 DOI: 10.1109/QoMEX.2011.6065686

A NEW DIMENSION-BASED FRAMEWORK MODEL FOR THE QUALITY OF SPEECH COMMUNICATION SERVICES

Sebastian Möller [1], Jens Berger [2], Alexander Raake [3], Marcel Wältermann [1], Benjamin Weiss [1]

[1] Quality and Usability Lab, Deutsche Telekom Labs, TU Berlin, Germany
[2] SwissQual AG, Zuchwil, Switzerland
[3] Assessment of IP-based Applications Lab, Deutsche Telekom Labs, TU Berlin, Germany

ABSTRACT

In this paper, we identify quality dimensions which are relevant for speech communication services, such as mobile telephony or Voice-over-IP. These include dimensions perceived when listening to degraded speech, talking against echoes, double-talk capabilities, interacting with delay, conversing over channels with time-varying characteristics, and service-related dimensions experienced during speech connection set-up and maintenance. For each dimension, we review subjective evaluation metrics and instrumental quality prediction models. We group these dimensions in a framework model which is able to diagnostically assess speech communication services, and may be used for monitoring and maintenance.

Index Terms **—** speech quality, speech communication, subjective evaluation, quality prediction model

1. INTRODUCTION AND MOTIVATION

Despite the advances made in telecommunication and signal processing, speech quality has become an issue for modern communication services again. Reasons are non-guaranteed and time-varying link characteristics, as well as a paradigm shift in the planning, set-up and maintenance of telecommunication networks, namely from a Quality of Service (QoS) viewpoint, in which individual transmission parameters were taken as criteria, to a Quality of Experience (QoE) viewpoint, in which the overall quality perceived by the user is used for optimization.

As a result, a large number of models emerged which try to predict different aspects of QoE on the basis of QoS parameters. The results of these models are usually not comparable amongst each other, as they address different aspects of QoE. The aim of our paper is to structure the speech quality space. The scenario we address is a speech communication service, in particular mobile telephony and Voice-over-IP (VoIP). The space we address contains all quality aspects related to listening, talking, conversation, stability, communication (link set-up and maintenance), situational and service-quality aspects. Aspects which we explicitly exclude are costs and contract conditions, requiring field-related evaluation methods which are currently out of our reach. For each quality aspect, we identify one or more perceptual quality dimensions, which frequently are linked to specific technical causes, see Section 2.

Although no new empirical data is given, we consider the analysis very helpful for planning, implementation, monitoring and maintenance. For this purpose, we define a new framework model which includes all mentioned dimensions. For each component of the framework model, we review available sub-models as building blocks. We propose to combine them to come up with more meaningful integral quality estimations, but also to externalize individual quality metrics for diagnosis. Section 3 provides the details of the proposed framework model. Some conclusions and plans for future work are given in Section 4.

2. SPEECH QUALITY DIMENSIONS

The idea of speech quality dimensions stems from the definition of quality. According to Jekosch [1], speech quality can be seen as the "result of appraisal of the perceived composition of an entity with respect to its desired composition". The perceptual space is thought to be composed of several inter-dependent factors, so-called dimensions. Perceptual dimensions do not need to be orthogonal, but an orthogonal description is interesting in order to avoid redundancy in the formalization. In the perceptual space, quality can be calculated from an (e.g. Euclidean) distance between the perceived (as a point in the multidimensional space) and the desired. Both perception and desire are influenced by the situation, as they "happen" at a discrete point in time; thus, we call them "events", following the notion of Blauert [2].

In order to determine orthogonal dimensions, two complementary methods are commonly used in psycho-physics: (1) the description of perceptual events on a number of bipolar "Semantic Differential" (SD) scales [3] and a subsequent factor analysis; (2) or a distance or similarity scaling of the perceptual events and a subsequent multidimensional analysis (MDS) [4]. Whereas the former aims at a better interpretability of the results (due to describing attributes), the latter aims at the completeness of the perceptual space, in terms of dimensions covered. If both methods lead to the same dimensions, we consider them as stable for a given stimulus space, situation and user group.

Not all perceptual dimensions contribute to quality in the same way. For example, there are dimensions of the type "the-more-the-better" or "the-less-the-better" for which the degree of presence/absence is decisive (such as noisiness, discontinuity). Such dimensions correspond to a "vector model" type. In turn, there are other dimensions which show an optimal point (such as loudness); this is called an "ideal-point model" [5].

Unfortunately, speech communication services are highly dynamic, and the perceived quality significantly depends on the behavior of the communication partners. Thus, there are no stable "stimuli", and no approach of an MDS or SD for identifying the dimensions of interactive speech communication services has been documented to the authors' knowledge. As a consequence, we have to combine different sources of knowledge, and hence the quality aspects and dimensions cited hereafter may be partial and incomplete, in a mathematical sense. In the following, we first address the aspects of the conversational situation (listening, talking, conversational), and then aspects which go beyond the direct perception (stability, communication set-up and maintenance, and service aspects).

2.1. Listening dimensions

Many analyses have been performed to extract the quality dimensions when listening to individual short (mostly 4-8 s) speech samples. Both SD and MDS approaches have been followed and – depending on stimuli and listener characteristics – different dimensions have been extracted, including the following ones (references in [6][7][8][9]):

- Intelligibility or clarity
- Naturalness or fidelity or speaker recognizability
- Interruptedness or discontinuity, or sub-dimensions such as clipping or bubbling
- Noisiness, or sub-dimensions such as hiss or whistling
- Loudness
- Color of Sound, or sub-dimensions such as brightness, height, spectral fullness, sharpness or nearness
- Distinction between background noise and signal distortions

The last dimension points at a side-effect which is not directly due to listening to the (transmitted and potentially noisy) speech signal, but may also be linked to the background noise situation, as these effects might overlap.

The question arises how many dimensions are actually necessary to fully describe the perceptual quality space associated with listening to transmitted speech. In a quiet environment, Wältermann et al. [7] came up with three vector-model dimensions (coloration, discontinuity, noisiness) capturing modern narrowband to wideband transmission equipment. Their later assessments include loudness as an additional (non-orthogonal) dimension, resulting in a 4-dimensional space. Sen [9] proposed coloration and discontinuity to be split into two sub-dimensions each (high-frequency absent and low-frequency absent, slowly-varying and rapidly-varying discontinuity), resulting in a 6-dimensional space.

2.2. Talking dimensions

Talking per se does not lead to a specific dimension; it only relates to quality when the listener perceives what s/he said, or when the talking process is otherwise compromised. Commonly, there is a feedback of the talker's own voice through the so-called sidetone channel in systems where the ear is partially shielded (e.g. by a handset), to compensate for the shielding loss. This sidetone is provided e.g. via a direct coupling of the microphone signal to the loudspeaker, and gives a feedback of the signal to the talker. Sidetone quality shows an ideal-point, and may be reduced when the sidetone signal is either too high (leading to improper placement of the handset/headset if the sidetone level persists), too low (leading to speaking up or to doubts about the proper functioning of the device), or otherwise strongly distorted (leading to the impression of a disguised voice).

Talking may be further compromised in case of acoustical or electrical echoes. Echoes should ideally be absent (i.e. a vector model applies), and echo degradations are characterized by their loudness and by their time lag. Subjective assessment methods for echoes and sidetone include "talking-and-listening tests" (ITU-T Rec. P.831), "third-party listening tests" (ITU-T Rec. P.831), and conversational tests (ITU-T Rec.s P.800, P.805). In order to avoid echoes, echo cancellation devices sometimes restrict the talking capability of one side in favor of the other, making double-talk (that is both parties speaking at the same time, which is a natural phenomenon in human conversations) impossible. Reduced double-talk capability can already be perceived in a pure talking-and-listening situation, when the voice of one communication partner is switched off or significantly attenuated by the own voice. It can be assessed in specific double-talk tests (ITU-T Rec. P.832), in "third-party listening tests" (ITU-T Rec. P.832), or in conversation tests (ITU-T Rec.s P.800, P.805).

2.3. Conversational dimensions

Conversational quality is of course impacted in case that the listening is degraded to an amount which makes it difficult to understand the other person, or which makes relevant paralinguistic or extralinguistic features of the communication partner difficult to grasp. On the other hand, conversational quality may also be impacted in case that talking behavior is restricted. In particular, if the communication channel characteristics depend on the communication behavior (e.g. in case of echo cancellation or level-switching), this might impact the natural flow of the conversation, as one partner has to stop if the other takes the floor, and it is difficult for that partner to provide feedback or acknowledgements.

Even in case of “transparent” transmission without specific listening or talking degradations, conversational quality may be impacted by transmission delay (cf. “interaction quality” in [11]). The effect of pure delay on conversational behavior has been the object of many investigations [5][6], without a conclusive result. It is now generally acknowledged that the quality impact of delay is difficult to capture in a standard conversation test, as it depends strongly on the range of other degradations in the system, the instructions to and expertise of the test participants, the personality of the communication partners, etc. Thus, conversation tests as described in ITU-T Rec.s P.800 and P.805 make use of different scenarios to address judgment differences.

2.4. Stability dimensions

We refer to the “stability” as the ability of the system to provide a stable transmission channel over the entire duration of the communication (usually a call), without excessive quality changes. Such quality changes may stem from transmission errors (frame loss in 2G and 3G mobile telephony, packet loss in VoIP), handovers between cells or technologies, changeovers of codecs (even changing audio bandwidth), as well as combinations hereof. There could even be systematic instabilities such as convergence effects at the beginning of a call. The associated audible effects range from interruptions (gaps), signal artifacts like noise bursts, clicks or tones, fading signal segments, to audible changes in the audio bandwidth. They may be assessed either via instantaneous ratings with a slider (ITU-T Rec. P.880), or with simulated conversations where participants are asked to listen to 8-16 s segments arranged in a logical order, answer questions after each segment, and judge the quality after 1-2 minutes (ETSI TR 102 506).

Several investigations addressed the effect of time-varying quality on the rating at the end of a call (see [10] and [5] for an overview). Averaging over typical segments is usually not sufficient, as negative segments have shown to have a higher impact on the call-final rating than positive ones. In addition, segments being temporally close to the judgment moment are more important than earlier segments. This effect can either be described via a constant proportion of time relative to the overall communication time span (so-called “recency effect”), or a fixed time period before the judgment (sometimes called an “end effect”).

2.5. Communication dimensions

The conversation quality experienced during an entire call is not the only determining factor for judging the communication situation. The call first has to be set up, and then maintained for the desired duration. Call set-up may be a determining factor especially when calls are frequent (business scenarios), telephone numbers are long and not memorized, the technical call setup time after dialing is unexpectedly long, or the network does not provide feedback within an expected time (e.g. via a ring-back tone). Sometimes several attempts have to be made before accessing the conversation partner. Degradation in call set-up is largely determined by the time it takes, but also other mental resources may be decisive (e.g. memory load for remembering long telephone numbers, cognitive effort in navigation through mobile phone interfaces before accessing the call, etc.). Once the call in established, technical reasons may interrupt the connection and force the user to re-dial. This is a frequent phenomenon in mobile situations (e.g. when calling from a train), and is particularly annoying to users, as it both interrupts the flow of information and requires additional effort and money.

2.6. Situational dimensions

The communication situation has an impact both on the perception process and on the user’s expectations. Background noise is apparently an influencing factor on the listening and talking quality and even informs about social components e.g. non-privacy. In addition, the user’s mobility contributes to time-varying transmission characteristics in terms of stability. With respect to the expectations, users seem to establish a trade-off between conversation quality and the communication situation they are in, frequently quantified in terms of an “advantage of access”. Such an advantage of access has been quantified e.g. for mobile vs. fixed telephony or hard-to-reach areas. The same trade-off is sometimes performed by mobile network operators in case of popular events (like the New Year’s Eve), when they reduce transmission bandwidth to increase the number of users per cell.

2.7. Service dimensions

A user’s perception of a service s/he is frequently using depends of course on the individual instances of service usage, however not necessarily in a one-to-one way. Duncanson [12] investigated the relationship between a user’s ‘general impression’ of his/her ordinary telephone service and the judgments given to particular (but not manipulated) calls. He found that asking the user about a particular call commonly resulted in more positive ratings than their ‘general impression’ of the service was. On the opposite, new investigations with time-varying and particularly manipulated VoIP calls [13] showed that averaging of individual call experiences led to more pessimistic judgments than they were obtained after a 12-days usage experience.

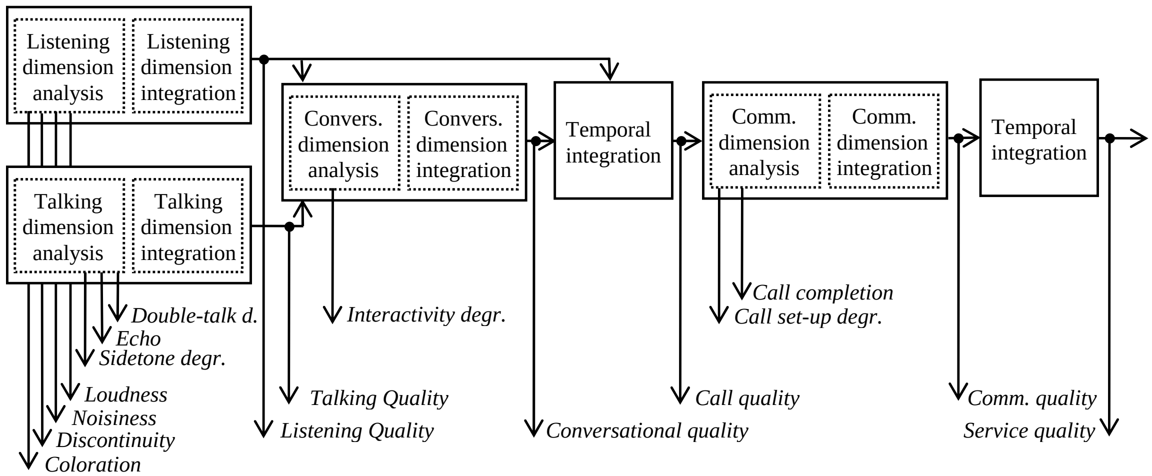

Figure 1: ***Dimension-based framework model for speech quality.***

3. DIMENSION-BASED FRAMEWORK MODEL

The mentioned dimensions of speech communication quality can be arranged in a model which provides both diagnostic and integral quality information at different levels. This framework model can be used for different purposes:

- Relationships between perceptual quality dimensions can be analyzed
- Framework components can be replaced by prediction models providing estimations of diagnostic or integral quality dimensions
- Consequences of individual degradations can be foreseen more easily
- Origins of bad quality can be diagnosed

In the following paragraphs, we review existing quality prediction approaches which can be seen as building blocks for the framework model.

3.1. Listening dimension models

Standard models for listening-only quality rely either on signals measured at the end of the transmission channel under study (ITU-T Rec.s P.862, P.863), on parameters describing the equipment used for setting up the channel (the so-called E-model in ITU-T Rec. G.107), or on packet information collected during service operation (ITU-T Rec.P.564). They all provide estimations of a listening-quality Mean Opinion Score, MOS.

Dimension-based models have also been proposed recently. The model developed by Côté [8] provides estimations of the 4 listening dimensions coloration, discontinuity, noisiness and loudness, and integrates them via a k-means regression with an overall quality estimation which is similar to the one from ITU-T Rec. P.862. Sen [9] provided estimations for his 5 listening dimensions, including two for time-varying and two for coloration sub-dimensions. Wältermann in [14] proposed a dimension-based parametric model similar to the one in ITU-T Rec. G.107.

3.2. Talking dimension models

The only model available so far for estimating talking-only quality is the PESQM proposed by Appel & Beerends [15]. It makes use of a P.862-like approach by comparing the input and return output signals of the system under study, the latter being compromised by echo and sidetone distortions. The output of the model is an estimation of the talking quality, comprising both echo and sidetone degradations. The E-model considers these degradations as well.

Double-talk capabilities have not yet been the subject of modeling approaches. ITU-T Rec. P.340 defines 3 classes of hands-free terminals depending on their double-talk capability: Full duplex capability (attenuation of the conversation partner unnoticeable, background noise transmitted in the send path), partial duplex capability (noticeable attenuation of the conversation partner), and no duplex capability (when one conversation partner is talking, the other is fully attenuated, no background transmission in sending direction). It is expected that even more fine-grained models can be defined for this dimension, depending on the attenuation and echo cancellation.

3.3. Conversational dimension models

A rough dimension-based model for conversational quality has been proposed by Gueguin et al. [16]. It combines a listening quality estimation from ITU-T Rec. P.862 with a talking quality estimation from PESQM and an evaluation of delay impact which is taken from ITU-T Rec. G.107. The latter defines a delay impairment factor which roughly reflects the delay degradation – in terms of MOS ratings – of highly interactive conversations.

3.4. Stability dimension models

Several researchers have addressed the question of how time-varying quality is integrated into a judgment for a longer speech stimulus, cf. [5] for an overview. First investigations, e.g. by Hansen et al [17], Rosenbluth [18] and Gros and Chateau [19] mainly addressed the listening situation, whereas Weiss et al. [10] focused on the conversation situation. The resulting integration models mainly contain three aspects: A first rough estimation of the final judgment can be obtained by averaging the individual judgments for short stretches of speech. This rough estimation can be improved by putting a higher weight to pronouncedly negative short-term judgments, as well as by putting a higher weight to the short-term judgments which are temporally close to the final judgment. The latter aspect is frequently modeled by a recency effect, although Weiss et al. found better correlations when a fixed time window with respect to the final judgment is emphasized. They also showed that the approach is applicable to the integration of both subjective short-term judgments as well as estimations of the judgments obtained using ITU-T Rec. P.862.

3.5. Communication dimension models

Models for predicting communication dimensions other than speech conversation quality or call quality are rare. Kort [20] provides formulae for the probability of customers to abandoning before the dial tone, while dialing (depending on the number of digits to be dialed), and before network response. He also provided probabilities to terminate a call early in case of excessively bad conversation quality, for re-dialing, and for operator complaints, using Gaussian error functions in conjunction with the transmission rating of the Bellcore Transmission rating model (a predecessor of the E-model) as the argument. Although these models are very interesting for network operators, they will be highly dependent on the network and the customer group they have been determined for. In case of Kort, these were users of the AT&T PSTN network in the late seventies, which will not be representative for today’s telephone users any more.

3.6. Situational dimension models

A few attempts have been made to describe the situational dimension of telephone service quality. A rule-of-thumb assumption is that the situational context can rule out half of the particular degradation which is typical for this context, in terms of a quality bonus. This quality bonus is included in the E-model as an “advantage of access” factor *A*. Values for *A* have been determined by halving the particular impairment which is typical for the service under consideration, namely the codec degradation in case of cordless and mobile telephony, or the delay degradation in case of satellite connections to hard-to-reach areas. It is acknowledged, though, that this modeling provides only a very rough estimation of this trade-off, and that the trade-off will vary with the habituation to the particular type of service, see a hypothetical behavior description in [6].

3.7. Service dimension models

First models for integrating individual communication service experiences into a rating for an entire service are currently being derived for audio-visual IP-based communication services, based on the data presented in [13]. As with a stability dimension model, simple averaging of the rating of individual calls is not sufficient for predicting the rating of the service. In contrast to the findings in [12], the rating after the 12-days trial period was found to be more positive than the arithmetic mean; this could be modeled with the help of a progressive slope. In addition, remembering effects after particularly bad experiences have to be taken into account. Further investigations are necessary to obtain stable models for such service dimensions.

3.8. Quality integration models

Besides the temporal aspect, the question arises how the multitude of dimensions can be integrated into one index of the “overall” or “integral” quality of the particular aspect (listening quality, talking quality, conversation quality, communication quality). This integration has been depicted at four instances of Figure 1, and can be obtained in different ways.

Following the definition of quality, we may assume that the perceived event is defined by an *N*-dimensional vector *p* in the perceptual space, and the expected/desired event by a corresponding vector *q*:

$$p = (p_1, p_2, \ldots, p_N) \quad q = (q_1, q_2, \ldots, q_N) \qquad (1)$$

with *N* the number of relevant quality dimensions. The integral quality *Q* can then assumed to be the Euclidean distance between *p* and *q*, i.e.

$$Q = d(p, q) = \sqrt{\sum_{i=1}^{N} \alpha_i (p_i - q_i)^2} \qquad (2)$$

where α_i determines the weighting of the particular dimension *i* for the overall quality. As it has been stated earlier, not all dimensions contribute in the same way to quality. Whereas ideal-point dimensions are directly applicable to the Euclidean distance approach, vector-model dimensions first have to be transformed before being applied in Eq. (2). For the integration of listening quality dimensions, this method has proven to provide quite accurate results [21].

Côté [8] used a k-means clustering algorithm to integrate his 4 dimension estimators for coloration, discontinuity, noisiness and loudness with an estimation of a P.862-like approach to form an overall estimation of listening quality.

Whereas the approach provided better performance than ITU-T Rec. P.862 on a large set of narrowband, wideband and super-wideband databases, it was outperformed by a predecessor of ITU-T P.863 in the ITU-T competition carried out between 2009 and 2010 which was not based on individual quality dimensions.

4. CONCLUSIONS AND FUTURE WORK

We identified quality dimensions which are relevant for speech communication services, and which cover different levels from listening, talking and conversational aspects up to communication, situational and service aspects. For each aspect and dimension, we reviewed available model components and aggregated them into a framework model which displays the pipeline of influences from listening to service quality. The framework model does not only provide one single output, but many, on different levels. These outputs correspond to perceptual quality dimensions which have proven to be important for service quality.

Although the framework model has not yet been implemented, we assume that it will provide reliable and robust estimations of quality. By separating the quality formation process into different aspects, we think that the prediction gets more reliable than it would be possible with a one-step procedure. Further separating quality aspects into perceptual quality dimensions increases the chances that today unknown transmission scenarios will also be captured, as we assume that the perceptual dimensions will be more stable than signal characteristics which are closely linked to the signal processing used. Depending on the type of input information available (e.g. signals, parameters, or protocol information), individual prediction modules of the framework can be exchanged without impacting other, e.g. integration modules.

The fact that our framework model has multiple outputs makes it particularly attractive for quality monitoring and diagnostics. We assume that stable diagnostics is linked to perceptual dimensions, and not necessarily to technical causes. However, for the listening quality aspect, it has been shown that there is a good congruence between perceptual dimensions and technical causes. For example, the coloration dimension has been associated with distortions of the linear frequency response, the discontinuity dimension with packet loss, and the noisiness dimension with the level of background and circuit noise. Despite this congruence, Study Group 12 of ITU-T has decided to launch two study items, one on the estimation of perceptual dimensions of listening quality (P.AMD) and one on the corresponding technical causes (P.TCA). Further work in this group is dedicated to signal-based modeling of conversation quality.

In order to improve the framework model, conversation quality dimensions have to be further elaborated. Besides the difficulties in interrupting and acknowledging, the overall behavior of the conversation partners was observed to change as a result of transmission delay; this has to be reflected in quality prediction models. Further work is expected for extending the framework model to audio-visual communication services, as well as to conferencing services involving more than two communication partners.

5. REFERENCES

[1] U. Jekosch, *Voice and Speech Quality Perception. Assessment and Evaluation*, Springer, Berlin, 2005.

[2] J. Blauert, *Spatial Hearing: The Psychophysics of Human Sound Localization*, MIT Press, Cambridge MA., 1997.

[3] C. Osgood, G. Suci, and P. Tannenbaum, *The Measurement of Meaning*, University of Illinois Press, Urbana IL, 1957.

[4] I. Borg, and P. Groenen, *Modern Multidimensional Scaling – Theory and Applications*, Springer, New York NY, 2005.

[5] A. Raake, *Speech Quality of VoIP – Assessment and Prediction*, John Wiley & Sons, Chichester, West Sussex, 2006.

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

[7] M. Wältermann, A. Raake, and S. Möller, "Quality Dimensions of Narrowband and Wideband Speech Transmission", *Acta Acustica united with Acustica* 96, pp. 1090-1103, 2010.

[8] N. Côté, *Integral and Diagnostic Intrusive Prediction of Speech Quality*, Doctoral Dissertation, TU Berlin, 2010.

[9] D. Sen, "Predicting Foreground SH, SL and BNH DAM Scores for Multidimensional Objective Measure of Speech Quality", *Proc. IEEE ICASSP'04*, Vol. 1, pp. 493-496, 2004.

[10] B. Weiss, S. Möller, A. Raake, J. Berger, and R. Ullmann, "Modeling Conversational Quality for Time-varying Transmission Characteristics", *Acta Acustica united with Acustica* 95, pp. 1140-1151, 2009.

[11] ITU-T Contr. COM 12-55, *A Subjective/Objective Test Protocol for Determining the Conversational Quality of a Voice Link*, Source KPN (Author: J. G. Beerends), ITU-T Study Group 12, Sept. 2003.

[12] J.P. Duncanson, "The Average Telephone Call Is Better than the Average Telephone Call", *The Public Opinion Quarterly* 33(1), pp. 112-116, 1969.

[13] S. Möller, C. Bang, T. Tamme, M. Vaalgamaa, B. Weiss, "From Speech Quality to Service Quality: A Study on Long-term Quality Integration in Audio-Visual Speech Communication Services", accepted for: *Proc. Interspeech 2011*, 27-31 Aug., IT-Firenze, 2011.

[14] ITU-T Del. Contr. D.071, *Perceptual Correlates of the E-Model's Impairment Factors*, Source: Federal Republic of Germany (Authors: M. Wältermann and S. Möller), ITU-T SG12 Meeting, Geneva, 2005.

[15] R. Appel, and J.G. Beerends, "On the Quality of Hearing One's Own Voice", *J. Audio Eng. Soc.*, Vol. 50(4), pp. 237-248, 2002.

[16] M. Guéguin, R. Le Bouquin-Jeannes, V. Gautier-Turbin, F. Faucon, and V. Barriac, "On the Evaluation of the Conversational Speech Quality in Telecommunications", *EURASIP Journal on Advances in Signal Processing*, 2008.

[17] M. Hansen, and B. Kollmeier, "Continuous Assessment of Time-varying Speech Quality, *J. Acoust. Soc. Am.* 106, pp. 2888-2899, 1999.

[18] ITU-T Del. Contr. D.064, Testing the Quality of Connections Having Time Varying Impairments, Source: AT&T (Author: J.H. Rosenbluth), ITU-T, Geneva, 1998.

[19] L. Gros, and N. Chateau, "Instantaneous and Overall Judgements for Time-varying Speech Quality: Assessments and Relationships", *Acta Acustica united with Acustica* 87, pp. 367-377, 2001.

[20] B.W. Kort, "Models and Methods for Evaluating Customer Acceptance of Telephone Connections", *Proc. IEEE GLOBECOM'83*, pp. 706-714, 1983.

[21] M. Wältermann, A. Raake, S. Möller, "Analytical Assessment and Distance Modeling of Speech Transmission Quality", *Proc. Interspeech 2010*, Makuhari, pp. 1313-1316, 2010.