Speech Bandwidth Extension Using Spectral Data Masking

Nizampatnam Prasad(✉)

Department of ECE, Faculty of Science and Technology (IcfaiTech), ICFAI Foundation for Higher Education, Hyderabad, Telangana, India
prasadnit@ifheindia.org

Abstract. Voice quality and understandability are both diminished by the telephone systems' limited voice bandwidth. The innovative speech bandwidth enhancement approach is suggested in this paper. To deliver a superior broad-band voice signal, the methodology employs a spectral data masking approach. The restricted band speech signal conceals the missing speech's code-excited linear prediction (CELP) parameters. At the other end, the missing speech is retrieved to create a broad-band voice signal. The suggested approach is resilient to quantization and channel disturbances, according to theoretical and simulation evaluations. The obtained findings support the suggested method's superior reconstructed broad-band speech quality to conventional approaches.

Keywords: Spectral data masking approach · Telephone networks · Speech quality · Speech bandwidth extension

1 Introduction

Due to the telephone network's constrained bandwidth, when a human speech signal is transferred across it, speech data is lost. Due to historical financial considerations, the telephone voice is confined to, around 0–4000 Hz, known as restricted band (RB) speech. This leads to a thin, unnatural-sounding voice and reduced speech signal's naturalness. 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 [1].

A broad-band (BB) speech which has a frequency range of 0-8kHz will improve the quality of the spoken signal as opposed to RB speech. As a result, new telephone networks (TNs) with higher bandwidths must be built, which will be costly and take significant time to establish [1]. By applying speech bandwidth enhancement (SBE) techniques, the quality of the received signal will be improved without having to make changes to the current TNs architecture.

Artificial SBE (ASBE) is one approach that reconstructs the BB signal only from the RB signal by predicting losing speech information (4 kHz above) [2]. The dependency between RB speech and missing band (MB) speech revealed by the speech production

P. Pareek et al. (Eds.): IC4S 2024, LNICST 598, pp. 84–93, 2025.
https://doi.org/10.1007/978-3-031-77078-4_7

mechanism is the basis for ASBE approaches. Several ASBE techniques have been presented to date. The ASBE is divided into two distinct subtasks: expanding the excitation and expanding the spectral envelope. In [3], a number of methods for excitation extension are explored. In [3–7], various approaches to find BB envelope are presented. The use of RB speech input to estimate MB speech information is supported by enough evidence. However, the intrinsic performance limitations of ASBE approach prevent them from regenerating high-quality BB speech [8].

MB information will be provided together with an RB signal to increase the quality of BB speech significantly compared to ASBE approaches [1]. Data concealing approaches are employed to hide supplementary MB information in RB signal in order to guarantee required backward compatibility with current TNs. In the literature, numerous solutions to this issue have already been put forth. An SBE technique was suggested in [9]. This method was shown to offer poor reconstructed BB (RBB) signal and mixed RB (MRB) signal quality. In order to more effectively encode the MB signal, phonetic categorization was utilized in [10] to enhance the quality of MRB and RBB signals over [9]. It was suggested in [11] that SBE use a least significant bit embedding approach. It was suggested in [12] to SBE using quantization. The idea of SBE by information concealing was put out in [13], where a secret channel is made by eliminating undetectable elements from the RB signal. Regenerating hidden audible components that are present in the RB allows for the reconstruction of high-quality BB speech. In [14–17], a number of information embedding-based methods for SBE are presented.

The existing processes [9–17] failed to deliver high-quality RBB signal along with less vigor towards channel and quantization noises (CQNs). To overcome these drawbacks in the contemporary SBE processes, a novel SBE using a spectral data masking (SDM) process for extending the bandwidth of TNs is proposed. SDM process is reported in [18] for entrenching the CELP parameters of secrete signal in cover signal detailed Wavelet coefficients. It was found that the technique generated a stego signal that was nearly identical to the cover signal and could precisely restore the secret signal. A novel robust SBE process using SDM [18] is proposed to insert the CELP parameters of MB within the RB signal. To provide a higher-quality BB signal, this embedded information is retrieved at the receiver side.

In this work, the impact of CQNs is examined. The effects of quantization noise are presented in [9] and [10]. However, [9] and [10] did not evaluate the effect of channel noise. The latest invention employs CDMA approach, which is marketed as being resilient against CQNs, to reproduce the secret information. Particularly, every bit of information ensconced in the RB signal is spread with a certain SS. Then, dispersed signals were summed together to produce the hidden information. Since there is little correlation between the spreading sequence (SSs), it is possible to reliably retrieve the secret information.

The remaining portions of the paper are divided into four groups. SDM technique for SBE was introduced in Sect. 2. In Sect. 3, a novel SBE via the SDM method is provided. In Sect. 4, the experimental analyses are covered. Conclusions are provided in Sect. 5.

2 Spectral Data Masking Technique for Speech Bandwidth Extension

Consider the MB signal $A_{mb}(n)$ as being masked by the RB signal $A_{lb}(n)$. The RB signal $A_{lb}(n)$ is first subjected to DWT to separate it into precise and approximate coefficients. The spectrum is then computed by applying FFT to the detailed coefficients, and the magnitude and phase spectra are then calculated. Take the MB signal $A_{mb}(n)$, which is represented by the bits, i.e., $E_b \in \{-1, 1\}, b = 0, 1, ..., B-1$.

The spread of the bits is accomplished by multiplying each bit with a specific PN code i.e.,$E_b p^b$. These spreading vectors are added together to create hidden data which is represented as

$$V(m) = \sum_{b=0}^{B-1} E_b p^b(m) \tag{1}$$

When $V(m)$ is inserted into the last six components of the first half of the magnitude spectrum, a modified magnitude spectrum is created [18]. Apply an inverse FFT and then an inverse DWT on the modified magnitude spectrum to convert it back to MRB signal. The receiver end receives the resulting MRB signal $A^1_{lb}(n)$ through the TNs. Here the channel introduces CQNs. Let $\hat{A}^1_{lb}(n)$ represent the received signal, i.e., $\hat{A}^1_{lb}(n) = A^1_{lb}(n) + \ddot{e}$. The combination of CQNs is symbolized by $\ddot{e}$. $\hat{A}^1_{lb}(n)$ will be regarded by a standard phone terminal as a regular signal. $A_{lb}(n)$ quality is not much diminished since the observed changes between $A_{lb}(n)$ and $A^1_{lb}(n)$ are relatively small.

Bringing back the MB signal $\hat{A}_{mb}(n)$ requires a receiver to compute the precise coefficients by applying DWT to the MRB signal $\hat{A}^1_{lb}(n)$. DFT is then applied to the precise coefficients. Then, the magnitude spectrum [18] is used to retrieve the information bits, which are subsequently decoded using a multiuser detector [19], i.e.,

$$\widetilde{E_b} = sign\left(\sum_{m=0}^{M-1} \breve{V}_m {p_m}^b\right) \tag{2}$$

In an environment with no noise, $\breve{V}_m = V_m$. Replace it in Eq. (2) and we get

$$\begin{aligned} \widetilde{E_b} &= sign\left(\sum_{m=0}^{M-1} V_m {p_m}^b\right) \\ &= sign\left(\sum_{m=0}^{M-1}\left(E_m {p_m}^b {p_m}^b + \sum_{g=0, g\neq b}^{B-1} E_g {p_m}^g {p_m}^b\right)\right) \\ &= sign\left(M E_m + \sum_{g=0, g\neq b}^{B-1} E_g \sum_{m=0}^{M-1} {p_m}^g {p_m}^b\right) \end{aligned} \tag{3}$$

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

$$\sum_{g=0, g\neq b}^{B-1} E_g \sum_{m=0}^{M-1} p_m^g p_m^b = 0 \tag{4}$$

This demonstrates that by using the CDMA approach, the parameters that reflect the MB signal are successfully retrieved.

3 Speech Bandwidth Extension Utilizing Spectral Data Masking

3.1 Transmitter

Figure 1 depicts the transmitter. With BB speech $A_{bb}(n)$ sampled at 16 kHz, an LPF and an HPF divide it into lower-band (LB) and upper-band (UB) signals, respectively. The speech information in the LB signal ranges from 0 to 4 kHz, while the speech information in the UB signal ranges from 4 kHz to 8 kHz. The output of the LPF is then decimated to create the RB signal $A_{lb}(n)$. A missing-band (MB) signal $A_{mb}(n)$ is subsequently created by decimating the high-pass filter's output.

In order to calculate CELP parameters, CELP analysis [20] is performed on MB signal $A_{mb}(n)$, and then these parameters are modified. Using the fuzzy c-means (FCM) technique, quantize the modified CELP parameters to the nearest vector quantization (VQ) codebook entry [21]. The six-bit binary entry index representation, $E_0, E_{1,\ldots}E_{B-1}$, is cloaked in an RB signal based on the SDM approach to produce an MRB signal $A^1_{lb}(n)$, which is then transmitted by TNC to the destination.

Every frame of $A^1_{lb}(n)$ is followed by the introduction of a synchronisation sequence (SYSE), such as 1111….111, in order to achieve frame synchronisation [22] among transmission and reception.

3.2 Receiver

Figure 2 depicts the receiver. Use the SDM technique to correctly recover the entry index, and then use the VQ codebook to accurately recover the CELP parameters. The MB speech signal is then synthesised using these CELP parameters. The MRB signal $A^1_{lb}(n)$ and the synthesised MB speech $A^1_{mb}(n)$ are currently being sampled at 8000 Hz, after which these signals will be interpolated. To create a high-quality BB signal $A^1_{bb}(n)$, the interpolated MRB $A^{11}_{lb}(n)$ and MB $A^{11}_{mb}(n)$ signals are combined.

4 Experimental Results

Eighty sentences were taken from the TIMIT database and spoken by fifty male talkers and fifty female talkers for the performance evaluation [23]. The RB signal is split into frames with a 10 ms frame overlap and a 20 ms frame duration. The frames are then handled one by one. Performance is judged using both subjective and objective tests. The proposed methodology's suitability is investigated by comparison with current methods, including contemporary telephony speech bandwidth enhancement (CTSE) by data hiding (CTSEDH) [9], CTSE by phonetic classification (CTSEPC) [10], CTSE using bit stream data hiding (CTSEBDH) [11], and CTSE using watermark side information (CTSEWSI) [16]. AWGN and the μ-law are the channel models that are taken into consideration here.

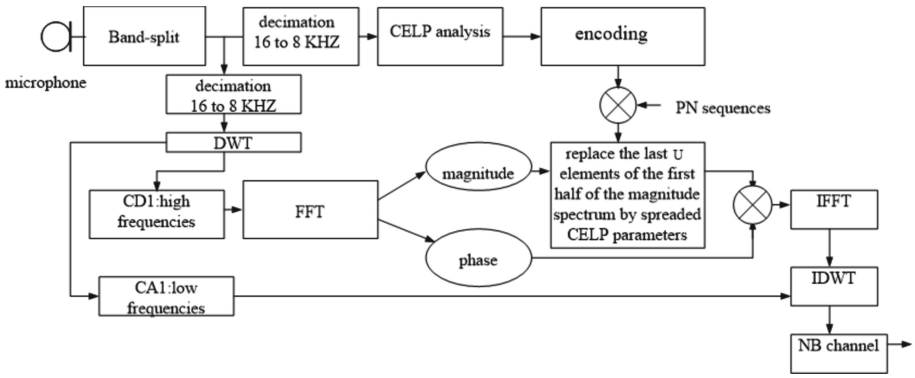

Fig. 1. Suggested transmitter

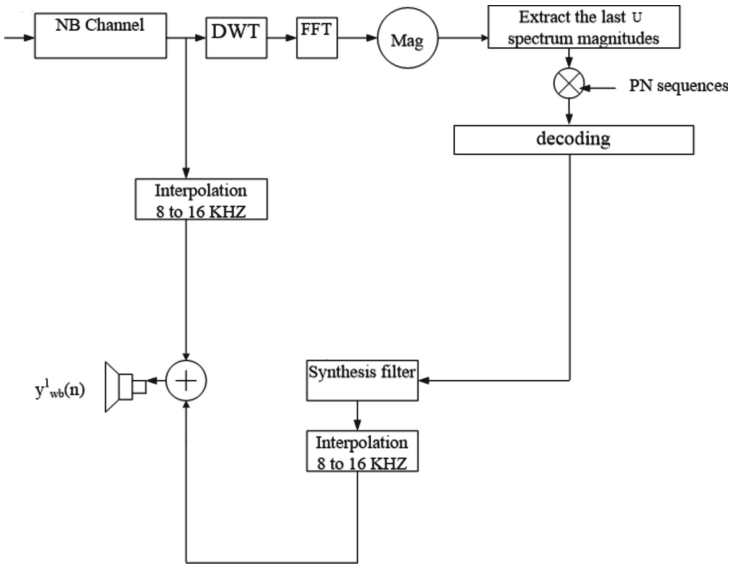

Fig. 2. Suggested receiver

4.1 Subjective Assessments

To assess the clarity of perception, the mean opinion score (MOS) test is performed [9, 10]. In the listening test, different speech signals including BB, RB, MRB, and RBB are compared [24]. In a silent space, headsets were used for these experiments. Seventy people are taken into consideration for each test.

4.1.1 Perceptual Clearness (PCL)

The information must be transparently concealed in the suggested method, meaning that RBand MRB signals cannot be told apart on a subjective level. Low perceptible

MRB signal degradation is indicated by a high PCL. The MOS test is used to evaluate PCL. Participants compare RB and MRB signals and then provide a MOS judgement, which is summarized in Table 1. In Table 2, the mean MOS results for the traditional [9–11, 16] and recommended approaches are shown. MOS data demonstrate the proposed approach's striking perceptual clarity when compared to the current methodologies in Table 2.

Table 1. MOS

Score	Instruction
1	RB and MRB signals are dissimilar
2	RB and MRB signals differ noticeably from one another
3	RB and MRB signals differ slightly from one another
4	RB and MRB signals are similar

Table 2. MOS assessment outcomes with μ-law coding

Technique	Mean opinion score
CTSEDH [9]	1.97
CTSEPC [10]	2.46
CTSEBDH [11]	2.87
CTSEWSI [16]	3.06
Proposed method	3.87

4.1.2 Subjective contrasts among BB, RB, MRB and RBB signals

Table 3 shows the BB signal, RB signal, MRB signal, and RBB signal as I, II, III, and IV, respectively. The subjects must determine whether the first signal is superior ($>$), deficient ($<$), or similar ($\approx$) to the second signal when doing a paired examination of signals from I to IV. The responses of I when compared pairwise to the other signals (II, III, and IV) are shown in Table 3. Arabic numbers are used in the table to indicate the number of participants who have a precise preference ($>$ or $<$ or $\approx$). For both established [9–11, 16] and new processes that are supported by Table 3, the BB signal outperforms the MRB signal. The proposed technique's RBB signal quality was further supported by Table 3 as being significantly better than that of the traditional techniques [9–11, 16].

4.2 Objective Assessments

LSD [9, 10]and BB-PESQ [25] tests are used to evaluate the RBB signal quality. The RB-PESQ test [26] is used to measure perceptual clarity. BER measure is used to assess how resilient concealed data is to CQNs.

Table 3. Subjective contrast with μ-law coding

Method	I	II	III	IV
CTSEDH [9]	$>$	70	70	18
	$<$	0	0	0
	$\approx$	0	0	52
CTSEPC [10]	$>$	70	70	15
	$<$	0	0	0
	$\approx$	0	0	55
CTSEBDH [11]	$>$	70	70	12
	$<$	0	0	0
	$\approx$	0	0	58
CTSEWSI [16]	$>$	70	70	9
	$<$	0	0	0
	$\approx$	0	0	61
Proposed method	$>$	70	70	1
	$<$	0	0	0
	$\approx$	0	0	69

4.2.1 RBB signal quality

In the BB-PESQ test, the BB and RBB signals are compared to assess the quality of RBB speech. Table 4 displays the mean BB-PESQ scores for the conventional [9–11, 16] and suggested approaches. The outcome of the suggested method is 4.34, which shows that the RBB signal quality achieved is remarkable. Thus, when compared to current processes, the suggested technique improved speech quality.

4.2.2 LSD for signal quality

LSD is a fairly accurate metric for evaluating the similarity between authentic and restored MB signals, and it is provided by

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

where $\frac{1}{a_s\left(e^{jw}\right)}$ and $\frac{1}{\hat{a}_s\left(e^{jw}\right)}$ are the spectrums, and g_p and $\hat{g}_p$ are gains of genuine and restored MB signals, respectively. The best MB signal reproduction generally has a low LSD value. The mean LSD scores for the current [9–11, 16] and suggested systems for μ-law coding are shown in Table 5. The numbers provided in Table 5 demonstrate a definite improvement in the proposed scheme's quality compared to current techniques [9–11, 16].

4.2.3 Perceptual Clearness

PCL is evaluated using the RB-PESQ test by comparing the RB signal to the MRB signal. The range of RB-PESQ is 0.5 to 4.5. The worsening PCL is represented by lower values like 0.5, while the best PCL is illustrated by higher values like 4.5. The replies of the mean scores for the current [9–11, 16] and suggested strategies are shown in Table 6. According to the results shown in Table 6, the proposed strategy outperforms more recent methods in terms of PCL.

4.2.4 Vigor of concealed data

MRB signal is used to sum up AWGN with SNR values between 15 and 35 dB [27]. Size 8 is the PN code. A high-quality RBB signal is indicated by a lower BER value. The BER values that were obtained is less than $5.67 * 10^{-4}$, which supports the outstanding quality of the RBB signal. The RBB signal is endorsed as being of high quality by the BER value of $3.75 * 10^{-4}$, which was achieved with μ-law coding.

Table 4. BB-PESQ test outcomes with μ-law coding

Technique	BB-PESQ
CTSEDH [9]	2.25
CTSEPC [10]	2.89
CTSEBDH [11]	3.03
CTSEWSI [16]	3.29
Proposed technique	4.34

Table 5. LSD test Outcomes with μ-law coding

Technique	Log spectral distortion
CTSEDH [9]	18.78
CTSEPC [10]	12.43
CTSEBDH [11]	9.54
CTSEWSI [16]	8.02
Proposed technique	1.76

Table 6. LB-PESQ test Outcomes with μ-law coding

Method	LB-PESQ
CTSEDH [9]	1.98
CTSEPC [10]	2.12
CTSEBDH [11]	2.34
CTSEWSI [16]	2.86
Proposed method	4.12

5 Conclusion

A brand-new SBE method based on SDM is suggested. The MB signal's spread spectral envelope parameters are masked by the RB signal's detailed coefficients. A top-notch BB signal is created by retrieving the secret data at the receiving end. The evaluation's findings support the suggested approach superior clear broad-band performance to the contemporary SBE processes.

References

1. Jax, P., Vary, P.: Bandwidth extension of speech signals: A catalyst for the introduction of wideband speech coding. IEEE Commun. Mag. **44**(5), 106–111 (2006)
2. Jax, P.: Enhancement of bandlimited speech signals: Algorithms and theoretical bounds. PhD Thesis. RWTH Aachen University, Aachen, Germany (2002)
3. Prasad, N., Kishore Kumar, T.: Bandwidth extension of speech signals: A comprehensive review. Int. J. Intell. Syst. Appl. **8**(2), 45–52 (2016)
4. Zhen-Hua, L., Yang, A., Yu, G., et al.: Waveform Modelling and Generation Using Hierarchical Recurrent Neural Networks for Speech Bandwidth Extension. IEEE/ACM Trans. Audio, Speech, and Language Process **26**(5), 883–894 (2018)
5. Bong-Ki, L., Kyoungjin, N., Joon-Hynk, C., et al.: Sequential Deep Neural Networks Ensemble for Speech Bandwidth Extension. IEEE ACCESS **6**, 27039–27047 (2018)
6. Abel, J., Fingscheidt, T.: A DNN Regression Approach to Speech Enhancement by Artificial Bandwidth Extension. In: Proceedings of the IEEE Workshop on Applications of Signal Processing to Audio and Acoustics, New Paltz, (Newyork, USA), pp. 219–223 (2017)
7. Yingwue, W., Shenghui, Z., Dan, Q., et al.: Using conditional restricted Boltzmann machines for spectral envelope modelling in speech bandwidth extension. In: Proceedings of the IEEE Int. Conf. Acoustics, Speech, and Signal Processing , Shanghai, China, pp. 5930–5934 (2016)
8. Jax, P., Vary, P.: An upper bound on the quality of artificial bandwidth extension of narrowband speech signals. In: Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing, Orlando, USA, pp. 237–240 (2002)
9. Chen, S., Leung, H.: Artificial bandwidth extension of telephony speech by data hiding. In: Proceedings of the IEEE International. Symposium on Circuits and Systems, Kobe, Japan, pp. 3151–3154 (2005)
10. Chen, S., Leung, H.: Speech bandwidth extension by data hiding and phonetic classification. In: Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing. Honolulu (Hawaii, USA), pp. 593–596 (2007)

11. Chen, Z., Zhao, C., Geng, G., et al.: An audio watermark based speech bandwidth extension method. EURASIP Journal on Audio, Speech and Music Processing **2013**(10), 1–8 (2013)
12. Sagi, A., Malah, D.: Bandwidth extension of telephone speech aided by data embedding. EURASIP Journal on Advances in Signal Processing **2007**(1), 37–52 (2007)
13. Chen, S., Leung, H., Ding, H.: Telephony speech enhancement by data hiding. IEEE Trans. Instrum. Meas. **56**(1), 63–74 (2007)
14. Geiser, B., Vary, P.: Speech bandwidth extension based on in band transmission of higher frequencies. In: Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing. Vancouver, Canada, pp. 7507–7511 (2013)
15. Geiser, B., Vary, P.: Backwards compatible wideband telephony in mobile networks: CELP watermarking and bandwidth extension. In: Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing. Honolulu, Hawaii, USA, pp. 533–536 (2007)
16. Bhatt, N., Kosta, Y.: A novel approach for artificial bandwidth extension of speech signals by LPC technique over proposed GSM FR NB coder using high band feature extraction and various extension of excitation methods. Int. J. Speech Technol. **18**(1), 57–64 (2015)
17. Bhatt, N.: 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. Int. J. Speech Technol. **19**(4), 881–893 (2016)
18. Rekik, S., Guerchi, D., Selouani, S.A., Hamam, H.: Speech steganography using Wavelet and Fourier transforms. EURASIP Journal on Audio, Speech, and Music Processing **2012**(20), 1–14 (2012)
19. Proakis, J.G.: Digital Communications, 2nd edn. McGraw-Hill, New York (1989)
20. Schroeder, MR., Atal, BS.: Code-excited linear prediction (CELP): high quality speech at low bit rates, In: Proceedings of ICASSP, Tampa, FL, USA, pp. 937–940 (1985)
21. Bezdek, J.C.: Pattern Recognition with Fuzzy Objective Function Algorithms. Plenum, New York (1981)
22. European Telecommunications Standards Institute (ETSI) Standard.: Speech Processing, Transmission and Quality Aspects (STQ); Distributed speech recognition; Front end feature extraction algorithm; Compression algorithms, ETSI ES 201 108 V1.1.2 (2000)
23. Garofolo, JS. Lamel, LF., Fisher, WM, et al.: Getting started with the DARPA TIMIT CD-ROM: An acoustic phonetic continuous speech database, National Institute of Standards and Technology (NIST). Gaithersburg, MD, USA. ISBN: 1–58563–019–5
24. Prasad, N., Kishore Kumar, T.: Speech bandwidth extension aided by spectral magnitude data hiding. Circuits Systems Signal Process. **36**(11), 4512–4540 (2017)
25. 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 super-wideband signals, ITU-T Recommendation P.863 (2011)
26. International Telecommunications Union. Perceptual evaluation of speech quality (PESQ): An objective method for end-to-end speech quality assessment of narrow-band telephone networks and speech codecs, ITU-T Recommendation P.862 (2001)
27. Keiser, BE., Strange, E.: Digital Telephony and Network Integration. Van Nostrand Reinhold, New York, ISBN 978–1–4615–1787–0 (1995)