwb}(n)$ is fed to the synthesis filter $1/\hat{A}(Z)$ to produce CWB speech signal of better quality $\hat{S}_{cwb}(n)$.

According to informal listening tests [14], ASBE systems are preferred over traditional LNB telephony, but their performance falls short of that of the original CWB speech. Systems trained first for a given speaker and later for a language have shown better performance for ASBE. But in both situations, the ASBE-processed speech is not of higher quality than the real CWB speech. These results are supported mainly by theoretical research into the mutual dependency between the speech features of the LNB signal and the SE parameters of the MHB signal [3].

1.7. Speech bandwidth extension applying data hiding

The MHB speech signal cannot be predicted more significantly from the LNB signal. SBE approaches have also been developed based on extra information about the MHB signal [15-22]. The added information allows for a more precise regeneration of the MHB than ASBE and results in a significantly higher-quality CWB voice than ASBE.

The backward compatibility with older LNB codecs isn't a concern. The embedded CWB speech codec (scalable codec) is an elegant way to increase an LNB's speech bandwidth. The SBE parameters must be subjected to proper (vector) quantization techniques before they can be used in conjunction with a codec. The resultant bits are transferred in the "add-on" bitstream layer. Such a multi-layer bitstream setup is a subset of the embedded CWB speech coding paradigm [23].

The SBE parameters with quantization and embedded encoding, however, cannot dependably ameliorate the issue. Because the legacy TNC discards any enhancement bits, the receiver cannot be guaranteed to receive a high-quality recovered voice.

Using parametric SBE techniques, a novel approach to this dilemma is proposed to communicate supplementary information about the MHB over a steganographic channel, i.e., the parameters of MHB or the related bits of MHB are embedded within the LNB speech signal or within the legacy, bitstream using data hiding or watermarking techniques. The legacy codec's bitstream design has not been altered. It is still possible to decode the changed bitstream using a regular LNB decoder, resulting in just a slight decrease in LNB voice quality. On the other hand, an improved decoder can output a CWB voice stream of far superior quality. Compatibility with existing LNB codecs may be done in three different methods [20]: by concealing the additional information in the speech signal itself [15-18], by modifying the encoded bitstream [19], or by jointly coding and data concealment inside the encoder [20-22]. This thesis focuses on speech SBE using data-hiding techniques.

1.8. Motivation

Human speech may have frequencies more than conventional telephone networks operating at 0.3-3.4kHz. When a human speech signal is transmitted through the telephone network leads to losing information due to the limited bandwidth of the telephone network. which results significantly low quality, less intelligibility, and lucidity of speech transmission.

This problem can be solved using a comprehensive wideband voice transmission whose spectrum ranges from 0.5-7kHz. As a traditional telephone network installed to operate at 0.3-3.4kHz, it is not feasible to work at a wideband spectrum. Hence, using a wideband spectrum needs to establish a new network, which turns out to be very expensive and time-consuming [1]. As a result, it is desirable to increase the receiving end's bandwidth utilizing SBE techniques [2] without changing the existing telephone network's infrastructure.

Existing telephone networks can benefit significantly from improved speech quality due to SBE technology. ASBE approaches are founded on the concept that reliance occurs between LNB and MHB, as indicated by the speech production model. The CWB signal may be reconstructed using ASBE, which estimates the MHB information just from the LNB signal. The relationship between the frequency bands justifies the estimate of the MHB information from LNB input. However, ASBE approaches suffer from fundamentally restricted performance, which is insufficient for regenerating high-quality CWB speech [3].

More Improved quality of CWB speech can be achieved when explicit and detailed information about the MHB is included in LNB signals [1].In this case, the backward compatibility with respect to the TNC can be maintained with data-hiding methods that hide the additional information in the LNB signal to form a CLNB signal. Several methods have been developed for this problem as a result of research efforts. Various experiments have demonstrated that the output quality of SBE techniques using data hiding is higher than that of the LNB speech and the output of ASBE methods. In addition, SBE techniques utilizing data concealing are beneficial for speech bandwidth extension. According to a literature review, traditional SBE approaches involving data concealing resulted in poor-quality CLNB and RCWB signals. Another issue is the lack of resistance to noise introduced by the channel and quantization.

In view of the above constraints and challenges, there is a need to search for novel techniques for the improved performance of the SBE algorithms. The quality of the CLNB signal and RCWB signal and the robustness of the SBE algorithms to the channel and

quantization disturbances may be used to quantify the effectiveness of the SBE algorithms. The Exploration of the same is the motivation for this research work.

The objectives of the work are as follows

- The output quality of Existing SBE techniques using data hiding is low. So, we need to improve the quality of the RCWB signal.
- Data-hiding approaches for conventional SBE suffer from low CLNB signal quality. That's why we need to increase the CLNB signal quality.
- When corrupted by CAQNs, conventional SBE techniques using data hiding provide poor SBE performance. So, SBE algorithms have to be developed that will improve SBE performance when corrupted by CAQNs.

1.9. Problem Statement

SBE methods that use data-hiding have to provide high-quality CLNB signal and RCWB signal, as well as being resilient to channel and quantization disturbances. However, most currently available approaches fail to deliver a high-quality LNB composite and reconstructed CWB signal. Another issue is the lack of resistance to noise introduced by the channel and quantization.

In order to increase the quality of the CLNB signal and the RCWB signal, as well as the effective management of channel and quantization disturbances, new SBE algorithms utilizing data-concealing methods are required.

1.10. Contributions of the thesis

The contributions of the thesis are as follows:

- The first contribution attempts to enhance CLNB signal and RCWB signal quality and is resilient to CAQNs disturbances; a unique SBE employed a Hybrid model Transform-domain based data hiding (HMTDBDH) [1*] is provided.

- The second contribution focuses on a novel SBE of the telephone speech algorithm using the DWT-DCT-BDH technique proposed to embed the SE parameters of the lost speech frequency components within the detailed coefficients of the LNB signal. These concealed parameters are retrieved at the receiver side to produce a better-quality CWB signal. The proposed scheme [2**] further improved the quality of the CLNB signal and RCWB signal. It is robust to channel and quantization noises.

- The third contribution focuses on developing a novel SBE algorithm aided by a frequency domain-based data hiding (FDBDH) technique proposed to embed the OOB spectral frequencies in the LNB signal. These embedded spectral frequencies are recovered steadily at the receiver side to produce a better-quality CWB signal [3**] to improve further the quality of the CLNB signal and RCWB signal and is robust to CAQNs.

- To further improve the quality of the CLNB signal and RCWB signal, a novel SBE using the DCTBDH technique is to embed the component's MHB speech signal parameters within the DCT coefficients of the LNB signal. These hidden parameters are retrieved at the receiver side to produce a better-quality CWB signal by combining the missing speech signal transmitted through the DCT coefficients with the LNB signal [4**].

The proposed algorithms are implemented in MATLAB.

1.11. Organization of Thesis Chapters

The thesis presents the development and evaluation of novel SBE algorithms using data-hiding techniques. The thesis is organized into seven chapters. The following section gives a summary of the chapters.

Chapter 1 presents an introduction to the work, motivation, problem statement, and thesis contributions.

Chapter 2 reviews the notable amount of the most updated literature, and a brief outline of the thesis is also presented.

Chapter 3 deals with the new SBE algorithm using a hybrid model transform-domain-based data hiding technique.

Chapter 4 discusses the novel SBE algorithm using discrete Wavelet transform-discrete Cosine transform-based data hiding technique.

Chapter 5 deals with the novel LNB speech SBE algorithm aided by the frequency domain-based data hiding technique.

Chapter 6 discusses the novel speech BWE algorithm using data hiding based on discrete Cosine transform.

Finally, in **Chapter 7,** the conclusions of the thesis are summarized from the contributions, and brief discussions on the direction for future work are given.

Capter-2

ASBE and SBE using data hiding are discussed in this chapter. The many signal-processing strategies that are implemented in ASBE are presented in detail. The limits on the achievable output quality of the ASBE techniques were also discussed. Finally, signal processing techniques utilized in SBE using data hiding are discussed in detail.

2.1. Introduction

Due to historical and economic reasons, the telephony speech frequency is confined to around 0.3–3.4kHz, known as LNB speech. Because of the telephone's limited audio capacity, customers hear muffled speech [2]. If the low end of speech (below LNB) is missing, it causes a thin voice and reduces the naturalness of the signal, while if the high end of speech (above LNB) is lacking, it causes fricative sounds like /s/ and /f/ to degrade, making for muffled speech. An inability to hear and understand what is being said on the other end results from a narrow frequency range. Additionally, several features that are exclusive to a speaker are removed.

The CWB speech, which has a frequency range of 0.5–7kHz, would improve the speech signal's quality, intelligibility, and perceived naturalness when sent across telephone networks. As a result, new telephone networks with higher bandwidths must be built, which will be costly and take some time to establish [1]. As a result, increasing the receiving end's bandwidth is desirable by utilizing SBE techniques [2] without changing the existing telephone network's infrastructure.

Existing telephone networks can benefit significantly from improved speech quality due to SBE technology. The CWB signal may be reconstructed using ASBE, which estimates the MHB information just from the LNB signal. From the speech production model, it can be seen how the LNB and MHB are mutually dependent on one other. Because of the strong correlation between LNB and MHB, it is reasonable to infer MHB information from LNB input alone.

ASBE approaches, on the other hand, have inherent limitations that make them unsuitable for generating high-quality CWB speech. As the MHB signal is combined with LNB signal information, the CWB speech quality is significantly improved compared to ASBE approaches [1]. Data concealing mechanisms would be employed to disguise the MHB information in the LNB signal in order to guarantee backward compatibility with current telephone networks.

2.2. Artificial Speech bandwidth extension

Separating the two components of the speech production process is possible. The vocal cords or vocal tract constrictions generate a periodic, mixed excitation signal or noise-like. The vocal tract is responsible for shaping the final sound.

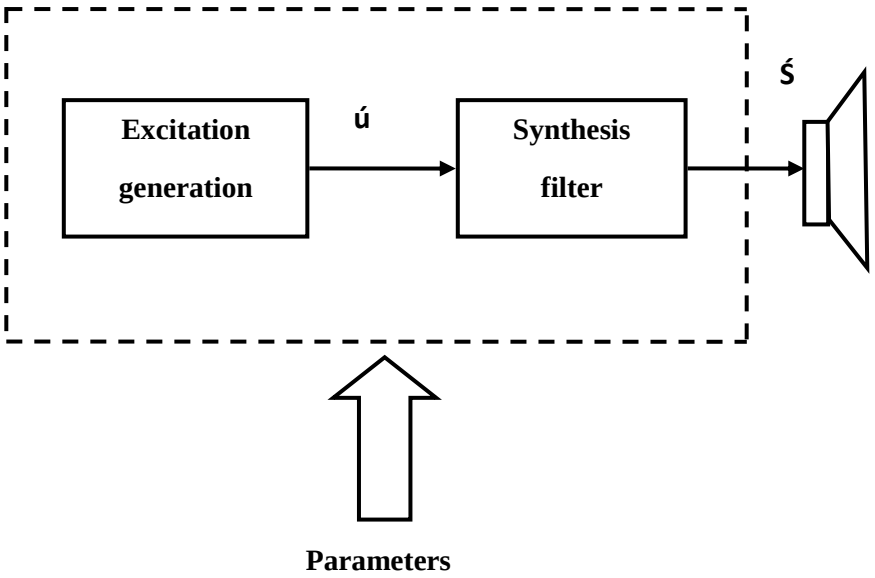


Figure 2.1: Source-filter model (SFM)

An SFM can approximate the workings of a speech production system. Figure 2.1 depicts SFM, which is divided into two components. A signal generator generates an ES, and a synthesis filter shapes the SE of the speech signal.

Conventional SBE methods typically utilize SFM as their basis. Both SE and ES are subjected to SBE methods immediately after spectral frequency-domain analysis[24, 25]. Due to their assumed independence, the ES and SE signals can be optimized independently. If you want the spectral envelope extension to work, you'll need an estimate method that can take the LNB input signal's characteristics and calculate the filter coefficients. The MHB signal is generated by feeding the extended excitation through the synthesis filter given by the filter coefficients. MHB and LNB signals are combined to create CWB speech. Figure 2.2 depicts a block diagram of this notion.

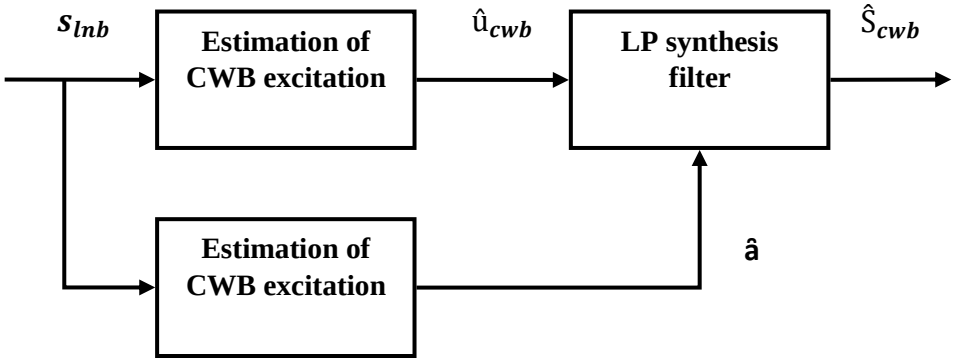


Figure 2.2: SBE with separate excitation signal extension and spectral envelope extension

Feature extraction (FE), spectral envelope estimate, and excitation extension will be discussed in further detail.

2.2.1 Feature Extraction

The spectral envelope extension needs an estimation technique that produces the coefficients of the filter from a set of features of every LNB input frame. It is the goal of FE to reduce the LNB frame to an extremely small set of values.

Features should be chosen so that they reveal a lot about the MHB, but the number of features should be kept low to save computational complexity.

Time-domain features (TDF) and Frequency-domain features may be categorized into two types. The TDF represents the temporal characteristics and is calculated directly from the signal samples. FDF is calculated from the FFT-based magnitude spectrum of an LNB input speech frame, and these features indicate the spectrum properties. The LNB input spectral envelope parameters are used only to estimate the MHB spectral envelope parameters in many early ASBE techniques. However, additional TDF and FDF are advantageous for the estimation [26].

A set of features used in ASBE techniques is given below:

Sub-band energy levels: The spectral structure of the input signal is represented by the energy levels in a few frequency bands as a whole [27].

Autocorrelation coefficients: The first ten autocorrelation coefficients represent the spectral envelope [9].

Linear predictive coding (LPC) filter coefficients: The linear prediction is used to get the filter coefficients, and the spectral envelope is also represented by these coefficients [28]. Convert the linear predictive coefficients (LPCs) into other representations to be used as features, such as line-spectral frequencies (LSFs) [29-32], Mel-scaled LSFs [33], or linear prediction cepstral coefficients [34].

Cepstral coefficients: The SE signal may also be represented by MFCCs (Mel-frequency cepstral coefficients), and ASBE makes use of these MFCCs [35-37]. Linear frequency cepstral coefficients can also be used to describe the spectral envelope [38].

Spectral Centroid (SC): The SC y_{spc} Corresponds to the magnitude spectrum's center of gravity is

$$y_{spc} = \frac{\sum_{h=0}^{\frac{M_h}{2}} |x(h)|}{(\frac{M_h}{2}+1)\sum_{h=0}^{\frac{M_h}{2}} |x(h)|} \quad (2.1)$$

Where $x(h)$is the h^{th} coefficient of the M_h-point FFT spectrum of the LNB input [2, 9]. This feature results in lower values for voiced sounds and higher for unvoiced sounds [2, 39].

Spectral flatness: The ratio of the geometric mean of the power spectrum to the arithmetic mean of the power spectrum is the spectral flatness y_{sf} [2, 40] and is specified as

$$y_{sf} = \frac{\sqrt[M_h]{\prod_{h=0}^{M_h-1}|x(h)|^2}}{1/M_h \sum_{h=0}^{M_h-1}|x(h)|} \quad (2.2)$$

This characteristic shows how uniform the power spectrum is.

Frame energy: The energy of a signal within a frame,E_{fre}, is defined as

$$E_{fre} = \sum_{j=0}^{M-1}(x_{lnb}(j))^2 \quad (2.3)$$

Where $x_{lnb}(j)$ is the LNB signal and $M$is the frame length. This feature results in lower values for unvoiced sounds and higher values for voiced sounds [2].

Normalized frame energy: The current frame energy relative to a reference value is normalized. The maximum possible energy of a frame is used as a reference in [28], whereas the noise floor and the average frame energy take into account the reference in [2].

Zero crossing rate: The number of times the signal crosses the zero level within a frame is the zero-crossing rate. This feature results in lower values for voiced sounds and higher values for noise-like unvoiced sounds [41].

Gradient index: Gradient index, y_{grix}, differentiates voiced and unvoiced sounds more efficiently [42] and is defined as

$$y_{grix} = \frac{\sum_{j-1}^{M-1} \Psi(j)|x_{lnb}(j) - x_{lnb}(j-1)|}{\sqrt{\sum_{j=0}^{M-1}(x_{lnb}(j))^2}} \quad (2.4)$$

Where $\Psi(j) = \frac{1}{2}|\Psi(j) - \Psi(j-1)|$, which indicates changes of direction, $M$is the length of the frame, and $\Psi$is the sign of the gradient$x_{lnb}(j) - x_{lnb}(j-1)$.

2.2.2 CWB spectral envelope Estimation

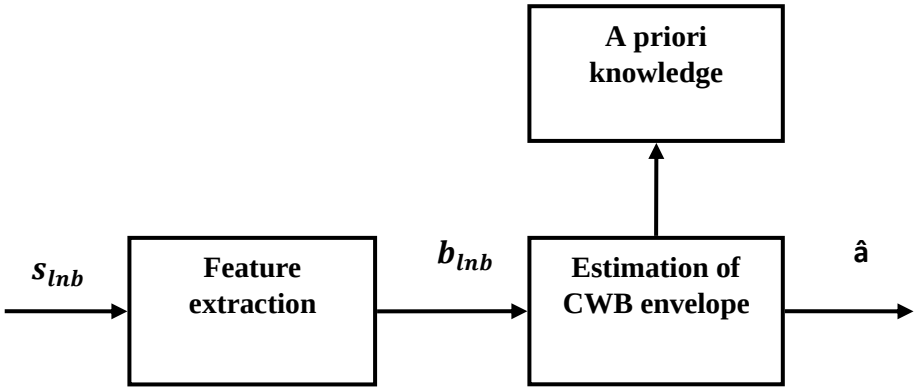


Figure 2.3: Spectral envelope estimation

Figure 2.3 shows a more detailed representation of the CWB signal's spectral envelope estimate. In most SBE algorithms, statistical estimating approaches that are at least in part related to pattern recognition techniques are used. The estimation technique is based on a set of featuresb_{lnb} of every LNB input signalS_{lnb} frame. This feature vector consists of LNB signal spectral envelope and features that differentiate voiced and unvoiced sounds.

A priori knowledge of the combined behavior of the observation (feature vector) and the quantity to be estimated is necessary for every estimating approach. Prior information in the form of an estimating approach may be found in the statistical model. Using the filter

coefficient vector, the estimate block creates the CWB-SE of the speech frame, which is characterized by the filter coefficient vector $\hat{a}$.

The state-of-the-art techniques for estimating the CWB-SE include:

2.2.2.1 Codebook mapping (CM)

The codebook mapping procedures employ two codebooks that are linked together. LPCs and LSFs parameters are maintained in two separate codebooks, one for the CWB spectral envelope (LPCs or LSFs) and the other for the corresponding LNB frame. In the training phase, vector quantization (VQ) is employed to create LNB and CWB codebooks. When the SBE approach is employed, the best matching item in the LNB codebook for each LNB input frame is found, and the corresponding entry in the CWB codebook is used to construct the MHB spectral envelope.

This technique was proposed in [43], and several refinements were later presented in [44-46]. Separate codebooks were made for voiced and unvoiced fricatives to improve the quality of spectral envelopes [44]. Furthermore, CM with interpolation was presented to improve the MHB signal quality in [45], and CM with memory was reported in [46].

2.2.2.2 Linear mapping (LM)

In [29, 45, 47], LM is used to estimate the MHB-SE. The LNB envelope in linear mapping is represented as a vector $p = [p_1, p_2, \ldots\ldots\ldots p_x]$ and the CWB envelope to be estimated is characterized by another vector of parameters $q = [q_1, q_2, \ldots\ldots\ldots q_y]$. The LPCs or LSFs reflect the linear mapping between LNB and CWB characteristics as

$$q = D_p \tag{2.5}$$

An offline training process with the least-squares method is used to obtain matrix D. This least-squares approach uses training data with LNB envelope parameters p and corresponding CWB parameters q to minimize the model error $q - D_p$:

$$D = (P^T P)^{-1} P^T q \tag{2.6}$$

It has been stated that the standard linear mapping method can be modified to reflect better the non-linear nature of the interaction between LNB and MHB envelopes. Instead of employing a single mapping matrix, many matrices can be used to implement the mapping. [29] employed a hard-decision clustering method in which four speech frames were grouped into four clusters, and each of these four groups had a different mapping matrix. In [47], clustering is performed based on the VQ of LNB vector p and CWB vector q, and soft-decision clustering is used.

The performance evaluation of ASBE techniques with linear mapping is rather concise. In [47], spectral distortion (SD) for piecewise LM was obtained to be smaller than the CM approach. Objective examination in [45] shows that CM is superior to LM in terms of performance.

2.2.2.3 Gaussian mixture model (GMM)

The LNB and MHB spectral envelope linear dependencies are the only ones exploited in linear mapping. There are several ways to include non-linearity in statistical models, such as GMM. Adding multiple multivariate Gaussian distributions in a GMM can approximate the probability density function (PDF). In ASBE, GMMs are used to model the combined

probability distribution of parametric representations of LNB and MHB. An LNB frame's feature vector determines the MHB parameters' minimal mean square error.

In ASBE techniques, the joint PDF between two random variables is modeled using the GMM. Represent the GMM PDF for $b = [b_0 \dots \dots b_{x-1}]$and $d = [d_0, \dots \dots d_{y-1}]$as a weighted sum of M Gaussian component densities f_{gau}

$$f_{gmm}(b,d) = \sum_{m=1}^{M} w_m f_{gau}(b,d; ev_{b,d,m}, cm_{b,d,m}) \quad (2.7)$$

Where w_m is the weight of the m^{th}mixture, $cm_{b,d,m}$ is the covariance matrix, $ev_{b,d,m}$ is the mean vector, and $M$is the number of individual Gaussian components. Training data is used to estimate the GMM parameters, which are then refined using an expectation-maximization method (EM).

To estimate CWB LPCs or LSFs from LNB characteristics, GMM is employed. By employing MFCCs instead of LPCs, the GMM-based envelope estimate increased its performance [50, 51]. To further boost performance in terms of log-spectral distortion (LSD) and perceptual evaluation speech quality (PESQ), the GMM with memory was used [52, 53].

The GMM provides a continuous approximation from LNB to CWB features in envelope estimation approaches. The subjective and objective analyses given in [48, 49] showed that the performance of GMM is better than CM.

2.2.2.4 Hidden Markov model (HMM)

The SE can also be estimated using a statistical model based on the HMM. Each HMM state corresponds to a code in the pre-trained CWB codebook. State-specific models for PDFs of input feature vectors and probabilities of transition between states and within states are all

included in the HMM. A GMM is used to approximate every PDF. The posterior probability of each state is determined using the state transition probabilities and the observation probabilities for a given sequence of input characteristics. The spectral envelope parameter's minimal mean square error (MSE) estimate is then calculated. It is important to note that state transitions consider prior frames' data.

HMM has the benefit of improving estimation quality by using previous frame information. Each state of HMM represents a typical MHB spectral envelope. Hence a change in the form of HMM identifies a difference in the speech envelope. The lower-order HMM provides better results than higher-order GMM.

HMM-based ASBE in [2, 9] further developed in [37], HMM training using phonetic transcriptions introduced in [54], and according to [36], a generic Baum-Welch training method used outperformed the approach provided in [2]. Several HMM-based ASBE approaches are also discussed in [55, 56].

2.2.2.5 Neural networks (NNs)

Neural processing methods are an inspiration for artificial neural networks [57]. A huge number of linked neurons carry out estimation. Each of these neurons generates a single output using a nonlinear or linear function on the weighted sum of inputs. Every neuron receives input from the layer above and sends output to the layer below it, forming a layered structure. A typical neural network (NN) has only a few layers. Networks having only forward connections between layers are called feed-forward networks, whereas those with backward connections are called recurrent networks. Neuronal network training entails determining the proper weights for connections within a predetermined network topology's connections between neurons. The back-propagation technique and a significant amount of training data are typically employed to train neural networks.

The NNs may also be trained using genetic algorithms. These algorithms are evolutionary-inspired approaches to optimization. Individuals, or candidate solutions to the optimization issue, can be evolved through a succession of generations to find better answers to the problem. A new generation can be created by randomly recombining and mutating the fittest people. Also, NN-specific genetic algorithms have been developed. Genetic algorithms may be used to change not just the weight of connections between neurons but also the structure of a network as it evolves [58].

Using an adaptive spline NNs, the CWB spectral envelope is estimated from the LNB input [59]. The multi-layer perceptron type NNs with only feed-forward connections are utilized for ASBE in [34, 60, 61]. In [27], an ASBE technique employing a partially recurrent network is used to estimate the CWB spectral envelope. A genetic algorithm is used to train the NN in [27], and the neuro evolution of augmenting topologies method was utilized for ASBE in [62].

NN and codebook mapping techniques were evaluated in [61]. According to [34, 63], NNs were employed to estimate the MHB envelope parameters in these studies. Even if the codebook mapping technique appears to be the better choice, NNs can also improve speech quality but at considerably lower computational complexity.

A recurrent neural network with long short-term memory cells was used to estimate the MHB log-power spectrum [185]. [186] uses DNN-based spectral envelope estimation for ASBE, although the network topologies and parameter selections are completely ad hoc. Various DNN structures are used to estimate the spectral envelope information of CWB in [187-189]. It is proposed in [190] to use hierarchical recurrent neural networks to model and generate waveforms, increasing the voice bandwidth. A recurrent temporal limited Boltzmann machine is presented in [191] for use in SBE. To further improve SBE performance, an ensemble of sequential deep neural networks is proposed in [192]. A DNN regression approach to estimate MHB spectral envelope is presented in [193]. The recurrent neural network-based SBE method is proposed in [194]. To recover missing spectral information by

using the sparsity of the spectrogram frames, [195] proposes a joint-dictionary training strategy in which the dictionaries for the CWB and related LNB spectrogram frames are learned in a linked way to learn the mapping from LNB to CWB frames. [196] proposes an ASBE method based on the constant Q transform.

2.2.3 Excitation signal extension

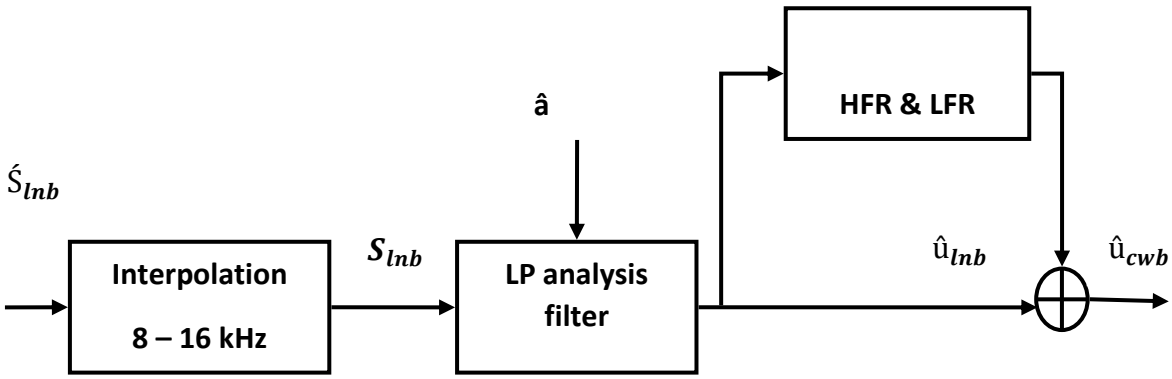


Figure 2.4: Excitation signal extension

Substituting the excitation signal's missing frequency components is the next stage in the SBE approach. Figure 2.4 describes the basic principle of the excitation signal extension. After f_s, is interpolated from 8 kHz to 16 kHz, the interpolated signal is applied to the CWB linear predictive analysis filter$1 - \hat{A}(Z)$ to obtain the LNB excitation signal $\hat{u}_{lnb}$. The extension of the excitation is done in the high frequency (above LNB) re-synthesis block labeled HFR and in the low frequency (below LNB) re-synthesis block labeled LFR. The CWB excitation signal $\hat{u}_{cwb}$ is obtained by combining the extended frequency components and the LNB excitation. Subjective tests have demonstrated that the ES extension has substantially less impact on the RCWB speech quality than the CWB spectral envelope estimate.

The state-of-the-art techniques for the extension of the excitation signal include:

2.2.3.1 Spectral folding (SF)

The excitation signal of an LNB signal can be extended using SF from 0 to 4 kHz to a range of 0 to 8 kHz. SF produces an MHB signal spectrum that mirrors the LNB signal spectrum [64]. When the aliased spectrum is used in the MHB, the effect is the same as up-sampling without an anti-aliasing LPF. The excitation extension can be performed in two ways: either by reflecting the FFT coefficients in the frequency domain or by adding a zero sample right after each input sample in the time domain (TD). Figure 2.5 depicts the impact of spectral folding.

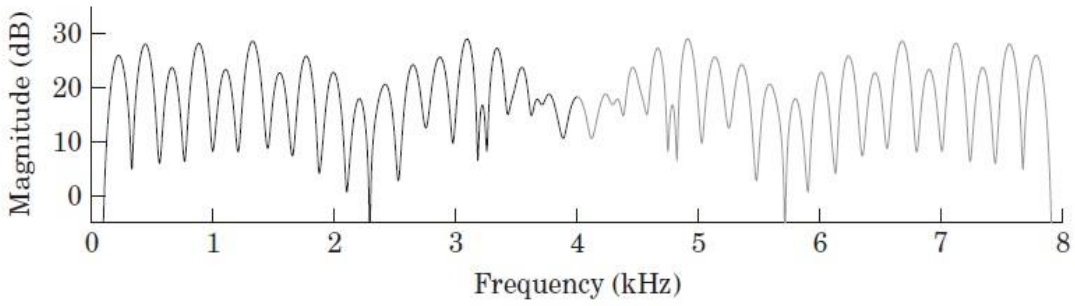


Figure 2.5: Spectral folding effect in the frequency domain

The computational complexity of spectrum-folding is low, and the resulting MHB temporal structure mirrors that of the input signal. However, the SF approach does have significant limitations [65]. Since the LNB signal goes up to 3.4 kHz, the mirrored MHB excitation can only fill a portion of MHB, leaving a gap at roughly 4kHz. When a harmonic spectrum is SF, the harmonic structure is thrown off. In spite of this, it is stated in [9] that the misalignment of the harmonic structure of MHB does not considerably decrease the subjective quality of the reproduced signal.

2.2.3.2 Modulation technique

The MHB spectrum is created by altering the spectrum of an interpolated LNB signal using a fixed frequency modulation technique. In order to keep the LNB signal spectrum from overlapping with the translated spectrum, filtering is utilized. In [64], spectral translation with a modulation frequency of 4kHz is explained. Besides the 4kHz band, other frequencies can be utilized and approved [9]. The ES signal harmonic structure is lost during fixed-frequency modulation, which is a disadvantage. When modulating with a fixed frequency, the ES signal loses some of its harmonic structure. The TD modulation of the LNB excitation signal is described in [9]. In the frequency-domain, the MHB excitation is constructed by repeatedly copying a part of the LNB spectrum to MHB [66, 67].

2.2.3.3 Pitch-adaptive modulation

Copy the LNB spectrum to MHB using an adaptive modulation frequency (AMF) which preserves the harmonic structure of voiced speech. An AMF is a multiple of f_0. This idea was used for the excitation extension in ASBE [2, 14, 68]. A TD modulation technique is described in [14], whereas an MHB excitation is constructed in the frequency domain by repeatedly copying part of the LNB spectrum to MHB [68].

An accurate estimate of $f_0$is required by the pitch-adaptive modulation technique. Several basic methods are available for pitch detection and have proposed a number of modifications to improve accuracy and robustness. As an example, see [65] for an advanced pitch estimate method. Pitch-adaptive modulation's benefit was determined to be insignificant in comparison to the technique's complexity.

2.2.3.4 Sinusoidal synthesis (SISY)

In SISY**,** the voiced speech excitation signal is generated as a sum of sine waves with frequencies equal to multiples of f_0. Also, an identical result is obtained by generating harmonic peaks in the frequency domain. Random harmonic phases or additive random noise provide a mixed excitation with varied harmonicity for various amounts of voicing [31, 69, 70].

The SISY[71] preserves the harmonic structure of the excitation signal. SISY does not require Spectral flattening since the envelope may be generated using sinusoidal amplitudes. Oscillators with amplitudes, phases, and frequencies derived from the LNB signal are used to construct the high-frequency harmonic structure.

For extending the excitation signal to the low-band (below LNB), SISY is particularly suitable since the low-band consists mostly of tiny harmonics of f_0. In [31, 60, 72, 73], SISY is used for low-frequency extension.

2.2.3.5 Non-linear processing

A non-linear function $f(y)$ is applied to the LNB excitation signal y to extend the excitation signal. Examples of non-linear functions used for SBE include a quadratic function, y^2[68], a cubic function, y^3[61], and a full wave rectification, $|y|$[38, 74]. Non-linear processing has the advantage of preserving the harmonic structure of the ES, i.e., non-linear processing of a voiced signal produces a spectrum having spectral peaks at integer multiples of f_0. Non-linearity makes managing the MHB's energy level challenging, necessitating further energy normalization [61]. It is common practice to employ non-linear functions for low-frequency extensions because they retain the harmonic structure. When listening at low frequencies, this harmonic structure is essential.

2.2.3.6 Noise modulation

The temporal envelope of a white noise signal is modulated in noise modulation to create an excitation signal. Pitch harmonics cannot be discerned independently in the human ear beyond 4kHz, yet voiced speech's temporal envelope exhibits pitch periodicity [75]. TD modulation of a noise excitation reconstructs pitch periodicity; therefore, the harmonic spectrum does not need to be rebuilt. The time envelope of the LNB signal sub-band is used for extracting the temporal modulation envelope [46, 48, 76, 77]. A combination of spectral folding and noise modulation techniques is used for MHB excitation in the GMM-ASBE method [78].

2.2.3.7 Noise excitation

To avoid an overly periodic excitation at high frequencies [66] or to give an excitation for unvoiced sounds [69, 79, 80], a noise signal is employed as an excitation in combination with other techniques. A noise excitation may be sufficient for the extension from CWB to super-CWB (0-14kHz) [81].

2.2.3.8 Voice source modeling

Extending the excitation of voiced speech was proposed in [37]. We discovered that it worked well in the lower frequency band. A lookup table of glottal pulse shapes is used in [82] to expand the voiced speech excitation.

2.2.4 MHB gain estimation

The estimation of the MHB gain forms a vital subpart of numerous ASBE techniques [14, 24, 28, 49, 61, 66, 70, 72, 73, 76]. In [14, 28, 30, 61], the matching of LNB energy in the estimated CWB spectrum with that of LNB to estimate the gain. The gain can also be calculated in [14, 24, 30, 49, 77] by using codebooks and in [66, 76] by using GMM. The SE continuity at the boundary between LNB and MHB is considered in [70].

2.2.5 Temporal envelope modeling

The MHB signal temporal envelope has also been found to be perceptually pertinent. In [83], regenerate the MHB effectively from a noise excitation utilizing a relatively coarse spectrum representation every 10 ms, but reconstruct the MHB signal energy contour more precisely using an MHB gain coefficient every 2.5 ms. Similarly, in [84], the temporal envelope is multiplied with the temporal fine structure to reconstruct the MHB signal.

The ASBE technique proposed in [37, 38, 85] uses temporal envelope shaping, which alters the MHB sub-bands amplitude envelopes. In [86], the MHB sub-band amplitude envelopes are altered with a temporal resolution of 2 ms. Also, the SBE technique proposed in [82] uses temporal envelope shaping.

2.2.6 Non-model-based techniques

Other ASBE approaches use methods different from the SBE techniques based on SFM. For example, the CWB spectral envelope is estimated using the vocal tract area function in [67, 87]. [88] Proposes an ASBE method that operates in the modified discrete

cosine transform domain [89]. A convolutional non-negative matrix factorization approach is described for increasing the bandwidth of the LNB signal. The CWB spectral envelope is reconstructed in [32] using a state-space model. The vocal tract front-most cavity model is used to produce the synthetic formants at estimated frequencies [73]. In [90], an error estimate algorithm was used to compute the MHB of the voice signal by combining N consecutive amplitude and frequency modulation signals. Finally, an ASBE method based on a speech recognition technique is proposed in [91]. This technique estimates the sequence of sounds from the input, and then this estimated sequence is used by a CWB speech synthesis technique to synthesize a CWB speech signal.

2.2.7 Limitations of ASBE

Subjective listening studies of [14, 71] demonstrate that listening tests favor ASBE systems, but their performance is poor than that of the original CWB speech. There have been better ASBE outcomes using systems that have been trained first on an individual speaker and subsequently on a particular language. However, in neither case does the ASBE approach provide speech superior to the original CWB. Theoretical investigations on the amount of reciprocal information between LNB and MHB speech characteristics support this [3, 92, 93].

2.3 SBE with additional information

Since MHB cannot be accurately estimated from LNB input, SBE methods using some additional transmitted information about the MHB along with the LNB signal have also been proposed [15-22, 60, 70, 75, 94-98]. The additional transmitted information allows more accurate reproduction of the CWB speech signal compared to traditional ASBE methods, and thus the quality of CWB speech is improved.

2.3.1 SBE technique using Embedded CWB coding

The embedded CWB speech codec is an elegant approach to enhance the speech bandwidth of a given LNB codec, provided that compatibility with existing LNB codecs is not required. The integrated CWB voice codec is a simple way to increase the speech bandwidth of a given LNB codec. It is necessary to apply proper quantization methods to SBE parameters in order to integrate them with a codec. The resultant bits are transferred in the "add-on" bitstream layer. The embedded CWB speech coding methodology is a specific example of this layered bit stream structure [23].

2.3.1.1. SBE information transmission

The encoder of an embedded CWB speech codec (ECWBSC) is shown in figure 2.6. Initially, the CWB input speech (0-7 kHz) S_{cwb}, with f_s of 16 kHz, is band-split using an LPF and a high pass filter (HPF), respectively. The LPF output (0-3.5 kHz) is down-sampled to 8kHz and then given as input to the LNB encoder. The embedded bit stream, labeled LNB info, is produced by the LNB encoder. Any conventional LNB decoder decodes this information.

The HPF output S_{mhb} is fed to an SBE encoder module. The time and spectral envelopes of the MHB signal S_{mhb} are analyzed by this module, and a set of corresponding parameters, labeled SBE parameters, are determined. The dashed arrow in figure 2.6 indicates that the spectral envelope information may be used to extract the time envelope characteristics.

As a result of quantizing the SBE parameters, a collection of quantizer indices known as SBE info are formed. The LNB signal features can be used to quantize SBE parameters to increase a quantizer's efficiency.

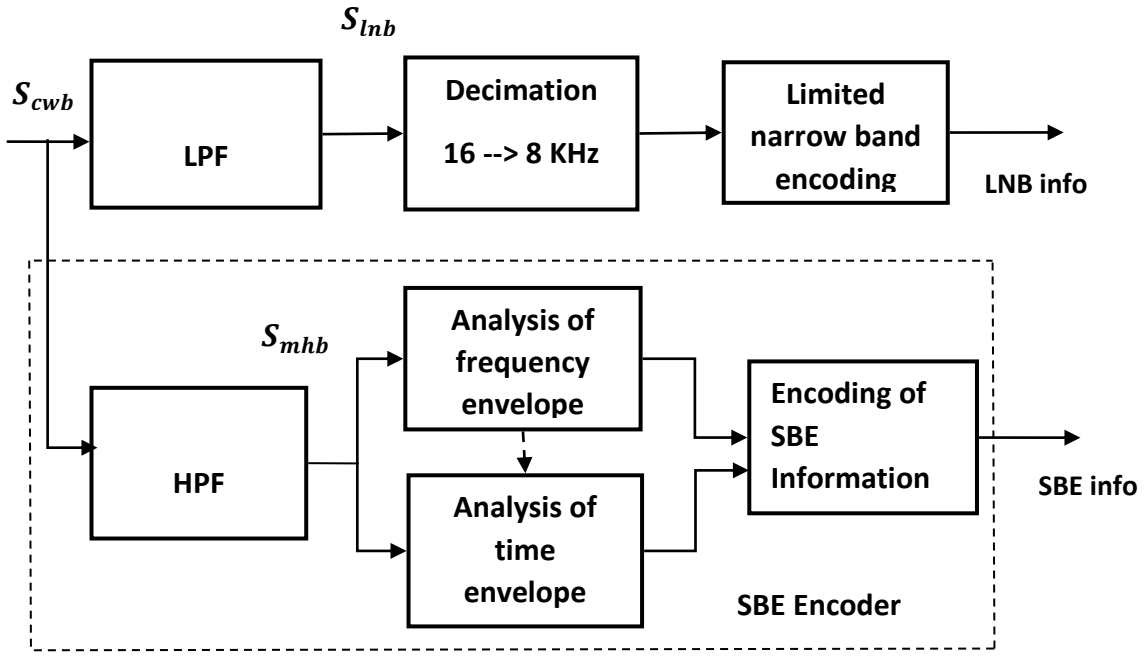


Figure 2.6: Embedded CWB encoding

2.3.1.2 SBE information reception

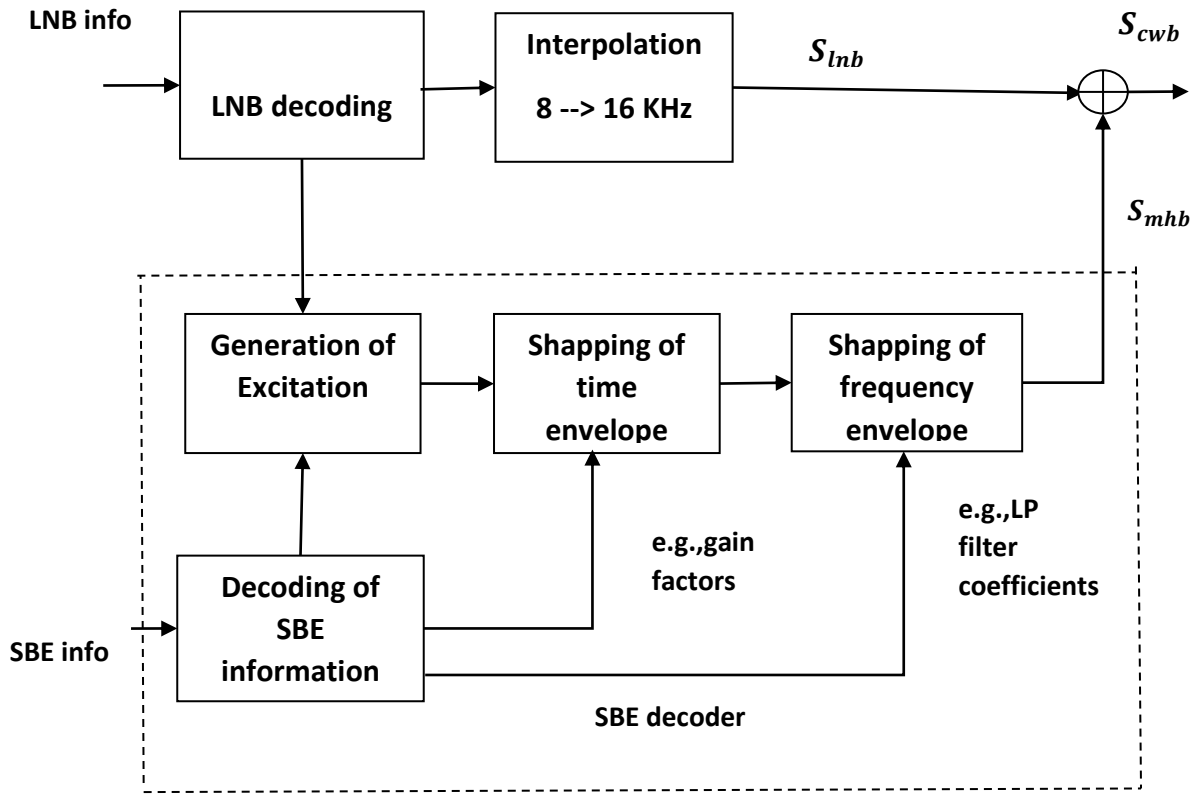


Figure 2.7: Embedded CWB decoding

The decoder of an ECWBSC is shown in figure 2.7. The LNB signal with a f_s 8 kHz is decoded using the embedded bit stream, and then this signal is down-sampled to 16 kHz. The hierarchical bit stream is decoded to provide the SBE parameters, which are then used in a three-step technique to reassemble the MHB signal, S_{mhb}. Initially, an ES is generated. Then, the excitation time envelope is shaped by applying gain factors. Finally, the MHB signal is reconstructed by applying a time-varying filter.

Several ECWBSC schemes that have been described in the literature are summarized below. In [99], an ECWBSC was proposed based on the observation that some particular fricatives, particularly /f/ and /s/, are hard to distinguish in ASBE. Hence, in [99], in order to differentiate fricatives, it was proposed to broadcast one additional bit every frame in addition to the LNB signal. The bit stream of quantized coefficients and gain factor of an MHB signal are transmitted in [100]. In [60, 70, 75, 83, 101], ECWBSC was proposed based on the parametric transmission of signal components of the MHB. The bit stream of the filter coefficients and gain factors are transmitted for every frame or for sub-frames in all of these proposals. In [75], proposed embedding of a conventional LNB codec. In order to successfully reconstruct the MHB signal in a brief manner using the time envelope in [75], white noise is modulated with the time envelope of the decoded LNB signal sub-band components (3-4 kHz). A pitch-dependent time envelope of the MHB signal is produced by this method, and then the SBE information is transmitted, in [82], proposed another ECWBSC. This method uses a finite impulse response (FIR) filter-bank equalizer to shape the SE. Adaptation of the filter coefficients by the decoder is based on comparing a target spectral envelope transmitted by the SBE encoder with the observed SE of the excitation signals.

A legacy telephone network can prevent high-quality CWB speech replication because it will discard enhancement bits even if both end-user terminals are adequately equipped. Unfortunately, ASBE approaches cannot deliver a steady CWB voice quality in all situations. Embedded coding with quantized SBE parameters, which is more resilient, will not help the problem much either.

A fresh approach to the problem is presented here. This explores an innovative approach to solving the issue at hand. It is proposed, on the basis of the parametric SBE techniques, to communicate information about the MHB frequencies over a steganographic channel. This would mean that the related parameters (LSFs and gain) or related bits would be embedded within the LNB signal or the legacy bitstream using data hiding or watermarking techniques. The old codec's bitstream format remains unchanged. Only a very small amount of quality loss may be tolerated when using a standard LNB decoder to decode the modified bitstream. On the other hand, an improved decoder can produce a CWB speech signal of far more excellent quality since it is aware of the concealed information.

2.3.2 Data Hiding

Data hiding or digital watermarking (DWM) methods create a secret communication channel inside the transmitted LNB voice. It is possible to hide the data (MHB information) with in the LNB signal samples in the form of coefficients or a parametric description of the LNB signal content.

2.3.2.1. Fundamentals

A basic model for a data-hiding system is depicted in figure 2.8. The general task of data hiding is to embed an MHB message m taken from a set of possible messages $M = \{0,1, \ldots\ldots, M-1\}$ into an LNB host signal (vector) $x \, \epsilon \, R^n$ by applying an embedding function $\dot{\mathrm{X}} = f(X, m)$. The modified signal (CLNB signal) $\dot{\mathrm{X}}$ has to be (in some sense) similar to the original LNB host signal X . At the same time, the MHB message m must remain recoverable from $\dot{\mathrm{X}}$ or even from a disturbed version $Y = \dot{\mathrm{X}} + \mathrm{N}$ of the signal with the (effective) additive noise term n. The decoded hidden message is denoted by $\acute{m} \, \epsilon \, M$. When it comes to data concealment, there are two primary sources of distortion: The first distortion is the so-called embedding distortion, which is caused by the embedding function itself. In addition, there's channel distortion to contend with in it.

Design a good data-hiding technique such that

- at the receiving end, we can retrieve hidden information reliably (possibly even after degradation of the modified signal by channel distortions),
- and the modified host signal (composite LNB signal) Ẋ should be subjectively indistinguishable from the original LNB host signal X.

Additional data (auxiliary data) is transmitted for improving the LNB host signal aims in this thesis. Transmission features like resilience to channel distortions are particularly critical in this scenario.

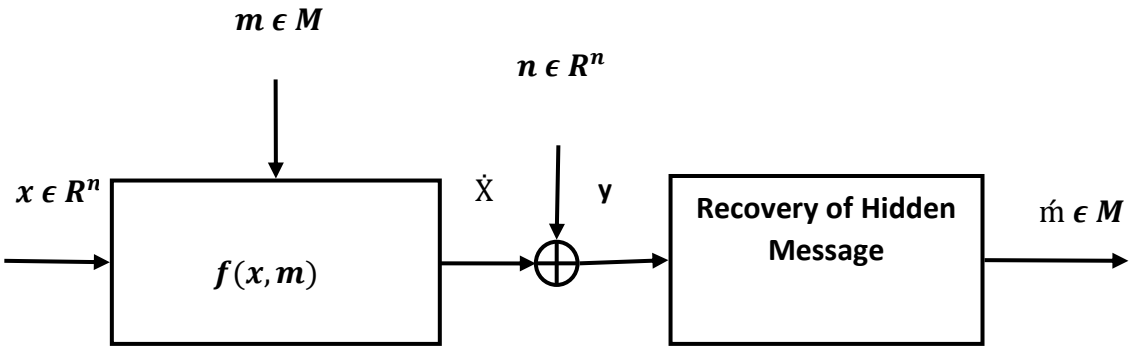


Figure 2.8: Generic model of a data hiding system

With existing codecs and networks, there are three methods to transmit data: by embedding additional information into the LNB signal itself [15–18, 96], by modifying encoded bitstream [19], or by combining source coding and hidden data inside the encoder [20–22]. Figure 2.9 illustrates three fundamentally distinct ways. In the following sections, we'll have a look at these systems:

I. A significant step towards a backward compatible CWB speech transmission may be taken by combining SBE with DWM techniques [15-18, 96]. The CWB speech transmission utilizing watermarking techniques is shown in figure 2.9 (a). The original CWB speech (0-7kHz) is first band-split using an LPF and an HPF with a sampling frequency of 16k Hz. The LPF output (0-3.5 kHz) is down-sampled to produce the LNB host signal. An SBE encoder module receives the HPF output in parallel. It is this module's responsibility to generate a collection of parameters known as the MHB message, m.

This MHB message m is fed to the watermarking embedded to be transmitted through the hidden channel. The MHB message m is hidden within the LNB host signal $X$and produces a CLNB speech signal Ẋ. This signal Ẋ is then sent over the TNC. Utilizing a well-designed watermarking technique Ẋ and X should be subjectively indistinguishable from each other. Because of this, an LNB speech decoder will provide results that are subjectively comparable to those results if coupled with any older speech encoder. This is one of the keys to enabling backward compatibility.

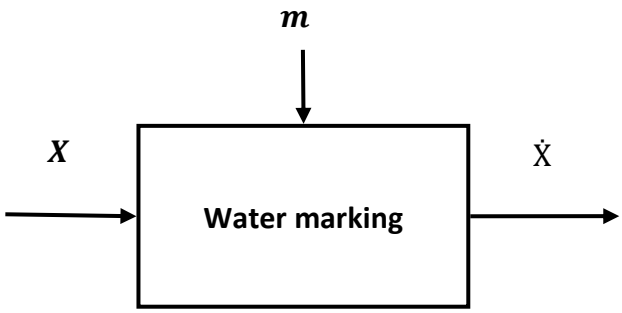


Figure 2.9: (a) Digital watermarking (DWM)

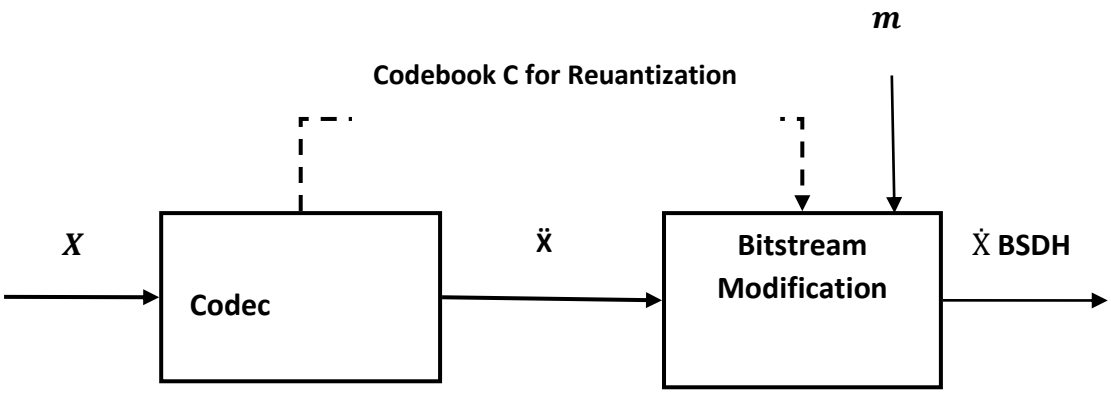


Figure 2.9: (b) Bitstream data hiding (BSDH)

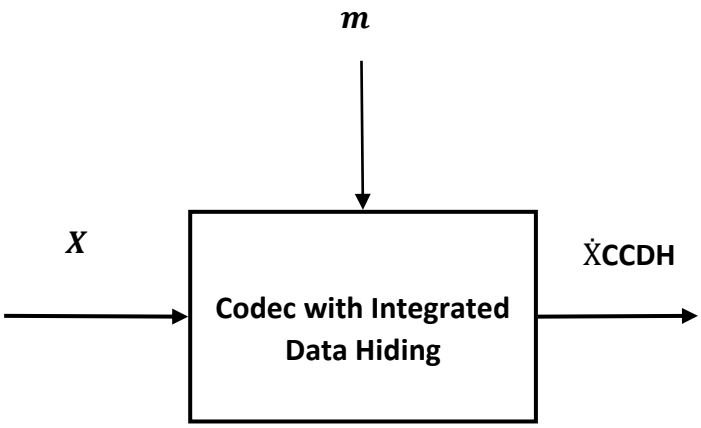


Figure 2.9: (c) combined coding and data hiding (CCDH)

II. It is also possible to include SBE data in an encoded or compressed version of the LNB signal. Bitstream data hiding (BSDH) or compressed domain data hiding is a common name for this approach. In this case, it is only relevant if the transmission system uses a speech codec (speech compression). It's only relevant if the taken-into-account transmission method uses signal compression (a speech codec). The next step is to embed the data and overwrite it directly into the bit stream. The least significant bits (LSBs) become complex using the previous quantization vector using the appropriate sub-codebook. Figure 2.9 (b) illustrates this system configuration.

III. Figure 2.9(c) illustrates a third strategy that uses the detailed data concealing and source encoding commonly found together. It is thus called "combined coding" or "data hiding" in this context (CCDH).DWM, BSDH, and CCDH versions and implementations for voice communication systems are discussed in the following sections.

2.3.2.2 Digital watermarking (DW)

Most of the time, data concealing for voice signals is done using PCM samples directly. In order to make the hidden watermark less noticeable, these changes frequently use masking techniques. Spread spectrum watermarking [102] and quantization-based approaches such as quantization index modulation [103] or the scalar costa scheme [104] are two common methods for implementing speech-specific DW. Alternatively, these methods can be applied in a modified environment. An inverse transform is also required in this scenario in order to recover the watermarked voice stream.

In the literature, several voice watermarking systems have been presented that employ one or more of the above-mentioned DW approaches [15-18, 96, 102-108]. MHB signal line spectrum pairs are concealed, encoded, and inserted into LNB speech to produce a CLNB speech signal [15]; this approach is referred to as SBE. A better CWB signal is reconstructed when the hidden information is retrieved and decoded at the receiving end. This method was shown to have a low-quality composite LNB signal. Phonetic categorization was used [16] to increase the quality of the CLNB signal and RCWB speech [15]. However, the methods in

[15, 16] have low SBE performance when tainted by channel noise. By removing undetectable LNB signal components, the authors of [17] suggested SBE with data concealing for LNB speech. The hidden audible components that exist outside of LNB are generated in order to reconstruct high-quality CWB speech. Only a certain number of audible missing frequency components may be inserted into the concealed channel, limiting the method's overall effectiveness.[106] and [18], for example, take into account the characteristics of signal equalization, synchronization, and noise in their suggestions. The data is embedded by altering the phase of LNB speech signals in [106]. The discrete Hartley transform domain is used by [18] to apply quantization-based watermarking methods to SBE. Several common telephone channels are used to transmit data reliably.

The system of [96] was meant to transmit digital speech. Lattice-based quantization index modulation incorporates the watermark signal into a subspace of linear prediction residual. We have demonstrated the use of digital waveform coders like ITU-T G.711 [109] and ITU-T G.726 [110] to insert secret data into LNB voice signals reliably. Voiced speech segments can be watermarked using a pitch-modification approach described in [105].

2.3.2.3 Bitstream data hiding (BDH)

BDH algorithms use the encoded bitstream in speech communication systems. The following is a list of BDH schemes in the literature. In most cases, the data-embedded methods are tailored to the individual codec they were created. The most common approaches are simple LSBs replacement, re-quantization, and reversible methods. These techniques use entropy coding on insignificant portions of the original bitstream to take advantage of redundant data left over from the compression process. In case the decoder is acknowledged of the data hiding, it will be able to fully recover the original (coded) host signal while the freed bits are used to inject a concealed message. By applying a parity condition for the entire bit group, BSDH also allows embedding one hidden bit within a group of source bits (for example, numerous LSBs). Even though just one byte is changed, the average embedding

distortion in the concrete is mitigated by permitting the embedding position to be variable. Appropriate parity constraints can be obtained using covering codes [111, 112].

Methods [113, [114], and [115] employ very simple LSB replacement for the ITU-T G.723.1 codec [116], for the conjugate structure-Algebraic code excited linear prediction (CS-ACELP) speech codec [117], and for the ITU G.711compander [109]. G.711 signal for SBE to CWB speech is achieved by exploiting concealed side information. LSB substitution is also used in [118]. There are two codecs used to disguise data: GSM enhanced full-rate (EFR) [120, 121] and GSM full-rate (FR) [119]. In these approaches, the exact placement of LSBs embedding is determined by the current speech frame's properties. The low-rate mixed excitation linear prediction speech vocoder [122] bitstream is embedded into the EFR codec's bitstream for the purpose of covert voice transmission.

A codec bitstream is overwritten with two LSBs of voice samples that have been quantized according to ITU-T G.711. This process significantly impacts the quality of the G.711 coded speech. However, decoding the changed G.711 stream is not generally expected. The ETSI standard for transporting, e.g., AMR-CWB or AMR-coded speech[123] across the TNC, can be viewed as a basic BSDH approach as another example.

For the GSM Full-Rate codec, [124] suggests a BSDH technique that uses re-quantization rather than LSB replacement. In [125],[126], and [127], the notation of covering codec is used for the ITU-T G.711 codec [109] and the ITU-T G.729 codec [117]. The ITU-T G.723.1 codec [116], respectively, [125], [126], and [127] follow the idea of embedding the hidden data in bit groups by using the concept of covering codes (e.g., parity constraints). G.729 codec is planned to have a reversible data concealing mechanism [128]. The original voice signal can be recovered by a decoder aware of the data concealing, i.e., there is no loss in quality. Telephony voice was watermarked using the LSB watermark approach in [19], which embeds MHB components within LNB speech to be retrieved at the receiver and reconstructed CWB signal with high quality.

2.3.2.4 Combined source coding and data hiding (CCDH)

Various ideas for CCDH in speech communication systems exist in the literature. Because the quantization or coding techniques must be tweaked directly, these solutions are particularly unique to the individual codec to which they are applied.

For example, [129] uses a hybrid BSDH and CCDH technique to modify the pitch lag of the third-generation Partnership Project adaptive multi-rate (AMR) codec [130, 131], resulting in an acceptable quality loss. ITU-T G.723.1 [116] and the internet low-bit-rate codec [133] are modified by [132]. The data is hidden inside the codecs' spectral envelope parameters. An optimum partitioning into two subsets is achieved with a chart-based quantization codebook representation. Even though such pseudo-random partitioning has a high memory expense, the results might be used as an upper bound for data hiding in the SE parameters.

In [134], another intriguing CCDH idea is offered. The quantizer for the GSM FR codec's prediction residual [119] has been updated for information embedding using the CCDH principle. The actual embedding was done using the covering code idea, which was a convolutional code in this case. In effect, the bitstream is subjected to a parity requirement once again. As a result, by recalculating this parity equation in the decoder, the secret data may be retrieved. The secret data might be encoded in the GSM FR coder's bitstream, with little impact on voice quality.

In [20], a more extensive range of LNB speech codecs was investigated and compared. The GSM FR codec [119], the ITU-T G.711 compander [109], the ITU-T G.726 adaptive delta pulse code modulation (ADPCM) codec [110], the ITU-T G.729 conjugate structure-Algebraic code excited linear prediction codec [117], and the GSM EFR codec [120] have all been investigated. For each codec, CCDH schemes were implemented. The subsequent degradation in LNB speech quality was minor to moderate based on the specific codec.

However, considerable quality enhancements may be gained when SBE is processed with concealed information on relevant parameters.[135] described an application of CCDH approaches to the fixed codebook of a code-excited linear prediction (CELP) voice codec. The codebook partitioning approach permitted overlapping partitions, and a Gaussian excitation codebook was assumed with the CELP model. ITU-T G.729 conjugate structure-Algebraic codec stimulated linear prediction codec [117] is addressed by [136]. [22] presented an improved CCDH approach for current Algebraic code excited linear prediction (ACELP) codecs, which was further investigated in [137] and [20]. CCDH was used in [20] to incorporate supplementary information into the LNB codec bit stream in order to create a backward-compatible CWB codec. [22] presented a backward compatible CWB telephony based on LNB coder and SBE with extra information contained in the LNB codec bit stream. When codec bits were damaged by channel noise, this strategy performed poorly. To construct a backward compatible CWB codec, [138,139] employed the CCDH approach to incorporate encoded SE and gain parameters of MHB into the bit stream of GSM FR 06.10 LNB speech coder.

2.4 Summary

It is typically considered desirable for a TNC to be able to transmit good-quality voice signals with a cut-off frequency of 7kHz. However, high-quality voice transmission in today's networks is hindered by expensive, and it takes more time to establish network equipment and communication protocols that must be modified.

To increase the quality of received band-limited (LNB) speech signals, an alternate yet the promising option is to use ASBE techniques, which extend the restricted frequency range of LNB speech at the receiving end.

This chapter started with ASBE to improve the quality of LNB speech and reduce the difference between LNB and CWB speech. The description of ASBE techniques comprised

the signal processing methods proposed for SBE of LNB speech. Following SFM, ASBE is performed separately for the SE and ES signals. The limitations on the achievable output quality of ASBE methods were also mentioned. A brief introduction to the embedded CWB speech coding was also provided. Then, several embedded CWB speech coding techniques presented in the literature have been described.

Unfortunately, ASBE approaches fail to provide a steady CWB speech quality under all conditions. Even the more robust CWB-embedded coding approach with quantized SBE settings cannot dependably improve the issue. This problem is then addressed with a novel approach (SBE with data concealing). The foundations of data concealing are also discussed. SBE strategies involving data concealing were described, as were the signal processing approaches proposed for LNB speech bandwidth extension. DW, BDH, and CCDH are three fundamentally distinct methods of the SBE using data hiding.

Chapter-3

This contribution describes the method proposed for the SBE of the LNB speech using a Hybrid Model Transform Domain-based data hiding (HMTDBDH). The chapter begins with the motivation for HMTDBDH in detail. Finally, the theoretical and simulation results of the proposed method show that it is robust to CAQNs.

3.1. Motivation

The existing methodologies failed to deliver high-quality CLNB and RCWB signals and had less vigor towards CAQNs. To overcome these drawbacks in the conventional SBE techniques, a novel SBE using an HMTDBDH technique for extending the bandwidth of TNC is proposed.

3.2. Introduction

Most traditional telephone networks allow only an LNB signal that is band-limited to 0.3-3.4kHz. Usually, human speech frequencies are confined far beyond the LNB frequency range. Transmission of human speech through telephone networks leads to muffled sound and poor-quality telephony speech. The transmission of CWB speech in the range of 0.5-7kHz would be desirable for better speech quality to solve this problem. To permit CWB speech in the network, the significant changes essential within the network architecture are quite expensive and time-taking [1]. This is happening to be a major hurdle for the transmission of high-quality speech in telephone networks. Therefore, other techniques are to be adopted to improve speech quality. To use the existing infrastructure and to improve the quality, SBE techniques can be implemented[2].

The existing telephone network can benefit significantly from improved speech quality due to SBE technology. Many SBE approaches have been proposed over the years. The ASBE is among various methods of SBE which can improve the intelligibility and quality of telephony speech. In the ABWE techniques, a CWB signal is generated by predicting the lost portion of the signal from the LNB speech alone. Most of the ASBE proposed in the literature is based on the SFM speech production system. The SFM system divides the SBE technique into ES extension and CWB speech signal SE estimation. Many methods for excitation enhancement are found in [197]. Many CWB spectral envelope approximation methods are illustrated in [189-193,197]. Even though ASBE has many advantages, there are a few limitations. Thus, it will not be able to reconstruct high-quality CWB [3].

The quality of CWB can be further improved when some supplementary information from out-of-band is communicated by hiding with the LNB signal [1]. When the embedded information is extracted at the receiver, a CWB signal with a much better speech quality can be reconstructed by combining the out-of-band signal transmitted by hiding within the LNB signal and the LNB signal. The speech bandwidth extension using data hiding approaches uses the original out-of-band information instead of its predicted signal, making the RCWB

speech signal more exact than the conventional ASBE. Several methods have been developed for this problem as a result of research efforts. An SBE technique has been stated in [15], accordingly that the encoded spectral envelop parameters (SEPs) of the missing spectral frequencies (MSFs) in the range of 4 to 8 kHz and known as MHB signal, is concealed into the LNB to generate a CLNB speech. A Technique for producing high-quality CWB over the above method was reported in [16], in which the MHB signal was encoded with high efficiency through phonetic classification. An SBE approach was reported in [19]. Accordingly, the SEPs of the MHB signal were concealed in the least significant bits of LNB. SBE based on the quantization-based data hiding technique has been stated in [18]. In [17], the noticeable components of the MHB signal are implanted within the hidden channel. The concealed data can be consistently reproduced at the destination. The better quality audio signal is retrieved in [4] using pitch-scaling. Enhancing the bandwidth using CCDH is introduced in [22]. A high-quality CWB signal is reproduced in [138, 139] based on the CCDH method.

The existing methodologies failed to deliver high-quality CLNB and RCWB signals along with vigor towards CAQNs. The SBE, using data hiding techniques, could deliver high-quality CLNB and RCWB signals and also be able to offer vigor towards CAQNs. Therefore, innovative SBE algorithms with data-hiding methods are vital for enhancing the quality of CLNB and RCWB signals and effectively managing CAQNs.

An HMTDBDH method is reported in [198] for embedding the secret signal in detailed constraints of the host speech signal in DWT coefficients of the cover signal without lowering the cover signal quality. It is observed that the HMTDBDH method could produce a stego signal which is indistinguishable from the cover signal and also be able to restore the secrete signal without lowering the quality. A novel robust SBE algorithm using HMTDBDH is proposed to insert the out-of-band spectral frequencies within the LNB signal. These embedded spectral frequencies are recovered steadily at the receiver side to produce a better-quality CWB signal.

3.3. Hybrid Model Transform Domain-based data hiding (HMTDBDH)

Consider an MHB signal $S_{mhb}(n)$ to be hidden within the LNB signal $S_{lnb}(n)$. Initially, DWT is performed on $S_{lnb}(n)$ to decompose $S_{lnb}(n)$ into high- and low-frequency components, then compute the spectrum by applying FFT on high-frequency Wavelet components, continued for generation of the magnitude spectrum $|S_{lnb}(k)|$and phase spectrum $\Phi_{lnb}(k)$.Consider that $S_{mhb}(n)$ is denoted with a representation vector, i.e., $C = [lsf_1, lsf_{2,\ldots\ldots\ldots}, lsf_{10}, g_r]$, where lSf denotes the line spectral frequencies and g_r represents the relative gain.

Every parameter of a vector C to be embedded and spread to distinct pseudo-random noise (PN) code, i.e.,$C_j \bullet q^{\rightarrow j}, 1 \leq j \leq R$. The PN code length is $q^{\rightarrow j}$is$R$. Adding all of these spreading vectors to produce concealed information is represented as

$$E(l) = \sum_{j=1}^{R} C_j q^j(l) \tag{3.1}$$

In this $q^j(l)$ is the l^{th} part of the $q^{\rightarrow j}$. The hidden information $E(l)$ is placed into the last 16 elements of the first half of $|S_{lnb}(k)|$ results in a modified magnitude spectrum $|S^1_{s\,lnb}(k)|$ and is given by

$$|S^1_{lnb}(k)| = \begin{cases} \left|S_{lnb}(k), k = 0, \ldots . \frac{M}{2} - 16\right| \\ E(l), k = \frac{M}{2} - 15, \ldots, \frac{M}{2} - 1 \\ . \\ E(l), k = \frac{M}{2}, \ldots . , \frac{M}{2} + 16 \\ |S_{lnb}(k)|, k = \frac{M}{2} + 17, \ldots, M - 1 \end{cases} \tag{3.2}$$

.

These changes in output and its spectrum of the composite signal represented as,

$$S(k) = |S^1_{s\,lnb}\,(k)|e^{j\Phi_{lnb}(K)}, k = 0, \ldots, M-1 \quad (3.3)$$

The inverse transform of the CLNB signal spectrum is converted back to the time representation of the CLNB signal by employing an Inverse FFT(IFFT) and then Inverse DWT(IDWT). The resultant CLNB signal $S^1_{lnb}(n)$ is transmitted through the telephone network to the receiver end. Here the TNC introduces CAQNs. Let $\hat{S}^{\,1}_{lnb}(n)$ denote received signal, i.e.,$\hat{S}^{\,1}_{lnb}(n) = S^1_{lnb}(n) + er$. The combination of CAQNs is represented by er. $\hat{S}^1_{lnb}(n)$ will be treated as an ordinary signal by a conventional phone terminal. $S_{lnb}(n)$ quality is not considerably degraded since the perceived differences between $S_{lnb}(n)$ and $S^1_{lnb}(n)$ is very low.

Restoration of the embedded information $\hat{S}_{mhb}(n)$ needs a receiver to compute the spectrum of the signal by applying DWT on $\hat{S}^{\,1}_{lnb}(n)$ then FFT is applied to high-frequency wavelet components, followed by the phase and magnitude spectrum calculation. The spread parameters of a vector C are then recovered from the magnitude spectrum of $\hat{S}^1_{lnb}(n)$ and are de-spread by using a correlator. Assuming a specific $\hat{C}_j$ represented as $\hat{C}_{jo}$to be recovered, the correlation can be expressed as

$$\hat{C}_{jo} = \frac{1}{R}\sum_{l=1}^{R} \hat{E}(l)\, q^{jo}(l) \quad (3.4)$$

where $\hat{E}(l)$ represents the corrupted version of $E(l)$ and is written as

$$\hat{E}(l) = E(l) + er(l) \quad (3.5)$$

Substitute (3.5) into (3.4), we have

$$\hat{C}_{jo} = \frac{1}{R}\sum_{l=1}^{R} \hat{E}(l)\, q^{jo}(l)$$

$$= \frac{1}{R}\sum_{l=1}^{R} q^{jo}(l)\left(\sum_{j=1}^{R} C_j\, q^{j}(l) + er(l)\right) = \frac{1}{R}\sum_{l=1}^{R} q^{jo}(l) \times \left(C_{jo} q^{jo}(l) + \sum_{j\neq jo} C_j q^{j}(l) + er(l)\right)$$

$$= C_{jo} + \frac{1}{R}\sum_{l=1}^{R}\sum_{j\neq jo} C_j q^{j}(l) q^{jo}(l) + \frac{1}{R}\sum_{l=1}^{R} q^{jo}(l) er(l) \qquad (3.6)$$

The orthogonal PN sequences. So, that

$$\sum_{l=1}^{R} q^{j}(l)\, q^{jo}(l) = 0,$$

Where j ≠ jo. Therefore,

$$\sum_{l=1}^{R}\sum_{j\neq jo} C_{jo} q^{j}(l) q^{jo}(l) = \sum_{j\neq jo} C_{jo} \sum_{l=1}^{R} q^{j}(l)\, q^{jo}(l) = 0 \qquad (3.7)$$

Also, since $q^{jo}(l)$and$er(l)$ are uncorrelated, that is

$$\frac{1}{R}\sum_{l=1}^{R} q^{jo}(l) er(l) \qquad (3.8)$$

when $R \to \infty$. Substitute (3.7) and (3.8) into (3.6), we have

$$\hat{C}_{jo} = C_{jo} \qquad (3.9)$$

3.4. Speech Bandwidth Extension Using HMTBDH Technique

3.4.1. Transmitter

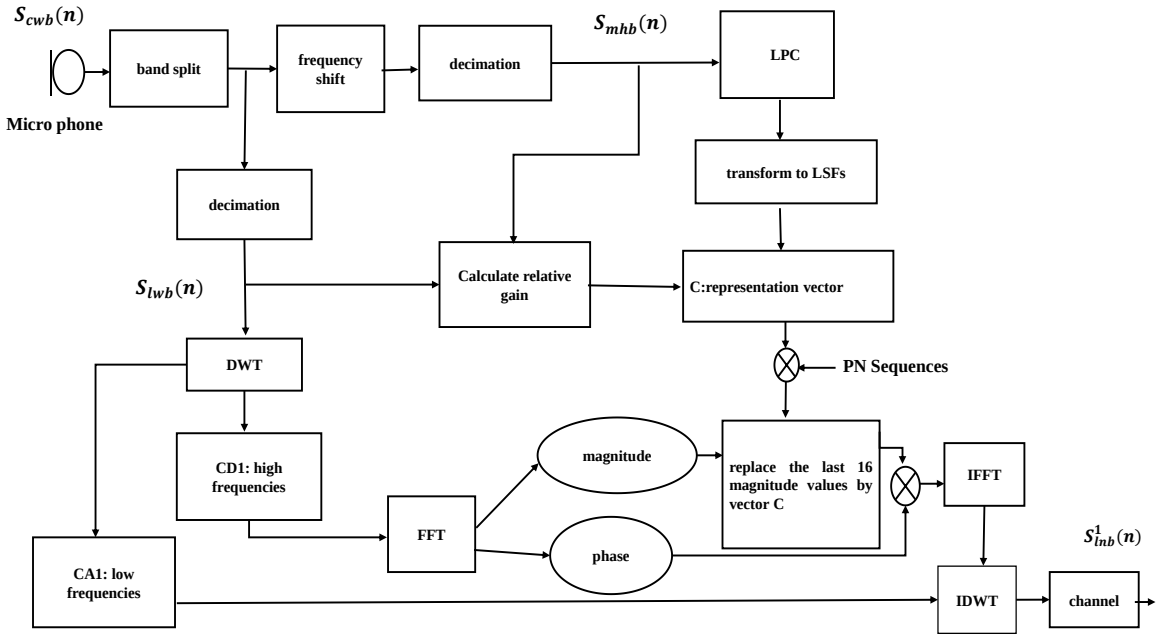


Fig. 3.1 Proposed HMTBDH transmitter

The HMTBDH transmitter is depicted in Fig.3.1. Initially, CWB speech $S_{cwb}(n)$ is passed through the LPF and the HPF separates the signal into a low and high band signal. Speech information is included in the low band signal between 0 and 4 kHz, whereas speech information is contained in the high band signal between 4 kHz and 8 kHz. The LNB signal $S_{lnb}(n)$ is produced by decimating LPF output by a factor of two. The output HPF is shifted to the LNB spectrum and then decimated to produce an MHB signal $S_{mhb}(n)$.

To imperceptibly embed the number of parameters that represent $S_{mhb}(n)$ into $S_{lnb}(n)$, minimize the number of parameters that represents $S_{mhb}(n)$. To produce LPCs, linear predictive analysis is carried out on $S_{mhb}(n)$ [170]. A small variation in LPCs results in substantial distortions when reconstructing $S_{mhb}(n)$; hence LPCs are modified into LSFs. Also, the gain of $S_{mhb}(n)$ has to be embedded to evade overestimation [171]. Thus, the

representation vector which represents $S_{mhb}(n)$ is formed by combining LSFs and gain as $g_r = \frac{g_{mhb}}{g_{lnb}}$, i.e., $C = [lsf_1, lsf_{2,\ldots\ldots,} lsf_{10}, g_r]$.The parameters which represent $S_{mhb}(n)$ are hidden using the HMTBDH technique in the LNB signal. Thus, a CLNB signal $S^1_{lnb}(n)$ is produced so that it can be communicated to the receiver on a TNC.

The ES has many parameters not implanted to lessen parameters that could be embedded. Since exceeding 3.4 kHz, the human ear remains insensitive to changes in ES [9]. Thus, the prediction of the MHB excitation signal from $S_{lnb}(n)$ at the receiver end assurances the retrieved signal performance.

A synchronization sequence (SYSE) like 1111….111 is introduced subsequently every frame of $S^1_{lnb}(n)$ to achieve frame synchronization [165] between the sender and receiver. The receiving of an SYSE designates the appearance of a novel frame.

3.4.2. Receiver

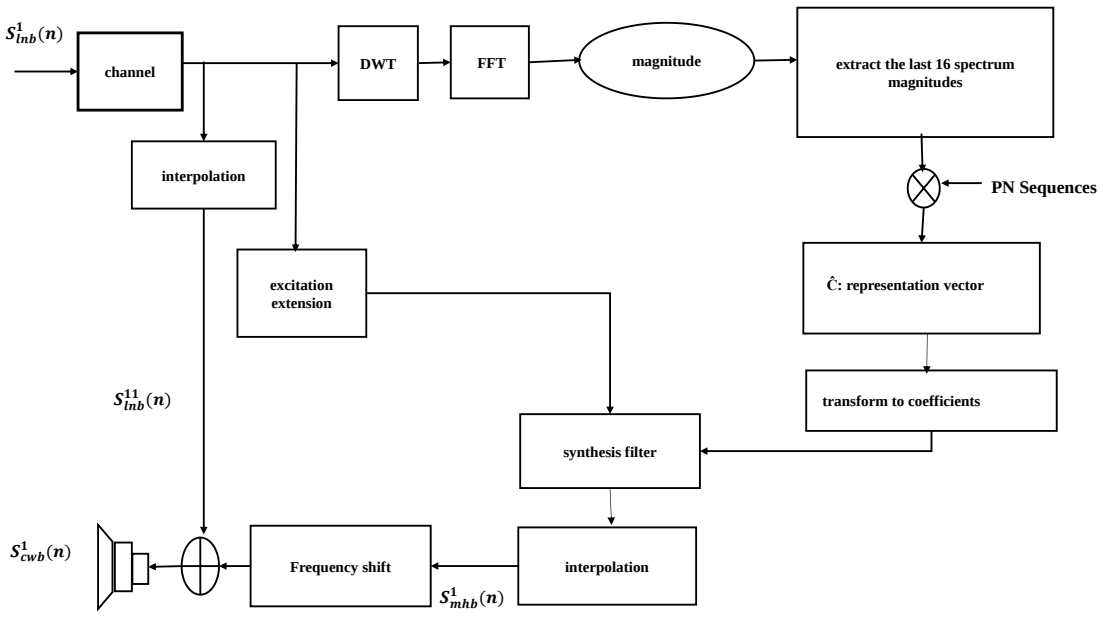


Fig. 3.2 Proposed HMTDBDH receiver

The HMTDBDH receiver is depicted in Figure 3.2. Using the HMTDBDH approach, accurately recover the representation vector and then generate LP coefficients from the retrieved lSf. Meanwhile, the inverse filtering is performed on $\hat{S}_{lnb}^{1}(n)$ which produces LNB residual signal. The residual signal is extended, which results in an MHB excitation signal as a consequence of this process. Synthesizing $\hat{S}_{mhb}(n)$is accomplished by stimulating the synthesis filter represented with the derived LP coefficients by an MHB excitation signal. At this stage, the sampling rate for both $\hat{S}_{lnb}^{1}(n)$ and $\hat{S}_{mhb}(n)$ is 8kHz. In order to increase the sampling rate to a wideband spectrum, these signals are interpolated.$S_{mhb}^{1}(n)$ denotes the interpolated $\hat{S}_{mhb}(n)$, resides in the LNB frequency band but has been relocated to the MHB band. The interpolated composite LNB ($S_{lnb}^{11}(n)$) and restored $S_{mhb}^{1}(n)$, signals are combined to create the CWB signal ($S_{cwb}^{1}(n)$) of high quality.

3.5. Evaluation

Twelve sentences spoken by ten male and ten female talkers were taken from the TIMIT database [166] for the performance evaluation. The LNB signal is decomposed into frames of 20ms duration with an overlap of 10ms among frames. The frames are then processed one by one. Subjective and objective tests are used to assess performance[140-149]. Speech quality can be better evaluated with subjective tests. The proposed methodology competency is explored by comparing with the traditional techniques, such as traditional telephony speech bandwidth extension (TTSBE) by data hiding (TTSBEDH) [15], TTSBE by phonetic classification (TTSBEPC) [16], TTSBE using bitstream data hiding (TTSBEBDH) [19], and TTSBE using watermark side information (TTSBEWTSI) [96]. The channel models considered here are μ-law and additive white Gaussian noise (AWGN).

3.5.1. Subjective Test assessments

The perceptual clearness is assessed based on the mean opinion score (MOS) test [15,16]. The listening test compares various speech signals like CWB, LNB, CLNB, and RCWB [96]. These tests were performed in a silent room using headsets. During each test, thirty participants are considered.

3.5.1.1 Perceptual Clearness (PCL)

In the proposed technique, the information must be transparently concealed, i.e., LNB and CLNB are subjectively indistinguishable. High PCL means low perceptible degradation in the CLNB signal. PCL is assessed based on the MOS test. Listeners participating in the test compare LNB and CLNB to provide a decision in terms of MOS, tabulated in Table 3.1. Table3.2. showcases the results of the averaged MOS for conventional [15,16,19,96] and proposed approaches. MOS values show the remarkable perceptual clearness of the proposed approach over the traditional approaches in Table 3.2.

Table 3.1. MOS

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

Table 3.2. MOS assessment outcomes

Technique	Mean opinion score
TSBWEDH [15]	2.97
TSBWEPC [16]	3.16
TTSBEBDH [19]	3.28
TTSBEWTSI [96]	3.64
Proposed method	3.88

3.5.1.2. Subjective contrasts among CWB, LNB, CLNB, and RCWB signals

I, II, III, and IV in Table 3.3 represent the CWB signal, LNB signal, CLNB signal, and RCWB signal. The subjects are asked to do a pairwise analysis of signals among I to IV and must tell whether the first signal is paramount (>), deprived (<), or alike (≈) to the second signal. Table 3.3 provides the responses of pairwise comparison of I, II, and III to the other signal, Table 3.4 provides the responses of pairwise comparison of II and III to the other signal, and Table 3.5 provides the responses of comparison among III and IV. The number of

subjects with an exact preference (> or < or ≈) is mentioned with Arabic digits in the table. The CWB signal outperforms the CLNB signal for conventional [15,16,19,96] and proposed methods which are endorsed by Table 3.3. Table 3.3 also endorsed a clearly enhanced RCWB signal quality of the proposed technique over the conventional methods. The remarkable perceptual clearness of the proposed approach over the traditional approaches is endorsed by Table 3.4. Compared to conventional methods, the RCWB signal is better than the LNB signal for the proposed approach which is endorsed in Table 3.4. Compared to conventional methods, the RCWB signal is better than the CLNB signal for the proposed approach endorsed in Table 3.5.

Table 3.3. Subjective contrast outcomes among I, II, III, and IV

Technique	I	II	III	IV
TSBWEDH [15]	>	30	30	10
	<	0	0	0
	≈	0	0	20
TSBWEPC [16]	>	30	30	8
	<	0	0	0
	≈	0	0	22
TTSBEBDH [19]	>	30	30	7
	<	0	0	0
	≈	0	0	23
TTSBEWTSI [96]	>	30	30	9
	<	0	0	0
	≈	0	0	21
Proposed method	>	30	30	4
	<	0	0	0
	≈	0	0	26

Table 3.4. Subjective contrast outcomes among II, III, and IV

Technique	II	III	IV
TSBWEDH [15]	>	6	5
	<	4	18
	≈	20	7
TSBWEPC [16]	>	7	2
	<	2	18
	≈	21	10
TTSBEBDH [19]	>	4	3
	<	4	19
	≈	22	8
TTSBEWTSI [96]	>	6	2
	<	3	22
	≈	21	6
Proposed method	>	4	0
	<	0	25
	≈	26	5

Table 3.5. Subjective contrast outcomes among III and IV

Technique	III	IV
TSBWEDH [15]	>	4
	<	18
	≈	8
TSBWEPC [16]	>	4
	<	17
	≈	9
TTSBEBDH [19]	>	4
	<	19
	≈	7
TTSBEWTSI [96]	>	5
	<	20
	≈	5
Proposed method	>	0
	<	25
	≈	5

3.5.2. Objective Quality Assessments

The RCWB signal quality is assessed with log spectral distortion (LSD) and CWB-perceptual evaluation of speech quality (CWB-PESQ), and CWB-POLQA tests [199]. The perceptual clearness is assessed with the LNB-PESQ and LNB-POLQA tests. The robustness of concealed data to CAQNs is evaluated with a bit error rate (BER) measure.

3.5.2.1. Perceptual Clearness (PCL)

Technique	LNB-PESQ
TSBWEDH [15]	2.97
TSBWEPC [16]	3.18
TTSBEBDH [19]	3.56
TTSBEWTSI [96]	3.58
Proposed method	3.63

Tab. 3.6. LNB-PESQ test Outcomes

The LNB-PESQ test assesses PCL by comparing the LNB signal with the CLNB signal. LNB-PESQ ranges from 0.5 to 4.5. Lower values, such as 0.5, represent the worsened

PCL, and higher values, like 4.5, represent the best PCL. Table 3.6 lists the responses of mean scores for the traditional [15,16,19,96] and proposed techniques. An apparent PCL enhancement of the proposed approach over the conventional methods is witnessed from the scores listed in Table 3.6.

The evaluation of perceptual clearness is done by providing LNB and CLNB signals as inputs and comparing them to rate speech quality. The LNB-POLQA value will range between 1 and 5; the higher the value, the superior the quality. The average LNB- POLQA values of conventional [15,16,19,96] and proposed methods are tabulated in table 3.7. The proposed technique gives an LNB-POLQA value of 4.05, which indicates that the proposed approach has excellent perceptual clearness over traditional techniques [15,16,19,96], which was already confirmed by subjective listening tests.

Tab. 3.7. Results of LNB-POLQA

Technique	LNB-POLQA
TSBWEDH [15]	2.54
TSBWEPC [16]	2.99
TTSBEBDH [19]	3.21
TTSBEWTSI [96]	3.34
Proposed method	4.05

The TD waveforms of LNB and CLNB signals are shown in Figures 3.3 (a) and 3.3 (b). It is clear from the figures that LNB and CLNB signals are almost indistinguishable.

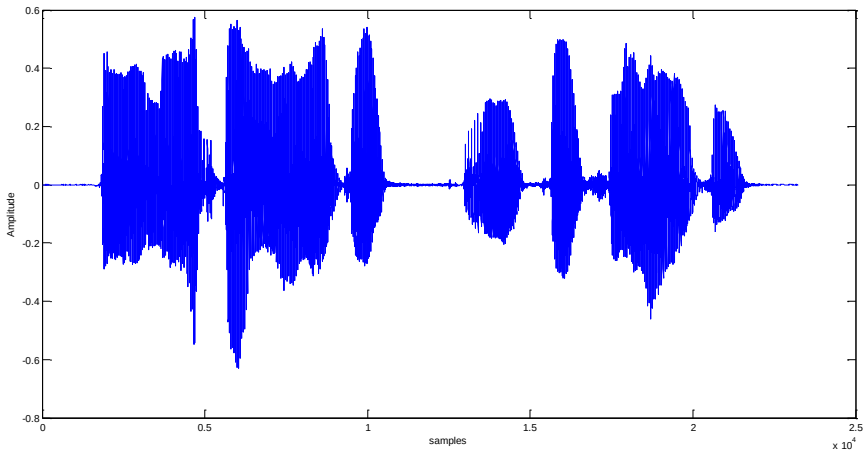


Figure 3.3: (a) LNB signal

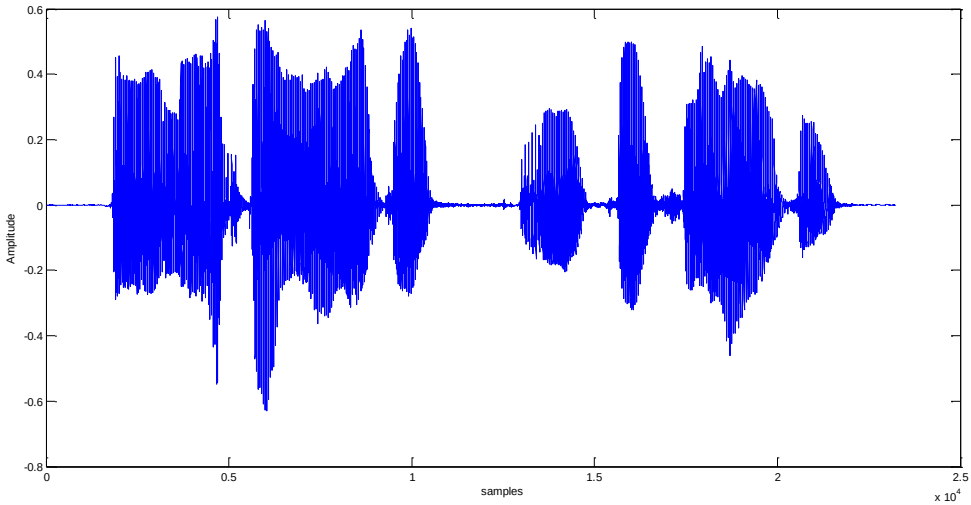


Figure 3.3: (b) CLNB signal

3.5.2.2. RCWB Signal Quality

The quality of RCWB speech is evaluated by comparing CWB and RCWB signals in the CWB-PESQ test. Table 3.8 presents the mean CWB-PESQ scores of the conventional [15,16,19,96] and proposed methods. The proposed technique produces a score of 4.01, which specifies that the RCWB signal quality attained is remarkable. Thus, the proposed approach improved the speech quality when compared to the traditional methods.

Tab.3.8 CWB-PESQ test Outcomes

Technique	CWB-PESQ
TSBWEDH [15]	2.46
TSBWEPC [16]	2.89
TTSBEBDH [19]	3.62
TTSBEWTSI [96]	3.73
Proposed method	4.01

The evaluation of the quality of RCWB speech is done by giving CWB and RCWB signals as inputs and comparing them in order to rate speech quality. The average CWB-POLQA values of the conventional [15, 16, 19, 96] and proposed methods are shown in table 3.9. A CWB- POLQA value of 4.02 confirms that the RCWB signal quality that was obtained by the proposed technique is excellent compared to traditional approaches [15, 16, 19, 96],

which was already confirmed by subjective listening tests on a set of participants. Thus, the speech quality was improved by using the proposed technique.

Tab. 3.9. Results of CWB-POLQA

Technique	CWB-POLQA
TSBWEDH [15]	2.08
TSBWEPC [16]	2.45
TTSBEBDH [19]	3.13
TTSBEWTSI [96]	3.34
Proposed method	4.02

3.5.2.3. Comparison of original and reconstructed MHB speech

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

$$LSD = \frac{1}{2\pi}\int_{-\pi}^{\pi}\left(20\log_{10}\frac{g_p}{a_s(e^{jw})} - 20\log_{10}\frac{\hat{g}_p}{|\hat{a}_s(e^{jw})|}\right)^2 dw \qquad (3.10)$$

where g_p and $\hat{g}_p$ are gains of true and restored MHB signals, respectively, $\frac{1}{a_s(e^{jw})}$ and $\frac{1}{\hat{a}_s(e^{jw})}$ are the spectral envelopes of true and restored MHB signals respectively. In general, the best-quality of reproduced MHB signal has a low value of LSD. Table 3.10 lists the mean LSD scores for the existing [15, 16, 19, 96] and proposed schemes under μ-coding. There is a clear enhancement in the quality of the proposed method over the traditional schemes [15, 16, 19, 96] witnessed from the values listed in Table 3.10.

Tab. 3.10 LSD test Outcomes

Technique	Log Spectral Distortion
TSBWEDH [15]	11.75
TSBWEPC [16]	9.57
TTSBEBDH [19]	4.98
TTSBEWTSI [96]	4.87
Proposed method	2.23

3.5.2.4. The vigor of concealed data

AWGN with SNR ranges from 15 to 35 dB is summed up with CLNB signal [183]. The proposed method's robustness is assessed based on BER. The PN code size is 8. The lower value of BER designates the RCWB signal of high quality. The BER values which were attained with SNR in the range of 15 to 35 dB are beneath $5.036 * 10^{-4}$which endorses the RCWB signal of high quality. The BER value, which is attained with μ-law coding, is $8.082 * 10^{-4}$which endorses the RCWB signal of high quality.

3.6. Results and Conclusions

A new SBE algorithm using HMTDBDH is proposed. The spread SEPs of the MHB signal is hidden in the high-frequency wavelet coefficients of the LNB signal. The hidden information is retrieved at the receiving end to produce a high-quality CWB signal. The subjective and objective assessment results confirm the excellent clear wideband performance of the proposed algorithm over the traditional SBE methods.

Chapter-4

The chapter begins with the motivation for LNB-SBE aided by Discrete Wavelet Transform-Discrete cosine transform-Based Data Hiding (DWT-DCT-BDH) described in detail. The performance of the proposed method under CAQNs is also analyzed. Finally, the proposed method's subjective and objective test results are discussed.

4.1. Motivation

When corrupted by CAQNs, conventional SBE approaches employing data hiding produced poor quality RCWB signal and CLNB signal. To further increase the quality of the RCWB signal and CLNB signal over the contribution 1 [1*] and conventional SBE techniques, a novel LNB SBE using DWT-DCT-BDH is proposed.

4.2. Introduction

Human speech may have frequencies more than conventional telephone networks operating at 300-3400Hz. When a human speech signal is transmitted through the telephone network leads to losing information due to the LNB of the telephone network. This results in significantly low quality and lucidity of speech transmission. This problem can be solved by using a CWB whose spectrum ranges from 50-7000Hz. As a traditional telephone network installed to operate at 0.3–3.4kHz, it is not feasible to work at a wideband spectrum. Hence, using a wideband spectrum must install a new network, which will be very expensive and take more time to establish [1]. Therefore, other techniques are to be adopted to improve speech quality. SBE techniques [2] can be implemented to use the existing infrastructure and improve the quality.

In the ASBE techniques, a clear wideband (CWB) signal is generated by predicting the lost portion of the signal from the LNB speech alone. Most ASBE methods proposed in the literature are based on the SFM speech production system. The SFM divides the SBE technique into ES extension and CWB speech signal SE estimation. Many methods for excitation enhancement are found in [197]. Many methods for CWB spectral envelope approximation are illustrated in [189-193,197]. Even though ASBE has many advantages, there are a few limitations, like its performance is limited. Thus, it will not be able to reconstruct high-quality CWB signals [3].

The quality of CWB can be further improved when some supplementary information from out-of-band is communicated by hiding with the LNB signal [1]. When the embedded information is extracted at the receiver, a CWB signal with a much better speech quality can be reconstructed by combining the out-of-band signal transmitted by hiding within the LNB signal and the LNB signal. The SBE using data hiding approaches uses the original out-of-band information instead of its predicted signal, making the RCWB speech signal more exact than the conventional ASBE. Several methods have been developed for this problem as a result of research efforts. An SBE technique has been stated in [15], accordingly that the encoded SEPs of the MSFs in the range of 4 to 8 kHz and known as MHB signal, is concealed

into the LNB to generate a CLNB speech. A Technique for producing high-quality CWB over the above method was reported in [16], in which the MHB signal was encoded with high efficiency through phonetic classification. A BE approach was reported in [19], accordingly that SEPs of MHB signal were concealed into the least significant bits of LNB. SBE based on the quantization-based data hiding technique has been stated in [18]. In [17], the noticeable components of the MHB signal are implanted within the hidden channel. The concealed data can be consistently reproduced at the destination[172-178]. The audio signal of better quality is regenerated in [4] using pitch-scaling. Enhancing the bandwidth using CCDH is introduced in [22]. A High-quality CWB signal is reproduced in [138, 139] based on CCDH Method.

SBE techniques with data hiding are expected to deliver high-quality CLNB alongside RCWB signals. Also, able to handle issues pertaining to CAQNs. However, these techniques fail to provide high-quality CLNB and RCWB signals. Also, they are less robust to CAQNs. Thus, developing a novel SBE technique using data hiding is essential to improve the quality of CLNB and RCWB signals and make them more robust to CAQNs.

An audio steganography technique is presented in [198], using the DWT-DFT-BDH technique to insert the secret message signal in detailed coefficients of a host speech signal without degrading the perceptual quality of the host signal. DCT is used here instead of FFT in the DWT-DFT-BDH technique. A novel SBE algorithm using the DWT-DCT-BDH technique is proposed to insert the parameters of the lost speech frequency components within the detailed coefficients of the LNB signal. These hidden parameters are retrieved at the receiver side to produce a better-quality RCWB signal by combining the missing speech signal transmitted through the detailed coefficients and the LNB signal. The proposed method uses the actual MHB speech signal instead of its prediction, making the reconstruction of the CWB speech much better quality than conventional SBE methods.

4.3. Discrete Wavelet Transform-Discrete cosine transform-Based Data Hiding

To embed the MHB signal $S_{mhb}(n)$into LNB signal $S_{lnb}(n)$, firstly, detailed and approximation coefficients are computed by applying DWT on $S_{lnb}(n)$ and then DCT coefficients are computed by applying DCT on detailed coefficients. Assume that the representation vector which represents $S_{mhb}(n)$ is $R = \left[LSF_1, LSF_{2,\ldots\ldots\ldots,}LSF_{10}, \bar{\mathrm{G}}_r\right]$, where line spectral frequencies are denoted by LSF and gain is denoted by $\bar{\mathrm{G}}_r$.

By multiplying R with a certain PN sequence, each parameter is spreaded. i.e.,$\check{\mathrm{D}}_i \bullet p^{\rightarrow i}, 1 \leq i \leq K$. Where K is the PN sequence $p^{\rightarrow i}$length. The hidden data is then produced by adding all of these spreading vectors and is given by

$$V(j) = \sum_{i=1}^{K} \check{\mathrm{D}}_i p^i(j) \qquad (4.1)$$

where j^{th} element of $p^{\rightarrow i}$ represented by $p^i(j)$. The last 16 DCT coefficients are replaced by $V(j)$ results frequency-domain CLNB signal spectrum [198]. IDCT and IDWT are applied to the CLNB signal spectrum to convert back the time-domain representation. Thus, a CLNB signal $S_{lnb}^1(n)$ is created to be transmitted to the receiver on a TNC, and the CAQNs are introduced by TNC. Assume that the received signal is represented by $\hat{\mathrm{S}}\,_{lnb}^{1}(n)$ i.e., $\hat{\mathrm{S}}_{lnb}^1(n) = S_{lnb}^1(n) + \bar{\mathrm{e}}$.Where $\bar{\mathrm{e}}$ represents the combination of CAQNs. The conventional phone terminal treats $\hat{\mathrm{S}}_{lnb}^1(n)$as an ordinary signal. The LNB signal quality is not noticeably degraded since there is a very small perceived difference between $S_{lnb}(n)$ and $S_{lnb}^1(n)$ [179-182].

At the receiver, retrieving the embedded data requires applying DWT on the CLNB signal and then applying DCT on detailed coefficients to obtain the DCT coefficients. The spread parameters are then obtained from the last 16 DCT coefficients, which are de-spread

using a correlator . Assuming a particular $\underset{\wedge}{D}_i$ to be retrieved is denoted by $\underset{\wedge}{D}_{io}$ and then the correlation is given by

$$\underset{\wedge}{D}_{io} = \frac{1}{K}\sum_{j=1}^{K} \underset{\cdot}{V}(j)\, p^{io}(j) \tag{4.2}$$

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

$$\underset{\cdot}{V}(j) = V(j) + \bar{e}(j) \tag{4.3}$$

Equation (4.3) is substituted into equation (4.2), we have

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

$$= \frac{1}{K}\sum_{j=1}^{K} p^{io}(j)\left(\sum_{j=1}^{K} \check{D}_i p^{i}(j) + \bar{e}(j)\right)$$

$$= \frac{1}{K}\sum_{j=1}^{K} p^{io}(j) \times \left(\check{D}_{io} p^{io}(j) + \sum_{i\neq io} \check{D}_i p^{i}(j) + \bar{e}(j)\right)$$

$$= \check{D}_{io} + \frac{1}{K}\sum_{j=1}^{K}\sum_{i\neq io} \check{D}_i p^{i}(j) p^{io}(j) + \frac{1}{K}\sum_{j=1}^{K} p^{io}(j)\bar{e}(j) \tag{4.4}$$

The PN sequences are orthogonal. i.e.$\sum_{j=1}^{K} p^{i}(j)p^{io}(j) = 0$, where *i≠ io.* Therefore,

$$\sum_{j=1}^{K}\sum_{i\neq io} \check{\mathrm{D}}_{io} p^{i}(j)p^{io}(j) = \sum_{i\neq io} \check{\mathrm{D}}_{io} \sum_{j=1}^{K} p^{i}(j)p^{io}(j) = 0 \qquad (4.5)$$

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

$$\frac{1}{K}\sum_{j=1}^{K} p^{io}(j)\,\bar{\mathrm{e}}(j) \qquad (4.6)$$

when $K \to \infty$. Equations (4.5) and (4.6) are substituted into equations (4.4);thus we have

$$\hat{\mathrm{D}}_{io} = \check{\mathrm{D}}_{io} \qquad (4.7)$$

This illustrates that the parameters which represent $\hat{\mathrm{S}}_{mhb}(n)$isrecovered more efficiently with the use of the spread spectrum.

4.4. SBE using Discrete Wavelet Transform-Discrete cosine transform-Based Data Hiding

4.4.1. Transmitter

The proposed DWT-DCT-BDH transmitter is depicted in Fig.4.1.The CWB speech signal is designated as $S_{cwb}(n)$with a sampling rate of 16 kHz. This signal is passed through an LPF and HPF filter to generate LNB and MHB signals. The LPF extracts speech signal information that is present between 0 and 4 kHz and is designated as a low-band signal. In contrast, the HPF extracts speech information that is present between 4 kHz and 8 kHz and designated as a high-band signal. The low-pass filter output is decimated by a factor of two in

order to produce an LNB signal $S_{lnb}(n)$. The high-band signal is decimated to produce an MHB signal $S_{mhb}(n)$. Therefore, 8 kHz is the sampling frequency of $S_{lnb}(n)$ and $S_{mhb}(n)$.

To imperceptibly embed $S_{mhb}(n)$ into $S_{lnb}(n)$, minimize the number of parameters that represents $S_{mhb}(n)$. To produce LPCs, LPA is carried out on $S_{mhb}(n)$ [170,182,184]. A small variation in LPCs results in substantial distortions when reconstructing $S_{mhb}(n)$; hence LPCs are modified into LSFs. Also, the gain of $S_{mhb}(n)$ has to be embedded to evade overestimation [171]. Thus, the representation vector which represents $S_{mhb}(n)$ is formed by combining LSFs and gain, $R = \left[LSF_1, LSF_{2,\ldots\ldots,} LSF_{10}, \bar{G}_r\right]$. The parameters which represent $S_{mhb}(n)$ are hidden using the DWT-DCT-BDH technique in the LNB signal. Thus, a CLNB signal $S_{lnb}^1(n)$ is created to transmit it to the receiver across a TNC.

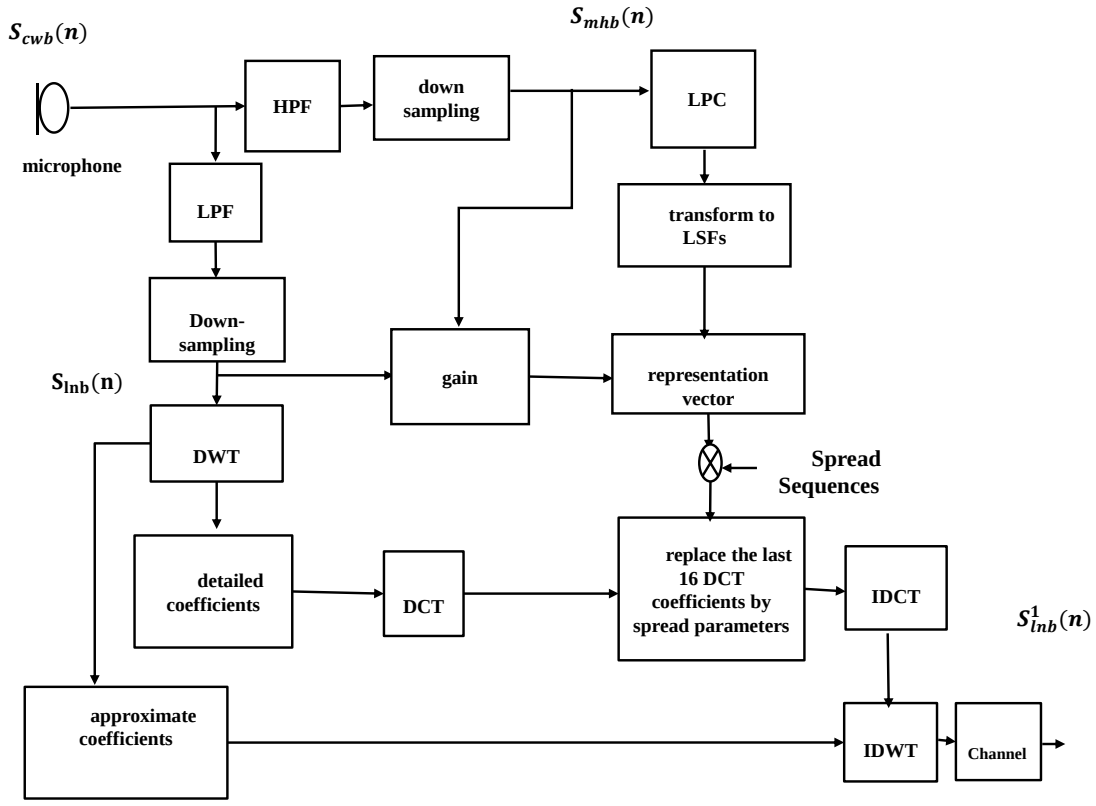


Fig. 4.1. Proposed DWT-DCT-BDH Transmitter.

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

4.4.2. Receiver

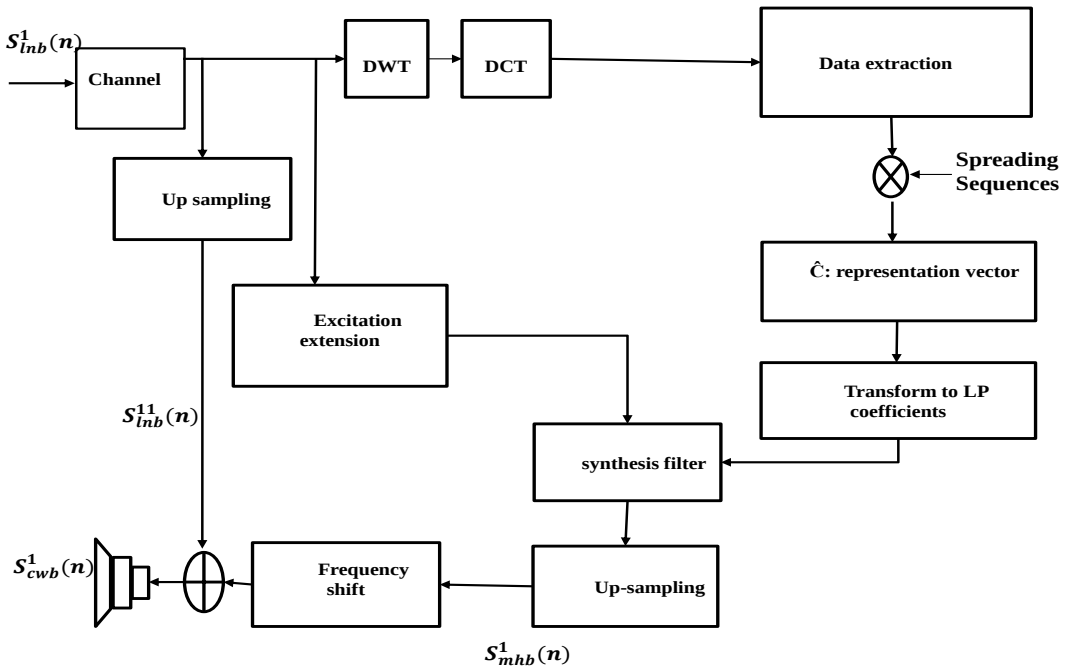


Fig. 4. 2. Proposed DWT-DCT-BDH Receiver.

The proposed DWT-DCT-BDH receiver is depicted in Fig. 4.2. The DWT-DCT-BDH technique properly recovers the representation vector, and then LPCs are obtained from LSFs. Meanwhile, LNB residual signal is obtained by inverse filtering $\hat{S}^1_{lnb}(n)$ using LPCs of $\hat{S}^1_{lnb}(n)$ and then obtain the MHB excitation signal by extending the LNB residual signal. The MHB signal $\hat{S}_{mhb}(n)$ that was embedded is synthesized by exciting the synthesis filter described by the recovered LPCs by an MHB excitation signal. An 8 kHz sampling rate is used to sample the recovered MHB signal and the received CLNB signal. These signals are

then interpolated by a factor of two. The interpolated CLNB ($S_{lnb}^{11}(n)$) and MHB ($S_{mhb}^{1}(n)$) signals are added up to reproduce a CWB signal ($S_{cwb}^{1}(n)$) of good quality.

4.5. Experimental Results

The speech samples used for the performance evaluations of traditional and proposed SBE techniques were obtained from the TIMIT database [166]. The evaluations were done by taking thirty different speech samples, of which thirty male and thirty female speakers spoke. Each speech signal was split to form frames 20 ms long, and an overlap of 10 ms was maintained between frames. Each frame was processed individually. The performance assessment of the methods was done by considering the subjective and objective measures. The proposed methodology competency is explored by comparing it with traditional techniques, such as TTSBEDH [15], TTSBEPC [16], TTSBEBDH [19], and TTSBEWTSI [96]. The channel models considered here are AWGN and μ-law model.

4.5.1. Subjective quality assessment

The obtained speech quality of the proposed and conventional SBE methods [15, 16, 19, 96] is assessed using an absolute category rating (ACR) listening test recommended by ITU-T [140,150-153]. The perceptual transparency is assessed with the mean opinion score (MOS) test [15,16]. The subjective comparison between CWB, CLNB, LNB, and RCWB signals is also employed [15]. During each test, thirty participants are considered.

4.5.1.1. ITU-T Test Results

The 100 speech samples from the TIMIT corpus database were used to prepare the listening test and compare the performance of the proposed method with traditional methods [15, 16, 19, 96]. The listening test samples are created in such a way as to mimic speech

carried over a cellular telephone network because the SBE method is mainly used in mobile communications. The sound level of each test sample was standardized to 26 dB below overloading [150] after the test samples were high-pass filtered with the mobile station input (MSIN) filter, which simulates the input response of a mobile station. These previously processed test samples were then down-sampled to an 8 kHz sampling rate and used as the LNB signal for the existing SBE techniques [15, 16, 19, 96] and the proposed technique.

The quality of speech signals generated by the proposed and the conventional SBE methods was compared using the ACR test. On a scale of 5 (excellent), 4 (good), 3 (fair), 2 (poor), and 1 (bad), Listeners were asked to rate in a quiet environment using headphones. Thirty subjects participated in the test. MOS values for the conventional SBE methods [15, 16, 19, 96] and the proposed method are presented in Table 4.1. An improved reconstructed CWB signal quality of the proposed method over the traditional methods is observed in Table 4.1.

Tab. 4.1. ACR listening test results

Technique	Mean opinion score
TSBWEDH [15]	2.45
TSBWEPC [16]	2.76
TTSBEBDH [19]	3.54
TTSBEWTSI [96]	3.61
Proposed method	4.57

4.5.1.2. Perceptual clearness (PCL)

The MHB signal in the proposed method should be transparently hidden. That is, the CLNB and LNB signals should be subjectively indistinguishable. High perceptual clearness means low noticeable LNB signal degradation. The perceptual clearness was assessed with the MOS test[151]. Listeners comparing CLNB and LNB signals decide in terms of MOS, as given in Table 4.2. The average MOS values of traditional [15, 16, 19, 96] and the proposed

techniques are shown in Table 4.3. The proposed method gives a MOS value of 3.97, which indicates that the proposed approach has excellent perceptual transparency over the traditional techniques [15, 16, 19, 96]. The proposed method gives a MOS value of 3.97, almost near the standard MOS value of 4, indicating that CLNB and LNB signals were more or less identical.

Tab. 4.2. MOS

score	Instruction
1	LNB and CLNBsignals sound different
2	Observable difference between LNB and CLNBsignals
3	Minute difference between LNB and CLNBsignals
4	LNB and CLNB signals sound alike

Tab. 4.3. Results of the MOS

Technique	Mean opinion score
TSBWEDH [15]	2.89
TSBWEPC [16]	3.07
TTSBEBDH [19]	3.18
TTSBEWTSI [96]	3.54
Proposed method	3.97

4.5.1.3. Subjective Comparisons between CWB, LNB, CLNB and RCWB Speech samples

A listening test was done to compare performances between the proposed and conventional methods [15,16,19,96]. Here, the CWB signal, LNB signal, CLNB signal, and RCWB signal were labeled I, II, III, and IV, respectively. Listeners are asked for pair wise comparison among the samples to tell whether the first sample was paramount to, deprived, or alike to the second. The corresponding responses after comparing I, II, and III with the other signals are tabulated in Tables 4.4, 4.5, and 4.6. Arabic numerals indicate the number of listeners with a specific preference in the table. It is observed that the CWB signal is superior to LNB and CLNB signals of traditional [15,16,19,96] and the proposed methods from Table 4.4 . Also, we observe that RCWB signal quality is far superior using the proposed method over traditional methods [15,16,19,96] from Table 4.4. Thus, the speech quality was enhanced by the proposed technique. Compared to conventional methods [15,

16, 19, 96], it is observed that the RCWB signal of the proposed method is superior to that of the LNB signal, as may be seen from Table 4.5. Compared to conventional methods [15, 16, 19, 96], it is observed that the RCWB speech of the proposed technique is better than CLNB speech from Table 4.6.

Tab.4.4. Subjective comparison test results between I, II, III,and IV

Technique	**I**	**II**	**III**	**IV**
TSBWEDH [15]	>	30	30	14
	<	0	0	0
	≈	0	0	16
TSBWEPC [16]	>	30	30	12
	<	0	0	0
	≈	0	0	18
TTSBEBDH [19]	>	30	30	11
	<	0	0	0
	≈	0	0	19
TTSBEWTSI [96]	>	30	30	9
	<	0	0	0
	≈	0	0	21
Proposed method	>	30	30	1
	<	0	0	0
	≈	0	0	29

Tab. 4.5. Subjective comparison test results between II, III, and IV

Technique	**II**	**III**	**IV**
TSBWEDH [15]	>	8	3
	<	4	18
	≈	18	9
TSBWEPC [16]	>	8	1
	<	2	19
	≈	20	10
TTSBEBDH [19]	>	5	2
	<	3	20
	≈	22	8
TTSBEWTSI [96]	>	5	2
	<	2	22
	≈	23	6
Proposed method	>	1	0
	<	0	28
	≈	29	2

Tab. 4.6. Subjective comparison results between III and IV

Technique	III	IV
TSBWEDH [15]	>	6
	<	18
	≈	6
TSBWEPC [16]	>	5
	<	17
	≈	8
TTSBEBDH [19]	>	3
	<	18
	≈	9
TTSBEWTSI [96]	>	4
	<	20
	≈	6
Proposed method	>	0
	<	29
	≈	1

4.5.2. Objective quality assessment

The perceptual clearness was assessed with LNB-POLQA and LNB-PESQ tests. RCWB speech quality was assessed with the LSD, CWB-PESQ, and CWB- POLQA measures [199]. The robustness of hidden data against quantization and channel noises is evaluated with the help of a mean square error (MSE) measure.

4.5.2.1. Perceptual clearness (PCL)

The evaluation of perceptual transparency is done by providing LNB and CLNB signals as inputs and comparing them to rate speech quality. The LNB-POLQA value will range between 1 and 5; the higher the value, the superior the quality. The average LNB-POLQA values of conventional [15,16,19,96] and proposed methods are tabulated in table 4.7. The proposed technique gives an LNB-POLQA value of 4.12, which indicates that the proposed approach has excellent perceptual transparency over traditional techniques [15,16,19,96], which was already confirmed by subjective listening tests.

Tab. 4.7. Results of the LNB-POLQA

Technique	LNB-POLQA
TSBWEDH [15]	2.54
TSBWEPC [16]	2.99
TTSBEBDH [19]	3.21
TTSBEWTSI [96]	3.34
Proposed method	4.12

LNB-PESQ test assesses PCL by comparing the LNB signal with the CLNB signal. LNB-PESQ ranges from 0.5 to 4.5. Lower values such as 0.5 represent the worsened PCL, and higher values like 4.5 represent the best PCL. Table 4.8 lists the responses of mean scores for the traditional [15,16,19,96] and proposed techniques. A clear PCL enhancement of the proposed approach over the conventional methods is witnessed from the scores listed in Table 4.8.

Tab. 4.8. Results of the LNB-PESQ

Technique	LNB-PESQ
TSBWEDH [15]	3.02
TSBWEPC [16]	3.25
TTSBEBDH [19]	3.65
TTSBEWTSI [96]	3.69
Proposed method	4.02

4.5.2.2. RCWB Signal Quality

The quality of RCWB speech is evaluated by comparing CWB and RCWB signals in the CWB-PESQ and CWE-POLQA tests. Table 4.9 presents the mean CWB-PESQ scores of the conventional [15,16,19,96] and proposed methods. The proposed method produces a score of 4.02, which specifies that the RCWB signal quality attained is remarkable. Thus, the proposed approach improved the speech quality compared to the traditional methods.

Table 4.10 presents the mean CWB-POLQA scores of the conventional [15,16,19,96] and proposed methods. The proposed method produces a score of 4.24, which specifies that

the RCWB signal quality attained is remarkable. Thus, the proposed approach improved the speech quality compared to the traditional methods.

Tab. 4.9. Results of the CWB-PESQ

Technique	CWB-PESQ
TSBWEDH [15]	2.43
TSBWEPC [16]	2.76
TTSBEBDH [19]	3.52
TTSBEWTSI [96]	3.69
Proposed method	4.02

Tab. 4.10. Results of the CWB-POLQA

Technique	CWB-POLQA
TSBWEDH [15]	2.08
TSBWEPC [16]	2.45
TTSBEBDH [19]	3.13
TTSBEWTSI [96]	3.34
Proposed method	4.24

The quality of RCWB speech is also evaluated using the LSD measure. An RCWB signal with the least value of LSD is said to be of good quality. The resultant LSD for conventional [15,16,19,96] and proposed techniques with a μ-law coding are presented in Table 4.11. It was evident that the RCWB signal quality of the proposed method was far superior to the signal quality generated using conventional techniques [15,16,19,96]. In addition, the proposed technique offers an LSD of 2.23, indicating that the RCWB speech of the proposed technique and original CWB speech qualities are almost equal. The better RCWB signal performance of the proposed technique, which was already found in the subjective tests, is now supported by these LSD values. The proposed technique offers an LSD of 2.41 with the AWGN channel model.

Tab. 4.11. Results of the LSD

Technique	Log Spectral Distortion
TSBWEDH [15]	12.83
TSBWEPC [16]	10.69
TTSBEBDH [19]	6.07
TTSBEWTSI [96]	5.94
Proposed method	2.23

The upper plot 4.3 (a) depicts the spectrogram of the RCWB speech of the proposed method, whereas the lower plot 4.3 (b) depicts the spectrogram of the original CWB speech. It is clear from the figures that the original and RCWB signals are almost the same.

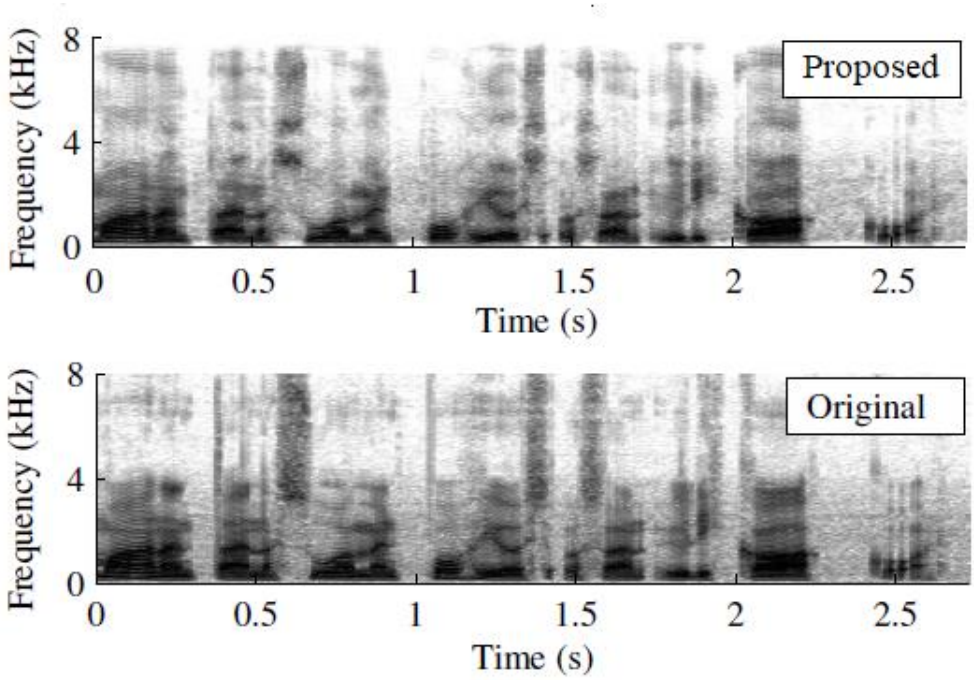


Fig. 4. 3 Spectrograms from top to bottom: (**a)** RCWB speech of the proposed method, (**b)** original CWB speech

4.5.2.3. Robustness of hidden information

AWGN with SNR ranges between 15 and 35 dB [183] is added to the CLNB signal. The evaluation of the robustness of the proposed technique is done by utilizing MSE and is calculated using the following formula

$$MSE = \frac{1}{N}\sum_{n=0}^{N-1}(S_{cwb}^{1}(n) - S_{wb}(n))^2 \qquad (4.8)$$

Where the RCWB signal is represented by $S_{cwb}^{1}(n)$ and the original CWB signal is represented by $S_{cwb}(n)$. The spreading sequence length is 16. An RCWB signal with a small value of MSE is said to be of good quality. The proposed technique gives MSE values as a

function of the SNR ranges between 15 and 35 dB, which are below 7.88×10^{-4}indicating that the RCWB signal quality obtained by the proposed technique is excellent. The proposed technique gives an MSE value after adding quantization noise (μ-law) to $S_{lnb}^{1}(n)$ is 6.07×10^{-4} which indicates RCWB signal quality that was obtained by the proposed technique is excellent.

4.6. Results and Conclusion

In this chapter, SBE utilizing the DWT-DCT-BDH technique has been proposed. The SEPs of the MHB signal is embedded within the LNB signal. The embedded information is used to reconstruct the CWB signal of good quality at the receiver end. The MSE test confirms the robustness of the proposed method. The proposed technique enhanced the RCWB signal quality over conventional techniques, evident through subjective and objective listening tests.

Chapter-5

This chapter describes the method proposed for the SBE of the LNB speech. The chapter begins with the motivation for LNB signal SBE aided by frequency domain-based data hiding using discrete Wavelet Transform-Discrete cosine transform-Based Data Hiding with Encoding (DWT-DCT-BDHWE). The performance of the proposed method under CAQNs is also analyzed. Finally, the proposed method's subjective and objective test results are discussed.

5.1. Motivation

The conventional SBE techniques using data hiding gave poor quality CLNB and RCWB signals when corrupted by CAQNs. To further improve the quality of the RCWB signal and CLNB signal over contribution 1[1*], contribution 2[2*], and the conventional SBE techniques, a novel DWT-DCT-BDHWE is proposed [3*].

5.2. Introduction

Human speech may have frequencies more than conventional telephone networks operating at 300-3400Hz. When a human speech signal is transmitted through the telephone network leads to losing information due to the LNB of the telephone network. This results in significantly low quality and lucidity of speech transmission. This problem can be solved using a CWB whose spectrum ranges from 0.5-7KHz. As a traditional telephone network installed to operate at 0.3–3.4kHz, it is not feasible to work at a wideband spectrum. Hence, the use of a wideband spectrum needs to establish a new network that is very expensive and time-consuming [1]. Therefore, other techniques are to be adopted to improve speech quality. SBE techniques [2] can be implemented to use the existing infrastructure and improve the quality.

In the ASBE techniques, a CWB signal is generated by predicting the lost portion of the signal from the LNB speech alone. Most ASBE techniques proposed in the literature are based on the SFM of speech production. The SFM divides the SBE technique into ES extension and CWB speech signal SE estimation. Many methods for excitation enhancement are found in [197]. Many methods for CWB spectral envelope approximation are illustrated in [189-193,197]. Even though ASBE has many advantages, there are a few limitations, like its performance is limited. Thus, it will not be able to reconstruct high-quality CWB signals [3].

The quality of CWB can be further improved when some supplementary information from out-of-band is communicated by hiding with the LNB signal [1]. When the concealed information is recovered at the receiver, a CWB signal with a much better speech quality can be reconstructed by combining the out-of-band signal transmitted by hiding within the LNB signal and the LNB signal. The SBE using data hiding approaches uses the real out-of-band information instead of its estimation, making the reconstruction of the CWB speech more accurate than the conventional ASBE. Several methods have been developed for this problem due to research efforts. An SBE technique has been stated in [15], accordingly that the encoded SEPs of the MSFs in the range of 4 to 8 kHz and known as MHB signal, are concealed into the LNB to generate a CLNB speech. A Technique for producing high-quality

CWB over the above method was reported in [16], in which the MHB signal was encoded with high efficiency through phonetic classification. An SBE approach was reported in [19], accordingly that SEPs of MHB signal were concealed into the least significant bits of LNB. SBE based on the quantization-based data hiding technique has been stated in [18]. In [17], the noticeable components of the MHB signal are implanted within the hidden channel. The concealed data can be consistently reproduced at the destination. The audio signal of better quality is regenerated in [4] using pitch-scaling. Enhancing the bandwidth using CCDH is introduced in [22]. A High-quality CWB signal is reproduced in [138, 139] based on CCDH Method.

The existing methodologies failed to deliver high-quality CLNB and RCWB signals along with vigor towards CAQNs. Therefore, innovative SBE algorithms with data-hiding methods development is vital for enhancing the quality of CLNB and RCWB signals and effectively managing CAQNs. The SBE using data hiding techniques, could deliver high-quality CLNB and RCWB signals and also be able to offer vigor towards CAQNs.

A DWT-FFT-DH method is reported in [198] for embedding the secrete signal in DWT coefficients of the cover signal without lowering the cover signal quality. It is observed that the DWT-FFT-DH method could produce a stego signal which is indistinguishable from the cover signal and also be able to restore the secrete signal without lowering the quality [154-156]. FFT is replaced with a discrete Cosine transform in the DWT-FFT-DH technique. A novel robust SBE algorithm using DWT-DCT-BDHWE is proposed to embed the out-of-band spectral frequencies within the LNB signal. These embedded spectral frequencies are recovered steadily at the receiver side to produce a better-quality CWB signal.

The effect of noises like CAQNs is discussed in this work. The effect of quantization noise is reported in [9] and [10]. The influence of the channel noise was not assessed in [9] and [10]. The current development uses a code division multiple access (CDMA) approach for reproducing the concealed information, which is appealed as robust towards noises like CAQNs. Especially, every information bit entrenched within the LNB signal is spread out as

the product of a definite spreading sequence. Then, the spread signals were summed to create concealed information. The concealed information could be consistently retrieved since the correlation among the SPSEs(spreading sequences) is low.

5.3. SBE aided by DWT-DCT-BDHWE

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

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

$$V(m) = \sum_{b=1}^{B-1} E_b p^b(m) \qquad (5.1)$$

The concealed data $V(m)$ is embedded into the last 8 DCT coefficients [18], resulting in a CLNB signal spectrum. The time-domain CLNB signal is obtained by applying an inverse DCT (IDCT) and then IDWT on the CLNB signal spectrum. The obtained CLNB signal $S_{lnb}^1(n)$ is transmitted through a TNC to the destination. The channel injects noises like CAQNs. Consider $\hat{S}_{lnb}^1(n)$ represent the received signal, i.e.,$\hat{S}_{lnb}^1(n) = S_{lnb}^1(n) + ë$. The mixture of CAQNsis represented by ë. The traditional phone terminal treats $\hat{S}_{lnb}^1(n)$ as a normal signal. $S_{nb}(n)$ quality is not significantly tarnished as the observed changes among $S_{lnb}(n)$ and $S_{lnb}^1(n)$ are very low.

Extraction of the concealed data $\hat{S}_{mhb}(n)$ needs a receiver that can calculate the DCT coefficients by performing DCT on $\hat{S}^{1}_{lnb}(n)$. The concealed data is then extracted from the last 8 DCT coefficients [198]. The data bits are decoded by employing a multiuser detector [198]. i.e,

$$\check{D}_b = sign(\sum_{m=1}^{M-1} \check{V}_m {p_m}^b) \tag{5.2}$$

In a noise-free environment, $\check{V}_m = V_m$ Substitute it into (5.2), we have

$$\check{D}_b = sign\left(\sum_{m=1}^{M-1} V_m {p_m}^b\right)$$

$$= sign\left(\sum_{m=1}^{M-1}\left(D_m {p_m}^b {p_m}^b + \sum_{g=0,g\neq b}^{B-1} D_g {p_m}^g {p_m}^b\right)\right)$$

$$= sign\left(MD_m + \sum_{g=0,g\neq b}^{B-1} D_g \sum_{m=0}^{M-1} {p_m}^g {p_m}^b\right) \tag{5.3}$$

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

$$\sum_{g=0,g\neq b}^{B-1} D_g \sum_{m=0}^{M-1} {p_m}^g {p_m}^b = 0 \tag{5.4}$$

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

5.4. DWT-DCT-BDHWE for speech Bandwidth Extension

5.4.1. Transmitter

The transmitter is depicted in Fig.5.1. Primarily, CWB speech $S_{cwb}(n)$ is sampled at 16kHz and is passed through an LPF and HPF to generate an LNB signal and an MHB signal. LNB signals have speech information ranging from 0 to 4k Hz, and MHB has speech information in the range of 4 to 8kHz. Then LNB signal $S_{lnb}(n)$ is generated by decimating the output of LPF. The HPF output is then decimated to generate an MHB signal $S_{mhb}(n)$.

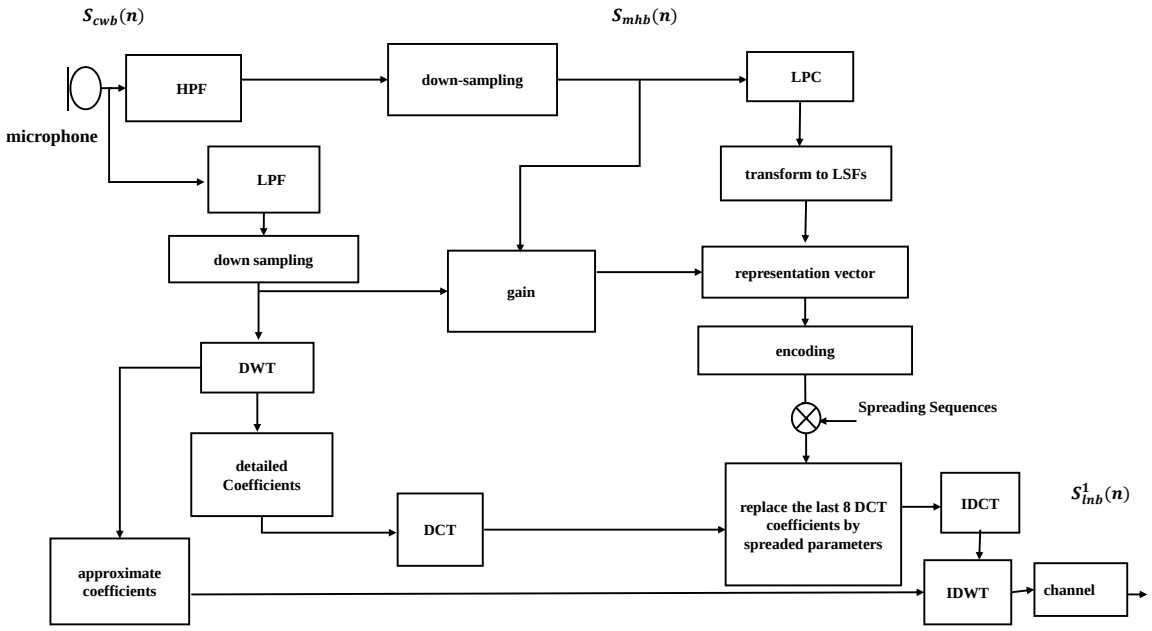


Fig.5.1. Proposed DWT-DCT-BDHWE Transmitter.

Reduce the parameters which characterize $S_{mhb}(n)$ to insert MHB signal imperceptibly in LNB signal, and LP analysis [170] is used here to fulfill. The LPCs are evaluated by applying the Levinson-Durbin method [170] on $S_{mhb}(n)$ and later, these are transformed into LSFs as there is a slight change in coefficients leading to distortions while reproducing. Furthermore, the gain of $S_{mhb}(n)$ needs to be hidden in order to evade over-approximation [171]. Hence, the gain is assessed as $g_r = \frac{g_{mb}}{g_{lb}}$ and pooled with LSFs to generate a representation vector, that is, $D = [lsf_1 lsf_2 \ldots . lsf_r, g_r]$. Quantize D to the nearby entry of a VQ codebook that is produced by the fuzzy c-means (FCM) algorithm [200]. The binary equivalent of entry index, i.e.,$D_0, D_{1,\ldots} D_{B-1}$ is concealed into the LNB signal based on

DWT-DCT-BDHWEtechnique, which results in a CLNB signal and is communicated through TNC to the destination.

The ES has many parameters which are not implanted to lessen parameters that could be implanted since exceeding 3.4 kHz the human ear remains insensitive to alterations of ES [9]. Thus, the prediction of the MHB excitation signal from $S_{lnb}(n)$ at the destination assurances the reproduction performance.

5.4.2. Receiver

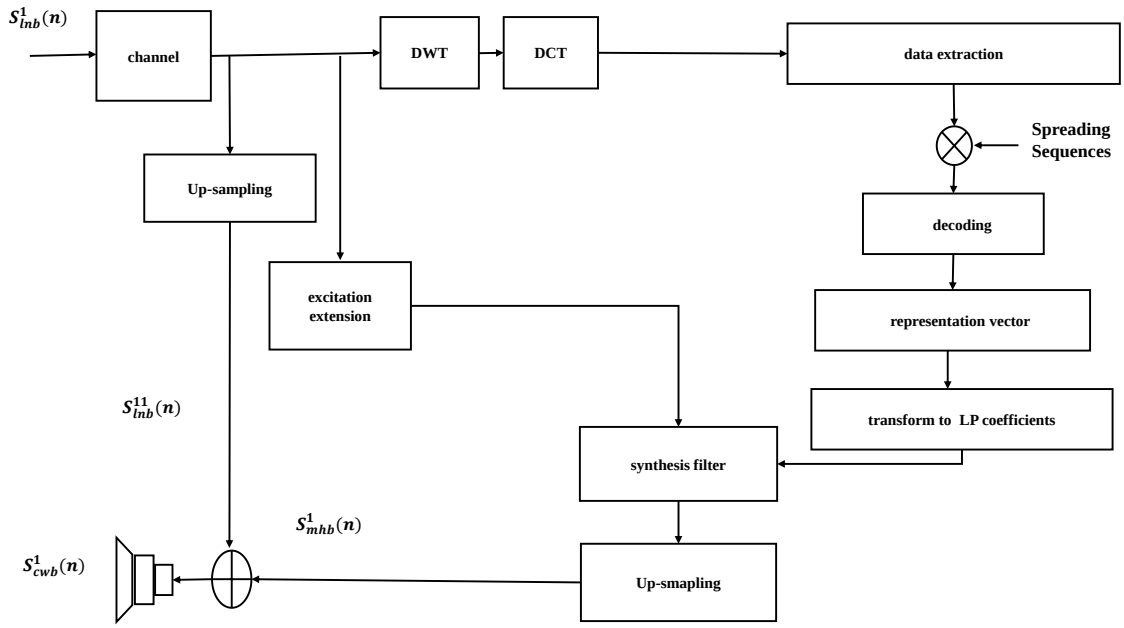


Fig. 5.2. Proposed DWT-DCT-BDHWE Receiver.

The receiver is shown in Fig.5.2. Recuperate the entry index appropriately using the DWT-DCT-BDHWE method, and then the VQ codebook is used to regain the corresponding quantized LSFs accurately. The recovered LSFs are then used for generating LPCs. Meanwhile, the inverse filtering is performed on $S^1_{lnb}(n)$ which will produces LNB residual signal. Then the residual signal is extended, which results in MHB excitation. The synthesis filter defined by the recovered LPCs is excited through UB excitation in order to reconstruct $S^1_{mhb}(n)$. At this instant, the sampling rate of $S^1_{lnb}(n)$ and $S^1_{mhb}(n)$ is 8kHz Hz and then these signals are interpolated. $S^{11}_{mhb}(n)$ represents the interpolated $S^1_{mhb}(n)$. The interpolated

CLNB $S_{lnb}^{11}(n)$ and recovered $S_{mhb}^{11}(n)$ signals are added in order to regenerate the high-quality CWB signal $S_{cwb}^{1}(n)$.

5.5. Experimental Results

To assess the performance of the proposed method, fifty sentences spoken by 40 talkers were collected from the TIMIT database [166]. The LNB signal is decomposed into frames of 20ms duration with an overlap of 10ms among frames. The frames are then processed one by one. Subjective and objective tests are used to assess performance. The proposed methodology competency is explored by comparing it with traditional techniques, such as TTSBEDH [15], TTSBEPC [16], TTSBEBDH [19], and TTSBEWTSI [96]. The channel models considered here are μ-law and AWGN.

5.5.1. Subjective assessments

The perceptual clearness is assessed based on MOS test [15-16]. The listening test compares various speech signals like CWB, LNB, CLNB, and RCWB. These tests were performed in a silent room using headsets. During each test, thirty participants are considered.

5.5.1.1. Perceptual Clearness (PCL)

In the proposed technique, the information must be transparently concealed, i.e., LNB and CLNB are subjectively indistinguishable. High PCL means low perceptible degradation in the CLNB signal. PCL is assessed based on the MOS test. Listeners participating in the test compare LNB and CLNB to provide a decision in terms of MOS, tabulated in Table 5.1. Table 5.2. showcases the results of the averaged MOS for conventional [15, 16, 19, 96] and proposed approaches. MOS values show the remarkable perceptual clearness of the proposed approach over the traditional approaches in Table .5.2.

Table 5.1. MOS

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

Table 5.2. MOS assessment outcomes

Technique	Mean opinion score
TTSBEDH [15]	2.81
TTSBEPC [16]	3.01
TTSBEBDH [19]	3.12
TTSBEWTSI [96]	3.46
Proposed method	3.98

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

I, II, III, and IV in Table 5.3 represent the CWB signal, LNB signal, CLNB signal, and RCWB signal, respectively. The subjects are asked to do a pairwise analysis of signals among I to IV and must tell whether the first signal is paramount (>), deprived (<), or alike (≈) to the second signal. Table 5.3 provides the responses of pairwise comparison of I, II, and III to the other signal, Table 5.4 provides the responses of pairwise comparison of II and III to the other signal, and Table 5.5 provides the responses of comparison among III and IV. The number of subjects with an exact preference (> or < or ≈) is mentioned with Arabic digits in the table. The CWB signal outperforms the CLNB signal for conventional [15, 16, 19, 96] and proposed methods which are endorsed by Table 5.3. Table 5.3 also endorsed a clearly enhanced RCWB signal quality of the proposed technique over the conventional methods. The remarkable perceptual clearness of the proposed approach over the traditional approaches is endorsed by Table 5.4. Compared to conventional methods, the RCWB signal is better than the LNB signal for the proposed approach endorsed in Table 5.4. Compared to conventional methods, the RCWB signal is better than the CLNB signal for the proposed approach which is endorsed in Table 5.5.

Table 5.3. Subjective contrast outcomes among I, II, III, and IV

Technique	I	II	III	IV
TSBEDH [15]	>	30	30	15
	<	0	0	0
	≈	0	0	15
TSBEPC [16]	>	30	30	14
	<	0	0	0
	≈	0	0	16
TSBEBDH [19]	>	30	30	13
	<	0	0	10
	≈	0	0	0
TSBEWTSI [96]	>	30	30	0
	<	0	0	20
	≈	0	0	21
Proposed method	>	30	30	2
	<	0	0	0
	≈	0	0	28

Table 5.4. Subjective contrast outcomes among II, III, and IV

Technique	II	III	IV
TSBEDH [15]	>	8	3
	<	4	18
	≈	18	9
TSBEPC [16]	>	8	1
	<	2	19
	≈	20	10
TSBEBDH [19]	>	5	2
	<	3	20
	≈	22	8
TSBEWTSI [96]	>	5	2
	<	2	22
	≈	23	6
Proposed method	>	1	0
	<	0	28
	≈	29	2

Table 5.5. Subjective contrast outcomes among III and IV

Technique	III	IV
TSBEDH [15]	>	6
	<	15
	≈	9
TSBEPC [16]	>	5
	<	16
	≈	9
TSBEBDH [19]	>	3
	<	17
	≈	10
TSBEWTSI [96]	>	4
	<	19
	≈	7
Proposed method	>	0
	<	28
	≈	2

5.5.2. Objective Quality Evaluation

The RCWB signal quality is assessed with LSD [15-16], CWB-PESQ and CWB-POLQAtests. The perceptual clearness is assessed with LNB-POLQA and LNB-PESQ tests [152]. The robustness of concealed data to CAQNs is assessed with the bit error rate (BER) measure.

5.5.2.1 Perceptual Clearness (PCL)

The LNB-PESQ test assesses PCL by comparing the LNB signal with the CLNB signal. LNB-PESQ ranges from 0.5 to 4.5. Lower values, such as 0.5, represent the worsened PCL, and higher values like 4.5, represent the best PCL. Table 5.6 lists the responses of mean scores for the traditional [15, 16, 19, 96] and proposed techniques. An apparent PCL enhancement of the proposed approach over the traditional techniques is witnessed from the scores as listed in Table 5.6.

Tab. 5.6. LNB-PESQ test Outcomes

Technique	LNB-PESQ
TSBEDH [15]	2.81
TSBEPC [16]	3.02
TSBEBDH [19]	3.33
TSBEWTSI [96]	3.40
Proposed method	4.43

The LNB-POLQA value will range between 1 and 5; the higher the value, the superior the quality. The average LNB- POLQA values of conventional [15,16,19,96] and proposed methods are tabulated in table 5.7. The proposed technique gives an LNB-POLQA value of 4.31, which indicates that the proposed technique has excellent perceptual transparency over traditional techniques [15,16,19,96], which was already confirmed by subjective listening tests.

Tab. 5.7. Results of LNB-POLQA

Technique	LNB-POLQA
TSBWEDH [15]	2.54
TSBWEPC [16]	2.99
TSBWEBDH [19]	3.21
TSBWEWTSI [96]	3.35
Proposed method	4.31

5.5.2.2 RCWB Signal Quality

The quality of RCWB speech is evaluated by comparing CWB and RCWB signals in the CWB-PESQ test. Table 5.8 presents the mean CWB-PESQ scores of the conventional [15, 16, 19, 96] and proposed methods. The proposed technique produces a score of 4.38, which specifies that the RCWB signal quality attained is remarkable. Thus, the proposed technique improved the speech quality when compared to the traditional methods.

Tab. 5.8 CWB-PESQ test Outcomes

Technique	CWB-PESQ
TSBEDH [15]	2.31
TSBEPC [16]	2.63
TSBEBDH [19]	3.54
TSBEWTSI [96]	3.62
Proposed method	4.38

The evaluation of the quality of RCWB speech is completed by giving CWB and RCWB signals as inputs and comparing them in order to rate speech quality. The average CWB-POLQA values of the conventional [15, 16, 19, 96] and proposed methods are shown in Table 5.9. A CWB- POLQA value of 4.57 confirms that the RCWB signal quality retrieved by the proposed method is excellent compared to traditional techniques [15,16,19, 96] which is already confirmed by subjective listening tests on a set of participants. Thus, the speech quality was improved by using the proposed technique.

Tab. 5.9. Results of CWB-POLQA

Technique	CWB-POLQA
TSBWEDH [15]	2.08
TSBWEPC [16]	2.45
TSBWEBDH [19]	3.13
TSBWEWTSI [96]	3.34
Proposed method	4.57

5.5.2.3 Comparison of original and reconstructed MHB speech

LSD is a reliable measure for assessing the resemblance between true and restored MHB signals. In general, the best-quality of reproduced MHB signal has a low value of LSD. Table 5.10 lists the mean LSD scores for the existing [15, 16, 19, 96] and proposed schemes under μ-law coding. There is a clear enhancement in the quality of the proposed scheme over the traditional schemes [15, 16, 19, 96] is witnessed from the values as listed in Table 5.10.

Tab. 5.10 LSD test Outcomes

Technique	Log Spectral Distortion
TSBEDH [15]	13.56
TSBEPC [16]	11.56
TSBEBDH [19]	7.12
TSBEWTSI [96]	6.67
Proposed method	2.31

5.5.3 Vigor of concealed data

AWGN with SNR ranges from 15 to 35 dB is summed up with CLNB signal [183]. The proposed method's robustness is assessed based on BER. The PN code size is 8. The lower value of BER designates the RCWB signal of high quality. The BER values which were attained with SNR in the range of 15 to 35 dB are beneath $7.7036 * 10^{-4}$which endorses the RCWB signal of high quality. The BER value, which is attained with μ-law coding, is $4.61 * 10^{-4}$which endorses the RCWB signal of high quality.

5.6 Results and Conclusion

A novel SBE based on the DWT-DCT-BDHWE technique is presented in this chapter for embedding spreaded SEPs of MHB signal within LNB signal DCT coefficients. The concealed information is extracted to generate an extraordinary-quality CWB signal at the receiver. The concealed information is vulnerable to CAQNs. Thus, the CDMA approach is employed for reproducing the concealed information that is appealed as robust towards CAQNs. Especially, every information bit entrenched within the LNB signal is spread out as the product by definite SPSEs. Then, the spread signals were summed to create concealed information. The concealed information could be consistently retrieved since the correlation among the SPSEs is low. The proposed approach was a robust solution for SBE. Subjective, CWB-PESQ, CWB-POLQA, and LSD test results confirmed excellent and improved RCWB signal performance using the proposed method over the conventional techniques. An apparent PCL enhancement of the proposed approach over the traditional methods is witnessed from MOS, LNB-POLQA, and LNB-PESQ tests.

Chapter-6

This chapter describes the method proposed for the SBE of the LNB speech. The chapter begins with the motivation for the SBE technique aided by Discrete cosine Transform domain-based data hiding(DCTBDH), described in detail. The performance of the proposed method under CAQNs is also analyzed. Finally, the proposed method's subjective and objective test results are discussed.

6.1. Motivation

The conventional SBE techniques using data hiding gave poor quality CLNB signal and RCWB signal and limited SBE performance when corrupted by CAQNs. To increase the quality of RCWB signal and CLNB signal over the conventional SBE techniques, a novel Discrete cosine transform-Based Data Hiding (DCT-BDH) is proposed [4*].

6.2. Introduction

Most traditional TNC allows only an LNB signal which is band-limited to 0.3–3.4kHz. Usually, Human speech has frequencies exceeding the bandwidth of the present TNC. At this instant, the transmission of voice-over TNC results in a loss of sections of the speech spectrum, producing a considerable drop in speech intelligibility and quality. The

transmission of CWB speech which lies in the range of 0.5–7kHz across TNC will boost the quality, intelligibility, and perceived naturalness of the speech signal compared to the transmission of LNB speech. It will be costly and take time to set up a new wide-spectrum TNC that can accommodate higher bandwidths [1]. As a result, it is desirable to increase the receiving end's bandwidth utilizing SBE techniques [2] without changing the existing TNC infrastructure.

The existing TNC can benefit significantly from improved speech quality due to SBE technology. Many SBE approaches have been proposed over the years. The ASBE is among various methods of SBE which can improve the intelligibility and quality of telephony speech. In the ABWE techniques, a CWB signal is generated by predicting the lost portion of the signal from the LNB speech alone. Most of the ASBE proposed in the literature is based on the SFM of speech production. The SFM system divides the SBE technique into ES extension and CWB speech signal SE estimation [2]. Different methods for estimating CWB SPEs are presented in[190-192]. A time-frequency network with channel attention and non-local modules is used for SBE. Latent representation learning for ASBE using a conditional variational auto-encoder is presented to enhance speech quality [201]. The time-domain multi-scale fusion neural network approach for improving the performance of SBE is presented in [202]. SBE using a conditional generative adversarial network with discriminative training is introduced in[203]. The audio signal of better quality is regenerated using audio bandwidth extension aided by the dilated convolutional neural network approach [204]. In[205],a deep neural network ensemble approach for reducing artificial noise in SBE is introduced. A waveform-based method for SBE that uses a deep three-way split summation FFTNet architecture is proposed in[206]. In[207], a time-domain ASBE towards a low-frequency band by a sinusoidal synthesis of missing harmonics is presented to enhance the quality of the reconstructed CWB signal. A Wave Net-based model conditioned on a log-mel spectrogram representation of LNB speech to reconstruct the better quality speech signal is proposed in [208]. However, traditional ASBE are suffering from rebuilding CWB speech with high quality under all conditions[2].

Compared to ASBE, a CWB speech signal quality is further improved when supplementary information from out-of-band is communicated by hiding with the LNB

signal[1]. Several techniques for SBE using data hiding are proposed in the state-of-the-art literature. An SBE technique is proposed in [86] to embed the encoded SPEs parameters of the lost speech frequency components within the LNB speech signal. The embedded information retrieves a better-quality CWB signal at the receiver end. A much better-quality CWB signal over [15] has been reconstructed in [16], where the SPEs are efficiently encoded using phonetic classification. The pitch-scaled frequencies of the OOB signal are hidden in the unused frequencies of traditional telephony speech to enhance the quality of RCWB speech [13]. The CWB signal of better quality is regenerated in [138-139] using the CCDH technique. High-quality CWB signal is reconstructed using various frequency-domain data hiding techniques [209-210].

SBE techniques with data hiding are expected to deliver high-quality CLNB alongside RCWB signals. Also, these methods must be able to handle issues pertaining to CAQNs. Nevertheless, most traditional approaches fail to provide high-quality CLNB and RCWB signals[16-19,20-22,138]. Also, they are less robust to CAQNs. Thus, developing a novel SBE technique using data hiding is essential to improve the quality of CLNB and RCWB signals and make them more robust to CAQNs.

An audio steganography technique is presented in [211], using the DCT-BDH method to insert the secret message signal in the DCT coefficients of a host speech signal without degrading the perceptual quality of the host signal. It was shown that this approach produces a stego signal that is indistinguishable from the host signal while reliably recovering the secret message signal at the receiver end without any degradation in quality.

A new SBE algorithm using the DCT-BDH technique is proposed to embed the parameters of the lost speech frequency components within the DCT coefficients of the LNB signal. These hidden parameters are retrieved at the receiver side to produce a better-quality CWB signal by combining the missing speech signal that was transmitted through the DCT coefficients and the LNB signal. The proposed scheme uses the real missing speech

information instead of its estimation, making the reconstruction of the CWB speech more accurate than the conventional ASBE.

The Techniques proposed in [1, 16, 22] for SBE are only quantization noise, ignoring the channel noise. The CAQNs effects are considered in this work. The spread spectrum technique is used in this work for retrieving the embedded information as it is claimed to be more robust against CAQNs [148]. In particular, each parameter to be inserted is spread by multiplying with a particular spreading sequence. The embedded information is then formed by adding the spread signals. Due to orthogonality among spreading sequences, the concealed information is retrieved reliably at the user end using a correlator.

Spreading sequences with low cross-correlations is preferred to minimize the interference caused by the other embedded components. Hadamard codes have an optimal cross-correlation performance, i.e., orthogonal to each other, whereas the m-sequences, Gold-codes, and Kasami-codes are with varying cross-correlation properties [157-159]. Because its optimal cross-correlation performance well recognizes the Hadamard codes, it is employed in this work to minimize the interference caused by the other embedded components [160-164].

6.3. Speech bandwidth extension using DCT Based data hiding

6.3.1. Transmitter

The proposed DCTBDH transmitter is depicted in Fig.6.1. The CWB signal $S_{cwb}(n)$ is split into LB (0-4kHZ) signal using an LPF and an MHB (4Khz-8kHz) signal using an HPF. The LPF output is down-sampled by a factor of two to create an LNB signal $S_{lnb}(n)$. The signal is also down-sampled to produce an MHB signal $S_{mhb}(n)$.

To imperceptibly embed $S_{mhb}(n)$ into $S_{lnb}(n)$, the number of parameters that represents $S_{mhb}(n)$is minimized. The LP analysis is used to complete this target [170]. LP analysis is based on the SFM of speech generation. The LPCs are the reciprocal of the AR filter coefficients. The LPCs represent the SE of $S_{mhb}(n)$ are denoted as $b_i(i = 1, \ldots\ldots, 10)$, where i is the order of the filter. The small variation in LPCs results in substantial distortions when reconstructing $S_{mhb}(n)$; hence LPCs are modified into LSFs. Also, the gain of $s_{mhb}(n)$, denoted with G_r, has to be embedded since synthesized MHB speech have to be scaled to appropriate energy to evade over-estimation. Thus, the representation vector which represents$S_{mhb}(n)$is formed by combining LSFs and gain, i.e.,$R = [LSF_1, LSF_{2,\ldots\ldots\ldots}, LSF_{10}, G_r]$.

The excitation parameters of $S_{mhb}(n)$ are not embedded to lessen the hidden parameters because the ear is not very sensitive to distortions of the ES above LNB [1]. Thus, estimating the excitation of $S_{mhb}(n)$ at the receiver from $S_{lnb}(n)$ is more compatible with the reconstruction performance.

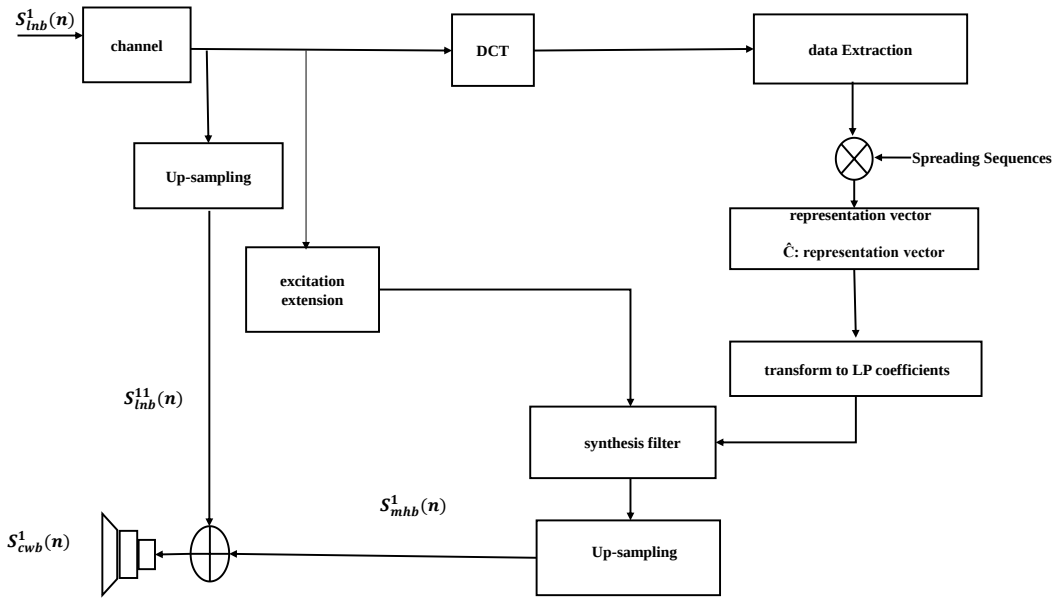


Fig. 6.1 Proposed DCTBDH Transmitter

All the parameters of $R$are denoted with D_i. One among all the parameters of R is then denoted with D_{i0}. Every parameter of R to be embedded is spread by multiplying it with a particular spreading sequence, i.e.,$D_i \bullet p^{\rightarrow i}, 1 \leq i \leq Q$. The hidden data is then produced by adding all of these spreading vectors and is given by

$$V(g) = \sum_{i=1}^{Q} D_i p^i(g) \tag{6.1}$$

where g^{th} element of $p^{\rightarrow i}$ represented by $p^i(g)$. DCT is then applied to the LNB signal$S_{nb}(n)$ and can be expressed as

$$S_{lnb}(k) = w(k) \sum_{n=0}^{N-1} S_{lnb}(n) \cos \frac{(2n+1)k\pi}{2N}, k = 0\ toN-1 \tag{6.2}$$

where

$$w(k) = \sqrt{\frac{1}{N}} if k = 0, w(k) = \sqrt{\frac{2}{N}} \qquad otherwise$$

The last 16 coefficients of the DCT coefficients are replaced by $V(g)$ resulting in a CNB signal spectrum. To transform back the CLNB signal spectrum to time-domain representation, IDCT is applied on the CLNB signal spectrum and can be expressed as

$$S_{lnb}(n) = \sum_{k=0}^{N-1} w(k) S_{lnb}(k) \cos \frac{(2n+1)k\pi}{2N}, n = 0\ toN-1 \tag{6.3}$$

Thus, a CLNB signal $S_{lnb}^{1}(n)$ is produced so that it can be communicated to the receiver on a TNC.

6.3.2. Receiver

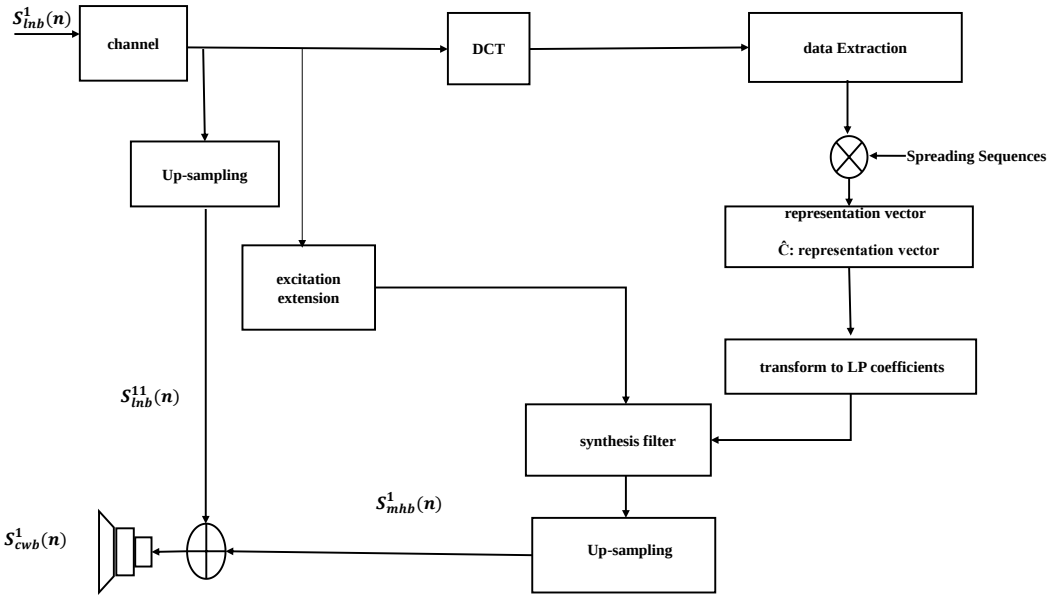


Fig.6. 2 Proposed DCTBDH receiver

The DCTBDH receiver is depicted in Fig. 6.2. The CLNB signal received through a TNC is noisy. Assume that the received signal is represented by $\hat{S}_{lnb}^{1}(n)$ i.e., $\hat{S}_{nb}^{1}(n) = S_{nb}^{1}(n) + e$.Where e represents the combination of CAQNs. The conventional phone terminal treats $\hat{S}_{lnb}^{1}(n)$as an ordinary signal. The LNB signal quality is not noticeably degraded since there is a very small perceived difference between$S_{lnb}^{1}(n)$ and $\hat{S}_{lnb}^{1}(n)$. Retrieval of the embedded data requires applying DCT on the CNB signal to obtain the DCT coefficients.

The spread parameters are then obtained from the last 16 DCT coefficients, and a correlator is used to de-spread these parameters. Assuming a particular $\underset{\hat{}}{D}_i$ is represented as $\underset{\hat{}}{D}_{io}$ to be retrieved, the correlation can be expressed as

$$\underset{\wedge}{\mathrm{D}}_{io} = \frac{1}{Q}\sum_{g=1}^{Q} \underset{\cdot}{\mathrm{V}}(g)\, p^{io}(g) \tag{6.4}$$

where $\underset{\cdot}{\mathrm{V}}(g)$ represents noisy $V(g)$and is given by

$$\underset{\cdot}{\mathrm{V}}(g) = V(g) + \bar{\mathrm{e}}(g) \tag{6.5}$$

Equation (6.5) is substituted into equation (6.4), so that we have

$$\underset{\wedge}{\mathrm{D}}_{io} = \frac{1}{Q}\sum_{g=1}^{Q} \underset{\cdot}{\mathrm{V}}(g)\, p^{io}(g)$$

$$= \frac{1}{Q}\sum_{g=1}^{Q} p^{io}(g)\left(\sum_{i=1}^{Q} \check{\mathrm{D}}_i p^{i}(g) + \bar{\mathrm{e}}(g)\right)$$

$$= \frac{1}{Q}\sum_{g=1}^{Q} p^{io}(g) \times \left(\check{\mathrm{D}}_{io} p^{io}(g) + \sum_{i\neq io} \check{\mathrm{D}}_i p^{i}(g) + \bar{\mathrm{e}}(g)\right)$$

$$= \check{\mathrm{D}}_{io} + \frac{1}{Q}\sum_{g=1}^{Q}\sum_{i\neq io} \check{\mathrm{D}}_i p^{i}(g) p^{io}(g) \quad + \frac{1}{Q}\sum_{g=1}^{Q} p^{io}(g)\bar{\mathrm{e}}(g) \tag{6.6}$$

The PN sequences are orthogonal. i.e.

$$\sum_{g=1}^{Q} p^{i}(g) p^{io}(g) = 0$$

where *i≠ io.* Therefore

$$\sum_{g=1}^{Q}\sum_{i\neq io}\check{\mathrm{D}}_{io}p^{i}(g)p^{io} = \sum_{i\neq io}\check{\mathrm{D}}_{io}\sum_{g=1}^{Q}p^{i}(g)p^{io}(g) = 0 \tag{6.7}$$

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

$$\frac{1}{Q}\sum_{g=1}^{Q}p^{io}(g)\,\bar{\mathrm{e}}(g) = 0 \tag{6.8}$$

when $Q \to \infty$. Equations (6.7) and (6.8) are substituted into equations (6.6), thus we have

$$\hat{\mathrm{D}}_{io} = \check{\mathrm{D}}_{io} \tag{6.9}$$

This reveals that the parameters which represent $\hat{\mathrm{S}}_{mhb}(n)$ can be effectively recovered by using the SS technique [169], and then the LPCs are obtained from LSFs. Meanwhile, LNB residual signal is received by inverse filtering $\hat{\mathrm{S}}^{1}_{lnb}(n)$ using LPCs of $\hat{\mathrm{S}}^{1}_{lnb}(n)$ and then obtain the MHB-ES signal by extending the LNB residual signal. The MHB signal $\hat{\mathrm{S}}_{mhb}(n)$ that was embedded is synthesized by exciting the synthesis filter described by the recovered LPCs by an MHB excitation signal. The received CLNB and reconstructed MHB signals are sampled at an 8 kHz sampling rate. These signals are then interpolated by a factor of two. $S^{1}_{mhb}(n)$, represents interpolated $\hat{\mathrm{S}}_{mhb}(n)$ signal. The interpolated CLNB ($S^{11}_{lnb}(n)$) and MHB $S^{1}_{mhb}(n)$ signals are added up to reproduce a CWB signal ($S^{1}_{cwb}(n)$) of good quality.

6.4. Experimental Results

The speech utterances used for the performance evaluations of traditional and proposed SBE techniques were obtained from the TIMIT database[166]. The evaluations were done by

taking thirty different speech utterances, of which thirty female and male speakers spoke. Each speech signal was split to form frames 20ms long, and an overlap of 10ms was maintained between frames. Each frame was processed individually. The performance assessment of the methods was done by considering the subjective and objective measures. The proposed methodology competency is explored by comparing it with traditional techniques, such as TTSBEDH [15], TTSBEPC [16], TTSBEBDH [19], and TTSBEWTSI [96].AWGN and μ-law channel models were used for analysis.

6.4.1. Subjective quality assessment

The perceptual clearness was assessed with the MOS test[150]. The subjective comparison between CWB, CLNB, LNB, and RCWB signals was also employed. An evaluation was done using a predefined scale by examining participant's views on speech sounds. Each person is made to hear the speech utterances through headphones in a silent chamber. Thirty persons participated in the tests.

6.4.1.1 Perceptual Transparency

The perceptual transparency was assessed with the MOS test. The CLNB and LNB signals have to be similar sounds. Compared to CLNB and LNB signals, the listener decides in terms of MOS, as shown in Table 6.1. The average MOS values of traditional [15, 16, 19, 96] and proposed techniques are given in Table 6.2. The proposed technique gave a MOS value of 3.99, which indicates that the proposed technique has excellent perceptual transparency over the traditional techniques [15, 16, 19, 96]. The proposed technique gave a MOS value of 3.99, almost near the standard MOS value of 4, indicating that CLNB and LNB signals were more or less identical.

Table 6.1 MOS

score	Instruction
1	LNB and CLNBsignals sound different
2	Observable difference between LNB and CLNBsignals
3	Minute differencebetweenLNB and CLNBsignals
4	LNB and CLNBsignals sound alike

Table 6.2 Result of MOS

Technique	Mean opinion score
TTSBEDH [15]	2.89
TTSBEPC [16]	3.07
TTSBEBDH [19]	3.18
TTSBEWTSI [96]	3.54
Proposed method	3.99

6.4.1.2 Subjective Comparisons between CWB, LNB, CLNB, and RCWB Speech samples

A listening test was done to compare performances between the proposed and conventional methods. Here, the CWB signal, LNB signal, CLNB signal, and RCWB signal were labeled I, II, III, and IV, respectively. Participants are asked to compare the samples pairwise to tell whether the first sample was superior to, inferior to, or equal to the second. The responses after comparing I, II, and III with the other signals are tabulated in Tables 6.3, 6.4, and 6.5. Arabic numerals indicate the number of participants with a specific preference in the table. It is observed that the CWB signal is superior to the LNB and CLNB signals of traditional [15, 16, 19, 96] and the proposed methods from Table 6.3. Also, we observe that RCWB signal quality is far superior using the proposed method over traditional methods [15, 16, 19, 96] from Table 6.3. Thus, the speech quality was enhanced by the proposed technique. Compared to traditional methods, it is observed that the RCWB signal of the proposed method is superior to that of the LNB signal, as may be seen in Table 6.4. Also, a clear perceptual transparency improvement of the proposed method over the conventional methods was observed in Table 6.4, which shows that the quality of the CLNB signal is almost identical to that of the LNB signal. The data embedding performed in the proposed method has very little impact on perception. Compared to conventional methods[15, 16, 19, 96], it is observed that the RCWB speech of the proposed technique is better than the CLNB speech

from Table 6.5. Thus, the proposed method produces a much better quality speech signal than the conventional methods[15, 16, 19, 96].

Table 6.3 Subjective comparison test results between I, II, III, and IV

Technique	**I**	**II**	**III**	**IV**
TTSBEDH [15]	>	30	30	14
	<	0	0	0
	≈	0	0	16
TTSBEPC [16]	>	30	30	12
	<	0	0	0
	≈	0	0	18
TTSBEBDH [19]	>	30	30	11
	<	0	0	0
	≈	0	0	19
TTSBEWTSI [96]	>	30	30	7
	<	0	0	0
	≈	0	0	23
Proposed method	>	30	30	2
	<	0	0	0
	≈	0	0	28

Table 6.4 Subjective comparison test results between II, III, and IV

Technique	**II**	**III**	**IV**
TTSBEDH [15]	>	8	3
	<	4	18
	≈	18	9
TTSBEPC [16]	>	8	1
	<	2	19
	≈	20	10
TTSBEBDH [19]	>	5	2
	<	3	20
	≈	22	8
TTSBEWTSI [96]	>	5	2
	<	2	22
	≈	23	6
Proposed method	>	2	0
	<	0	27
	≈	28	3

Table 6.5 Subjective comparison test results between III and IV

Technique	III	IV
TTSBEDH [15]	>	6
	<	18
	≈	6
TTSBEPC [16]	>	5
	<	17
	≈	8
TTSBEBDH [19]	>	3
	<	18
	≈	9
TTSBEWTSI [96]	>	4
	<	20
	≈	6
Proposed method	<	0
	≈	29
	>	1

6.4.2. Objective Quality Assessment

6.4.2.1. RCWB speech quality

The quality of RCWB speech is evaluated using the LSD measure. An RCWB signal with the least value of LSD is said to be of good quality. The resultant LSD for conventional [15, 16, 19, 96] and proposed techniques with a μ-law channel model are presented in Table 6.6. It was evident that the proposed technique's RWB signal quality was far superior to the signal quality generated using conventional [15, 16, 19, 96], and proposed techniques with a μ-law channel model are presented in Table 6.6. It was evident that the proposed technique's RCWB signal quality was far superior to the signal quality generated using conventional techniques [15, 16, 19, 96]. In addition, the proposed technique offers an LSD of 2.2248, indicating that the RCWB speech of the proposed technique and original CWB speech qualities are almost equal. The good RCWB signal performance of the proposed technique, which was already found in the subjective tests, is now supported by these LSD values. The proposed technique offers an LSD of 2.35 with the AWGN channel model.

Table 6.6LSD test results.

Technique	Log Spectral Distortion
TTSBEDH [15]	12.83
TTSBEPC [16]	10.69
TTSBEBDH [19]	6.07
TTSBEWTSI [96]	5.94
Proposed method	2.248

6.4.2.2 Perceptual transparency

The evaluation of perceptual transparency is done by providing LNB and CLNB signals as inputs and comparing them to rate speech quality. The LNB-PESQ value will range between 0.5 and 4.5; the higher the value, the superior the quality. The average LNB-PESQ values of conventional [15, 16, 19, 96] and proposed methods are tabulated in table 6.7.The proposed technique gives the LNB-PESQ value of 4.47, indicating that the proposed technique has excellent perceptual transparency over traditional techniques [15, 16, 19, 96]which was already confirmed by subjective listening tests.

Table 6.7 LNB-PESQ test results.

Technique	LNB-PESQ
TTSBEDH [15]	2.87
TTSBEPC [16]	3.07
TTSBEBDH [19]	3.42
TTSBEWTSI [96]	3.45
Proposed method	4.47

The LNB-POLQA value will range between 1 and 5; the higher the value, the superior the quality. The average LNB- POLQA values of conventional [15,16,19,96] and proposed methods are tabulated in table 6.8. The proposed technique gives an LNB-POLQA value of 4.03, which indicates that the proposed technique has excellent perceptual transparency over traditional techniques [15,16,19,96], which was already confirmed by subjective listening tests.

Table 6.8 LNB-POLQA Test Results.

Technique	LNB-POLQA
TTSBEDH [15]	2.54
TTSBEPC [16]	2.99
TTSBEBDH [19]	3.21
TTSBEWTSI [96]	3.35
Proposed method	4.03

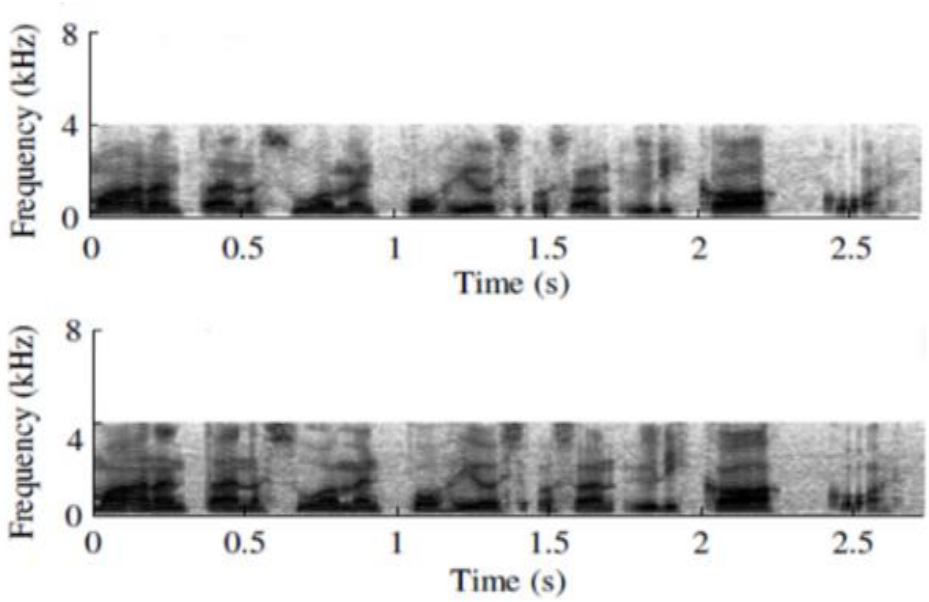


Fig. 6.3 Spectrograms from top to bottom: (a) CLNB speech, (b) LNB speech.

In Fig. 6.3, the upper plot depicts the spectrogram of LNB speech $S_{lnb}(n)$, whereas the lower plot b depicts the spectrogram of the CLNB speech $S^{1}_{lnb}(n)$. It is clear from the figures that $S_{lnb}(n)$ and $S^{1}_{lnb}(n)$ are almost indistinguishable [1,9,140,167,168].

6.4.2.3 Robustness of embedded information

AWGN with SNR ranges between 15 and 35 dB is added to the CLNB signal [17,18,151,159,183]. The evaluation of the vigor of the proposed method is done by utilizing MSE and is calculated using the formula.

$$MSE = \frac{1}{N}\sum_{n=0}^{N-1}(S_{cwb}^{1}(n) - S_{cwb}(n))^2 \quad (6.10)$$

Where the RCWB signal is represented by $S_{cwb}^{1}(n)$ and the original CWB signal is represented by $S_{cwb}(n)$. The spreading sequence length is 16. An RCWB signal with a small value of MSE is said to be of good quality. The proposed technique gives MSE values as a function of SNR ranges between 15 and 35 dB, which are below 7.7083×10^{-4}, indicating that the RCWB signal quality obtained by the proposed technique is excellent. The proposed technique gives an MSE value after adding quantization noise (μ-law) to $s_{lnb}^{1}(n)$ is 5.78×10^{-4}indicating that the RCWB signal quality obtained by the proposed technique is excellent.

6.4.2.4.CWB speech Quality

The evaluation of the quality of RCWB speech is done by giving CWB and RCWB signals as inputs and comparing them in order to rate speech quality. The average CWB-PESQ values of the conventional [15,16,19,96], and proposed methods are shown in table 6.9. A CWB-PESQ value of 4.45 confirms that the RCWB signal quality that was obtained by the proposed technique is excellent compared to traditional techniques [15, 16, 19, 96], which was already confirmed by subjective listening tests on a set of participants. Thus, the speech quality was improved by using the proposed technique.

Tab. 6.9. Results of the CWB-P**ESQ**

Technique	CWB-P**ESQ**
TTSBEDH [15]	2.49
TTSBEPC [16]	2.73
TTSBEBDH [19]	3.64
TTSBEWTSI [96]	3.71
Proposed method	4.45

The evaluation of the quality of RCWB speech is done by giving CWB and RCWB signals as inputs and comparing them in order to rate speech quality. The average CWB-POLQA values of the conventional [15, 16, 19, 96]and proposed methods are shown in table

6.10. A CWB-POLQA value of 4.15 confirms that the RCWB signal quality that was obtained by the proposed technique is excellent compared to traditional techniques [15, 16, 19, 96], which was already confirmed by subjective listening tests on a set of participants. Thus, the speech quality was improved by using the proposed technique.

Tab. 6.10. Results of the CWB-POLQA

Technique	CWB-POLQA
TTSBEDH [15]	2.08
TTSBEPC [16]	2.45
TTSBEBDH [19]	3.13
TTSBEWTSI [96]	3.34
Proposed method	4.15

6.5.Results and Conclusions

In this Chapter, SBE is proposed using the DCTBDH technique for extending the bandwidth of the existing LNB telephone networks. The spread SEPs of the MHB signal is embedded within the DCT coefficients of the LNB signal at the transmitter. The embedded information is extracted at the receiver end to reconstruct the CWB signal of good quality.

The spread spectrum technique is employed to increase the robustness of the embedded MHB signal to CAQNs by spreading the SEPs by multiplying them with spreading sequences and then adding them up together to provide the embedded information. The embedded information can be reliably recovered by using a correlator. The MSE test confirms the robustness of the proposed method to CAQNs. The MOS, LNB-PESQ, and LNB-POLQA test values obtained for the proposed method indicate that the method embeds the MHB information more transparently than conventional methods. The proposed technique enhanced the RWB signal quality over conventional techniques, and it was evident through subjective listening, CWB-POLQA, and LSD tests. The proposed method produces a much better-quality speech signal than the conventional techniques. Hence it is suitable for extending the bandwidth of the existing telephone networks without making changes to the telephone networks.

Chapter- 7

7.1. Conclusions

In this thesis, a significant focus is on creating and assessing innovative SBE algorithms that use data masking strategies. At frequencies above LNB, new SBE approaches are introduced. CLNB speech and RCWB speech quality can be improved using new techniques that are more robust to CAQNs. SBE methods utilizing hybrid transform-based data hiding, frequency-domain data hiding, discrete Wavelet transform-discrete Cosine transform-based data hiding with encoding, and discrete Cosine transform-based data hiding have all been developed and analyzed in this report.

Chapter 3 proposes an evaluation of SBE using a Hybrid Model Transform Domain speech bandwidth extension using data hiding(HMTDBWE).The subjective, CWB-POLQA and LSD test results show that the proposed method improves speech quality than traditional SBE techniques. The MOS, LNB-PESQ, and LNB-POLQA test results show that the proposed method hides the MHB information more transparently than traditional SBE techniques.

In Chapter 4, A novel speech BWE technique using DWT-DCT-BDH is proposed. The spread parameters of the MHB signal are hidden inside the low-amplitude, high-spectral frequency part of the LNB signal. The hidden information is retrieved at the other end to recreate a better-quality CWB speech signal. The subjective, CWB-POLQA, and LSD test results show that the proposed method improves speech quality more than traditional SBE techniques. The MSE test results show that the proposed method is robust to CAQNs. The MOS, LNB-PESQ, and LNB-POLQA test results show that the proposed method hides the MHB information more transparently than traditional SBE techniques.

Chapter 5 proposes a novel SBE algorithm aided by frequency-domain data hiding technique for embedding spread SEPs of MHB signal within LNB signal DCT coefficients. The concealed information is extracted to generate an extraordinary-quality CWB signal at the receiver end.The concealed information is vulnerable to CAQNs. Thus, the CDMA approach is employed for reproducing the concealed information thatis appealed as robust towards CAQNs. Especially, every information bit entrenched within the LNB signal is spread out as the product by definite SPEs. Then, the spread signals were summed to create concealed information. The concealed information could be consistently retrieved since the correlation among the SPEs is low. It is evident that the proposed approach was a vigorous solution for SBE. Subjective, CWB-PESQ, LSD, and CWB-POLQA test results confirmed excellent and improved RCWB signal performance using the proposed method over the traditional techniques. An apparent PCL enhancement of the proposed approach over the conventional methods is witnessed from MOS, LNB-PESQ, and LNB-POLQA tests.

In Chapter 6, a novel SBE algorithm utilizing the DCTBDH technique has been proposed. The spread SEPs of the MHB signal is kept embedded within the DCT coefficients of the LNB signal at the transmitter. The hidden information is retrieved at the receiver end to recreate the CWB signal of good quality. The spread spectrum technique is employed to increase the robustness of the embedded MHB signal to CAQNs by spreading the spectral envelope parameters by multiplying them with spreading sequences and then adding them up together to provide the embedded information. The embedded information can be reliably recovered by using a correlator. The MSE test confirms the robustness of the proposed method to CAQNs. The MOS, LNB-PESQ and LNB-POLQA test values obtained for the

proposed method indicate that the method embeds the MHB information more transparently compared to the conventional methods. The proposed technique enhanced the RCWB signal quality over conventional techniques, and it was evident through subjective listening, LSD CWB-PESQ, and CWB-POLQA tests. The proposed method produces a much better-quality speech signal than the conventional techniques.

7.2. Future scope

This section provides future directions for work on SBE using data hiding. The main issues to be considered include robustness to channel and quantization noises, improved perceptual transparency, and improvement in reconstructed CWB signal quality. In this thesis, four different SBE algorithms have been developed and evaluated, which include a novel SBE algorithm using hybrid transform domain-based data hiding technique,a novel SBE algorithm using DWT-DCT based data hiding technique, a novel SBE algorithm aided by frequencydomain data hiding technique and,a novel SBE algorithm using Discrete cosine transform technique. In the future, other data-hiding techniques for SBE can be used to improve the performance of SBE systems.

In addition, multi-languages with different speaking styles vary from user to user. Therefore, studying the dependence of SBE using data hiding on multi-languages with varying speaking styles is mandatory to make sure the international applicability of the SBE using data hiding.

Finally, many researchers have contributed to the development and evaluation of SBE using data hiding. Since the early 2000s, a multitude of approaches and complete algorithms have been proposed. Unfortunately, comparisons between methods from different authors have rarely been reported. A comprehensive comparison of state-of-the-art methods would be an interesting topic of future study and beneficial for SBE using data-hiding research.

References

[1] P. Jax and P. Vary, “Bandwidth extension of speech signals: A catalyst for the introduction of wideband speech coding?,” *IEEE Communications Magazine*, vol. 44, no. 5, pp. 106–111, 2006.

[2] P. Jax, *Enhancement of bandlimited speech signals: Algorithms and theoretical bounds*, Ph. D. dissertation, RWTH Aachen University, 2002.

[3] P. Jax and 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, pp. 237–240, May 2002.

[4] B. Geiser and P. Vary, “Speech bandwidth extension based on in-band transmission of higher frequencies,” in *Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing (ICASSP)*, Vancouver, Canada, pp. 7507–7511, May 2013.

[5] D. O’Shaughnessy, *Speech Communications: Human and Machine*. Wiley-IEEE Press, Second edition, November 1999.

[6] M. Ashby and J. Maidment, *Introducing Phonetic Science*, Cambridge University Press, 2005.

[7] E. Zwicker and H. Fastl, *Psychoacoustics, Facts and Models*, Springer, Second edition, 1999.

[8] H.Pulakka, P. Alku, L. Laaksonen and P. Valve. “The effect of highband harmonic structure in the artificial bandwidth expansion of telephone speech,” In *Proc. Interspeech*, Antwerp, Belgium, pp. 2497–2500, 2007.

[9] P. Jax and P. Vary, “On artificial bandwidth extension of telephone speech,” *Signal Processing*, vol. 83, no. 8, pp. 1707–1719, 2003.

[10] S. Voran, “Listener Ratings of Speech Passbands,” in *Proc. IEEE Workshop on Speech Coding*, Pocono Manor, Pennsylvania, USA, pp. 81–82, 1997.

[11] F. T. Andrews, “Early T-carrier history,” *IEEE Communications Magazine*, vol. 49, no. 4, pp. 12–17, 2011.

[12] Third Generation Partnership Project (3GPP), *AMR wideband speech codec; general description, 3GPP TS 26.171*, 2001, version 5.0.0.

[13] ITU-T Rec *Wideband Coding of Speech at around 16 kb/s Using Adaptive Multirate Wideband (AMR-WB)*. G.722.2, , Jan. 2002.

[14] Jason A. Fuemmeler, Russell C. Hardie and William R. Gardner, "Techniques for the regeneration of wideband speech from narrowband speech," *EURASIP Journal on Applied Signal Processing*, vol. 2001, no. 4, pp. 266–274, 2001.

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

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

[17] S. Chen, H. Leung and H. Ding, "Telephony speech enhancement by data hiding," *IEEE Transactions on Instrumentation and Measurement*, vol. 56, no. 1, pp. 63-74, 2007.

[18] A. Sagi and D. Malah, "Bandwidth extension of telephone speech aided by data embedding," *EURASIP Journal on Advances in Signal Processing*, vol. 2007, Article ID 64921, pp. 1- 16, 2007.

[19] Z. Chen, C. Zhao, G. Geng and F. Yin, "An audio watermark based speech bandwidth extension method," *EURASIP Journal on Audio, Speech, and Music Processing*, vol. 2013, no. 10, pp. 1-8, 2013.

[20] P. Vary and B. Geiser, "Steganographic wideband telephony using narrowband speech codecs," in *Proc. Asilomar Conf. Signals, Syst., Comput.*, Pacific Grove, CA, USA, pp. 1475–1479, Nov. 2007.

[21] A. Nishimura, "Steganographic bandwidth extension for the AMR codec of low bit-rate modes," in *Proc. Interspeech*, Brighton, UK, pp. 2611–2614, Sept. 2009.

[22] B. Geiser and 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, USA, pp. 533–536, Apr. 2007.

[23] R. Martin, U. Heute and C. Antweiler, *Advances in Digital Speech Transmission*, John Wiley & Sons Ltd., Chapter 8, pp. 201–247, 2008.

[24] H. Carl and U. Heute, "Bandwidth Enhancement of Narrow-Band Speech Signals," in *Proc. of European Signal Processing Conference (EUSIPCO)*, Edinburgh, Scotland, pp. 1178–1181, 1994.

[25] Y. M. Cheng, D. O'Shaughnessy and P. Mermelstein, "Statistical Recovery of Wideband Speech from Narrowband Speech," *IEEE Transactions on Speech and Audio Processing*, vol. 2, no. 4, pp. 544–548, 1994.

[26] P. Jax and P. Vary, “Feature selection for improved bandwidth extension of speech signals,” in *Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing (ICASSP)*, Montreal, QC, Canada, pp. 697–700, May 2004.

[27] J. Kontio, L. Laaksonen and P. Alku, “Neural network-based artificial bandwidth expansion of speech,” *IEEE Transactions on Audio, Speech and Language Processing*, vol. 15, no. 3, pp. 873–881, 2007.

[28] U. Kornagel, “Techniques for artificial bandwidth extension of telephone speech,” *Signal Processing*, vol. 86, no. 6, pp. 1296–1306, 2006.

[29] S. Chennoukh, A. Gerrits, G. Miet and R. Sluijter, “Speech enhancement via frequency bandwidth extension using line spectral frequencies,” in *Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing (ICASSP)*, Salt Lake City, UT, USA, vol. I, pp. 665–668, May 2001.

[30] S. Vaseghi, E. Zavarehei and 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, France, vol. III, pp. 844–847, May 2006.

[31] G. Miet, A. Gerrits and J. C. Valiere, “Low-band extension of telephone-band speech,” in *Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing (ICASSP)*, Istanbul, Turkey, vol. III, , pp. 1851–1854, June 2000.

[32] Sheng Yao and Cheung-Fat Chan, “Speech bandwidth enhancement using state space speech dynamics,” in *Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing (ICASSP)*, Toulouse, France, vol. I, pp. 489–492, May 2006.

[33] C. Liu, Q.-J. Fu and S. S. Narayanan, “Effect of bandwidth extension to telephone speech recognition in cochlear implant users,” *Journal of the Acoustical Society of America*, vol. 125, no. 2, pp. EL77–EL83, 2009.

[34] A. Shahina and B. Yegnanarayana, “Mapping neural networks for bandwidth extension of narrowband speech,” in *Proc. Interspeech*, Pittsburgh, PA, USA, pp. 1435–1438, Sept. 2006.

[35] M. L. Seltzer, A. Acero and J. Droppo, “Robust bandwidth extension of noise corrupted narrowband speech,” in *Proc. Interspeech*, Lisbon, Portugal, pp. 1509–1512, Sept. 2005.

[36] G. B. Song and P. Martynovich, “A study of HMM-based bandwidth extension of speech signals,” *Signal Processing*, vol. 89, no. 10, pp. 2036–2044, 2009.

[37] M. R. P. Thomas, J. Gudnason, P. A. Naylor, B. Geiser and P. Vary, “Voice source estimation for artificial bandwidth extension of telephone speech,” in *Proc. IEEE Int.*

Conf. Acoustics, Speech, and Signal Processing (ICASSP), Dallas, TX, USA, pp. 4794–4797, Mar. 2010.

[38] Kyung-Tae Kim, Min-Ki Lee and Hong-Goo Kang, "Speech bandwidth extension using temporal envelope modeling," *IEEE Signal Processing Letters*, vol. 15, pp. 429–432, 2008.

[39] D. A. Heide and G. S. Kang, "Speech enhancement for bandlimited speech," in *Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing (ICASSP)*, Seattle, WA, USA, vol. I, pp. 393–396, May 1998.

[40] J. D. Johnston, "Transform coding of audio signals using perceptual noise criteria," *IEEE Journal on Selected Areas of Communication*, vol. 6, no. 2, pp. 314–323, 1988.

[41] B. S. Atal and L. R. Rabiner, "A pattern recognition approach to voiced-unvoiced-silence classification with applications to speech recognition," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 24, no. 3, pp. 201–212, June 1976.

[42] J. W. Paulus, "Variable bitrate wideband speech coding using perceptually motivated thresholds," in *Proc. IEEE Workshop Speech on Coding (SCW)*, Annapolis, MD, USA, pp. 35–36, Sept. 1995.

[43] Y. Yoshida and M. Abe, "An algorithm to reconstruct wideband speech from narrowband speech based on codebook mapping," in *Proc. International Conference on Spoken Language Processing*, pp. 1591–1594, 1994.

[44] Y. Qian and P. Kabal, "Wideband speech recovery from narrowband speech using classified codebook mapping," in *Proc. Australian International Conference on Speech Science and Technology*, Australia, pp. 106–111, 2002.

[45] J. Epps and W. H. Holmes, "A new technique for wideband enhancement of coded narrowband speech," in *Proc. IEEE Workshop on Speech Coding (SCW)*, Porvoo, Finland, pp. 174–176, June 1999.

[46] R. Hu, Venkatesh Krishnan and David V. Anderson, "Speech bandwidth extension by improved codebook mapping towards increased phonetic classification," in *Proc. Interspeech*, Portugal, pp. 1501–1504, 2005.

[47] Y. Nakatoh, M. Tsushima and T. Norimatsu, "Generation of broadband speech from narrowband speech using piecewise linear mapping," in *Proc. EUROSPEECH*, Greece, pp. 1643-1646, 1997.

[48] Y. Qian and P. Kabal. "Dual-mode wideband speech recovery from narrowband speech," in *Proc. European Conference on Speech Communication and Technology (EUROSPEECH)*, Geneva, Switzerland, pp. 1433–1436, 2003.

[49] K. Y. Park and H. S. Kim, “Narrowband to wideband conversion of speech using GMM based transformation,” in *Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing (ICASSP)*, Turkey, pp. 1843–1846, 2000.

[50] D. G. Raza and C. F. Chan, “Quality enhancement of CELP coded speech by using an MFCC based Gaussian mixture model,” in *Proc. European Conference on Speech Communication and Technology (EUROSPEECH)*, Geneva, Switzerland, pp. 541–544, 2003.

[51] Amr H. Nour-Eldin and P. Kabal, “Mel-frequency cepstral coefficient-based bandwidth extension of narrowband speech,” in *Proc. Interspeech*, Australia, pp. 53–56, 2008.

[52] Amr H. Nour-Eldin and P. Kabal, “Combining frontend based memory with MFCC features for bandwidth extension of narrowband speech,” in *Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing (ICASSP)*, Taiwan, pp. 4001–4004, 2009.

[53] Amr H. Nour-Eldin and P. Kabal, “Memory-based approximation of the Gaussian model framework for bandwidth extension of narrowband speech,” in *Proc. Interspeech*, vol.1, Florence, Italy, pp. 1185–1188, 2011.

[54] P. Bauer and 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)*, USA, pp. 4589–4592, 2008.

[55] S. Yao and Cheung-Fat. Chan, “Block-based bandwidth extension of narrowband speech signal by using CDHMM,” in *Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing (ICASSP)*, vol. I, Philadelphia, PA, USA, pp. 793–796, Mar. 2005.

[56] C. Yaˇglı and E. Erzin, “Artificial bandwidth extension of spectral envelope with temporal clustering,” in *Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing (ICASSP)*, Prague, Czech Republic, pp. 5096–5099, May 2011.

[57] S. Haykin, *Neural Networks: A Comprehensive Foundation*, Second edition, Prentice Hall, 1999.

[58] K. O. Stanley and R. Miikkulainen, “Evolving neural networks through augmenting topologies,” *Evolutionary Computation*, vol. 10, no. 2, pp. 99–127, 2002.

[59] A. Uncini, F. Gobbi and F. Piazza, “Frequency recovery of narrow-band speech using adaptive spline neural networks,” in *Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing (ICASSP)*, vol. II, Phoenix, AZ, USA, pp. 997–1000, Mar. 1999.

[60] Jean-Marc. Valin and R. Lefebvre, “Bandwidth extension of narrowband speech for low bit-rate wideband coding,” in *Proc. IEEE Workshop on Speech Coding (SCW)*, Delavan, WI, USA, pp. 130–132, Sept. 2000.

[61] B. Iser and G. Schmidt, “Neural networks versus codebooks in an application for bandwidth extension of speech signals,” in *Proc. European Conference on Speech Communication and Technology(EUROSPEECH)*, Geneva, Switzerland, pp. 565–568, Sept. 2003.

[62] H. Pulakka and P. Alku, “Bandwidth extension of telephone speech using a neural network and a filter bank implementation for highband Mel spectrum,” *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 19, no. 7, pp. 2170–2183, 2011.

[63] D. Zaykovskiy and B. Iser, “Comparison of neural networks and linear mapping in an application for bandwidth extension,” in *Proc. SPECOM*, Greece, 2005.

[64] J. Makhoul and M. Berouti, “High-frequency regeneration in speech coding systems,” in *Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing (ICASSP)*, Washington, DC, USA, pp. 428–431, 1979.

[65] A. de Cheveigne and H. Kawahara, “YIN, a fundamental frequency estimator for speech and music,” *J. Acoust. Soc. Am.*, vol. 111, no. 4, pp. 1917–1930, Apr.2002.

[66] M. Nilsson and W. B. Kleijn, “Avoiding over-estimation in bandwidth extension of telephony speech,” in *Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing (ICASSP)*, vol. II, Salt Lake City, UT, USA, pp. 869–872, May 2001.

[67] I. Katsir, I. Cohen and D. Malah, “Speech bandwidth extension based on speech phonetic content and speaker vocal tract shape estimation,” in *Proc. European Signal Processing Conference (EUSIPCO)*, Barcelona, Spain, pp. 461–465, Sept. 2011.

[68] U. Kornagel, “Spectral widening of the excitation signal for telephone-band speech enhancement,” in *Proc. Int. Workshop on Acoustic Echo and Noise Control (IWAENC)*, Darmstadt, Germany, Sept. 2001, pp. 215–218.

[69] J. Epps and W. H. Holmes, “Speech enhancement using STC-based bandwidth extension,” in *Proc. International Conference on Spoken Language Process (ICSLP)*, vol. 2, Sydney, Australia, pp. 519–522, Nov. 1998.

[70] J. Epps, *Wideband extension of narrowband speech for enhancement and coding*, Ph.D. dissertation, The University of New South Wales, Australia, Sept. 2000.

[71] Cheung-Fat Chan and Wai-Kwong Hui, “Wideband re-synthesis of narrowband CELP-coded speech using multiband excitation model,” In *Proc. International Conference on Spoken Language Processing*, Philadelphia, PA, USA, pp. 322–325, 1996.

[72] J. S. Park, M. Y. Choi and H. S. Kim, “Low-band extension of CELP speech coder by harmonics recovery,” in *Proc. Int. Symposium on Intelligent Signal Processing And Communication Systems (ISPACS)*, Seoul, South Korea, pp. 147–150, 2004.

[73] H. Gustafsson, U. A. Lindgren and I. Claesson, “Low-complexity feature-mapped speech bandwidth extension,” *IEEE Transactions on Audio, Speech and Language Processing*, vol. 14, no. 2, pp. 577–588, Mar. 2006.

[74] T. V. Pham, F. Schaefer and G. Kubin, “A novel implementation of the spectral shaping approach for artificial bandwidth extension,” in *Proc. International Conference on Communications and Electronics (ICCE)*, Nha Trang, Vietnam, Aug. 2010, pp. 262–267.

[75] McCree, “A 14 kb/s wideband speech coder with a parametric highband model,” in *Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing (ICASSP)*, vol. II, Istanbul, Turkey, pp. 1153–1156, June 2000.

[76] Y. Qian and P. Kabal, “Combining equalization and estimation for bandwidth extension of narrowband speech,” in *Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing (ICASSP)*, vol. I, Montreal, QC, Canada, pp. 713– 716, May 2004.

[77] T. Unno and A. McCree, “A robust narrowband to wideband extension system featuring enhanced codebook mapping,” in *Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing (ICASSP)*, vol. I, Philadelphia, PA, USA, pp. 805–808, Mar. 2005.

[78] H. Pulakka, U. Remes, K. Palomaki, M. Kurimo and P. Alku, “Speech bandwidth extension using Gaussian mixture model-based estimation of the highbandmel spectrum,” in *Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing (ICASSP)*, Prague, Czech Republic, pp. 5100–5103, May 2011.

[79] J. P. Cabral and L. C. Oliveira, “Pitch-synchronous time-scaling for high frequency excitation regeneration,” in *Proc. Interspeech*, Lisbon, Portugal, pp. 1513–1516, Sept. 2005.

[80] T. Ramabadran and M. Jasiuk, “Artificial bandwidth extension of narrow-band speech signals via high-band energy estimation,” in Proc. European Signal Processing Conference (EUSIPCO), Lausanne, Switzerland, pp. 1-5, Aug. 2008.

[81] B. Geiser and P. Vary, “Beyond wideband telephony – bandwidth extension for super-wideband speech,” in *Proc. Deutsche Jahrestagung für Akustik (DAGA)*, Dresden, Germany, pp. 635–636, Mar. 2008.

[82] B. Geiser, P. Jax, P. Vary, H. Taddei, S. Schandl, M. Gartner, C. Guillaume and S. Ragot, “Bandwidth extension for hierarchical speech and audio coding in ITU-T Rec. G.729.1,”

IEEE Transactions on Audio, Speech and Language Processing, vol. 15, no. 8, pp. 2496–2509, Nov. 2007.

[83] R. Taori, R. J. Sluijter and A. J. Gerrits, "Hi-BIN: An alternative approach to wideband speech coding," in *Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing (ICASSP)*, vol. II, Istanbul, Turkey, pp. 1157–1160, June 2000.

[84] K. T. Kim, J.-Y. Choi and H. G. Kang, "Perceptual relevance of the temporal envelope to the speech signal in the 4–7 kHz band," *Journal of the Acoustical Society of America*, vol. 122, no. 3, pp. EL88–EL94, Sept. 2007.

[85] B. Geiser, H. Taddei and P. Vary, "Artificial bandwidth extension without side information for ITU-T G.729.1," in *Proc. Interspeech*, Antwerp, Belgium, pp. 2493–2496, Aug. 2007.

[86] P. Jax, B. Geiser, S. Schandl, H. Taddei and P. Vary, "An embedded scalable wideband codec based on the GSM EFR codec," in *Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing (ICASSP)*, vol. I, Toulouse, France, pp. 5–8, May 2006.

[87] K. Kalgaonkar and M. Clements, "Vocal tract area based artificial bandwidth extension," in *Proc. IEEE International Workshop on Machine Learning for Signal Processing (MLSP)*, Cancun, Mexico, pp. 480–485, Oct. 2008.

[88] N. I. Park, Y. H. Lee and H. K. Kim, "Artificial bandwidth extension of narrowband speech signals for the improvement of perceptual speech communication quality," in *Proc. International Conference Future Generation Communication And Networking (FGCN)*, Jeju Island, Korea, Dec. 2011, pp. 143–153.

[89] D. Bansal, Bhiksha Raj and Paris Smaragdis, "Bandwidth expansion of narrowband speech using non-negative matrix factorization," in *Proc. Interspeech*, Portugal, pp. 1505–1508, 2005.

[90] H. Tolba andDouglas O'Shaughnessy, "On the application of the AM-FM model for the recovery of missing frequency bands of telephone speech," in *Proc. International Conference on Spoken Language Processing*, Australia, 1998.

[91] S. Kuroiwa, M. Takashina, S. Tsuge and R. Fuji, "Artificial bandwidth extension for speech signals using speech recognition," in *Proc. Interspeech*, Antwerp, Belgium, pp. 2501–2504, Aug. 2007.

[92] M. Nilsson, H. Gustafsson, S. V. Andersen and W. B. Kleijn, "Gaussian mixture model based mutual information estimation between frequency bands in speech," in *Proc.IEEE Int. Conf. Acoustics, Speech, and Signal Processing (ICASSP)*, vol. I, Orlando, FL, USA, pp. 525–528, May 2002.

[93] M. Nilsson, S. V. Andersen and W. B. Kleijn, "On the Mutual Information between Frequency Bands in Speech," in *Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing (ICASSP)*, Istanbul, Turkey, vol. 3, pp. 1327–1330, 2000.

[94] Y. Agiomyrgiannakis and Y. Stylianou, "Combined estimation/coding of highband spectral envelopes for speech spectrum expansion," in *Proc.IEEE Int. Conf. Acoustics, Speech, and Signal Processing (ICASSP)*, vol. I, Montreal, QC, Canada, pp. 469–472, May 2004.

[95] Y. Agiomyrgiannakis and Y. Stylianou, "Conditional vector quantization for speech coding," IEEE Transactions Audio, Speech, and Language Processing, vol. 15, no. 2, pp. 377–386, Feb. 2007.

[96] B. Geiser, P. Jax and P. Vary, "Artificial bandwidth extension of speech supported by watermark-transmitted side information," in *Proc. Interspeech*, Lisbon, Portugal, pp. 1497–1500, Sept. 2005.

[97] G. Fuchs and R. Lefebvre, "A new post-filtering for artificially replicated highband in speech coders," in *Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing (ICASSP)*, vol. I, Toulouse, France, pp. 713–716, May 2006.

[98] V. Berisha and A. Spanias, "Wideband speech recovery using psychoacoustic criteria," *EURASIP Journal on Audio, Speech, and Music Processing*, vol. 2007, Article ID 16816, pp. 1-18, Aug. 2007.

[99] P. J. Patrick, *Enhancement of Bandlimited Speech Signals*, PhD thesis, Loughborough University of Technology, 1983.

[100] C. McElroy, B. Murray and A. D. Fagan, "Wideband Speech Coding in 7.2 kb/s," in*Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing (ICASSP)*, Minneapolis, MN, USA, vol. 2, pp. 620–623, 1993.

[101] G. Aguilar, Juim-Hwey Chen, R. B. Dunn, R. J. McAulay, X. Sun, W. Wang and R. Zopf, "An Embedded Sinusoidal Transform Codec with Measured Phases and Sampling Rate Scalability," in *Proc.IEEE Int. Conf. Acoustics, Speech, and Signal Processing (ICASSP)*, Istanbul, Turkey, vol. 2, pp. 1141–1144, 2000.

[102] Q. Cheng and J. Sorensen, "Spread spectrum signaling for speech watermarking," in *Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing (ICASSP)*, Salt Lake City, UT, USA, pp. 1337–1340, 2001.

[103] B. Chen and G. W. Wornell, "Quantization index modulation: a class of provably good methods for digital watermarking and information embedding," *IEEE Transactions on Information Theory*, vol. 47, no. 4, pp. 1423–1443, 2001.

[104] J. J. Eggers, R. Bauml, R. TzschoppeandB.Girod, “Scalar costa scheme for information embedding,” *IEEE Transactions on Signal Processing*, vol. 51, no. 4, pp. 1003–1019, 2003.

[105] M. Celik, G. Sharma and M. A. Tekalp, “Pitch and duration modification for speech watermarking,” in *Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing (ICASSP)*, Philadelphia, PA, USA, pp. 17–20, 2005.

[106] K. Hofbauer, G. Kubin and W. B. Kleijn, “Speech watermarking for analog flat-fading bandpass channels,” *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 17, no. 8, pp. 1624–1637, 2009.

[107] A. Gurijala, *Speech Watermarking through Parametric Modeling*, Ph. D. dissertation, Michigan State University, East Lansing, MI, USA, 2007.

[108] A. Gurijala and J. R. Deller, “On the robustness of parametric watermarking of speech,” in *Proc. International Conference on Multimedia Content Analysis and Mining (MCAM)*, Weihai, China, pp. 501–510, 2007.

[109] International Telecommunications Union-Telecommunication Sector (ITU-T), *ITU-T Rec. G.711: Pulse code modulation (PCM) of voice frequencies*, 1972.

[110] International Telecommunications Union-Telecommunication Sector (ITU-T), *ITU-T Rec. G.726: 40, 32, 24, 16 kbit/s adaptive differential pulse code modulation (ADPCM)*, 1990.

[111] G. Cohen, I. Honkala, S. Litsyn and A. Lobstein, *Covering Codes*, Elsevier, North-Holland Mathematical Library, Volume 54, 1997.

[112] F. Galand and G. Kabatiansky, “Information hiding by coverings,” in *Proc. IEEE Information Theory Workshop (ITW)*, Paris, France, pp. 151–154, 2003.

[113] Oscal T.-C. Chen and Chia-Hsiung Liu, “Content-dependent watermarking scheme in compressed speech with identifying manner and location of attacks,” *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 15, no. 5, pp. 1605–1616, 2007.

[114] H. Tian, K. Zhou, H. Jiang, J. Liu, Y. Huang and D. Feng, “An M-sequence based steganography model for Voice over IP,” in *Proc. IEEE International Conference on Communications (ICC)*, Dresden, Germany, 2009.

[115] Aoki, N. Improvement of a band extension technique for G.711 telephony speech by using steganography, in *Proc. International Conference on Intelligent Information Hiding and Multimedia Signal Processing (IIH-MSP)*, Kyoto, Japan, pp. 487–490, 2009.

[116] International Telecommunications Union-Telecommunication Sector (ITU-T), *ITU-T Rec. G.723.1: Dual rate speech coder for multimedia communications transmitting at 5.3 and 6.3 kbit/s*, 1996.

[117] International Telecommunications Union-Telecommunication Sector (ITU-T), *ITU-T Rec. G.729: Coding of speech at 8 kbit/s using conjugate structure algebraic-code-excited linear prediction (CS-ACELP)*, 1996.

[118] A. Shahbazi, A.H. Rezaie and R. Shahbazi, "MELPe coded speech hiding on enhanced full rate compressed domain," in *Proc. Fourth Asia International Mathematical/Analytical Modelling and Computer Simulation Conference (AMS)*, Bornea, Malaysia, pp. 267–270, 2010.

[119] ETSI, *ETSI EN 300 961: Digital cellular telecommunications system (phase 2+); full rate speech; transcoding (GSM 06.10 version 8.1.1 release 1999)*, 1990.

[120] ETSI, *ETSI EN 300 726: Digital cellular telecommunications system (phase 2+); enhanced full rate (EFR) speech transcoding (GSM 06.60 version 8.0.1 release 1999)*, 1998.

[121] K. Jarvinen, J.Vainio, P. Kapanen, T. Honkanen, P. Haavisto, R. Salami, C. Laflamme and J.-P. Adoul, "GSM enhanced full rate speech codec," in *Proc.IEEE Int. Conf. Acoustics, Speech, and Signal Processing (ICASSP)*," Munich, Germany, pp. 771–774, 1997.

[122] T. Wang, K. Koishida, V. Cuperman, A. Gersho and J. S. Collura, "A 1200/2400 bps coding suite based on MELP," in *Proc. IEEE Workshop on Speech Coding (SCW)*, Tsukuba, Japan, pp. 90–92, 2002.

[123] ETSI, *ETSI EN 128 062: Digital cellular telecommunications system (phase 2+); universal mobile telecommunications system (UMTS); LTE; inband tandem free operation (TFO) of speech codecs; service description; stage 3 (3GPP TS 28.062 version 9.0.0 release 9)*, 1999.

[124] H. Licai and W. Shuozhong, "Information hiding based on GSM full rate speech coding," in *Proc. International conference on Wireless Communications, Networking and Mobile Computing*," Wuhan, China, pp. 1–4, 2006.

[125] Heping Ding, "Wideband audio over narrowband low-resolution media," in Proc. *IEEE Int. Conf. Acoustics, Speech, and Signal Processing (ICASSP)*, Montreal, QC, Canada, pp. 489–492, 2004.

[126] L. Liu, M. Li, Q. Li and Y. Liang, “Perceptually transparent information hiding in G.729 bitstream,” in *Proc. International Conference on Intelligent Information Hiding and Multimedia Signal Processing (IIH-MSP)*, Harbin, China, pp. 406–409, 2008.

[127] T. Xu and Z. Yang, “Simple and effective speech steganography in G.723.1 low-rate codes,” in *Proc. International Conference on Wireless Communications & Signal Processing (WCSP)*, Nanjing, China, 2009.

[128] M. Li, Y. Jiao and X. Niu, “Reversible watermarking for compressed speech,” in *Proc. 8th International Conference on Intelligent Systems Design and Applications (ISDA)*, Kaohsiung, Taiwan, pp. 197–201, 2008.

[129] A. Nishimura, “Data hiding in pitch delay data of the adaptive multi-rate narrow-band speech codec,” in *Proc. International Conference on Intelligent Information Hiding and Multimedia Signal Processing (IIH-MSP)*, Kyoto, Japan, pp. 483–486, 2009.

[130] ETSI, *ETSI EN 301 704: Digital cellular telecommunications system (phase 2+) (GSM); adaptive multi-rate (AMR) speech transcoding (GSM 06.90 version 7.2.1 release 1998)*, 2000.

[131] E. Ekudden, R. Hagen, I. Johansson and J. Svedberg, “The adaptive multirate speech coder,” in *Proc. IEEE Workshop on Speech Coding (SCW)*, Porvoo, Finland, pp. 117–119, 1999.

[132] B. Xiao, Y. Huang and S. Tang, “An approach to information hiding in low bit-rate speech stream,” in *Proc. IEEE Global Telecommunications Conference (GLOBECOM)*, New Orleans, LA, USA, pp. 1–5, 2008.

[133] S. Andersen, A. Duric, H. Astrom, R. Hagen, W. Kleijn and J. Linden, *Internet low bit rate codec (iLBC)*, IETF RFC 3951, 2004.

[134] N. Chetry and M. Davies, “Embedding side information into a speech codec residual,” in *Proc. European Signal Processing Conference (EUSIPCO)*, Florence, Italy, 2006.

[135] Zhe-Ming Lu, Bin Yan and Sheng-He Sun, “Watermarking combined with CELP speech coding for authentication,” *IEICE Transactions on Information and Systems*, vol. E88-D, no. 2 , pp. 330–334, 2005.

[136] M. Iwakiri and K. Matsui, “Embedding a text into conjugate structure algebraic code excited linear prediction audio codes,” in *Proc. IPSJ Computer System Symposium*, Shizuoka, Japan, pp. 2623–2630,1999.

[137] B. Geiser and P. Vary, “High rate data hiding in ACELP speech codecs,” in *Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing (ICASSP)*, Las Vegas, NV, USA, pp. 4005–4008, 2008.

[138] N. Bhatt and Y. Kosta, “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*, vol. 18, no. 1, pp. 57–64, 2015.

[139] Y. Kosta, “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*, vol. 19, no. 4, pp. 881–893, 2016.

[140] H. Pulakka, L. Laaksonen, M. Vainio, J. Pohjalainen and P. Alku, “Evaluation of an Artificial speech bandwidth extension method in three languages,” *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 16 , no. 6, pp. 1124–1137, 2008.

[141] S. Voran, “Objective estimation of perceived speech quality—part I: development of the measuring normalizing block technique,” *IEEE Transactions on Speech and Audio Processing*, vol. 7, no. 4, pp. 371–382, Jul. 1999.

[142] G. Fant, *Speech Sounds and Features*, Cambridge, MIT Press, 1973.

[143] M. R. Schroeder, B. S. Atal and J. L. Hall, “Optimizing digital speech coders by exploiting masking properties of the human ear,” *Journal of the Acoustical Society of America*, vol. 66, pp. 1647–1652, Dec. 1979.

[144] E. Zwicker and E. Terhardt, “‘Analytical expression for critical-band rate and critical bandwidth as a function of frequency,” *Journal of the Acoustical Society of America*, vol. 68, , pp. 1523–1525, Nov. 1980.

[145] B. C. J. Moore, *An Introduction to the Psychology of Hearing*, Academic, 1989.

[146] J. G. Beerends and J. A. Stemerdink, “A perceptual audio quality measure based on a psychoacoustic sound representation,” *Journal of Audio Engineering Society*, vol. 40, no. 2, pp. 963–978, Dec. 1992.

[147] *Information Technology—Coding of Moving Pictures and Associated Audio for Digital Storage Media at up to About 1.5 Mbit/s—Part 3:Audio*, ISO/IEC JTC 1/SC 29/WG 11, ISO/IEC 11172-3, May 20, 1993.

[148] A. A. Hassan, J. E. Hershey and G. J. Saulnier, *Perspectives in Spread Spectrum*, Springer, 1998.

[149] HannuPulakka, *Development and evaluation of artificial bandwidth extension methods for narrowband telephone speech*, Ph. D. dissertation, Aalto University, Finland, 2013.

[150] International Telecommunications Union, *Methods for subjective determination of transmission quality*, ITU-T Recommendation P.800, August 1996

[151] Siyue Chen and Henry Leung, "Concurrent data transmission through analog speech channel using data hiding," *IEEE Signal Processing Letters*, vol. 12, no. 8, pp. 581-584, 2005.

[152] International Telecommunications Union-Telecommunication Sector (ITU-T), *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

[153] International Telecommunications Union-Telecommunication Sector (ITU-T), *Wideband extension to recommendation P.862 for the assessment of wideband telephone networks and speech codecs*, ITU-T Recommendation P.862.2, November 2005

[154] T. Rabie and D. Guerchi, "Magnitude spectrum speech hiding," in *Proc. IEEE International Conference on Signal Processing and Communications (ICSPC 2007)*, Dubai, pp. 1147–1150, November 2007.

[155] E. Hansler and G. Schmidt, *Speech and Audio Processing in Adverse Environments*, Springer, 2008.

[156] B. Iser, W. Minker and G. Schmidt, *Bandwidth extension of speech signals*, Springer, 2008.

[157] E. H. Dinan and E. H. Jabbari, "Spreading codes for direct sequence CDMA and wideband CDMA cellular networks," *IEEE Communications Magazine*, vol. 36, no. 9, pp.48–54, 1998.

[158] A. Goldsmith, *Wireless communications*, Cambridge University Press, 2005.

[159] Siyue Chen and Henry Leung, "A bandwidth extension technique for signal transmission using chaotic data hiding," *Circuits Systems and Signal Processing*, vol. 27, no. 6, pp. 893–913, 2008.

[160] W. Strange, T. R. Edman and J. J. Jenkins, "Acoustic and phonological factors in vowel identification," *Journal of Experimental Psychology: Human Perception and Performance*, vol. 5, no. 4, pp. 643–656, 1979.

[161] S. Andreas, P. T. Ed and A. Venkatraman, *Audio Signal Processing and Coding*, Wiley-Interscience Publication, 2006.

[162] C. Erdmann, P. Vary, K. Fischer, W. Xu, M. Marke, T. Fingscheidt, I. Varga, M. Kaindl, , C. Quinquis, , B. Kovesi and D. Massaloux, "A Candidate Proposal for a 3GPP Adaptive Multi-Rate Wideband Speech codec," in *Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing (ICASSP)*, Salt Lake City, UT, USA , pp. 757–760, 2001.

[163] J. W. Paulus and J. Schnitzler, "16 kbit/s Wideband Speech Coding Based on Unequal Subbands," in *Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing (ICASSP)*, Atlanta, GA, USA, pp. 651–654, 1996.

[164] M. R. Schroeder and B. Atal, "Code-Excited Linear Prediction (CELP): High Quality speech at Low Bit Rates," in *Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing (ICASSP)*, Tampa, FL, USA, pp.937-940, 1985.

[165] ETSI ES 201 108 V1.1.2, *Speech Processing, Transmission and Quality aspects (STQ); Distributed speech recognition; Front-end feature extraction algorithm; Compression algorithms*, 2000.

[166] 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.

[167] P. Jax, *Audio Bandwidth Extension: Application of Psychoacoustics, Signal Processing and Loudspeaker Design*, John Wiley & Sons Ltd, 2004.

[168] P. Vary and R. Martin, *Digital speech transmission: Enhancement, Coding and Error Concealment*, John Wiley & Sons Ltd, 2006.

[169] J. G. Proakis, *Digital Communications*, Second edition, McGraw-Hill, 1989.

[170] L. Hanzo, F. C. A. Somerville and J. P. Woodard, *Voice Compression and Communications: Principles and Applications for Fixed and Wireless Channels*, IEEE Press, 2001.

[171] M. Nilsson and W. B. Kleijn, "Avoiding overestimation in bandwidth extension of telephony speech," in *Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing (ICASSP)*, vol. 2, Salt Lake City, UT, pp. 869–872, May 2001.

[172] Y. Linde, A. Buzo and R.M. Gray, "An algorithm for vector quantizer design," *IEEE Transactions on Communications*, vol. 28, no. 1, pp. 84-95, 1980.

[173] W. Feller, *An Introduction to Probability Theory and Its Applications*, Third edition, Wiley, 1970.

[174] N. Bhatt, "Implementation and overall performance evaluation of CELP based GSM AMR NB coder over ABE," in *Proc. International Conference on Communication Systems and Network Technologies*, Gwalior, India, pp. 402–406, Apr. 2015.

[175] G. h. Alipoor and M. H. Savoji, “Wideband speech coding using ADPCM and a new enhanced bandwidth extension method,” in *Proc. International Symposium on Intelligent Signal Processing*, Floriana, Malta, pp. 1–4, September 2011.

[176] G. h. Alipoor and M. H. Savoji, “Wideband speech coding based on bandwidth extension and sparse linear prediction,” in Proc. International Conference on Telecommunications and Signal Processing, Prague, Czech Republic, pp. 454-459, July 2012.

[177] A. Delforouzi and M. Pooyan, “Adaptive Digital Audio Steganography Based on Integer Wavelet Transform,” *Circuits, Systems, and Signal Processing*, vol. 27, no. 2, pp. 247-259, 2008.

[178] Stephane G. Mallat, “A theory for multiresolution signal decomposition: The wavelet representation,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 11, no. 7, pp. 674–693, 1989.

[179] N. Cvejic and T. Seppanen, “Channel capacity for high bit rate audio data hiding algorithms in diverse transform domains,” in Proc. International Symposium on Communications and Information Technologies, Sapporo, Japan, pp. 84-88, October 2004.

[180] N. Baranwal and K. Datta, “Comparative Study of Spread Spectrum Based Audio Watermarking Techniques,” in Proc. International Conference on Recent Trends in Information Technology, Chennai, India, pp. 896–900, June 2011.

[181] Chin-Su Ko, Ki-Young Kim, Rim-Wo Hwang, YoungSeop Kim and Sang-Burm Rhee “Robust audio watermarking in Wavelet domain using pseudorandom Sequences,” in *Proc. International Conference on Computer and Information Science*, Jeju Island, South Korea, pp. 397–401, July 2005.

[182] Y. H. Chen and J. C. Chen, “A new multiple audio watermarking algorithm applying DS-CDMA,” in *Proc. International Conference on Machine Learning and Cybernetics*, Baoding, pp. 2205–2210, July 2009.

[183] B. E. Keiser and E. Strange, *Digital Telephony and Network Integration*, Springer, 1995.

[184] A. V. Oppenheim and J. S. Lim, “The importance of phase in signals,” *Proceedings of the IEEE*, vol. 69, no. 5, pp. 529–541, May 1981.

[185] Y. Gu, Z.-H. Ling and L.-R. Dai, “Speech Bandwidth Extension Using Bottleneck Features and Deep Recurrent Neural Networks,” in *Proc. of Annual Conference of the*

International Speech Communication Association (Interspeech), San Francisco, USA, pp. 297–301, Sep. 2016.

[186] J. Abel, M. Strake and T. Fingscheidt, "Artificial Bandwidth Extension Using Deep Neural Networks for Spectral Envelope Estimation ," in *Proc. of International Workshop on Acoustic Signal Enhancement (IWAENC)*, Xi'an, China, pp. 1–5, Sep. 2016.

[187] Y. Li and S. Kang, "Artificial bandwidth extension using deep neural network-based spectral envelope estimation and enhanced excitation estimation," *IET Signal Process.*, vol. 10, no. 4, pp. 422427, Jun. 2016.

[188] A. Johannes and F. Tim, "Artificial speech bandwidth extension using deep neural networks for wideband spectral envelope estimation," *IEEE/ACM Trans. Audio, Speech, and Lang. Process.*, vol. 26, no. 1, pp. 71-83, Jan. 2018.

[189] W. Yingwue, Z. Shenghui, Q. Dan and K. Jingming, "Using conditional restricted boltzmann machines for spectral envelope modeling in speech bandwidth extension," in *Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing (ICASSP)*, Shanghai, China, pp. 5930-5934, Apr. 2016.

[190] L. Zhen-Hua, A. Yang, G. Yu and D. Li-Rong, "Waveform Modeling and Generation Using Hierarchical Recurrent Neural Networks for Speech Bandwidth Extension*,"* *IEEE/ACM Trans. Audio, Speech, and Lang. Process.*, vol. 26, no. 5, pp. 883-894, May. 2018.

[191] Y.Wang, S. Zhao, J. Li, and J. Kuang, "Speech bandwidth extension using recurrent temporal restricted Boltzmann machines," *IEEE Signal Process. Lett.*, vol. 23, no. 12, pp. 1877–1881, Dec. 2016.

[192] L. Bong-Ki, N. Kyoungjin, C. Joon-Hyuk, C. Kihyun and O. Eunmi, "Deep Neural Networks Ensemble for Speech Bandwidth Extension," *IEEE ACCESS*, pp. 1–8, 2018.

[193] J. Abel and T. Fingscheidt "A DNN Regression Approach to Speech Enhancement by Artificial Bandwidth Extension," in *Proc. IEEE Workshop on*Applications of Signal Processing to Audio and Acoustics (WASPAA), New Paltz, NY, USA, pp. 219-223, Oct. 2017.

[194] W.Yingxue, Z. Shenghui, L. Jianxin, K. Jingming and Z. Qiang"Recurrent neural network for spectral mapping in speech bandwidth extension," in *Proc. IEEE Global Conference on* Signal and Information Processing,Washington, DC, USA**,** pp. 242-246, Dec. 2016.

[195] S. Jishnu, M. Subhadip and S. Chandra Sekhar, “Joint Dictionary Training For Bandwidth Extension Of Speech Signals” in *Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing (ICASSP)*, Shanghai, China, pp. 5925–5929, 2016.

[196] Pramod B. Bachhav , T. Massimiliano , M. Moctar, B. Christophe and E. Nicholas “Artificial bandwidth extension using the constant Q transform,” in *Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing (ICASSP)*, New Orleans, LA, USA, pp. 5550-5554, March. 2017.

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

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

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

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

[201] B.Pramod, T.Massimiliano, E. Nicholas(2019).latent representation learning for artificial bandwidth extension using a conditional variational auto-encoder In Proceedings of IEEE international conference on acoustics, speech, and signal processing (ICASSP).PP.7010-7014.

[202] H.Xiang,X.Chenglin, H.Nana, X.Lei , Ch.EngSiong, and L.Haizhou (2020) time-domain neural network approach for speech bandwidth extension.In Proceedings of IEEE international conference on acoustics, speech, and signal processing (ICASSP).PP.866-870.

[203] S.Jonas, F.Friedrich, B. Markus ,S. Gerhard, (2020) artificial bandwidth extension using a conditional generativeadversarial network with discriminative training . ICASSP.PP.7005-7009.

[204] L.Mathieu, & G.Felix (2020) bandwidth extension of musical audio signals with no side informationusing dilated convolutional neural networks *IEEE international conference on acoustics, speech, and signal processing (ICASSP)*.PP. 801-805.

[205] N.Kyoungjin & Ch.Joon-Hyuk, (2020). Deep neural network ensemble for reducing artificial noise in bandwidth extension` *ELSIVER Digital Signal Processing*, 102, 102760. https://doi.org/10.1016/j.dsp.2020.102760

[206] F.Berthy, J.Zeyu, S.Jiaqi, and F.Adam (2019) learning bandwidth expansion using perceptually-motivated loss*In Proceedings of IEEE international conference on acoustics, speech, and signal processing (ICASSP)*.PP. 606-610.

[207] A.Johannes & F.Tim (2019), Sinusoidal-Based Lowband Synthesis for Artificial Speech Bandwidth Extension, *IEEE/ACM transactions on audio, speech, and language processing*, (27) 4, p.765-776.

[208] G.Archit, S.Brendan. A.Yannis, & C.W.Thomas (2019).speech bandwidth extension withwavenet. *IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*.PP.205-208.

[209] N.Prasad, & T. Kishore Kumar (2017) Speech bandwidth extension aided by spectral magnitude data hiding, *Circuits, Systems, and Signal Processing*, 36,(11), pp. 4512-4540.

[210] K.Sunil Kumar, & T.Kishore Kumar (2019) Speech Bandwidth Extension Aided by Hybrid Model Transform Domain Data Hiding, Proceedings of *IEEE International Symposium on Circuits and Systems (ISCAS)*, pp. 1-5.

[211] A.Kanhe, & Aghila, (2016) DCT based Audio Steganography in Voiced and Un-voiced Frames', Proceedings on*International Conference of Information and Analytics*, pp. 1-4.

LIST OF PUBLICATIONS

International Journals

1. 2**. Sunil Kumar Koduri and T. Kishore Kumar, "DWT-DCT-Based Data Hiding for Speech Bandwidth Extension'' Radio engineering **(SCI),** DOI: 10.13164/re.2021.0435 VOL.30,N0.2 June 2021.

2. 3**. Sunil Kumar Koduri and T. Kishore Kumar, "Hybrid transform -Based Data Hiding for Speech Bandwidth Extension''Traitment Du Signal **(SCI),** https://doi.org/10.18280/ts.390324.

3. 4**. Sunil Kumar Koduri and T. Kishore Kumar, "DCT-Based Data Hiding for Speech Bandwidth Extension '' IJST **(ESCI/SCOPUS),** DOI: /10.1007/s10772-022-09980-x June 2022.

International Conferences

1. 1*. Sunil Kumar Koduri and T. Kishore Kumar, "Speech Bandwidth Extension Aided By HybridModel Transform Domain Data Hiding" IEEE International Symposium on circuits and systems (ISCAS 2019), at Sapporo Japan ,26-29 May 2019.**(IEEE Xplore)**