Bandwidth Extension of Speech Signals: A Comprehensive Review

2 authors, including:

Prasad Nizampatnam
National Institute of Technology, Warangal
11 PUBLICATIONS   **30** CITATIONS

Bandwidth Extension of Speech Signals: A Comprehensive Review

N.Prasad
National Institute of Technology Warangal, Warangal-506004, India
E-mail: prasad.niz@gmail.com

T. Kishore Kumar
National Institute of Technology Warangal, Warangal-506004, India
E-mail: kishorefr@gmail.com

Abstract—Telephone systems commonly transmit narrowband (NB) speech with an audio bandwidth limited to the traditional telephone band of 300-3400 Hz. To improve the quality and intelligibility of speech degraded by narrow bandwidth, researchers have tried to standardize the telephonic networks by introducing wideband (50-7000 Hz) speech codecs. Wideband (WB) speech transmission requires the transmission network and terminal devices at both ends to be upgraded to the wideband that turns out to be time-consuming. In this situation, novel Bandwidth extension (BWE) techniques have been developed to overcome the limitations of NB speech. This paper discusses the basic principles, realization, and applications of BWE. Challenges and limitations of BWE are also addressed.

Index Terms—Speech bandwidth extension, Data hiding, Model-based techniques, Non-Model- based techniques, Speech quality, Speech intelligibility, wideband speech coding.

I. INTRODUCTION

Due to historical and economic reasons, the audio frequency range of telephone speech is typically limited to approximately the traditional telephone band of 300–3400 Hz. The mobile users are experiencing the muffled sound of speech due to its narrow audio bandwidth [1]. The lack of low frequencies results in a thin voice and degrades the naturalness and presence, whereas the absence of high frequencies deteriorates especially fricative sounds such as [s] and [f] and causes the characteristic muffled speech quality.This constraint limits the quality of speech received by the listener on either side of the connection. Furthermore, some of the speaker-specific characteristics of speech are lost.

The technology and standardized methods exist for the transmission of speech of a wider bandwidth. Wideband speech typically refers to the acoustic bandwidth of 50–7000 Hz, which gives a substantially better speech quality with brighter and fuller sound compared to narrowband speech. Until recently, the deployment of wideband speech transmission has been slow, and wideband speech has been mainly used in specific applications such as video conferencing and internet telephony.WB speech transmission requires the transmission network and terminal devices at both ends to be upgraded to the wideband that turns out to be a costly for all the users across the communication system.

Presently, majority of the network operators are offering wideband speech services, but it requires the transmission network and terminal devices at both the ends to be upgraded to the wideband that turns out to be prolonged changeover time.To improve the quality during the changeover, novel BWE techniques have been developed using digital signal processing techniques. One such method is artificial bandwidth extension (ABWE) in which the missing spectra is estimated from narrowband signal. Alternatively, the additional information about the missing frequency regions can be transmitted as side information to support the bandwidth extension. In general, this provides better output quality than artificial bandwidth extension, but the transmission of side information requires special arrangements that may be difficult to implement in practice.

The speech bandwidth can be extended to frequencies lower than or higher than the narrow band or both [2]. The missing frequency band higher than the narrow band typically ranges from 3.4 kHz (or 4 kHz) up to 7 kHz (or 8 kHz) and is referred to as the highband (HB). The missing frequency band below the narrow band typically covers frequencies below 300 Hz and is referred to as the lowband (LB). Bandwidth extension methods towards high frequencies create new content in the frequency band from 3.4 kHz (or 4 kHz) to 7 kHz (or 8 kHz). There are also bandwidth extension methods towards low frequencies, 50-300 Hz. The LB effects the quality and speech naturalness but has only a minor influence on intelligibility, whereas the HB influences the speech intelligibility and quality.

Bandwidth extension can be implemented in the frequency domain or in the time domain. Frequency domain processing commonly computes the spectrum of the input frame using Fast Fourier transform (FFT), constructs the spectrum of a frame in the spectral domain, and applies the inverse FFT to produce the corresponding time domain signal. Time domain processing typically involves adaptive filters or a filter bank for the shaping of

the extension band spectrum. Frequency domain processing has the benefit of accessing the spectral representation directly, whereas time-domain processing has the benefit of yielding a lower overall delay and avoid potential degradations due to the limited accuracy of FFT computation especially on fixed point platforms.

The Paper is organized as follows. Sect II discusses ABWE techniques. Sect III presents the limitations of ABWE approach. Sect IV describes the BWE with additional information. BWE with embedded additional information is describes in Sect V. Sect VI presents the challenges of BWE with embedded additional information. Sect VI presents the applications of BWE techniques. Finally, Sect VII presents the conclusion remarks.

II. Artificial Bandwidth Extension

The ABWE methods can be classified into the following categories

- Model based techniques
- Non-Model- based techniques

A. Model based Techniques

A generalized model for ABWE proposed in the literature is shown in fig.1.Here, initially up sample the NB signal through interpolation. A feature set is computed from the up sampled NB signal. Then, the extracted features set are utilized to estimate parameters for highband shaping. These estimated parameters are used for analysis filtering of the interpolated narrowband signal to get NB linear predictive coding (LPC) residual signal. The extended excitation signal is obtained by extending the NB LPC residual signal to the missing frequency range (LB and HB). The LB and HB signals are obtained after shaping the excitation utilizing the estimated shaping parameters. Finally, the original speech signal (output signal) contains the bandlimited signal (NB) and an artificially generated missing frequency range signals.
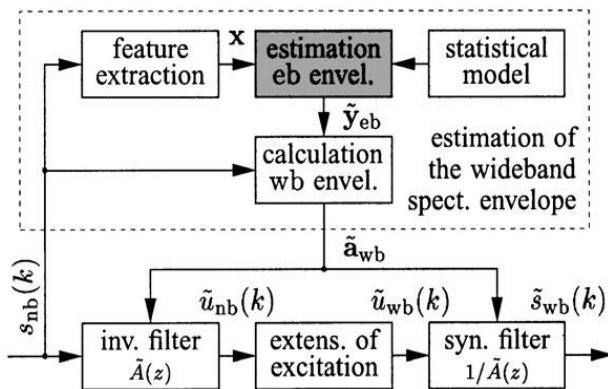

Fig.1. ABWE according to [1]

The model based techniques are based on a source filter model of speech production. This model consists of two parts, an excitation signal and a vocal tract filter. The

ABWE is divided into two individual tasks based on this model as:

1. Spectral envelope extension
2. Excitation signal extension

1. Spectral Envelope Extension

The extension of the spectral envelope from the narrowband to the highband is usually made based on some model that maps a narrowband feature vector onto the wideband envelope. The feature vector may consist of features that describe the shape of the envelope directly or some other features that are related to the temporal waveform.The existing techniques [2, 3] for estimating the highband spectral envelope are discussed here.

1.1 Codebook Mapping

The codebook mapping techniques make use of two codebooks and are constructed in training phase utilizing vector quantization to find a representative set of narrowband and wideband speech frames. A wideband codebook contains a number of wideband spectral envelopes that are parameterized and stored as, e.g., LPC coefficients or line spectral frequencies (LSFs). A narrowband codebook contains the parameters of the corresponding narrowband speech frames. When the bandwidth extension system is used, the best matching entry for each narrowband input frame is identified in the narrowband codebook and the spectral envelope of the extension band is generated using the corresponding entry in the wideband codebook.

This technique was proposed in [4], and presented in [5] with refinements. Separate codebooks were constructed for unvoiced and voiced fricatives in [5], which enhance the quality of spectral envelope. In addition, a codebook mapping with interpolation was reported to enhance the extension quality in [6]. Furthermore, in [7], codebook mapping with memory was implemented by interpolating the current envelope estimate with the envelope estimate of the previous frame.

1.2 Linear Mapping

Linear mapping is utilized in [8] to estimate the highband spectral envelope. The linear mapping between the input and output parameters (LPCs or LSFs) is denoted as

$$Y = WX \qquad (1)$$

Where the matrix W is obtained through an off-line training procedure with the least-squares approach that minimizes the model error (Y−WX) using a training data with NB parameters is denoted by a vector x = [x1, x2, . . . , xn] and corresponding wideband envelope to be estimated by another vector Y = [y1, y2 . . . , ym].

To better reflect the non-linear relationship between the NB and HB envelopes, modifications for the basic linear mapping technique have been presented. The mapping can be implemented by several matrices instead of using a single mapping matrix. In [9], input data are

clustered into four clusters, and for every cluster a separate mapping matrix is created. In [8], Objective analysis was showing that spectral distortion (SD) for codebook mapping and neural network was larger than for piecewise linear mapping. On the other hand, the objective comparison given in [6] indicates the performance of codebook mapping is better than linear mapping.

1.3 Gaussian Mixture Model

In linear mapping, only linear dependencies between the NB spectral envelope and the HB envelope are exploited. Non-linear dependencies can be included in the statistical model by utilizing Gaussian mixture model (GMM). A GMM approximates a probability density function as a sum of several multivariate Gaussian distributions, a mixture of Gaussians. GMMs are used in ABWE to model the joint probability distribution of the parametric representations of the NB input and the extension band. Given the input feature vector of a speech frame, the parameters representing the extension band envelope are determined using an error estimation rule.

The GMM is utilized directly in envelope extension to estimate WB LPC coefficients or LSFs from corresponding NB parameters [10]. The performance of the GMM based spectral envelope extension was then enhanced by using Mel frequency cepstral coefficients (MFCCs) instead of LPC coefficients [11]. The GMM mapping with memory further results in better performance in terms of log spectral distortion (LSD) and perceptual evaluation of speech quality (PESQ) [12].The advantage of the GMM in envelope extension methods is that they offer a continuous approximation from NB to WB features compared to the discrete acoustic space resulting from vector quantization (VQ). Better results were reported for GMM-based methods compared to codebook mapping in [10] in terms of SD, cepstral distance, and a paired subjective comparison.

1.4 Hidden Markov Model

Hidden Markov model (HMM) is used to estimate the high band spectral envelope. Each state of the HMM represents a typical extension band spectral envelope. The model also contains state probabilities, transition probabilities between the states, and state specific models for the probability density function of the input feature vectors. Each probability density function is approximated with a GMM. Given a sequence of input features, the aposteriori probability of each state is calculated from the observation probabilities and the state transition probabilities. The parameters representing the extension band envelope are determined using an error estimation rule. In particular, information in preceding frames is taken into account by the state transitions.

The advantage of HMM is its capability of implicitly exploiting information from the preceding frames to improve the estimation quality. Each state of the HMM represents a typically highband spectral envelope, so the change of the state of the HMM defines the envelope change of speech. HMM offers better results with lower order in comparison with GMM with even higher order.

HMM based ABWE was proposed in [13].Training of the HMM aided by phonetic transcriptions was introduced in [14].The Baum-Welch training algorithm was utilized in [15] and shown to outperform the one presented in [13].

1.5 Neural Networks

In [16], proposed neural networks to estimate the highband spectral envelope. The principle of an artificial neural network is inspired by the mechanisms of neural processing in the brain. An estimation task is performed by a large number of interconnected neurons, each of which computes a single output from several inputs by applying a simple linear or nonlinear function to the weighted sum of the inputs. The neurons are typically arranged in a layered structure where each neuron receives inputs from the previous layer and provides outputs to the following layer. The neural network is usually composed of a small number of such layers. Networks with only forward connections between layers are called feedforward networks, whereas networks containing feedback loops to preceding layers are known as recurrent networks. The training of a neural network involves finding suitable weights for the connections between the neurons in a fixed network topology. Neural networks are commonly trained using a large set of training data and the back-propagation algorithm. In [17], a neural network technique was compared with a codebook mapping technique, and attained better cepstral distortion and log spectral distortion. The computational complexities of code book methods are higher than those of neural network methods.

2. Excitation Signal Extension

The existing techniques [2, 3] for excitation signal extension are discussed here.

2.1 Spectral Folding

Spectral folding is a simple way to extend the excitation signal of a narrowband speech signal from the frequency range 0–4 kHz to 0–8 kHz. In spectral folding, a mirror image of the bandlimited (NB) signal spectrum in the missing frequency range is generated [18]. It has the same effect as up sampling the signal by two. Excitation extension can be implemented by mirroring the FFT coefficients in the frequency domain or by inserting a zero sample after each input sample in the time domain. Spectral folding which preserves the input spectrum is computationally simple, and the resulting temporal structure in the highband follows that of the input signal. However, the spectral folding technique has some drawbacks [19].Due to the limited telephone bandwidth of 300–3400 Hz there is a spectral gap at around 4 kHz. Also, the harmonic structure is not preserved in the highband.

2.2 Modulation Technique

A translated spectrum of the up sampled NB spectrum

can be produced in the extension band. High pass filtering is used to prevent the overlapping between the both spectrums (NB spectrum and the translated spectrum). The spectral translation with a modulation frequency can be equal to 4 kHz described in [18]. Other modulation frequencies can also be used and allow, e.g., filling the gap in the spectrum at 4 kHz [13]. A drawback of the modulation techniques with a fixed frequency is that the harmonic structure is not preserved in the HB.

2.3 Pitch Adaptive Modulation

Adaptive modulation frequency is used to preserve the excitation signal harmonic structure .The modulation frequency should be multiple of the fundamental frequency of speech (pitch). Pitch adaptive modulation requires an accurate estimate of fundamental frequency. Pitch detection is a non-trivial task for which there are several basic methods and a number of modifications proposed to improve robustness and accuracy. For example, an advanced pitch estimation algorithm was proposed in [19]. The benefit of the pitch-adaptive modulation approach has been found to be small compared to the complexity of the technique

2.4 Sinusoidal Synthesis

The harmonic structure of the excitation signal can be preserved by using a sinusoidal synthesis [20]. The harmonic structure for the HB is produced by a bank of oscillators with amplitudes, frequencies, and phases that are found from the NB speech. Spectral flattening is not needed in sinusoidal synthesis, since the spectral envelope can be created directly through sinusoidal amplitudes. Sinusoidal synthesis is especially suitable for the excitation extension to low frequencies below 300 Hz because the lowband signal primarily consists of a small number of harmonics of fundamental frequency and the ear is sensitive to the harmonic components in this frequency range.

2.5 Non-Linear Processing

The excitation signal can be extended by applying a nonlinear function to the narrowband excitation. The non-linear functions include a half wave rectification, a full wave rectification, a quadratic function, and a cubic function. Nonlinear processing has the advantage of maintaining the harmonic structure of the excitation signal, i.e., nonlinear processing of a periodic signal generates a spectrum with spectral peaks at integer multiples of fundamental frequency. However, due to the non-linearity, the energy level generated in the extension band is difficult to control, and required subsequent energy normalization [17]. Nonlinear functions are often used for the low-frequency extension because they generate a harmonic spectrum, which is perceptually important at low frequencies.

2.6 Noise Modulation

An excitation signal can be extended by applying noise modulation is motivated by the fact that the harmonic structure becomes more noisy above 4 kHz. Furthermore,

above 4 kHz the frequency resolution of human ear is poor and the pitch harmonics are not resolved individually, but the pitch periodicity is maintained by band pass speech signal temporal envelope [21].Therefore, the harmonic spectrum does not need to be reproduced and the excitation can be extended by modulating HB noise with the temporal envelope of the band pass speech signal (2.5–3 kHz) [22].

2.7 Noise Excitation

A noise signal has also been used as an excitation in combination with other techniques to avoid an overly periodic excitation at high frequencies [23] or to provide an excitation for unvoiced speech sounds [24]. For the extension from the wideband frequency range to the super-wideband range, a noise excitation may be sufficient, especially if the temporal envelope of the excitation is adjusted.

2.8 Voice Source Modeling

In [25], proposed estimating the wideband voice source signal from the narrowband signal to extend the excitation of voiced speech. The technique was found to be especially effective in the low frequency range. Furthermore, the bandwidth extension layer in the ITU-T G.729.1 codec utilizes a lookup table of glottal pulse shapes to reconstruct the excitation for voiced speech

B. Non-Model-Based Techniques

BWE without make use of source filter model was proposed in the literature. In [26], the speech signal was modelled as the addition of N consecutive amplitude and frequency Modulation (AM-FM) signals. The lost frequencies are determined using error estimation rule. In [27], proposed a method based on convolutive non-negative matrix factorization. A set of wideband non negative bases determined by using recordings of original speech signal (WB) and bandlimited (NB) speech signals from a target talker. In [28], proposed a method that maps each telephony band speech feature to the same feature of missing frequency band and uses a model for missing frequency band. By using the statistical characteristics of missing frequencies, the correlation between telephony band speech features and missing frequency band model is determined. In [29], proposed a method that operates in the modified discrete cosine transform domain (MDCT).

III. Limitations of Abwe

Though the informal listening tests [30] show preference of ABWE systems compared to the conventional NB telephony, its performance is less compared to that of the original WB speech. The better results for ABWE have been obtained for systems trained for an individual speaker and then for a specific language. But, in both the cases, the quality of ABWE technique is not better than the original WB speech. An information theoretic measure on the correlation between the telephonyband and the HB supports this observation [31].

IV. BWE WITH ADDITIONAL INFORMATION

BWE techniques utilizing a small amount of additional transmitted information about the missing frequency range, the process is known as BWE with additional (side) information [32, 33]. The additional information transmitted over a separate channel enables a more accurate estimation of the missing frequencies than ABWE. A generalized model for BWE with additional information is illustrated in fig.2. The additional information when transmitted through a separate channel compromises the backward compatibility of the communication network. To solve the backward compatibility problem by embedding the additional information as a digital watermark into the NB speech.
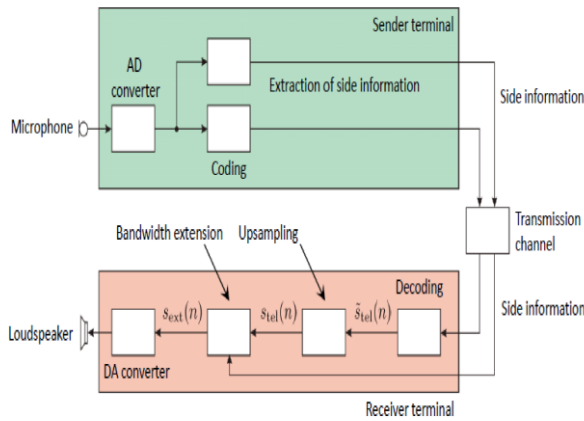

Fig.2. BWE with additional information

V. BWE WITH EMBEDDING ADDITIONAL INFORMATION

A backward compatible WB speech communication system can be achieved using BWE with watermarking technology. Digital watermarking is art of hiding data in the certain host signal. The system architecture for BWE with embedding additional information transmission is shown in Fig.3.Here, after passing the signal through low pass filter, the speech signal is down sampled through decimation. The host signal for the watermark is the down sampled signal. In parallel, BWE encoder analyzes the missing frequency band (HB) spectral envelope of
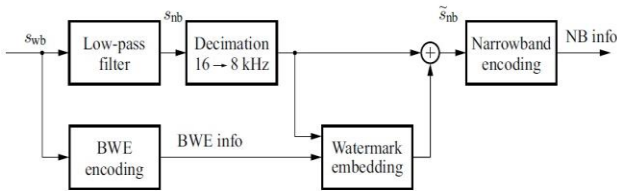

Fig.3. BWE with embedded additional information transmission according to [34]

wideband input signal and determines a set of corresponding parameters, and produces BWE information. The input to the watermarking encoder is the

BWE information. This BWE information is encoded by watermarking encoder and added to the host signal.

The system architecture for BWE with embedding additional information reception is shown in Fig.4. Here, the watermarked NB speech signal is decoded by NB decoder. The telephony band (NB) signal is up sampled through interpolation. In parallel, the watermark detector reconstructs the hidden message (BWE information) by analysing the signal mixture. The BWE information is decoded, and produces the linear prediction filter coefficients and gain factor of missing frequency band (HB) signal. The WB signal is produced by adding the missing frequency band (HB) signal to telephony band (NB) signal.
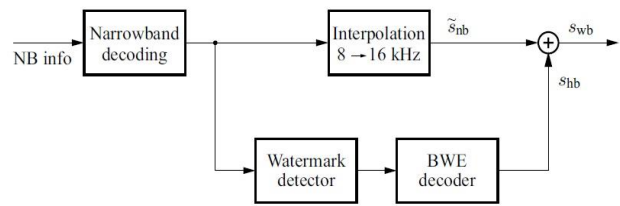

Fig.4. BWE with embedded additional information reception according to [34]

If interoperability with existing standard codecs is required, data transmission of BWE with embedding additional information can be accomplished in three different ways by embedding the additional (side) information in the samples of speech signal itself, by modifying the encoded bit stream, or most efficiently by joint coding and data hiding within the encoder. All these approaches maintain interoperability with the existing networks and codecs but may cause a minor degradation in speech quality. The signal domain approach is problematic because it is not sufficiently robust when coding with code excited linear prediction (CELP) codecs. The third approach of including the data hiding within the encoding process provides the best results but requires codec dependent processing.

In [35], according to auditory masking principle, the imperceptible components of the NB signal can be removed. A concealed transmission medium provided by this removal process. The audible components outside the telephone bandwidth are embedded into this hidden channel by using spread spectrum technique and then transmitted. The hidden signal can be extracted at the receiver and results a better perceptual speech signal.In [36], a frame of missing frequency band (HB) signal is encoded into 23 bits and embed these bits unnoticeably into 80 telephone band (NB) speech bit streams. Due to the distortion introduced by 23 bits, there is deterioration in the NB speech quality. In [37], a frame of missing frequency band (HB) signal is encoded into 12 bits by using acoustic phonetic classification and results a significant improvement in NB speech quality than in [36].

In [38], the missing frequency band (HB) components parameters are extracted and compressed. Embed these parameters unnoticeably into telephone band (NB) speech

bit steam. The missing frequency band (HB) components are extracted in the decoder. Finally, the original speech signal (output signal) contains the bandlimited signal (NB) and an artificially generated HB signal.In [39], the signal spectrum is divided into sub bands. The data embedding parameters are determined for each adaptively selected sub band. According to the principles of the scalar costa scheme (SCS), alter the discrete Hartley transform (DHT) coefficients in order to embed data.

VI. CHALLENGES OF BWE WITH EMBEDDING ADDITIONAL INFORMATION

One of the challenges of a watermarking system is to provide a virtually inaudible watermark with high data rate. For consistent results, the embedding additional information requires a   data rate of at least 100–300 bits/sec for every frame of input speech signal. Higher data rates produce better results. Another challenge is that the watermarking system must be robust to withstand the strong signal distortion introduced by transmission medium.

VII. APPLICATIONS

The most important application of speech BWE is to improve the telephone speech quality and intelligibility. Speech BWE can be utilized in a hands-free system for the car environment [40]. The recognition of WB speech is easier for cochlear implant users than the recognition of NB telephone speech [41]. In [42], it was shown that significant enhancement in speech recognition for cochlear implant users by utilizing highband BWE. The other application of speech BWE is to enhance the WB speech. In [43], proposed a method where the traditional speech enhancement methods [44] are used to process the NB and the HB is constructed with BWE.

Narrowband speech features of BWE can be utilized in automatic speech recognition (ASR). In [45], it was shown that the BWE of ASR features can be successfully used to train wideband ASR systems with narrowband speech, which is beneficial if the amount of WB training speech is insufficient. In this application, no audio signal needs to be generated from the bandwidth extended feature representation. The BWE techniques can also be utilized in WB speech coding.

VIII. CONCLUSIONS

ABWE is used to enhance the speech quality and intelligibility of bandlimited signal (NB speech) thus the enhanced speech signal reducing the listening effort. One of the earliest ABWE techniques estimates the parameters of a source filter model utilizing only the information available in NB signal. The theoretical performance upper bound on this estimation of ABWE is improved by using additional information extracted from the original speech signal. The additional information when transmitted through a separate channel compromises the

backward compatibility of the communication network. Further works done using BWE with embedding additional (side) information solves the backward compatibility problem by embedding the additional information as a digital watermark into the NB speech.

REFERENCES

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

[2] Laura Laaksonen, "Artificial bandwidth extension of narrowband speech - enhanced speech quality and intelligibility in mobile devices," Ph.D. dissertation, Aalto University, Finland, 2013.

[3] Hannu Pulakka, "Development and evaluation of artificial bandwidth extension methods for narrowband telephone speech," Ph.D. dissertation, Aalto University, Finland, 2013.

[4] Y.Yoshida and M.Abe, "An algorithm to reconstruct wideband speech from narrowband speech based on codebook mapping," In *Pro. of ICSLP*, pages 1591–1594, 1994.

[5] Y.Qian and P.Kabal, "Wideband speech recovery from narrowband speech using classified codebook mapping,"In *Pro.of AICSST*, pages 106–111, Australia, 2002.

[6] J.Epps and W.H.Holmes, "A new technique for wideband enhancement of coded narrowband speech," In *Pro. of the IEEE Workshop on Speech Coding*, pages 174–176, Finland, 1999.

[7] R.Hu et al., "Speech bandwidth extension by improved codebook mapping towards increased phonetic classification," In *Pro.of Interspeech*, pages 1501–1504, Portugal, 2005.

[8] Y.Nakatoh et al., "Generation of broadband speech from narrowband speech using piecewise linear mapping," In *Proc.of EUROSPEECH, Pages 1643-1646, Greece, 1997.*

[9] S.Chennoukh et al., "Speech enhancement via frequency bandwidth extension using line spectral frequencies," In *Proc.of ICASSP*, Pages 665-668, USA, 2001.

[10] K.Y.Park and H.S.Kim, "Narrowband to wideband conversion of speech using GMM based transformation," In *Proc. of ICASSP*, pages 1843–1846, Turkey, 2000.

[11] A.H.NourEldin and P.Kabal, "Mel frequency cepstral coefficient based bandwidth extension of narrowband speech," In *Proc.of Interspeech*, pages 53–56, Australia, 2008.

[12] A.H.NourEldin and P.Kabal, "Combining frontend based memory with MFCC features for bandwidth extension of narrowband speech," In *Proc.of ICASSP*, pages 4001–4004, Taiwan, 2009.

[13] P.Jax and P.Vary, "On artificial bandwidth extension of telephone speech," Signal Process., pages 1707–1719, 2003.

[14] P.Bauer and T.Fingscheidt, "An HMM based artificial bandwidth extension evaluated by cross-language training and test," *In Proc.of ICASSP*, pages 4589–4592, USA, 2008.

[15] G.B.Song and P.Martynovich, "A study of HMM based bandwidth extension of speech signals," Signal Process., pages 2036–2044, 2009.

[16] D.Zaykovskiy and B.Iser, "Comparison of neural networks and linear mapping in an application for bandwidth extension," *In pro.of SPECOM*, Greece, 2005.

[17] B.Iser and G.Schmidt, "Neural networks versus codebooks in an application for bandwidth extension of

speech signals," In *Proc. of EUROSPEECH, pages* 565–568, Switzerland, 2003.

[18] J.Makhoul and M.Berouti, "High-frequency regeneration in speech coding systems," in *Proc.of ICASSP*, pages 428–431, USA, 1979.

[19] A.deCheveigne and H.Kawahara, "YIN, a fundamental frequency estimator for speech and music," *J. Acoust. Soc. Am.*, pages 1917–1930, Apr.2002.

[20] C.F.Chan and W.K.Hui, "Wideband resynthesis of narrowband CELP coded speech using multiband excitation model," In *Proc.of ICSLP*, pages 322–325, USA, 1996.

[21] J. Epps, "*Wideband Extension of Narrowband Speech for Enhancement and Coding*," Ph.D. dissertation, School of Electrical Engineering and Telecommunications, The University of New South Wales, 2000.

[22] Y.Qian and P.Kabal, "Dual-mode wideband speech recovery from narrowband speech," In *Pro.of EUROSPEECH*, pages 1433–1436, Switzerland, 2003.

[23] M. Nilsson and W. B. Kleijn, "Avoiding over-estimation in bandwidth extension of telephony speech," in *Proc.of ICASSP*, pp. 869–872,USA, May 2001.

[24] J. Epps and W. H. Holmes, "Speech enhancement using STC based bandwidth extension," *in Proc. of ICSLP*, pp. 519–522, Australia, Nov. 1998.

[25] 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.of ICASSP*, pp. 4794–4797, USA, Mar. 2010.

[26] H.Tolba, "On the application of the AM-FM model for the recovery of missing frequency bands of telephone speech," In *Proc.of ICSLP*, Australia, 1998.

[27] D.Bansal et al., "Bandwidth expansion of narrowband speech using non-negative matrix factorization," *In Proc. Interspeech*, pages 1505–1508, Portugal, 2005.

[28] H.Gustafsson et al., "Low-complexity feature mapped speech bandwidth extension," *IEEE Trans. Audio, Speech, Language Process.*, pages 577–588, 2006.

[29] N.I.Park et al., "Artificial bandwidth extension of narrowband speech signals for the improvement of perceptual speech communication quality," *In Proc.of FGCN*, pages 143–153, Korea, 2011.

[30] J.A.Fuemmeler et al., "Techniques for the regeneration of wideband speech from narrowband speech," *EURASIP J. Appl.Signal Process.*, pages 266–274, 2001.

[31] P.Jax and P.Vary, "An upper bound on the quality of artificial bandwidth extension of narrowband speech signals," *In Proc. of ICASSP*, pages 237–240, USA, 2002.

[32] P. Jax and P. Vary, "Bandwidth extension of speech signals: A catalyst for the introduction of wideband speech coding?," *IEEE Commun.Mag.*, vol. 44, no. 5, pp. 106–111, 2006.

[33] B. Geiser et al., "Artificial Bandwidth Extension of Speech Supported by Watermark-Transmitted Side Information," *in Proc. Of INTERSPEECH*, Lisbon, Portugal,Sept. 2005.

[34] P.Jax, "Backwards Compatible Wideband Telephony," Chapter 9, Advances in Digital Speech Transmission, Eds., Wiley, 2008.

[35] Siyue Chen et al., "Telephony speech enhancement by data hiding," *IEEE Trans.instrumentation and measurement*, pages 63-74, 2007.

[36] S Chen and H Leung, "Artificial bandwidth extension of telephony speech by data hiding," in *pro.of ISCAS*, pages 3151–3154, Kobe, May 2005.

[37] S Chen and H Leung, "Speech bandwidth extension by data hiding and phonetic classification," in *Proc.of*

ICASSP, pages 593-596, 2007.

[38] Zhe Chen et al., "An audio watermark based speech bandwidth extension method," *in pro.of EURASIP J.audio, speech and music processing,* 8 pages, 2013.

[39] A.Sagi and D.Malah, "Bandwidth extension of telephone speech aided by data embedding," *in pro.of EURASIP J. Advances in Signal Process.*, 16 pages, 2007.

[40] B.Iser and G.Schmidt, "Receive side processing for automotive hands free systems," in *Proc. of HSCMA*, pages 236–239, Italy, 2008.

[41] Y.Hu and P.C.Loizou, "Effects of introducing low frequency harmonics in the perception of vocoded telephone speech," *J. Acoust. Soc. Am.*, pages 1280–1289, 2010.

[42] C.Liu et al., "Effect of bandwidth extension to telephone speech recognition in cochlear implant users," *J. Acoust. Soc. Am.*, pages EL77–EL83, 2009.

[43] F.Mustiere et.al, "Bandwidth extension for speech enhancement," in *Proc.of CCECE,* 4 pages*, Canada*, 2010.

[44] V.Sunny dayal et al., "A Survey On Statistical Based Single Channel Speech Enhancement Techniques," In *Proc.of IJISA* ,pages 69–85, 2014.

[45] D.Macho, "Narrowband to wideband feature expansion for robust multilingual ASR," *In Proc.of Interspeech*, pages 1118–1121, Belgium, 2007.

Authors' Profiles

N.Prasad received the B.E Degree in E.C.E from Andhra University, Vizag, India, in 2006, and M.Tech degree in Electronic instrumentation from NIT Warangal, India in 2009. He is currently a research scholar with the department of E.C.E, NIT Warangal, India .His research interest is speech processing.

T. Kishore Kumar received the B.Tech degree in E.C.E from Sri Venkateswara University, Tirupati, India, in 1992, M. Tech degree in Digital Systems and Computer Electronics from J.N.T.U.Hyderabad, India in 1996, and Ph.D. degree in E.C.E from J.N.T.U.Hyderabad in 2004. He is currently Associate Professor and Head of the department of E.C.E, NIT Warangal, India. His research interests include Real Time Embedded Systems, Real Time signal processing and speech processing. He is currently working on two R&D projects sponsored by DST and DRDO worth about 28 lakhs. He is also developing Course Material for UG under Pedagogy in collaboration with IIT Kharagpur. He has awarded three Ph.D scholars in NITW, at present 4 students are pursuing PhD under him. He holds 22 publications in last five years. He is a member for Central Council Board (CCB), AIEEE. He has been invited as Chair person for International Conference on Signal Processing held at Czech Republic in 2008 and deputed for R & D Project on Real Time Embedded Systems in Signal Processing at University of Escigelec, Rouen, FRANCE in 2009. He is well travelled and has visited Tel Aviv, Singapore and Dubai to attend international conferences. He served the Tallinn University of Technology, Estonia as Visiting Professor during October and November 2013 under Heritage (INDO EUROPE Exchange Program).He has organized several MHRD Programs

on Signal Processing at NIT WARANGAL for Faculty Development Program. He has delivered Guest Lectures at JNTUH and other institutes in AP. He is a member of IEEE and fellow of IETE.