International Telecommunication Union

ITU-T

TELECOMMUNICATION
STANDARDIZATION SECTOR
OF ITU

Series P

Supplement 27

(01/2017)

SERIES P: TELEPHONE TRANSMISSION QUALITY, TELEPHONE INSTALLATIONS, LOCAL LINE NETWORKS

Application of ITU-T P.863 and ITU-T P.863.1 for speech processed by blind bandwidth extension approaches

ITU-T P-series Recommendations – Supplement 27

ITU-T P-SERIES RECOMMENDATIONS

TELEPHONE TRANSMISSION QUALITY, TELEPHONE INSTALLATIONS, LOCAL LINE NETWORKS

Vocabulary and effects of transmission parameters on customer opinion of transmission quality	Series	P.10
Voice terminal characteristics	Series	P.30 P.300
Reference systems	Series	P.40
Objective measuring apparatus	Series	P.50 P.500
Objective electro-acoustical measurements	Series	P.60
Measurements related to speech loudness	Series	P.70
Methods for objective and subjective assessment of speech quality	Series	P.80
Methods for objective and subjective assessment of speech and video quality	Series	P.800
Audiovisual quality in multimedia services	Series	P.900
Transmission performance and QoS aspects of IP end-points	Series	P.1000
Communications involving vehicles	Series	P.1100
Models and tools for quality assessment of streamed media	Series	P.1200
Telemeeting assessment	Series	P.1300
Statistical analysis, evaluation and reporting guidelines of quality measurements	Series	P.1400
Methods for objective and subjective assessment of quality of services other than speech and video	Series	P.1500

For further details, please refer to the list of ITU-T Recommendations.

Supplement 27 to ITU-T P-series Recommendations

Application of ITU-T P.863 and ITU-T P.863.1 for speech processed by blind bandwidth extension approaches

Summary

Supplement 27 to the ITU-T P-series of Recommendations provides a method for the application of ITU-T P.863 to speech signals processed by a blind bandwidth extension (BBE), which is complementary to the existing procedures given in Recommendation ITU-T P.863.1. When bandwidth extension techniques are used, not only does the reference bandwidth need to be set but ITU-T P.863 also has a limited ability to discriminate small bit rate and bandwidth improvements. These quality differences are clearly distinguishable in subjective tests. For ITU-T P.863 tests, a complementing bandwidth requirement check is needed and detailed.

History

Edition	Recommendation	Approval	Study Group	Unique ID*
1.0	ITU-T P Suppl. 27	2017-01-19	12	11.1002/1000/13242

Keywords

Bandwidth extension, ITU-T P.863, ITU-T P.863.1.

---

* To access the Recommendation, type the URL http://handle.itu.int/ in the address field of your web browser, followed by the Recommendation's unique ID. For example, http://handle.itu.int/11.1002/1000/11830-en.

FOREWORD

The International Telecommunication Union (ITU) is the United Nations specialized agency in the field of telecommunications, information and communication technologies (ICTs). The ITU Telecommunication Standardization Sector (ITU-T) is a permanent organ of ITU. ITU-T is responsible for studying technical, operating and tariff questions and issuing Recommendations on them with a view to standardizing telecommunications on a worldwide basis.

The World Telecommunication Standardization Assembly (WTSA), which meets every four years, establishes the topics for study by the ITU-T study groups which, in turn, produce Recommendations on these topics.

The approval of ITU-T Recommendations is covered by the procedure laid down in WTSA Resolution 1.

In some areas of information technology which fall within ITU-T's purview, the necessary standards are prepared on a collaborative basis with ISO and IEC.

NOTE

In this publication, the expression "Administration" is used for conciseness to indicate both a telecommunication administration and a recognized operating agency.

Compliance with this publication is voluntary. However, the publication may contain certain mandatory provisions (to ensure, e.g., interoperability or applicability) and compliance with the publication is achieved when all of these mandatory provisions are met. The words "shall" or some other obligatory language such as "must" and the negative equivalents are used to express requirements. The use of such words does not suggest that compliance with the publication is required of any party.

INTELLECTUAL PROPERTY RIGHTS

ITU draws attention to the possibility that the practice or implementation of this publication may involve the use of a claimed Intellectual Property Right. ITU takes no position concerning the evidence, validity or applicability of claimed Intellectual Property Rights, whether asserted by ITU members or others outside of the publication development process.

As of the date of approval of this publication, ITU had not received notice of intellectual property, protected by patents, which may be required to implement this publication. However, implementers are cautioned that this may not represent the latest information and are therefore strongly urged to consult the TSB patent database at http://www.itu.int/ITU-T/ipr/.

Table of Contents

Introduction

Recently the industry has started to move from narrowband speech coders (NB) to wideband (WB) or super-wideband (SWB) coders. However, until complete coverage has been achieved, a significant proportion of calls will still use legacy narrowband. Even then, calls from landlines will likely still be narrowband for some time.

Blind bandwidth extension (BBE) technology aims to solve this problem, by transforming NB speech into WB or SWB speech. A requisite to successful deployment of BBE technology is having a good evaluation methodology. In this document, we propose that ITU-T P.863 in conjunction with a bandwidth requirement is a suitable methodology for BBE performance evaluation.

Supplement 27 to ITU-T P-series Recommendations

Application of ITU-T P.863 and ITU-T P.863.1 for speech processed by blind bandwidth extension approaches

1 Scope

Supplement 27 to the ITU-T P-series of Recommendations provides a method for the application of [ITU-T P.863] to speech signals processed by a blind bandwidth extension (BBE), which is complementary to the existing procedures given in [ITU-T P.863.1]. This Supplement provides an evaluation of speech quality using [ITU-T P.863] for bandwidth extension, when the bandwidth speech under evaluation is wider than the original speech content and cannot be directly related to the input signal.

2 References

[ITU-T P.501] Recommendation ITU-T P.501 (2012), *Test signals for use in telephonometry*.

[ITU-T P.800] Recommendation ITU-T P.800 (1996), *Methods for subjective determination of transmission quality*.

[ITU-T P.863] Recommendation ITU-TP.863 (2014), *Perceptual objective listening quality assessment*.

[ITU-T P.863.1] Recommendation ITU-T P.863.1 (2014), *Application guide for Recommendation ITU-T P.863*.

3 Definitions

3.1 Terms defined elsewhere

None.

3.2 Terms defined in this Supplement

None.

4 Abbreviations and acronyms

This Supplement uses the following abbreviations and acronyms:

ACR Absolute Category Rating

BBE Blind Bandwidth Extension

DCR Degradation Category Rating

HD High Definition

NB Narrowband

WB Wideband

SWB Superwideband

5 Conventions

None.

6 BBE and objectives

Blind bandwidth extension (BBE) technology aims to transform NB speech into WB or SWB speech. For simplicity, the focus of this Supplement is the WB case only. Typically using some form of either spectral folding or statistical modelling, the 4-8 kHz part of a speech signal is predicted from the 0-4 kHz part, to generate a signal which has the general characteristics of wideband speech [b-Carl], [b-Pulakka]. While perfect prediction cannot be expected reasonably good quality speech can be obtained.

There are two ways to view the objectives of BBE. It can either be seen as a way to improve NB, or as a way to make NB closer to WB. While these may seem like very similar objectives, in practice they are quite different, and apply to different scenarios. The first case is that of a network that is currently NB only, while the second case is encountered when a network has a mix of NB and WB calls. Both of these scenarios are encountered across mobile phone networks, but as networks move towards deploying more HD voice codecs, the second scenario will become more common. The user will likely experience a mix of wideband and narrowband calls, or possibly even experience both bandwidths during the same call. The lack of uniformity of experience will be a problem, as some calls will appear muffled or of lower quality, which in turn will lead to user dissatisfaction.

7 BBE quality evaluation

7.1 Challenge: bandwidth vs quality

BBE algorithms are not perfect and the process of predicting a high band introduces artefacts. There is a trade-off between bandwidth of the signal and overall noisiness of the BBE extended speech, which can be controlled easily by attenuating the overall high-band energy.

This can lead to confusion during comparative evaluations, where listeners might prefer an algorithm because it shows fewer artefacts when this is in fact due to it having less high-band energy, rather than being intrinsically a better algorithm. Therefore, it is important that different BBE algorithms are compared at the same operating point. This is illustrated in Figure 1.

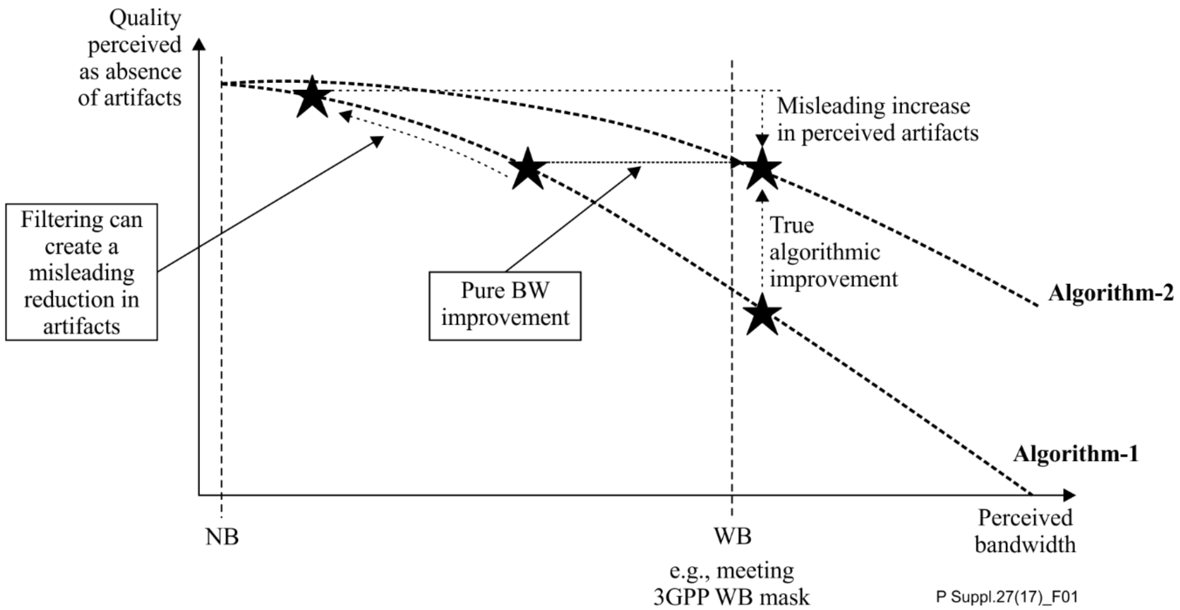

Figure 1 – Bandwidth vs absence of artefacts trade-off

In Figure 1, two BBE algorithms are represented. Algorithm-2 is clearly better than Algorithm-1. This is easily seen when fixing one dimension, either bandwidth or quality: Algorithm-2 is superior in the other dimension. The problem occurs when comparing Algorithm-1 at low bandwidth (the operating point furthest to the left), to Algorithm-2 at high bandwidth (the operating point furthest to the right). In this situation, Algorithm-1 has fewer artefacts than Algorithm-2, even though the algorithm itself is not as good, only the operating points are different. This shows the necessity of considering both dimensions when comparing BBE algorithms.

Additionally, as bandwidth is reduced, all BBE algorithms converge to the input narrowband signal, and are indistinguishable. Therefore, for maximum resolution, it is best to evaluate BBE algorithms at a high bandwidth, even if it might not be the bandwidth at which the algorithm is intended to be used for deployment.

7.2 Defining bandwidth

Frequency response of BBE technologies is undefined, as the predicted high band is not a function of the original high band. This can be resolved by defining a reference wideband input. The speech material defined in [ITU-T P.501] is a good choice since it is broadly used across the wireless industry for testing compliance for voice services.

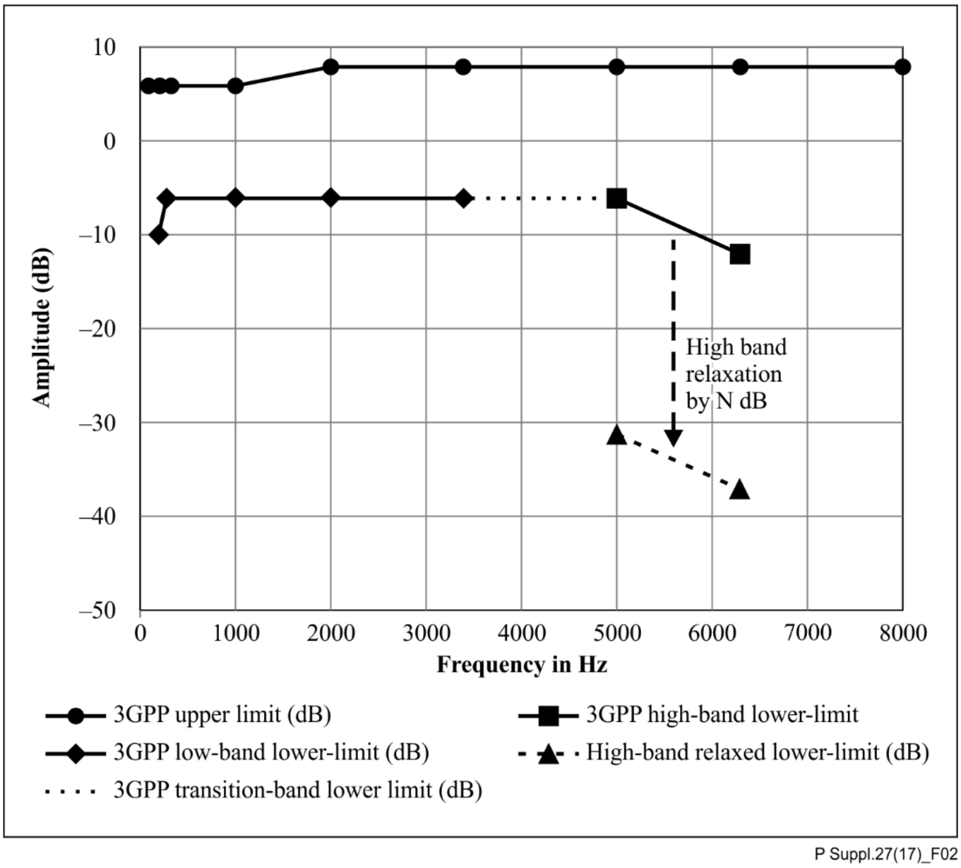

Figure 2 – Frequency mask for bandwidth estimation

The 3GPP WB Rx mask defined in [b-3GPP TS 26.131] is a good mask to use with WB BBE, as it ensures that the bandwidth of the BBE output is similar to that of a coded, wideband output meeting the same mask.

However, to allow for a different operating point at lower bandwidth, a series of masks can be defined as modifications to the 3GPP WB Rx mask wherein its lower limit is relaxed by N dB in the high band. This is illustrated in Figure 2. Note that the 3.3-5 kHz transition-band has been left undefined, to allow for classic frequency extension techniques such as spectral folding, which can lead to a frequency dip around 4 kHz without adversely affecting speech quality.

7.3 Subjective and objective evaluation methods

The most commonly used techniques for subjective quality evaluation of vocoders are the ITU-T P.800 DCR (degradation category rating) and ACR (absolute category rating) tests [ITU-T P.800]. Both are suitable for BBE evaluation, the main difference being that DCR measures degradation from the WB reference input, whereas ACR does not present a reference. Interestingly, these two cases match the two deployment scenarios described above, with DCR corresponding to the NB/WB mixed network case, and ACR to the NB-only case.

However, subjective tests are costly and time-demanding. An increasingly popular alternative is to use objective evaluation methods, in particular ITU-T P.863, also known as POLQA [ITU-T P.863]. While it is not perfect, [ITU-T P.863] claims to handle a wide range of input degradations, and when used appropriately, can give a good indication of subjective speech quality [ITU-T P.863.1]. Additionally, it is already widely used in the industry for speech quality evaluation, often with ITU-T P.501 source material. For BBE, the source material should be transcoded by an appropriate narrowband vocoder. If cellular wireless transmission is under consideration, this most commonly means the 3GPP AMR codec operating at 12.2 kbps [b-3GPP TS 26.090], as this is the narrowband speech codec used in the vast majority of today’s mobile communication networks.

7.4 Proposed BBE objective evaluation methodology

We propose the following objective evaluation methodology for BBE.

- Bandwidth requirement:
  - Measure bandwidth by testing the response to verify whether it passes a frequency mask derived from the 3GPP WB Rx mask, as per Figure 2, and using ITU-T P.501 British English speech material as the input.
  - We recommend using N=0 dB (i.e., no relaxation of the mask) as the operating point.
- Quality requirement:
  - Measure quality using ITU-T P.863 with ITU-T P.501 British English coded by AMR at 12.2 kbps.
  - A good quality reference is the ITU-T P.863 MOS-LQO score of the input NB signal, up-sampled to 16 kHz.

Note that commercial implementations of [ITU-T P.863] have a number of options and versions. In this document, the so-called "POLQA v2.4", in high-accuracy mode, and a WB reference are used. Other options change the absolute ITU-T P.863 MOS-LQO scores, but generally have little impact on the relative scores, and do not change the overall conclusions.

8 BBE algorithm evaluation

8.1 Algorithms used

To illustrate the various evaluation techniques, we have evaluated four BBE algorithms according to the above-proposed methodologies. **BBE1** is a simple noise addition algorithm which is included for illustrative purposes, as a simplistic form of bandwidth extension with poor quality. **BBE2, BBE3** and **BBE4** are proprietary blind bandwidth extension technologies. The input is narrowband PCM transcoded by AMR at 12.2 kbps. [b-3GPP TS 26.090]

8.2 Objective performance

Figure 3 shows the ITU-T P.863 MOS-LQO scores for these BBE algorithms versus their bandwidth, from 0 to 25 dB attenuation from the 3GPP WB Rx mask. The scores for AMR NB at 12.2 kbps and AMR-WB at 8.85 kbps are shown as references.

As expected, ITU-T P.863 MOS-LQO scores drop as the bandwidth of the signal gets closer to WB. In effect, the model described in [ITU-T P.863] heavily penalizes over-predicting high-band energy, and reducing the amount of predicted high-band energy overall helps to improve the ITU-T P.863 MOS-LQO score, even as the subjectively perceived bandwidth decreases. As the mask is relaxed further, the scores flatten out. This is expected, as the lower limit of the mask is a minimum requirement for high-band energy, but the signal does not have to follow the mask attenuation.

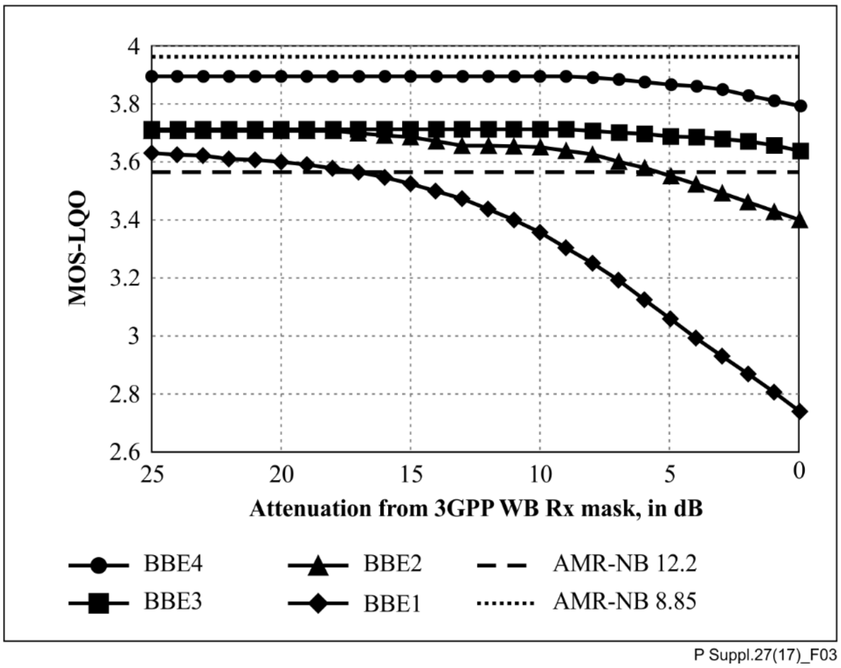

Figure 3 – ITU-T P.863 MOS-LQO vs bandwidth

There are several interesting points shown in Figure 3. Firstly, it can be seen that BBE technology can provide a significant objective quality advantage over narrowband, and it can approach the quality of AMR-WB at 8.85 kbps. Indeed BBE4 scores up to 0.35 MOS-LQO higher than the narrowband reference.

Secondly, even BBE1, a very basic BBE algorithm with poor audio quality, can outperform the original narrowband, up to approximately the 18 dB attenuation point. It can also outperform a good BBE solution such as BBE2 and BBE3, when these are used with high levels of bandwidth. This clearly indicates that the ITU-T P.863 MOS-LQO score is not a reliable indicator by itself, and must be considered in conjunction with the bandwidth.

Finally, even though BBE2 and BBE3 achieve similar ITU-T P.863 MOS-LQO scores at high attenuation, BBE3 is able to maintain that performance much better than BBE2 as bandwidth increases. Therefore, for reliable discrimination between BBE algorithms, the most interesting measurements are the attenuation at the crossover point with the narrowband reference, and the ITU-T P.863 MOS-LQO score at 3GPP mask level (i.e., the 0 dB point on the curve).

8.3 Subjective performance

The subjective performance of the various BBE algorithms presented here was evaluated using the ITU-T P.800 methodology. Both a DCR (degradation category rating) and an ACR (absolute category rating) test were run at an independent test laboratory. Both the DCR and ACR tests were run using 32 listeners, 36 conditions and 192 votes per condition.

The results from the DCR test are shown in Figure 4, with error bars indicating 95% confidence intervals. Note that BBE1 was not included in the test, as its subjective performance is very poor. It can be seen that the scores are consistent with the ITU-T P.863 MOS-LQO scores shown in

Figure 3. The rank order of the BBE algorithms is maintained, and BBE4 is again equivalent to AMR-WB at 8.85 kbps.

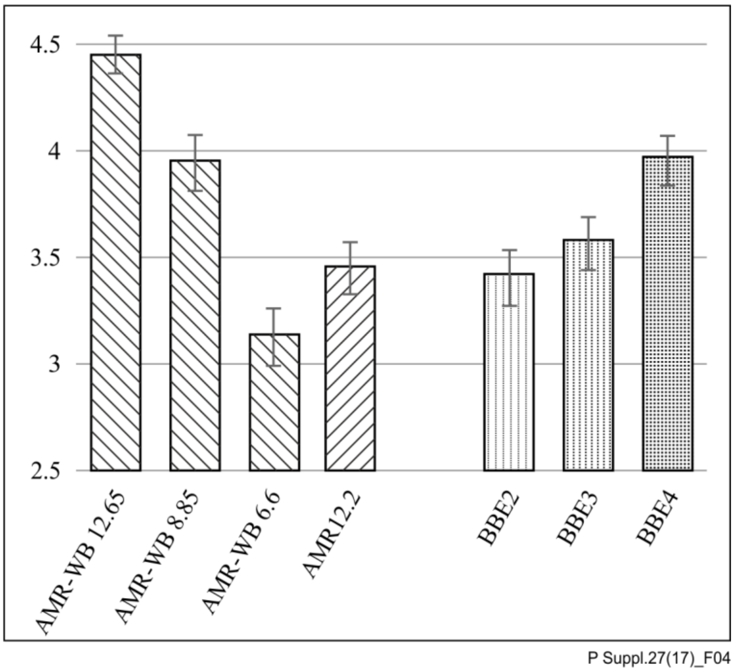

Figure 4 – ITU-T P.800 DCR MOS-LQS at 3GPP mask level

The test results for the ACR are shown in Figure 5. It can be seen that the results are consistent with both the ITU-T P.863 MOS-LQO results and the DCR results. Again BBE4 matches AMR-WB 8.85's level of quality. The scores are shown in Table 1.

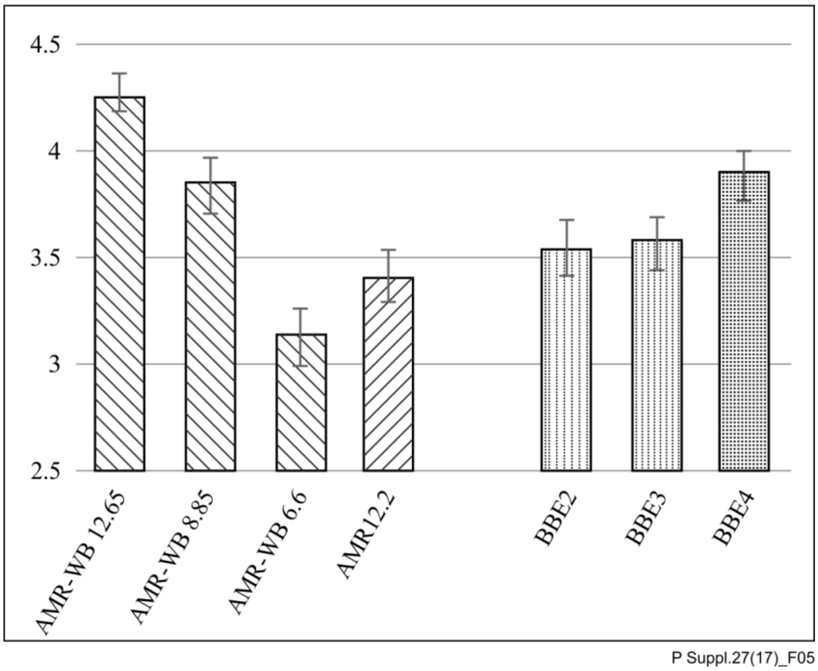

Figure 5 – ITU-T P.800 ACR MOS-LQS at 3GPP mask level

Table 1 – ACR vs DCR scores

Condition	DCR	ACR
AMR-WB 12.65	4.46	4.26
AMR-WB 8.85	3.95	3.86
AMR-WB 6.6	3.13	3.15
AMR 12.2	3.45	3.41
BBE2	3.41	3.54
BBE3	3.57	3.57
BBE4	3.96	3.90

8.4 Effect of high-band attenuation on subjective performance

In previous clauses, it was suggested that BBE algorithms should be compared at a given bandwidth, and we suggest using the 3GPP WB Rx mask as the evaluation point for maximum discrimination. However, it is not clear that this is the bandwidth that should be used in real-world deployments.

To establish this, the best performing algorithm, BBE4, was taken, tuned to meet the 3GPP WB Rx mask level and applied several attenuations to the high band, from 5 to 15 dB. This attenuation is denoted as N, as per Figure 2. Figure 6 shows the ITU-T P.800 ACR and DCR scores for these conditions.

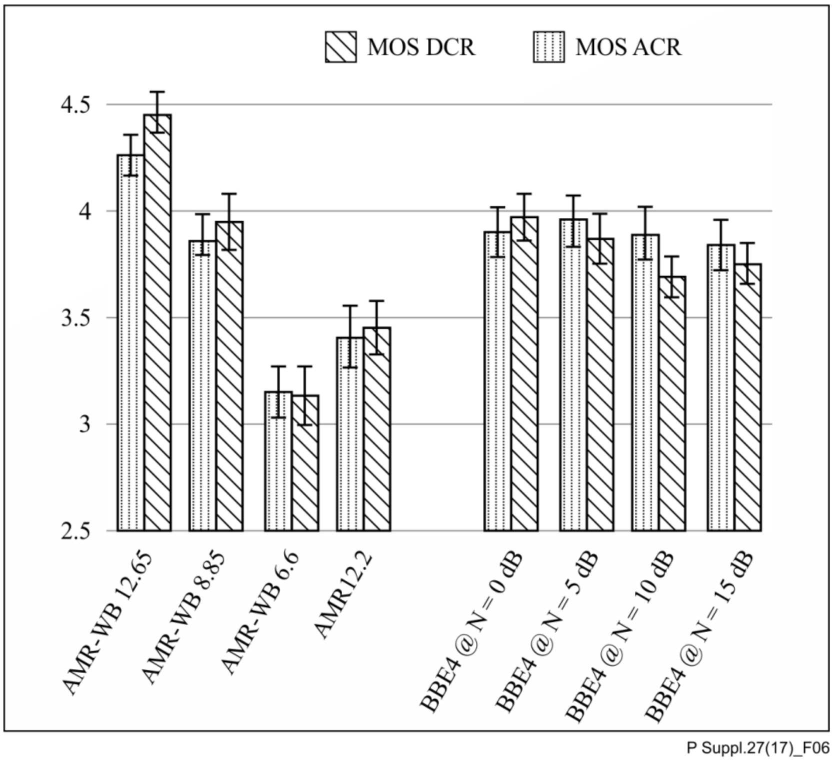

Figure 6 – BBE4 ACR and DCR MOS vs bandwidth.
N is the attenuation from the 3GPP WB Rx mask

Several observations can be made. Firstly, there appears to be an optimal operating point. For DCR, 0 dB attenuation seems best. For ACR, 5 dB attenuation seems optimal. Note that these results must be taken with some degree of caution, as the differences observed are small, and not all statistically significant with 95% confidence.

This difference between ACR and DCR is expected: the DCR methodology presents the original wideband signal as a reference, therefore the results tend to weight bandwidth more, compared to an ACR test where the samples are presented without a reference. This can be tied to the observations from clause 7.3: the optimal operating point of BBE will probably be at a higher bandwidth if the network has both NB and WB, compared to a NB-only network.

Secondly, Figure 3 suggests that an optimal operating point for BBE4 would be around 5 dB below the 3GPP level, as the ITU-T P.863 MOS-LQO score starts to drop above this point. This result matches the result of the ACR test, which is reasonable as the ITU-T P.863 model is designed to predict ACR scores. Again, the objective methodology matches well with the subjective results.

8.5 Summary

Overall, results show that the proposed objective evaluation methodology, combining an ITU-T P.863 MOS-LQO score with a bandwidth requirement, works well. The results correlate well with both ACR and DCR testing, and in our testing clearly identify which BBE algorithm performs best. In addition, it gives a good indication of the optimal level of bandwidth of a given algorithm.

It can also be noted that the best BBE algorithm we tested achieves a quality equivalent to AMR-WB 8.85 when operating on AMR 12.2 transcoded inputs and meeting the 3GPP WB mask. This is consistent across testing methodologies, objective and subjective.

It can be argued that we have only tested a small number of BBE algorithms, and there is no guarantee that results will extend to all BBE algorithms. This is impossible to disprove, and is unavoidable considering the current limited number of BBE solutions commercially available in devices. However, even though the 4 BBE algorithms presented here use very different signal processing techniques, the conclusions have been consistent for all of them, giving confidence that they will extend to other BBE algorithms.

Previously, several papers have attempted to tackle the issue of objective versus subjective quality evaluation [b-Möller], [b-Pulakka2], but concluded that while there is reasonable correlation between objective and subjective scores, it is not reliable as a means to compare different BBE technologies. We believe that this may have been caused by not taking the bandwidth aspects into account. When considering the bandwidth, a reasonably reliable estimation of quality can be obtained.

Bibliography

[b-3GPP TS 26.090] 3GPP TS 26.090 (2012), *Adaptive multi-rate (AMR) speech codec; Transcoding functions*, Sept., version 11.0.0.

[b-3GPP TS 26.131] 3GPP TS 26.131 (2016), *Terminal acoustic characteristics for telephony; Requirements*, Jun., version 13.3.0.

[b-3GPP TS 26.190] 3GPP TS 26.190 (2017), *Adaptive multi-rate wideband (AMR-WB) speech codec; Transcoding functions*, Mar., version 14.0.0.

[b-3GPP TS 26.441] 3GPP TS 26.441 (2015), *Codec for Enhanced Voice Services (EVS); General overview*, Dec., version 13.0.0.

[b-Carl] H. Carl and U. Heute (1994), *Bandwidth Enhancement of Narrow-Band Speech Signals*, in Proc. EUSIPCO, vol. 2, Edinburgh, UK, Sept., pp. 1178-1181.

[b-Möller] Sebastian Möller et al (2013), *Speech Quality Prediction for Artificial Bandwidth Extension Algorithms*, in INTERSPEECH 2013, 14th Annual Conference of the International Speech Communication Association, September 8-12, pp. 3439-3443.

[b-Pulakka] H. Pulakka and P. Alku (2011), *Bandwidth Extension of Telephone Speech Using a Neural Network and a Filter Bank Implementation for Highband Mel Spectrum*, IEEE Trans. Audio, Speech, Language Process., vol. 19, no. 7, pp. 2170-2183, Sept.

[b-Pulakka2] Pulakka, Hannu, Ville Myllylä, Anssi Rämö, and Paavo Alku (2015), *Speech Quality Evaluation of Artificial Bandwidth Extension: Comparing Subjective Judgments and Instrumental Predictions*, in Interspeech 2015, 16th Annual Conference of the International Speech Communication Association.

SERIES OF ITU-T RECOMMENDATIONS

Series A	Organization of the work of ITU-T
Series D	Tariff and accounting principles and international telecommunication/ICT economic and policy issues
Series E	Overall network operation, telephone service, service operation and human factors
Series F	Non-telephone telecommunication services
Series G	Transmission systems and media, digital systems and networks
Series H	Audiovisual and multimedia systems
Series I	Integrated services digital network
Series J	Cable networks and transmission of television, sound programme and other multimedia signals
Series K	Protection against interference
Series L	Environment and ICTs, climate change, e-waste, energy efficiency; construction, installation and protection of cables and other elements of outside plant
Series M	Telecommunication management, including TMN and network maintenance
Series N	Maintenance: international sound programme and television transmission circuits
Series O	Specifications of measuring equipment
Series P	**Telephone transmission quality, telephone installations, local line networks**
Series Q	Switching and signalling, and associated measurements and tests
Series R	Telegraph transmission
Series S	Telegraph services terminal equipment
Series T	Terminals for telematic services
Series U	Telegraph switching
Series V	Data communication over the telephone network
Series X	Data networks, open system communications and security
Series Y	Global information infrastructure, Internet protocol aspects, next-generation networks, Internet of Things and smart cities
Series Z	Languages and general software aspects for telecommunication systems

Printed in Switzerland
Geneva, 2017