# QoE Assessment for Broadcast Audio Contribution over IP (ACIP)

Maxim Graubner[1,2], Parag S. Mogre[1], and Thorsten Lorenzen[2]

[1]*Multimedia Communications Lab (KOM), Dept. of Electrical Engineering and Information Technology, Technische Universitaet Darmstadt, Rundeturmstr. 10, 64283 Darmstadt, Germany*

[2]*IT Systemservice - Radio, Hessischer Rundfunk, Betramstr. 8, 60320 Frankfurt a.M., Germany*

Correspondence should be addressed to Maxim Graubner (`mgraubner@hr-online.de`)

## ABSTRACT

Broadcasters will increasingly use audio contribution over Internet protocol (ACIP). So far, no dedicated Quality of Experience (QoE) prediction framework exists. In this paper, we present a novel non-intrusive parametric QoE rating framework for such a professional broadband audio communication using Voice over IP (VoIP) technology, based on the extended E-model for telephone networks (ITU-T Rec. G.107). For this, we propose an R-factor scale extension to a maximum value of 157, instead of 129 as is used for wideband. Finally, our QoE rating model provides separated impairment factors for the delay, loss, coding and bandwidth impairment. The model was developed based on our proposal for instrumental evaluation as well as subjective experience by experts.

## 1. INTRODUCTION

Broadcast audio contribution generally refers to an exchange of professional audio material, normally between remote sites and broadcasting stations. The contribution use cases range from simple outside broadcasts (e.g. sports commentary) to complex interactive audio communication, which are explicitly for broadcasting to the radio listener.

Presently, public radio broadcasters established their audio contribution links mainly using the synchronous circuit-switched Integrated Services Digital Network (ISDN). ISDN is set to disappear in the future, being replaced by packet-switched Voice over IP (VoIP) technology run on managed networks [19] as well as on the best effort Internet, which provide no more than statistical Quality of Service (QoS) guarantees. In contrast, the ISDN QoS satisfies the broadcaster's audio contribution needs by providing *guaranteed services* meeting the requirements with high availability. In VoIP networks, these guarantees disappear and QoS parameters such as data rate and network delay can vary over time. Additionally, packet loss can occur, resulting in loss of audio data.

So far, only intelligibility-optimized speech conversation over IP respective delay-tolerant audio streaming has been extensively researched. However, many Audio Contribution over IP (ACIP) use cases require a much lower end-to-end delay compared to VoIP in order to ensure high interactivity. The main difference in terms of listening-only quality between speech and audio communication over IP is the required audio bandwidth. The audio bandwidth for VoIP of 3 kHz narrowband and 7 kHz wideband extends to above 15 kHz for broadband audio communication. Moreover, the desire for a most accurate sound reproduction when using ACIP leads to less acceptance of coding artifacts or packet loss, which would then need to be concealed by complex reconstruction.

Available objective QoE assessment methods for speech and audio quality evaluation are not directly usable for a quality rating of ACIP applications [6]. To fill the gap, we designed a dedicated non-intrusive parametric QoE model for conversational quality rating, based on the E-model [9] approach. The applications for such a QoE assessment model for ACIP are versatile. They include e.g. transmission planning and set-up optimization before starting a connection, real-time quality monitoring and transmission rating after a conversation using ACIP technology. All these tasks can further be used for a QoS improvement at the application layer (especially for real-time control) incorporating a cross-layer impair-

ment trade-off [5]. While ACIP ideally requires a low delay connection where impairments due to IP packet loss are possible, estimates of the conversational quality are necessary for finding a perceptually optimal trade-off between delay and audio signal quality, which represent the associated interaction and listening QoE respectively [21].

Our contributions can be summarized as:

1. We propose a broadband extension of the wideband $R$-factor from the extended E-model to enable a conversational quality prediction for ACIP. The approach relies on an extrapolation of the bandwidth impairment model for wideband [15] to fullband.

2. We present a delay impairment factor for ACIP as well as a derivation methodology for fullband equipment impairment factors from instrumental models, based on a novel listening-only quality evaluation approach for broadband, which we presented in detail in [6]. Furthermore, we introduce a necessary set of reference conditions for ACIP, based on our own objective QoE evaluations [5].

3. We propose a non-intrusive parametric QoE assessment model for ACIP quality rating, providing separated impairment factors for the delay, loss, coding and bandwidth impairment.

## 2.  AUDIO CONTRIBUTION OVER IP

The European Broadcasting Union (EBU) N/ACIP project group recently published a technical specification for professional audio contribution over networks using IP technology [4]. The recommendation developed by the EBU N/ACIP project group includes well known coding algorithms (see Section 2.1) as well as signaling standards (SIP/SDP[1]) and transport protocols (RTP/UDP[2]) specified by the Internet Engineering Task Force (IETF) for real-time communication services over IP, e.g. VoIP. The N/ACIP project group also provides QoS parameters for different ACIP use cases. In Table 2, these operational requirements are presented while we added VoIP QoS specifications [15] in order to compare the different communication applications.

---

[1]SIP: Session Initiation Protocol (IETF RFC 3261), SDP: Session Description Protocol (IETF RFC 4566).

[2]RTP: Real-time Transmission Protocol (IETF RFC 3550), UDP: User Datagram Protocol (IETF RFC 768).

### 2.1.  ACIP Audio Codecs

The N/ACIP project group made four common audio coding formats mandatory in order to create a common ground for compatibility among ACIP equipment. The mandatory audio coding algorithms are ITU-T G.711, ITU-T G.722, ISO MPEG-1/2 Layer 2 and Linear PCM. Hence, we defined usual configurations of these as anchors leading to reference conditions for the fullband equipment impairment factor derivation (see Section 5.3). In addition to the mandatory set, the EBU also specified recommended and optional coding algorithms. The recommended formats are ISO MPEG-1/2 Layer 3, MPEG-4 AAC-LC (low complexity Advanced Audio Coding) and MPEG-4 AAC-LD (low delay). Optionally it further mentions amongst others the proprietary Enhanced APT-X (Eapt-X) algorithm, which we also used for our model design.

### 2.2.  Quality Impairments

The most important impairment on the interaction quality in ACIP is the value of the overall end-to-end delay $d_{\mathrm{e2e}}$, where the impairment worsens the higher the latency becomes. Possible listening-only quality degradations in ACIP can be categorized into two classes: the *coding algorithm impairments* and possible *network impairments* due to packet loss [5]. The coding algorithm impairments due to coding artifacts depend mainly on the performance of the algorithm type $c$, the coding bitrate $r_c$ as well as possible restrictions of the audio bandwidth $B$.

|  | $B$ [kHz] | $F_s$ [kHz] | Example Codec |
|---|---|---|---|
| Narrowband (NB) | 300-3400 | 8 | ITU-T G.711 |
| Wideband (WB) | 50-7000 | 16 | ITU-T G.722 |
| Super-WB (SB) | 70-12000 | 24 | Skype SILK [22] |
| Ultra-WB (UB) | 40-15000 | 32 | Eapt-X |
| Fullband (FB) | 20-22000 | 48 | MPEG [20] |

**Table 1:** Bandwidths in audio communication.

Generally, the audio bandwidth $B$, which depends on the used audio sampling frequency $F_s$, is an important factor determining the audio quality in terms of acoustics and human perception [24]. In Table 1, an overview of different audio bandwidths used in digital audio transmission and specifically for IP-based communication services is presented with examples of coding algorithms for the respective bandwidths. The classification here is based on definitions for speech communication [15] but we extended these for broadband audio communication.

| QoS Parameter | Outside Broadcasts | | | Interviews | Discussions | VoIP |
|---|---|---|---|---|---|---|
| | Concerts | Sports | News | | | |
| End-to-end Delay [ms] | $< 500$ contribution, $< 50$ talkback | | | $< 100$ | $< 100$ | 150-400 |
| Audio Bandwidth [kHz] | 15-20 | 7-15 | 3.5-12 | 7-12 | 12-20 | 3.5-7 |

**Table 2:** Operational requirements for different audio contribution use cases and VoIP.

The perceived audio quality based on coding impairments, assessed with a Mean Opinion Score (*MOS*) metric, can be described in general[3] as $MOS_{coding} = f(\mathbf{c})$ with the vector of coding parameters for each coding algorithm, $\mathbf{c} = [c, r_c, F_s, Q, B, ch]^T$. For the coding impairment analysis, the packet loss ratio $\rho$ needs to be $\rho = 0$.

The impairments to the perceived audio quality due to network packet loss are determined by the amount of packets lost, the burstiness of the loss process as well as the packet size $L$ of the continuous bit rate (CBR) audio stream because the larger the packet is the more consecutive audio information gets lost. These impairments can also be described by a QoE metric, with respect to loss now as $MOS_{loss} = f(\mathbf{P}_{loss}, \mathbf{c}, L)$, while the packet loss process here is characterized by the state transition matrix $\mathbf{P}_{loss}$ of a 2-state Markov model [18]. This relatively simple two-parameter model fully describes a short-term loss process allowing for loss correlation [15]. The parameters of the Markov model can be directly related to the packet loss ratio $\rho$ and the mean loss period $\mu$ in packets [5].

## 3. MOS TERMINOLOGY

In the last Section, the Mean Opinion Score was introduced. To better identify the quality category and assessment method to which a *MOS* value refers, a *MOS* terminology was defined in ITU-T Rec. P.800.1 [11]. Table 3 lists the identifiers. The different *MOS* types receive a suffix with respect to the quality category ("LQ" for listening, "CQ" for conversational quality) and assessment method ("S" for subjective assessment, "O" for objective assessment and "E" for an estimate based on a quality prediction model as the E-model presented in Section 4).

---

[3]We did not regard quantization distortions due to the sampling resolution $Q$ or the impairment to the perceived audio quality due to different channel configurations $ch$ (e.g. mono or stereo transmission).

The different metrics in Table 3 can further be differentiated with respect to the audio bandwidth of the transmission system. Therefore, the ITU-T recommends to use of another *MOS* suffix to differentiate between a narrowband ("n") or wideband ("w") transmission system under test. This additional suffix is neglected in this work for convenience even more because mostly the quality assessment with respect to fullband transmission systems is regarded, which is not standardized yet.

| | Listening-only | Conversational |
|---|---|---|
| Subjective | *MOS LQS* | *MOS CQS* |
| Objective | *MOS LQO* | *MOS CQO* |
| Estimated | *MOS LQE* | *MOS CQE* |

**Table 3:** Mean Opinion Score (*MOS*) terminology.

## 4. THE E-MODEL

In the following, we present the parametric E-model for conversational quality assessment which we used as a basis framework for our broadband QoE rating model for ACIP. The E-model is currently the ITU-T recommendation for a mathematical model with which to perform a speech transmission rating (ITU-T Rec. G.107 [9]). It gives a QoE estimate based on instrumentally measurable characteristics of the system [15]. Its main advantage is the ability to predict the overall *conversational quality*, among others due to the incorporation of both equipment impairments on listening quality, as well as impairments due to transmission delay. The main characteristics of the E-model are that it is parameter-based and non-intrusive.

The E-model is usable for estimating the transmission quality during network planning and set-up optimization [13], but also for operational monitoring [2] or for perceptually-driven QoS optimization [21]. Until now however, it is only suitable for speech communication up

to wideband transmission and therefore not directly applicable for fullband quality rating.

The E-model outputs the transmission rating factor $R$ ($\in [0,100]$ for narrowband), where 100 indicates a user satisfaction of "very satisfied". The $R$-factor is obtained by summing up different impairment factors. Herewith different scalar input parameters (i.e. signal-to-noise ratio, packet loss or transmission delay) are grouped into different classes of impairments. The assumption behind this approach is an *impairment additivity* [15]. The $R$-factor is comprised of

$$R_0 - I_s - I_d - I_{e,eff} + A = R \,, \tag{1}$$

where the basic transmission rating factor $R_0$ reflects the signal-to-noise ratio, the simultaneous impairment factor $I_s$ accounts for degradations which occur in the transmitted speech signal (e.g. signal-correlated noise) and the delay impairment factor $I_d$ includes the delay and echo impacts in a bidirectional transmission. The advantage factor $A$ stands for some "advantage in access", for example if a wireless system is used and the mobile access is able to compensate some of the quality degradation. $I_{e,eff}$ is the effective equipment impairment factor which is a function of the packet loss ratio $\rho$ and the equipment impairment factor $I_e = I_{e,eff}(\rho = 0)$, i.e. of degradations due to lossy coding. The E-model accounts for independent "random" packet loss since its revision in 2002. Following a proposal by [15], the rational function model was later extended for *dependent* packet loss based on the 2-state Markov model [18]. Another approach is used in [21], where the E-model compatibility was not an explicit demand. Here a non-linear least-squares curve fitting for a logarithmic model is proposed, which so far has only been used for the description of independent loss impairments.

The E-model for a narrowband transmission rating has $R_0 = 93.2$. For wideband transmission, the $R$-factor range is extended to 129, based on subjective evaluations [15], and by definition, the basic transmission rating factor $R_{0,WB} = 129$. The $R$-factor can be mapped to a *MOS CQE* $\in [1;4.5]$. The inverse mapping from *MOS CQE* to the $R$-factor can be achieved with the complicated Candono's Formula [9]. In [21], a 3rd-order polynomial fitting is proposed for a simplification of this mapping. For the wideband case, the *MOS*-to-$R$-factor formula is stretched by a linear or non-linear model, proposed in [14] and [15], i.e. for the linear model

$R_{WB} = 1.29\,R_{NB}$. This approach was exploited for the fullband case.

When narrowband results for $I_e$ exist (i.e. from the "provisional planning values" list in the ITU-T Rec. G.113 [10]), the respective wideband value $I_{e,WB}$ can be obtained by simple shifting, $I_{e,WB} = 35.8 + I_{e,NB}$.

### 4.1. Considering Bandwidth Impairment

The consideration of an impairment factor for linear distortion of narrowband and wideband speech transmission was proposed in [23], based on findings in [15]. The so called bandwidth impairment factor $I_{bw}$ represents linear distortions due to band limitations, until now contained in the equipment impairment factor $I_e$. This now consists of $I_{e,WB} = I_{bw} + I_{res}$, where $I_{res}$ is the residual portion of $I_e$ and reflects the coding algorithm impairment (i.e. the non-linear distortions). The idea behind this approach is, that with wideband consumer telephony transmission, a narrowband codec is perceived as distorted in comparison with a wideband codec.

To obtain $I_{bw}$, the resulting *MOS* values from subjective auditory tests were mapped on the $R$-factor scale. A formula for $I_{bw}$ was then found by curve fitting in [15] as

$$I_{bw} = 3.5 \cdot 10^{-2}\,|s| - 6.7 \cdot 10^{-3}s - 7.4\,z_{bw} + 129.2 \tag{2}$$
$$\text{with } s = f_c - 9.9\,(z_{bw} + 101.8)\,, \tag{3}$$

where $f_c$ is the center frequency in Hz (calculated by the geometric mean of the lower frequency $f_{low}$ and the upper frequency $f_{up}$), and $z_{bw}$ is the transmission bandwidth in Bark [24], also calculated from the limiting frequencies $f_{low}$ and $f_{up}$.

### 4.2. Equipment Impairment Factor Derivation

While the provisional planning values in ITU-T Rec. G.113 [10] are only specified for a limited set of narrowband and wideband codecs (e.g. ITU-T G.711 and G.722), sometimes new $I_e$ results for new coding algorithms have to be derived. This can be achieved based on subjective evaluations. Because the subjective approach requires a greater effort, a methodology for the derivation of equipment impairment factors from instrumental models is also proposed in ITU-T Rec. P.834 for narrowband codecs using the Perceptual Evaluation of Speech Quality (PESQ) algorithm (ITU-T Rec. P.862) and in ITU-T Rec. P.834.1 [12] for wideband codecs using WB-PESQ (ITU-T Rec. P.862.2) [17]. The latter was proposed in [14].

# 5. PROPOSED QOE RATING MODEL

We chose the E-model as the basic framework for our broadband QoE rating model for ACIP because it is inherently extendable to fullband quality rating and allows for the incorporation of both equipment related impairments, such as coding characteristics and network packet loss, as well as the impairment due to transmission delay. First, we simplified the general $R$-factor formula of the E-model focusing on coding and IP transmission impairments, resulting in

$$R_{FB} = R_{0,FB} - I_{e,eff,FB} - I_{d,\text{ACIP}} \qquad (4)$$

with the fullband basic transmission rating factor $R_{0,FB}$ and the delay impairment factor $I_{d,\text{ACIP}}$, dedicated to audio contribution (see Section 5.2). The fullband equipment impairment factor $I_{e,eff,FB}$ can further be separated into

$$I_{e,eff,FB} = I_{bw,FB} + I_{res,FB} + I_{loss,FB} , \qquad (5)$$

with the bandwidth impairment factor $I_{bw,FB}$ representing linear bandwidth distortions for the fullband case and $I_{res,FB}$ representing non-linear coding distortions. The loss impairment factor $I_{loss,FB}$ includes a continuous QoE model for the loss-only perception. It is obtained through non-linear least-squares curve-fitting of discrete experiment outcomes using a logarithmic model dependent on the packet loss ratio $\rho$ [21],

$$I_{loss,FB} = a \, ln((1+b)\rho) . \qquad (6)$$

Hence, our ACIP QoE rating framework includes separate models for the delay, loss, coding and bandwidth impairments, which are the most important factors for audio contribution applications. We excluded the advantage factor for now and the simultaneous impairment factor $I_s$ of the E-model is also neglected, because a perfect audio source is assumed.

The challenge then was to find a reasonable fullband basic transmission rating factor $R_{0,FB}$ as well as the mathematical models for the delay impairment factor $I_{d,\text{ACIP}} = f(d_{e2e})$ and the fullband equipment impairment factor $I_{e,eff,FB} = f(\mathbf{P}_{\text{loss}}, \mathbf{c}, L)$.

## 5.1. The Fullband R-Factor

First, we present the necessary $R$-factor scale extension for a fullband quality rating exploiting the bandwidth impairment model approach (see Section 4.1). For this, we extrapolated the model for wideband speech communication to the full audio bandwidth in order to derive $R_{0,FB}$. This was justified by a comparison of the model outcomes for different audio bandwidths up to fullband with subjective results from such literature as from the EBU research [3], objective measurements from the PEMO-Q algorithm [8] and subjective appraisal from experts.

In [15] it is stated that the wideband advantage over narrowband is in the range of 1.3-1.5 points *MOS*. Using this as groundwork, we compared this result with objective difference grade (*ODG*) results from the PEMO-Q algorithm [8] for bandwidth-only impairments, since it is the best suited instrumental method for evaluating high impairments on audio quality [1], as with the bandwidth restriction by downsampling for example from 48 to 16 kHz. For the PEMO-Q evaluations, a 16 bit linear PCM testfile was evaluated with $F_s = 48, 32, 16$ and 8 kHz, which result in respective audio bandwidth restrictions with cutoff frequencies of approx. 22, 15, 7.5 and 3.5 kHz [5].

Results from EBU subjective listening tests [3], obtained with the MUSHRA method (Multi Stimulus Test with Hidden Reference and Anchors), were attained for low bitrate codec evaluation. An examination of this showed that one property of the MUSHRA method is especially interesting: it uses a fullband hidden reference and hidden anchors which are low-pass filtered versions of the reference. The QoE evaluations in [3] used two anchors limited to 3.5 kHz and 7 kHz respectively. The overall quality gradings for the hidden reference and anchors are also given next to the codec results. We assume that these results can be used for the comparison in our work, even if the Continuous Quality Scale ($CQS \in [0, 100]$) of the MUSHRA method is not directly relatable to the *MOS* scale.

The averaged results of the EBU listening tests are noted in Table 4, compared with respective $I_{bw}$ results and the *ODG* scores obtained with the PEMO-Q algorithm. In Table 5, the relative advantages of wideband to narrowband and ultra-wideband respective fullband to wideband are depicted.

In general, all methods give results with the same trend. Wideband transmission has a high advantage over narrowband while for ultra-wideband compared to wideband a similar but slightly higher advantage can be read out of the table, but it is necessary to assume a non-linear dependence between small MUSHRA *CQS* results and

| Bandwidth | $F_s$ [kHz] | $I_{bw}$ | ODG | MUSHRA CQS |
|-----------|------------|----------|------|------------|
| Narrowband | 8 | 35.4 | -2.15 | 27 |
| Wideband | 16 | 6.7 | -1.19 | 52 |
| Ultra-WB | 32 | -25.47 | -0.10 | $\approx 95$ [4] |
| Fullband | 48 | -27.89 | 0 | 100 |

**Table 4:** Different bandwidth impairment measures.

| Rel. advantage in $B$ | $\Delta MOS\,LQS$ | $\Delta ODG$ | $\Delta I_{bw}$ | $\Delta CQS$ |
|-----------------------|-------------------|--------------|-----------------|--------------|
| WB to NB | $\approx 1.4$ | 1 | 28.7 | 25 |
| Ultra-WB to WB | | 1.1 | 32.2 | $\approx 43$ |
| FB to WB | | 1.2 | 34.6 | 48 |

**Table 5:** Relative bandwidth impairments.

respective results of the other measures, which is reasonable. For the fullband advantage with respect to ultra-wideband, only a small improvement results for all methods. This result is meaningful, and has been confirmed by the subjective impression of consulted quality experts. It is furthermore in accordance with the psychoacoustical phenomenon that the sensibility of the human ear decreases for higher frequencies (exponential decrease for frequencies above 10 kHz [24]). The so called hearing area for audio signals has its limits for most subjects between 16 and 18 kHz and the limit decreases with age.

Finally, we assumed an applicability of the wideband bandwidth impairment model extrapolation based on the assumption of a linear dependence between the $I_{bw}$ and the PEMO-Q ODG values (squared 2-norm of the residual: 0.22). This allows us to use the difference in $I_{bw}$ between the best possible wideband transmission (audio bandwidth 200-7000 Hz [15] and linear PCM coding) with $I_{bw} = 0$ and the result for undistorted fullband transmission (linear PCM with $F_s = 48$ kHz) for extending the wideband $R$-factor scale to fullband using $\Delta I_{bw} \approx 28$. We therefore propose the fullband basic transmission factor to be

$$R_{0,FB} = R_{0,WB} + \Delta I_{bw} = 129 + 28 = 157 \ . \quad (7)$$

### 5.2. A Delay Impairment Factor for ACIP

In the following, the formulation of the delay impairment factor for ACIP is addressed. Usually, for formulating a delay impairment factor $I_d$ for a communication service's use case, extensive subjective conversational tests are necessary (refer [7] for telephony) which require an even larger effort compared to subjective listening-only tests. Furthermore, the perception of delay strongly depends on the conversational situation, since conversation situations are strongly influenced by the degree of interaction between the participants [7]. Hence, a delay perception model must be geared to the use case it is dedicated for. While ACIP has various use cases with differ-

ent expectations on the delay (resulting in different delay models), the focus in the following is on the applications requiring high interactivity, e.g. "live discussions". For a delay impairment factor dedicated to this ACIP application, a conceptually simple but promising approach was chosen to enable an incorporation of the delay impairment on the perceived QoS in the QoE prediction framework. It must be stated here that it cannot be assumed, that this method reaches the accuracy of conversational tests.

In the operational requirements for different audio contribution use cases, presented in Table 2, a maximum one-way delay of 100 ms is specified for interviews or discussions. This was used as anchor for the delay model derivation. Different approaches for $I_d$ are available. Our decision to use the AT&T simplified model [21] is based on the fact that it is relatively simple, has a linear increase even for higher delays which are even less tolerated for interactive ACIP and most important: it has a precise turning point where the slope increases, easy realized using the step function. For the dedicated ACIP approach, the turning point of the AT&T simplified model was shifted to the left to introduce the stricter delay requirements. The result was furthermore rescaled from the ordinary narrowband design to the fullband framework. In Fig. 1 the resulting characteristic can be ob-
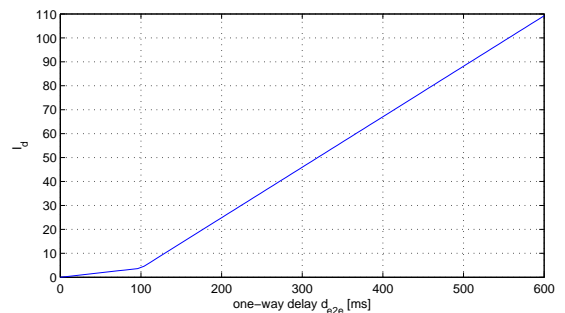


**Fig. 1:** Model for $I_{d,\text{ACIP}}$ versus one-way delay $d_{e2e}$.

---

[4]Result from non-linear curve fitting between ODG values and MUSHRA CQS results.

served. The underlying delay impairment factor is given by

$$I_d = 0.024 d_{\mathrm{e2e}} + 0.11 (d_{\mathrm{e2e}} - d_0) \cdot H(d_{\mathrm{e2e}} - d_0) , \quad (8)$$

which is the original AT&T simplified model but with a turning point at $d_0 = 100$ ms instead of using $d_0 = 177.3$ ms. $H(x)$ is the step function. The ACIP delay impairment factor results after the linear conversation to fullband as $I_{d,\mathrm{ACIP}} = 1.57 \, I_d$.

### 5.3. Fullband Equipment Impairment Factors

Because with this work a new fullband quality rating model based on the E-model is proposed, until now no provisional planning values such as specified by the ITU-T for the narrowband and wideband cases exist (see Section 4.2). Furthermore, a methodology for the derivation of equipment impairment factors from instrumental models for the fullband case is missing so far. Fortunately, the principle approach in the ITU-T Rec. P.834.1 [12] (for wideband) can be modified for the fullband case, since our approach is based on the E-model concept. Therefore, the challenge was to find an equipment impairment factor derivation methodology for the fullband case. We hereby assume that the WB-PESQ algorithm can also be used for the fullband transmission quality rating, if results from the Perceptual Evaluation of Audio Quality (PEAQ) algorithm (ITU-R Rec. BS.1387-1) [1] are considered as well (see below). The goal was to find a reasonable mapping $I_{e,eff,FB} \Longleftrightarrow MOS$. This is addressed in Section 5.3.2.

### 5.3.1. Objective QoE Evaluations

Primarily, for the design of the desired equipment impairment model, a basis of QoE rating results with respect to different values of quantitative model parameters must exist (e.g. parameters from the audio stream). Therefore, normally, subjective tests are performed. Instead we decided to focus on objective evaluation because subjective auditory testing is expensive and slow,

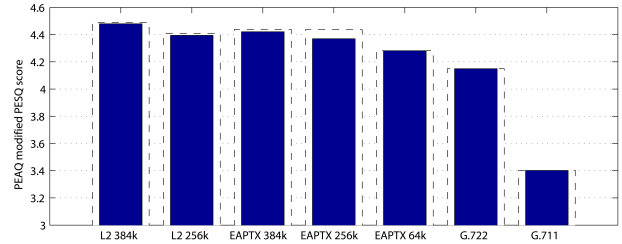| Coding Algorithm | Channels | $r_c$ [kbit/s] | $F_s$ [kHz] |
|---|---|---|---|
| MPEG Layer2 (1) | 2 | 384 | 48 |
| MPEG Layer2 (2) | 2 | 256 | 48 |
| Eapt-X (1) | 2 | 384 | 48 |
| Eapt-X (2) | 2 | 256 | 32 |
| Eapt-X (3) | 1 | 64 | 16 |
| ITU-T G.722 / G.711 | 1 | 64 | 16/8 |

**Table 6:** Selected ACIP coding algorithms.



**Fig. 2:** PEAQ modified WB-PESQ results in comparison to the raw WB-PESQ scores (dashed).

which makes it unsuitable for day-to-day quality evaluations which we required. However, we cross-checked the suitability of our model and results by individual expert testing.

We used the WB-PESQ as well as the PEAQ algorithm for objectively quantifying the user perceived audio quality for different equipment impairments due to the usage of VoIP technology. We combined the method's scores to a new QoE metric, based on a linear combination of the used objective methods, presented in [6]. Thereby, we exploited the ability of the PEAQ algorithm to describe small impairments on fullband audio signals, while the WB-PESQ algorithm more ideally rates stronger impairments and accounts for possible impairments due to packet loss. Furthermore, some evaluations in [16] showed, that it can also be used for the evaluation of wideband mono audio signals. This trade-off approach was chosen because so far no unified fullband method for objective audio quality evaluation is available, which would also regard impairments possibly introduced by IP-based transmissions.

Finally, we calculated WB-PESQ scores with respect to coding and loss impairments while we took necessary results of the PEAQ algorithm with respect to coding impairments from already existing evaluations[5]. For the loss evaluation, a loss process characterization based on a 2-state Markov process [18] has been derived, which allows for taking bursty loss into account. We performed our experiments in a dedicated application-oriented experimental environment for the coding algorithm configurations specified in Table 6. Our experimental methodology is illustrated in depth in [5], where we also

---

[5]These PEAQ results are recorded in a measurement report from M. Karle (Hessischer Rundfunk, 2006)

present results with respect to coding and packet loss, furthermore considering different IP packet lengths. In Fig. 2, the PEAQ modified WB-PESQ results of our QoE metric in comparison to the raw WB-PESQ scores are demonstrated for $\rho = 0$.

### 5.3.2. Methodology for $I_{e,FB}$ Derivation

The focus in the following is on clean transmission (i.e. no packet loss). Nevertheless, we also used the $I_{e,FB}$ derivation methodology for computing impairment factor anchors for results with respect to packet loss ratios $\rho \neq 0$ and mean loss periods $\mu \geq 1$ to later interpolate the discrete results for different loss ratios to an comprehensive $I_{e,eff,FB}$ using a non-linear least-squares curve fitting strategy [5].

In principle, the procedure for the derivation of equipment impairment factors is analogous to the narrowband and wideband case (see Section 4.2): the derived *MOS LQO* values based on the WB-PESQ and PEAQ evaluations are mapped to a *MOS LQS* which is then transformed on the fullband $R$-factor scale. From the obtained $R$-factor estimates $\hat{R}_{FB}$ a raw equipment impairment factor $K$ is calculated with $\hat{R}_{FB} = R_{0,FB} - K$. $K$ is finally mapped to the desired $I_{e,FB}$ value with an adjusting function derived with reference conditions. This ensures that the fullband results are consistent with the results of the E-model for narrowband and wideband transmission. An overview of the procedure is given in Fig. 3 (for details, confer [5]). It can be summarized as:

$$MOS_{LQO} \Longrightarrow MOS_{LQS} \Longrightarrow \hat{R}_{FB} \Longrightarrow K \Longrightarrow I_{e,FB} . \quad (9)$$

In the following, we describe the mapping of the *MOS LQS* to the $\hat{R}_{FB}$ as well as the transformation of the $K$ value to a stable $I_{e,FB}$.

For the mapping of the *MOS LQS* to the $\hat{R}_{FB}$, the calculated *MOS* estimations based on the objective evaluations (see Section 5.3.1) are first transformed to the non-extended $R_{NB}$-scale (range $[0, 100]$). We hereby used the

| Codec name | Channels | $F_s$ [kHz] | $r_c$ [kbit/s] | $I_{e,FB}$ |
|---|---|---|---|---|
| 16 bit lin. PCM | 2 | 48 | 1536 | 0 |
| MPEG L2 | 2 | 48 | 384 | 0.2 |
| Eapt-X | 2 | 32 | 256 | 6.5 |
| Eapt-X | 1 | 16 | 64 | 36.7 |
| ITU-T G.722 | 1 | 16 | 64 | 41 |
| ITU-T G.711 | 1 | 8 | 64 | 63.8 |

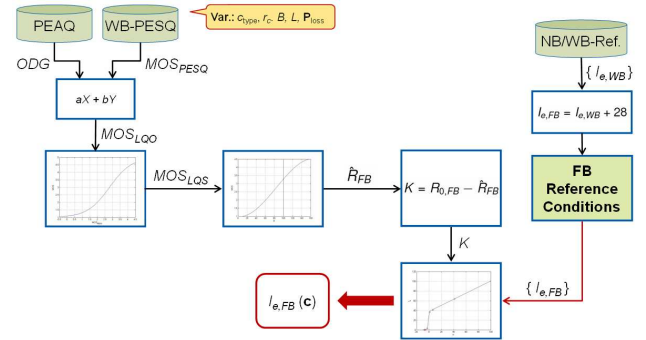**Table 7:** $I_{e,FB}$ reference conditions.



**Fig. 3:** Methodology for the derivation of $I_{e,FB}$.

relationship of *MOS* and $R$-factor from the narrowband E-model, which still reflects the narrowband use of the *MOS* scale assumed by the original E-model [12]. For obtaining results which reflect the superior quality of fullband transmission to the narrowband and wideband cases, the $R$-factor scale is linearly extended to the fullband case with $R_{FB} = 1.57\, R_{NB}$.

In the last step, the obtained $K$ value has to be transformed to a stable $I_{e,FB}$ value. This required the design of a transformation function. This $K$ to $I_{e,FB}$ mapping was found by using reference conditions, which firstly had to be defined because until now, no reference conditions for $I_{e,FB}$ are available. For this, we derived 6 reference conditions, 3 of which were derived during the methodology design process for audio coding algorithms to ensure a high correlation of the results with the underlying QoE results of the WB-PESQ and PEAQ algorithm, as well as subjective impressions. Conditions 4-6 are: the undistorted linear fullband case, the narrowband speech codec ITU-T G.711 and the wideband codec ITU-T G.722. For all these coding algorithms we computed raw $K$ values following the steps of the methodology described before. Then, the reference conditions for ITU-T G.711 and ITU-T G.722 were obtained by shifting the wideband values from the provisional planning values in ITU-T Rec. G.113 [10] to fullband with $I_{e,FB} = I_{e,WB} + 28$, following the wideband principle, while for the undistorted linear fullband case we defined $I_{e,FB} = 0$. In Table 7 the 6 reference conditions are listed with their basic configuration, and finally fullband equipment impairment factors are given by $I_{e,FB}$. The interrelationship of raw $K$ and defined $I_{e,FB}$ values were then used to derive the normalization

function by section-wise linear and non-linear curve fitting.

### 5.4. Bandwidth Impairment Factor

We finally incorporate the concept for a separate bandwidth impairment factor (see Section 4.1) in our fullband QoE model. For this, We simply shifted $I_{bw}$ for wideband to the fullband R-factor scale,

$$I_{bw,FB} = I_{bw,WB} + \Delta I_{bw} = I_{bw,WB} + 28 , \qquad (10)$$

using the wideband to fullband rating conversion presented in Section 5.3.2. Therefore, the best performance without perceived bandwidth impairments is now shifted to full audio bandwidth with $I_{bw,FB} = 0$. The previously best rating for the bandwidth 200-7000 Hz which resulted in $I_{bw,WB} = 0$, is now rated with $I_{bw,FB} = 28$.

Analogous to [23], the obtained final fullband equipment impairment factors for ACIP algorithms are plotted against their resulting bandwidth impairment and coding-only impairment in Fig. 4. The reasonable results consolidate the proposed solutions and give a nice picture of the different impairment relations to the total equipment impairment value.

### 6. CONCLUSIONS AND FURTHER WORK

In this paper, we proposed a non-intrusive parametric QoE assessment model for ACIP quality rating, providing separated impairment factors for the delay, loss, coding and bandwidth impairment. For this, we derived a broadband extension of the wideband *R*-factor of the extended E-model to enable the conversational quality prediction for ACIP. Moreover, we presented a delay impairment factor for ACIP, as well as a derivation methodology for fullband equipment impairment factors from instrumental models, based on our novel listening-only quality evaluation approach for broadband. We also proposed a necessary set of reference conditions for ACIP.

Our findings correlate well with subjective expert opinions, while the validation of the framework with sophisticated subjective testing is desirable. Alternatively, the accuracy of our approach may be investigated in the future with the broadband P.OLQA model (Objective Listening Quality Assessment), which is still in standardization process in the ITU-T [17].

For a comprehensive quality rating, the framework should be further extended. Until now, only the packet
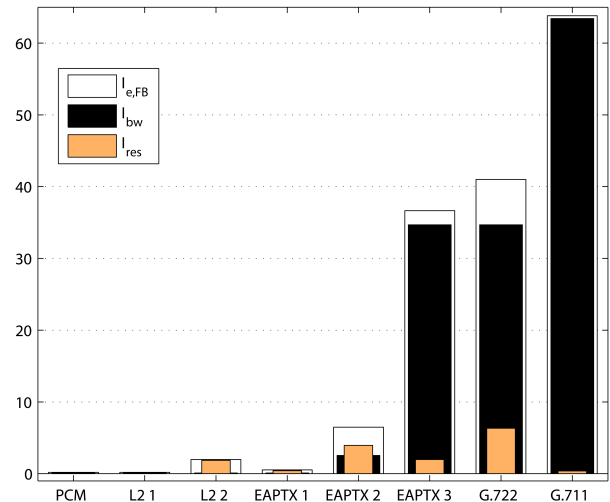


**Fig. 4:** $I_{e,FB}$ (white), $I_{bw,FB}$ (black) and $I_{res,FB}$ (beige) for the examined coding algorithms.

loss and delay impairments were parametrized, while the model can be enhanced if also parametric models for the QoE dependence on coding bitrate, packet size and loss burstiness are incorporated. Moreover, the advantage of using stereo instead of mono transmission was not considered so far. Also the E-model advantage factor *A* could be incorporated for describing an access advantage. For example, in an extreme case where someone is exclusively at a catastrophe location and has mobile communication access via mobile phone, so that he can give some live comments for a broadcasting station, then *A* will be high because the audio quality is then not the main issue, it prevails the advantage of actuality combined with the access possibility. Hence, actuality and exclusiveness of transmitted information may be considered in this measure. Furthermore, a greater set of audio coding algorithms potentially useful for professional audio communication should be evaluated (e.g. AMR-WB+, AAC ELD). Thereby, also their packet loss concealment methods should be examined with respect to the QoE for quantifying their advantage and capabilities.

### 7. REFERENCES

[1] D. Campbell, E. Jones, and M. Glavin. Audio Quality Assessment Techniques - A Review, and Recent Developments. *Signal Processing*, 89(8):1489–1500, 2009.

[2] R. G. Cole and J. H. Rosenbluth. Voice over IP Performance Monitoring. *ACM SIGCOMM Computer Communication Review*, 31(2):9–24, 2001.

[3] EBU Project Group B/AIM. EBU Subjective Listening Tests on Low-Bitrate Audio Codecs. EBU TECH 3296, Jun. 2003.

[4] EBU Project Group N/ACIP. Audio Contribution over IP - Requirements for Interoperability. EBU TECH 3326rev, 2008.

[5] M. Graubner. QoE Assessment and a Perceptually-Driven QoS Optimization Model for Audio Contribution over IP. Diploma thesis, TU Darmstadt, Germany, Sep. 2009.

[6] M. Graubner, P. Mogre, R. Steinmetz, and T. Lorenzen. A New QoE Model and Evaluation Method for Broadcast Audio Contribution over IP. submitted to: *ACM NOSSDAV '10*, Jun. 2010.

[7] J. Holub, M. Kastner, and O. Tomiska. Delay Effect on Conversational Quality in Telecommunication Networks: Do We Mind? In *IEEE WTS '07*, pages 1–4, Apr. 2007.

[8] R. Huber and B. Kollmeier. PEMO-Q - A New Method for Objective Audio Quality Assessment Using a Model of Auditory Perception. *IEEE Trans. Audio Speech Language Process.*, 14(6):1902–1911, 2006.

[9] ITU-T Rec. G.107. The E-model, a Computational Model for Use in Transmission Planning, Mar. 2005.

[10] ITU-T Rec. G.113. Transmission Impairments due to Speech Processing, Nov. 2007.

[11] ITU-T Rec. P.800.1. Mean Opinion Score (MOS) Terminology, Apr. 2003.

[12] ITU-T Rec. P.834.1 (prepublished). Extension of the Methodology for the Derivation of Equipment Impairment Factors from Instrumental Models for Wideband Speech Codecs, Apr. 2009.

[13] S. Möller. *Assessment and Prediction of Speech Quality in Telecommunications*. Kluwer Academic Publ., 2000.

[14] S. Möller, A. Raake, N. Kitawaki, A. Takahashi, and M. Waltermann. Impairment Factor Framework for Wide-Band Speech Codecs. *IEEE Trans. Audio Speech Language Process.*, 14(6):1969–1976, Nov. 2006.

[15] A. Raake. *Speech Quality of VoIP: Assessment and Prediction*. Wiley, 2006.

[16] A. Rix. Perceptual Wideband Speech and Audio Quality Measurement. In *ETSI STQ/ITU Workshop on Wideband Speech Quality in Terminals and Networks: Assessment and Prediction*, Mainz, Germany, Jul. 2004.

[17] A. Rix, J. Beerends, K. Doh-Suk, P. Kroon, and O. Ghitza. Objective Assessment of Speech and Audio Quality - Technology and Applications. *IEEE Trans. Audio Speech Language Process.*, 14(6):1890–1901, Nov. 2006.

[18] H. Sanneck and G. Carle. A Framework Model for Packet Loss Metrics based on Loss Runlengths. In *SPIE/ACM SIGMM MMCN '00*, pages 177–187, 2000.

[19] N. Seitz. ITU-T QoS Standards for IP-based Networks. *IEEE Communications Magazine*, 41(6):82–89, Jun. 2003.

[20] A. Spanias, T. Painter, and V. Atti. *Audio Signal Processing and Coding*. Wiley-Interscience, 2007.

[21] L. Sun and E. C. Ifeachor. Voice Quality Prediction Models and their Application in VoIP Networks. *IEEE Trans. on Multimedia*, 8(4):809–820, Aug. 2006.

[22] K. Vos, S. Jensen, and K. Soerensen. SILK Speech Codec. IETF Standards Track Internet-Draft, Jul. 2009. Skype Technologies S.A.

[23] M. Waltermann and A. Raake. Towards a New E-Model Impairment Factor for Linear Distortion of Narrowband and Wideband Speech Transmission. In *IEEE ICASSP '08*, pages 4817–4820, Apr. 2008.

[24] E. Zwicker and H. Fastl. *Psychoacoustics: Facts and Models*. Springer, 1999.