A New QoE Model and Evaluation Method for Broadcast Audio Contribution over IP

Maxim Graubner[1,2], Parag S. Mogre[1], Ralf Steinmetz[1] and Thorsten Lorenzen[2]
[1]Multimedia Communications Lab (KOM), Technische Universitaet Darmstadt, Germany
[2]IT Systemservice - Radio, Hessischer Rundfunk, Frankfurt a.M., Germany
contact: mgraubner@hr-online.de, parag.mogre@kom.tu-darmstadt.de

ABSTRACT

Available objective Quality of Experience (QoE) assessment methods for speech and audio quality evaluation are not directly usable for quality rating of professional broadband audio communication applications over IP such as audio contribution links for broadcasting. To fill this gap, we designed a dedicated non-intrusive parametric QoE model for conversational quality rating based on the E-model approach. With this the QoE of Audio Contribution over IP (ACIP) can be monitored. Moreover, the estimated QoE scores can be used for a perceptually-driven Quality of Service (QoS) optimization for ACIP, which has different requirements and characteristics compared to Voice over IP (VoIP). In this paper, we present our ACIP QoE model and propose an objective QoE metric for assessing the listening-only quality in ACIP. The latter is used by us for intrusive QoE evaluations, which are necessary for the derivation of a parametric QoE model. Our experimental methodology is illustrated in depth and we give exemplified results. Finally, we demonstrate the application of our model in perceptually-driven QoS optimization.

Categories and Subject Descriptors

C.4 [**Performance of Systems**]: Measurement techniques; H.4.3 [**Information Systems Applications**]: Communications Applications—*Computer conferencing, teleconferencing, and videoconferencing*; H.1.2 [**Models and Principles**]: User/Machine Systems

General Terms

Human Factors, Measurement, Performance

Keywords

ACIP, Audio Communication, AoIP, Quality of Experience, QoE, User Satisfaction, QoS, PQoS, WB-PESQ, PEAQ, PEMO-Q, E-Model, MOS

NOSSDAV'10, June 2–4, 2010, Amsterdam, The Netherlands.

Table 1: QoS requirements: ACIP versus VoIP

QoS Parameter	ACIP: discussion	VoIP
End-to-end delay	< 100 ms	150-400 ms
Audio bandwidth	12-20 kHz	3.5-7 kHz
IP data rate	80-2000 kBit/s	20-80 kBit/s
Packet loss	<< 0.2 %	< 0.2 %
Coding	MPEG/Eapt-X	G.7XX/AMR

1. INTRODUCTION

Radio broadcasters require versatile yet reliable ways of professional audio transport, capable of carrying out interactive communications, which are explicitly for broadcasting to the radio listener. *Audio contribution* in this context refers to an exchange of audio material, normally between remote sites and main studio centers. The contribution use cases range from simple outside broadcasts (e.g. sports commentary) to complex live discussions (numerous contributors at multiple locations). We focused our investigations on the latter because they are the most challenging applications in Audio Contribution over IP (ACIP). Live discussions require the provision of broadband audio quality conversational services, which must allow for high interactivity. Presently, public broadcasters establish their audio contribution links mainly using synchronous circuit-switched ISDN systems with Quality of Service (QoS) guarantees. ISDN is set to be phased out in the future, being replaced by packet-switched Voice over IP (VoIP) technology run on managed networks as well as on the best effort Internet, which both provide no more than statistical 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 vary over time. Additionally, packet loss can occur, resulting in loss of audio data.

The VoIP technology is optimized for *intelligibility*, audio contribution applications on the other hand aim for the *most accurate sound reproduction* [3]. This results in greater network bandwidth demands due to the usage of wider audio bandwidths and a possible multichannel transmission (e.g. stereo). Moreover, it results in less acceptance of coding artifacts or packet loss, which would need to be concealed by complex reconstruction. VoIP on the other hand allows for packet loss if the intelligibility is maintained. In order to ensure high interactivity many ACIP use cases require a lower end-to-end delay than VoIP allows for. This also reduces the applicability of concepts from *delay-tolerant* broadband audio streaming. In table 1 some important QoS parameters for the "live discussion" ACIP use case and VoIP are compared. QoS parameters for different ACIP use cases are provided by the European Broadcasting Union (EBU) N/ACIP project group, which also published a technical specification for ACIP defining standardized VoIP protocols such as RTP with SIP for future contribution feeds[1].

To meet the ACIP requirements, network adaptive application layer methods [2] are essential but need to be optimized with regard to the user satisfaction in terms of Quality of Experience (QoE). It is the final measure of interest and hence is the parameter to be maximized with *perceptually-driven* QoS optimization concepts [10]. In order to derive QoE metrics a mathematical QoE model needs to be available, which generates appropriate concept input. Thereby an automated QoE monitoring is possible, enabling the required QoS optimization through cross-layer QoS parameter trade-offs leading to maximized QoE.

Our contributions can be summarized as:

1. We discuss available objective QoE assessment methods for speech and audio quality evaluation and show that they are not directly usable for the quality rating of fullband ACIP (section 2).

2. We propose a non-intrusive parametric QoE model for broadband ACIP quality rating (audio bandwidth up to 22 kHz), based on the extended E-model (ITU-T rec. G.107) [7] and recent findings for wideband speech quality evaluation [11] as well as on our audio QoE experiments (section 3).

3. For the derivation of the model, we conducted QoE evaluations using several audio coding algorithms and configurations commonly used for ACIP using the objective listening quality assessment method WB-PESQ (ITU-T rec. P.862.2) [9]. We propose a new QoE metric based on a calibration of the WB-PESQ score for broadband audio signal rating with the outcome of the PEAQ method (ITU-R rec. BS.1387-1) [1], presenting our experiment methodology (section 4).

4. The obtained output can be used to derive two-dimensional *conversational Mean Opinion Score (MOSc)* surfaces [10] for conversational quality prediction, depending on packet loss and end-to-end delay. We present $MOSc$ results and discuss their usability for a perceptually-driven QoS optimization (section 5).

2. RELATED WORK

The International Telecommunication Union (ITU) provides several recommendations for the assessment of the QoE in communications. The most commonly used quality measure is the *Mean Opinion Score (MOS)* which was standardized in the ITU-T rec. P.800 for subjective determination of listening-only quality as well as the entire conversational quality [7]. It is also used for ratings that originate from objective models which automatically process measurements into estimates of a subjective opinion [9]. The P.800 MOS uses a five-grade rating scale based on the philosophy that the test subjects perceive quality between "bad" and "excellent" ($MOS \in [1;5]$). Similar opinion scales used for audio quality evaluation are defined in ITU-R rec. BS.1284-1 [1].

[1]Available: http://www.ebu-acip.org

2.1 Signal-based QoE Models

The *listening-only quality* is assessable by intrusive objective methods which compare the undistorted reference signal and the corresponding test signal to obtain a degradation rating. For speech signals, the Perceptual Evaluation of Speech Quality (PESQ) algorithm from ITU-T rec. P.862 and its wideband extension WB-PESQ (ITU-T rec. P.862.2) are available [9]. The PESQ algorithm outputs a MOS-like *PESQ score* ($\in [-0.5; 4.5]$) which can be mapped to the P.800 MOS with a transformation operation described in the standards. The Perceptual Evaluation of Audio Quality (PEAQ) algorithm from ITU-R rec. BS.1387-1 [1] is used for audio signal quality assessment. The PEAQ gives an *Objective Difference Grade (ODG)*, which can be classified as a *Difference Mean Opinion Score (DMOS)* with respect to the relative impairment rating from "imperceptible" to "very annoying" ($ODG \in [0; -5]$). A further proposal is the PEMO-Q algorithm, which is able to perform an evaluation of speech signals as well as audio signals but is not standardized. It also outputs an ODG value [1]. The PEAQ algorithm was only developed for evaluating the potential of audio coding algorithms which introduce small impairments. PEMO-Q matches better with subjective results for a higher amount of impairments, but also does not account for loss impairment. In contrast, the PESQ algorithm is able to objectively assess the quality of purely narrowband speech signals and also consider packet loss impairment. The wideband extension of PESQ can rate speech signals up to an audio bandwidth of 7 kHz, which unfortunately is still too low for broadband audio signal evaluation. Furthermore, it was not extensively tested on perceptual transform coding algorithms (e.g. MPEG) or with non-speech signals. Nevertheless, some evaluations in Ref. [8] showed, that it can also be used for the quality assessment of wideband mono audio signals.

2.2 The Parameter-based E-model

For non-intrusive parameter based speech transmission rating in bidirectional telephone services, the E-model is defined in ITU-T rec. G.107 [7]. This mathematical model estimates the quality using quantitatively measurable parameters, i.e. ordinary packet-header information. Its main advantage is the ability to predict the overall *conversational quality* due to the incorporation of both equipment impairments on listening quality as well as impairments due to transmission delay. The E-model outputs a scalar rating grade in terms of user satisfaction, the *transmission rating factor* $R \in [0, 100]$, where 100 indicates a user satisfaction of "very satisfied". The R-factor is obtained by summing up different impairment factors. The assumption behind this approach is an *impairment additivity*. The R-factor can be mapped to an estimated conversational quality MOS. For wideband transmission, the R-factor range was extended to 129, based on subjective evaluations [7]. Hence, the E-model is only suitable for rating speech communication up to wideband so far and therefore is not directly applicable to broadband ACIP applications with stricter delay requirements (see table 1).

3. 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 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 transmission impairments. The simultaneous impairment factor I_s of the E-model is neglected, because a perfect audio source is assumed to be present at the input of the audio codec which is normally the case for professional audio over IP. The useful advantage factor A is also excluded from this initial framework for now. The resulting fullband R-factor formula is then

$$R_{FB} = R_{0,FB} - I_{e,eff,FB} - I_{d,\mathrm{ACIP}} \qquad (1)$$

with the fullband basic transmission rating factor $R_{0,FB}$, the fullband equipment impairment factor $I_{e,eff,FB}$ and the delay impairment factor $I_{d,\mathrm{ACIP}}$ for audio contribution as a function of the end-to-end delay d_{e2e}. We chose a modified version of the AT&T simplified model [10] for $I_{d,\mathrm{ACIP}}$. It is a linear model with two slopes: first, the impairment only slightly increases up to a turning point at $d_{\mathrm{e2e}} = 100$ ms. Thereafter, the slope rises strongly. Additional details about this are presented in Ref. [4]. A possible echo is neglected because mostly clean-feed return lines are used.

In Ref. [5], we propose the necessary R-factor scale extension for fullband quality rating using recent findings from telecommunication quality research. In Ref. [7], a bandwidth impairment model for wideband speech communication is presented, which we extrapolated to full audio bandwidth in order to derive $R_{0,FB}$. We justified the validity of our approach by a relation comparison with different subjective results from EBU research with respect to bandwidth restriction-only impairments and similar objective measures obtained with the PEMO-Q algorithm. This results in our proposal of a fullband basic transmission rating factor of $R_{0,FB} = 157$ [5].

As in Ref. [11] for the wideband case, we defined a bandwidth impairment factor $I_{bw,FB}$, describing the perceptual effect of linear frequency distortions, by separation of the equipment impairment factor to

$$I_{e,eff,FB} = I_{bw,FB} + I_{res,FB} + I_{loss,FB}, \qquad (2)$$

where $I_{res,FB}$ represents non-linear coding distortions for the fullband case. The loss impairment factor $I_{loss,FB}$ includes a continuous QoE model for the loss-only perception, obtained by non-linear least-squares curve-fitting of discrete experiment outcomes with a logarithmic model [10]. Hence, our ACIP QoE rating framework includes separate models for the delay, loss, coding and bandwidth impairments, which are the most important ones for audio contribution applications.

We derived the necessary set of $I_{e,eff,FB}$ reference conditions during the formulation of a methodology for the derivation of fullband equipment impairment factors $I_{e,eff,FB}$, based on the objective evaluations presented in the next section 4. Thereby, the principle of recent proposals for the wideband case (ITU-T rec. P.834.1) has been extended to fullband. This enables the transformation of the objectively obtained MOS values to an $I_{e,eff,FB}$ model. Finally, a set of configuration-related functions which are non-reference models substitute the expensive full-reference quality assessment if the used coding algorithm was sometimes evaluated based on the proposed methodology, which is presented in detail in Ref. [5].

4. OBJECTIVE QOE EVALUATION

For the construction of the desired equipment impairment model, a basis of QoE rating results related to quantitative model parameters must exist, such as from the audio stream and from particular network properties. Therefore, normally, subjective tests are performed. We decided to focus on objective evaluation because subjective auditory testing is expensive and slow, 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 with the absolute category rating (ACR) method [7].

We used the WB-PESQ algorithm 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 and combined the scores of these methods to obtain a new QoE metric. We used this approach as both these algorithms have different individual advantages (see section 4.1). Ref. [6] states that: *"composite objective measures are obtained by combining basic objective measures to form a new measure"*, this approach also seems to generally apply to our QoE metric. Finally, we calculated WB-PESQ results while necessary results of the PEAQ algorithm were taken from already existent evaluations[2].

Before presenting our QoE metric, we define the most important listening-only quality degradations. These can be categorized into two classes: the *coding algorithm impairments* and possible *network impairments* due to packet loss. The coding algorithm impairments due to coding artifacts depend mainly on the performance of the algorithm type (referred to as variable c), the coding bitrate r_c as well as possible restrictions of the audio bandwidth B, which depends on the used audio sampling frequency F_s. We did not regard quantization distortions due to the sampling resolution Q and the impairment to the perceived audio quality due to different channel configurations ch (i.e. mono or stereo transmission) in particular.

The perceived audio quality based on coding impairments, assessed with some MOS_{coding} metric can be described in general as $MOS_{\mathrm{coding}} = f(\mathbf{c})$ 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 QoE analysis, the packet loss ratio ρ needs to be $\rho = 0$. The following equations show the principal dependence of the perceived audio quality assessed by a MOS on the principal coding parameters of a specific coding algorithm where only one parameter is varied,

$$MOS_{\mathrm{coding}}(r_c) \propto r_c, \qquad (3)$$

$$MOS_{\mathrm{coding}}(B) \propto B. \qquad (4)$$

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 of the CBR audio stream L, because the longer the packet size is the more consecutive audio information gets lost. These impairments can also be described by a QoE

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

metric, now with respect to loss as $MOS_{\text{loss}} = f(\mathbf{P}_{\text{loss}}, \mathbf{c}, L)$ while the packet loss process here is characterized by the state transition matrix $\mathbf{P}_{\text{loss}}$ of a 2-state Markov model [7]. This relatively simple two-parameter model fully describes a short-term loss process allowing for loss correlation. For our evaluations with respect to packet loss, we chose the following parameter sets determining the two parameters of the Markov model [3],

$$\text{Set for } \rho : \mathcal{P} = \{\ 0,\ 0.2,\ 0.5,\ 1.5,\ 3,\ 5,\ 10\ \}\ [\%] \quad (5)$$

$$\text{Set for } \mu : \mathcal{Q} = \{\ 1,\ 1.18,\ 1.43,\ 2\ \} \quad (6)$$

where ρ is the packet loss ratio and μ the mean loss period (MLP) in packets.

4.1 Proposed QoE Metric

For our approach of an objective QoE metric for ACIP listening quality based on the outputs of the WB-PESQ and the PEAQ algorithm, 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. In contrast to PEAQ, WB-PESQ accounts also for possible impairments due to packet loss but is limited to wideband quality rating. Hence, we considered the outcomes of both models for the MOS_{coding} and only the WB-PESQ scores for the MOS_{loss} in a $DMOS_{\text{loss}}$, excluding the WB-PESQ scores MOS_{A} for the packet loss ratio $\rho = 0$ case, characterizing the coding-only impairments,

$$DMOS_{\text{loss}}(\mathbf{P}_{\text{loss}}, \mathbf{c}, L) = MOS_{\text{A}} - MOS_{\text{A}}(\rho = 0)\ . \quad (7)$$

The total objective QoE metric $MOS_{\text{obj}}(\mathbf{P}_{\text{loss}}, \mathbf{c}, L)$ is then formulated as

$$MOS_{\text{obj}}(\mathbf{P}_{\text{loss}}, \mathbf{c}, L) = MOS_{\text{coding}} + DMOS_{\text{loss}}\ , \quad (8)$$

where the MOS_{coding} remains the WB-PESQ score $MOS_{\text{A}}(\rho = 0)$ if the used sampling frequency is $F_s \leq 16$ kHz. For higher sampling frequencies, we propose to use a linear combination between the MOS-like WB-PESQ score MOS_{A} and the transformed ODG of the PEAQ algorithm MOS_{B} as

$$MOS_{\text{coding}}(\mathbf{c}) = \frac{1}{\alpha + \beta}[\alpha MOS_{\text{A}}(\mathbf{c}) + \beta MOS_{\text{B}}(\mathbf{c})] \quad (9)$$

Thereby the PEAQ MOS_{B} results from the transform

$$MOS_{\text{B}} = 4.5 + ODG \quad (10)$$

based on the justifiable assumption that the $ODG = 0$ case (no degradations) corresponds to an objectively obtained MOS value of 4.5 as in Ref. [1], similar to the *WB-PESQ score* to *MOS* mapping in the ITU-T rec. P.862.2. We chose the variables α and β from the linear combination empirically as $\alpha = 3$ and $\beta = 1$ during our metric and model design, details for the choice are provided in Ref. [3] and [4]. Finally, the resulting $MOS_{\text{obj}}(\mathbf{P}_{\text{loss}}, \mathbf{c}, L)$ can be transformed to an equipment impairment factor $I_{e,eff,FB}$ for the fullband E-model using the methodology mentioned in section 3.

4.2 Experimental Test Setup

The setup principle with its main components is depicted in fig. 1. We performed the experiments with the network emulator NetDisturb by ZTI[3], which forces packets to be

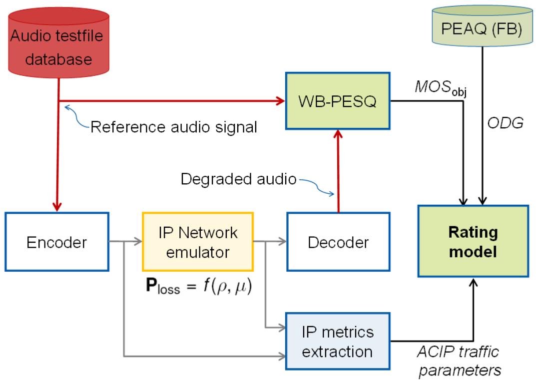

Figure 1: QoE evaluation environment.

lost based on the desired loss process emulation. NetDisturb is used because this tool has the capability of also including selected deterministic packet loss traces in textfiles. We generated the desired Markov model based packet loss process realizations using the numerical computing environment Matlab[4]. NetDisturb can filter the RTP audio stream, with this the packet loss impairments based on the Markov model are only applied to the media flow and other IP packets in the network such as control messages are ignored for the packet loss process as desired. The parameter sets for the loss model presented above led to 25 network packet loss processes, which were evaluated 4 times on each of the 6 different test signals[5] for each coding algorithm configuration. We chose the test signals carefully to reflect the range of possible audio signals in audio contribution [4]. We use dedicated professional hardware audio codecs[6] for the audio coding and decoding as well as for the IP stream generation from the digital audio signals. The IP stream is monitored using the Wireshark[7] network analyzers before and after the IP network emulator to extract important IP performance metrics which are delivered to the QoE model. The audio playout and recording is done simultaneously with Matlab and a professional digital audio card. The later QoE analysis is also done with Matlab.

4.2.1 *Studied Coding Algorithms*

For our evaluations, we selected a set of algorithms theoretically suitable for ACIP. Table 2 summarizes the chosen algorithms and their configurations. All coding algorithms only used silence insertion for their packet loss concealment (PLC) strategy. For the QoE analysis with respect to the packet size, we selected appropriate values, which enable a comparison with respect to the impact of the packet size on a lossy link for a given coding algorithm. Because of the necessary high effort for evaluating an algorithm configuration for all chosen packet loss process characterizations, we chose only extreme values for the packet size.

[3] Available: http://www.zti-telecom.com

[4] Available: http://www.mathworks.com

[5] The signals were taken from the set of PEAQ audio test signals from the professional QoE evaluation software Opera by Opticom. Available: http://www.opticom.de

[6] We used three different devices from APT (http://www.aptcodecs.com), MAYAH (http://www.mayah.com) and ORBAN (http://www.orban.com)

[7] Available: http://www.wireshark.org

Table 2: Set of coding algorithms selected for the QoE evaluation and their basic configurations.

Coding Algorithm	Channels	r_c [kbit/s]	F_s [kHz]	IP Packet Size [Byte]	Inter-packet Time [ms]
MPEG Layer2 (1)	2	384	48	1196	24
MPEG Layer2 (2)	2	256	48	812	24
Eapt-X (1) a/b	2	384	48	1100/332	21.25/5.25
Eapt-X (2) a/b	2	256	32	828/336	24/8
Eapt-X (3)	1	64	16	208	16
ITU-T G.722 / G.711	1	64	16/8	200	20

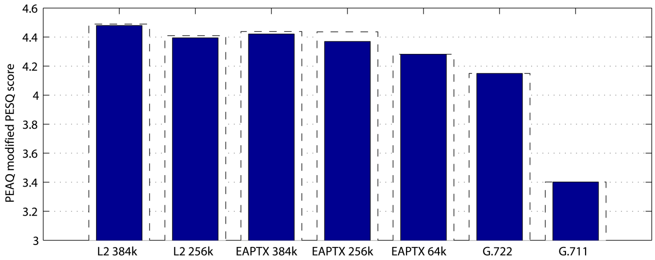

Figure 2: Our QoE metric (blue) in comparison to the raw WB-PESQ scores (dashed).

4.3 Selected Results

In fig. 2, the PEAQ modified WB-PESQ results of our QoE metric proposal in comparison to the raw WB-PESQ scores for the evaluated coding algorithms are depicted for the $\rho = 0$ case. For the Eapt-X coding algorithm it can clearly be seen that the WB-PESQ algorithm could not deal with the differences of Eapt-X with $F_s = 48$ kHz and coding rate 348 kbit/s ("EAPTX 384k") and another Eapt-X coding with $F_s = 32$ kHz and coding rate 256 kbit/s ("EAPTX 256k"). WB-PESQ is only evaluating at $F_s = 16$ kHz, therefore it can not determine impairments on audio frequencies above 8 kHz. Hence, also the advantage of using broadband transmission can not be reflected in the results. The WB-PESQ scores for the mentioned Eapt-X algorithms indicates, that both codecs have similar impairments on the lower frequencies, which is reasonable because of their subband structure. The results of our proposed QoE metric principally have the desired relations and are well correlated with subjective impressions of the experts. More results in the objective listening MOS domain especially for the different loss process conditions can be found in Ref. [3].

5. SIMPLIFIED MODEL FOR MOSC

We were able to use our objective results in order to build simplified two-dimensional parametric models of the *conversational MOS* ($MOSc$) [10]. We obtained the $MOSc$ results by transformation of the objective QoE evaluations $MOS_{\mathrm{obj}}(\mathbf{P}_{\mathrm{loss}})$ for a specific coding algorithm and configuration $\mathbf{c}$ as well as packet size L to a fullband equipment impairment factor $I_{e,eff,FB}$ in the R-factor domain enabled by the QoE rating model framework for ACIP described in section 3. The resulting R_{FB} from eq. (1) can then be transformed to an estimate of the $MOSc$, depending continuously on the overall packet loss ratio ρ and the end-to-end delay d_{e2e}, which are the most important QoS parameters besides the available bandwidth a_{bw}. The latter one is not considered directly in the QoE model, but it limits the maximum coding bitrate which can be used, $r_c = f(a_{bw})$.

The principal dependencies of a continuous $MOSc$ model can be described as

$$MOSc = f(\rho, d_{\mathrm{e2e}})|_{\mu=\mu_k,\mathbf{c}=\mathbf{c}_i,L=L_j(\mathbf{c}_i)} \ . \qquad (11)$$

Here the codec configuration $\mathbf{c}_i$ is taken from the set of coding algorithm configurations $\mathcal{C} = \{\mathbf{c}_1, \mathbf{c}_2, \ldots, \mathbf{c}_i, \ldots \mathbf{c}_I\}$ of size I, while the packet size L_j is taken from the respective set of possible packet length $\mathcal{L} = \{L_1, L_2, \ldots, L_j, \ldots, L_J\}$ of size J. Moreover, for the different discrete mean loss periods μ_k ($k = 1 \ldots 4$) from the set $\mathcal{Q}$ of size $K = 4$ defined in eq. (6), dedicated parameter models can be accessed. The $MOSc$ surfaces can finally be least-squares fitted using a general polynomial model [10].

Fig. 3 and fig. 4 depict exemplary $MOSc$ surfaces for $\mu = 1$ of dedicated ACIP algorithms. Fig. 3 shows the resulting model for the MPEG Layer 2 coding algorithm with a 384 kbit/s coding bitrate and $F_s = 48$ kHz, while fig. 4 shows the one for the Eapt-X coding algorithm with 64 kbit/s and $F_s = 16$ kHz. In both figures a QoE result is marked, which corresponds to a typical non-optimized ACIP operation, assuming a network delay of 70 ms and a playout buffer delay of 50 ms, which represent minimal VoIP contribution values to the overall delay [7]. For the packet loss $\rho = 0$ is assumed. Even if the MPEG algorithm in fig. 3 has the desired audio quality, the high coding delay reduces the conversational quality rapidly. If loss were present, we could not recommend its usage with IP communication, even if it was specified by the EBU.

The $MOSc(\rho, d_{\mathrm{e2e}})$ models can be used for estimating the conversational quality in ACIP systems based on non-intrusive passive measurements of the network QoS parameters. This enables a perceptually-driven QoS optimization, e.g. by choosing a perceptually optimal rate control at the sender-side or a perceptually optimal playout buffer size at the receiver-side. For example, the optimum playout buffer size is a trade-off between buffer delay d_b and late loss ρ_b, which is the possible loss due to buffer constraints additional to the network packet loss ρ_n, $\rho = \rho_n + \rho_b$. Sometimes accepting a late loss (ideally concealed thereafter) may be perceptually more meaningful than a higher playout buffer size to cope a wider range of network delay variation Δd_n (jitter), because this would increase the overall end-to-end delay which is desired to be as small as possible to ensure interactivity. In fig. 4 such a trade-off is depicted (thick arrow on the surface). Allowing for late losses, the conversational MOS value is increased while the delay is decreased because the playout buffer delay d_b is lowered. Hence, a QoS optimization is achieved.

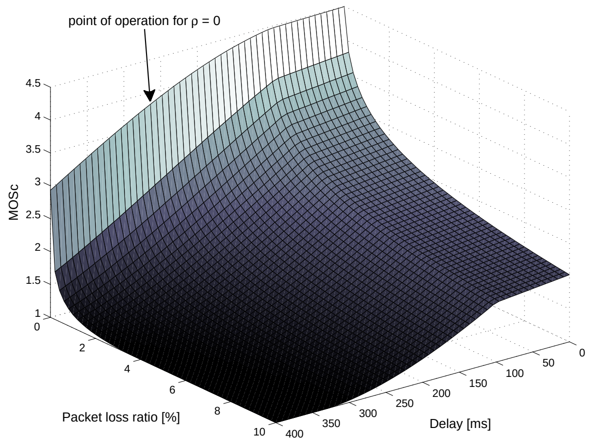

Figure 3: *MOSc* surface for MPEG Layer 2 (1).

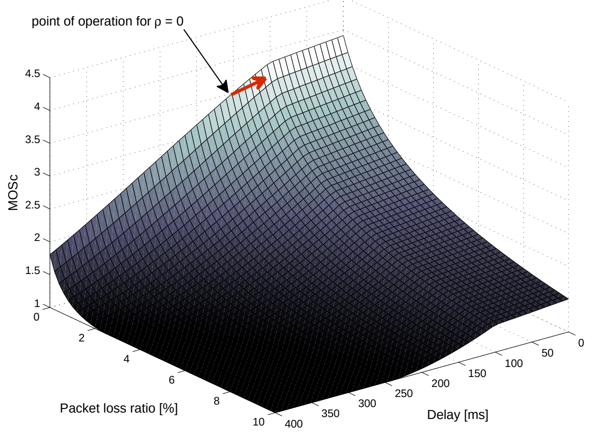

Figure 4: *MOSc* surface for Eapt-X (3).

6. CONCLUSIONS AND FURTHER WORK

In this paper, we showed that the available objective QoE assessment methods for speech and audio quality evaluation are not directly usable for the quality rating of broadband ACIP. We proposed a dedicated non-intrusive parametric QoE model for conversational quality rating, based on the E-model approach, as well as a QoE metric and experiment methodology for assessing the listening quality in ACIP. Moreover, we derived $MOSc$ surfaces and gave an example for their application in perceptually-driven QoS optimization.

Our findings correlate well with the opinion of our experts, 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 P.OLQA model (Objective Listening Quality Assessment) for broadband quality rating, which is still in standardization process in the ITU-T.

The proposed QoE prediction framework and the simplified model for perceptually-driven QoS optimization purposes provides a basis for the future fulfillment of ACIP requirements. For a comprehensive quality rating, the framework should be further extended. Until now, only the packet loss, audio bandwidth restriction and delay impairments were parametrized, while the model can be further simplified if also parametric models for the QoE dependency on coding bitrate, packet size and loss burstiness are incorporated by curve and surface fitting. Also more factors of the E-model such as the promising advantage factor A may be taken into account to include a possible access advantage. Furthermore a greater set of audio coding algorithms potentially useful for professional audio communication should be included (e.g. MPEG AAC ELD). Thereby, the provided 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] P. A. Chou and M. v. d. Schaar, editors. *Multimedia over IP and Wireless Networks: Compression, Networking, and Systems*. Academic Press, 2007.

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

[4] M. Graubner et al. A New QoE Model, Evaluation Method and Experiment Methodology for Broadcast Audio Contribution over IP. Techn. rep., TU Darmstadt, Germany, Mar. 2010. available: ftp://ftp.kom.tu-darmstadt.de/TR/KOM-TR-ACIP.pdf.

[5] M. Graubner, P. Mogre, and T. Lorenzen. QoE Assessment for Audio Contribution over IP (ACIP). submitted to: *AES 38th International Conference on Sound Quality Evaluation*, Jun. 2010.

[6] S. Quackenbush, T. Barnwell, and M. Clements. *Objective Measures of Speech Quality*, volume 16. Prentice Hall, 1988.

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

[8] A. Rix et al. Perceptual Wideband Speech and Audio Quality Measurement. In *ETSI STQ/ITU Workshop on Wideband Speech Quality*, Jul. 2004.

[9] A. Rix. Objective Assessment of Speech and Audio Quality - Technology and Applications. *IEEE Trans. Audio Speech Language Process.*, 14(6):1890–1901, Nov. 2006.

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

[11] 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.