How to Measure Interactivity in Telecommunications

Florian Hammer and Peter Reichl
Telecommunications Research Center Vienna (ftw.)
Donau-City-Strasse 1, A-1220 Vienna, Austria.
Email: {hammer | reichl}@ftw.at

Abstract

Most recently, the notion of interactivity has gained rapidly increasing interest within the telecommunications community. Especially in the context of network quality planning (E-Model) it is suspected that interactivity might be of crucial impact on the perceived Quality-of-Service of bi- and multilateral communication over delayed transmission channels. In order to explore this in depth, an instrumental (objective) metric for conversational interactivity is an indispensable prerequisite. However, it is interesting that already the task of providing a concise definition of interactivity is a surprisingly hard job, as we demonstrate in this paper through a comprehensive survey on the various proposals made in the related work. As a preliminary conclusion, this demonstrates that current approaches do not allow for deriving a metric which is both meaningful and easy to be technically implemented. Therefore, we propose to use an axiomatic approach based on the relationship between the stateful activity of the speakers and the resulting interactivity of their conversation. We demonstrate that our resulting concept of interactivity is closely related to the notion of physical temperature as known from statistical thermodynamics. Finally, we back our approach using results from dedicated user tests.

1 Introduction and Motivation

It is quite obvious that the notion of interactivity reflects a very general and broad concept which by no means is restricted to a specific research area (and certainly not to telecommunications). Thus, it is not astonishing that related work provides lots of different approaches to define what could be termed the "100 faces of interactivity". However, introducing and exploring our topic in a telecommunications environment immediately raises a variety of questions like: What is the level of interactivity achieved during an email exchange compared to an online chat? How interactive is a telephone conversation over the Internet? What about Web services or so-called "interactive gaming applications"? Already from these few examples, we see that talking about a definition of interactivity first of all requires to narrow down the scope to the application under discussion .

In this paper, we address the application of telephony. Here, the transition from PSTN towards Voice-over-IP (VoIP) has led to a completely new set of transmission impairements which have caused a sudden rise of the importance of Quality-of-Service (QoS) as perceived by the user. Note that this implies a fundamental change in the QoS framework, as quality is no longer measured in terms of technical network parameters like packet loss and transmission delay, but also in terms of perceived speech quality at the receiver side. Of course, evaluating perceived QoS through user tests is quite an expensive and time consuming enterprise. Therefore, suitable user modelling, as fundamental e.g. to the "E-Model" standardized by the ITU-T [1], provides a useful tool for predicting the perceived quality from the network QoS parameters. However, the E-model has shown potential for improvements with respect to the impact of transmission delay on the perceived quality. Evidently, different conversational tasks which test persons need to accomplish in subjective speech quality tests result in different sensitivity with respect to the impact of the transmission delay on the quality rating of the subject. A defined measure of conversational interactivity would allow to distinguish the tasks with regard to their conversational structure and may serve as a useful parameter for both quality tests and quality prediction. This may serve as a sufficient explanation for the increasing interest of ITU-T in the relation between end-to-end transmission delay, speech quality as perceived by the user, and the interactivity of telephone conversations. Since the E-Model behaves very conservative with regard to the prediction of the impact of delay on the perceptual quality, monitoring the actual conversational situation may allow increased delay tolerance, and thus more flexibility regarding the adaptation of delay jitter buffers, to mention but the most evident advantage..

In this paper, we are interested in distinguishing situations by means of their interactivity. In order to be able to distinguish the interactivity of conversations, we need an instrumental (objective) metric which can easily be implemented into the technical framework of network QoS. Definitions of interactivity as given by the related work do not allow a quantitative comparison of conversations. Hence, we describe a completely different approach to formalize "conversational interactivity", starting from our intuitive insight *what interactivity is not*.

The remainder of this paper is structured as follows. Section 2 gives an overview on the related work on definitions of interactivity. In Section 3, we approach an instrumental metric for conversational interactivity. Section 4 describes the set-up of user tests we have carried out and Section 5 presents the results of these tests. Finally, Section 6 summarizes our conclusions.

2 Related Work

Surveying the literature, we find that interactivity is a concept used in a wide range of disciplines such as communication science, new media research, human-to-computer interaction, and web-design among others. In the following, we present an overview on the different perspectives from which interactivity is defined.

In his explication of the concept of interactivity, Kiousis [2] came up with a classification of interactivity as shown in Figure 1. He clusters the definitions of interactivity into the major groups: structure of technology, communication context and user perception.

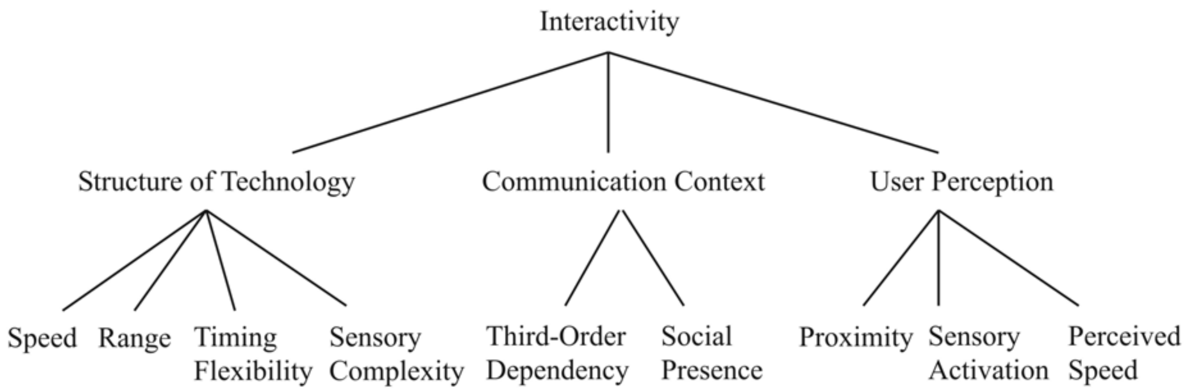

Figure 1. Operationalization of Interactivity (From [2]).

All interactive services are based on some underlying technology, thus, we will first discuss definitions of interactivity which refer to the structure and capabilities of that technology.
Within the context of telepresence, Steuer [3] defines interactivity as

> *"the extent to which users can participate in modifying the form and content of a mediated environment in real time".*

He identifies three factors which contributes to interactivity: Speed, which refers to the rate at which input can be assimilated into the mediated environment; range, which refers to the number of possibilities for action at any given time; and mapping, which refers to the ability of a system to map its controls to changes in the mediated environment in a natural and predictable manner.
In close relation to the speed feature as identified by Steuer, Novak et al. [4] defines interactive speed as follows:

> *Interactive speed is a construct that contributes to flow and is based on measures such as waiting time, loading time, and degree to which interacting with the Web is "slow and tedious".*

McMillan and Downes [5] give a five-dimensional definition of interactivity:

> *1. Direction of communication*
> *2. Timing flexibility*
> *3. Sense of place*
> *4. Level of control*
> *5. Responsiveness and the perceived purpose of communication.*

From these dimensions, the flexibility of message timing is a key dimension for both real-time communication and asynchronous communication such as email and newsgroups. Note that there is a difference between speed and timing flexibility. Speed represents the rate at which information flows through a system while timing flexibility refers to the degree to which users can control that rate.
As an additional technical aspect, sensory complexity, the amount of devices employed by a system to activate the five senses, contributes to the level of interactivity that the system can provide.

In the group of definitions related to the *context* of communication, Rafaeli [6] gives a clear definition of interactivity towards third-order dependency, i.e. related to the complete sequence of previous messages in contrast to simple re-actions to a stimulus. Thus, he defines interactivity as

> *"an expression of the extent that in a given series of communication exchanges, any third (or later) transmission (or message) is related to the degree to which previous exchanges referred to even earlier transmissions" .*

Furthermore, Williams et al. [7] state that interactivity is

> *"the degree to which participants in a communication process have control over, and can exchange roles in, their mutual discourse".*

Zack [8] suggests that the following key factors emerge from the literature as elements of interactivity: the simultaneous and continuous exchange of information; the use of multiple nonverbal cues; the potentially spontaneous, unpredictable, and emergent progression of remarks; the ability to interrupt or preempt; mutuality; patterns of turn-taking; and the use of adjacency pairs.
Note that these definitions of interactivity neither provide any formalisation allowing a quantitative comparison of any type of communication, nor do they allow the development of an instrumental measurement method with reasonable effort.

One of the first studies which operationalise interactivity as a *perception* of the user is Newhagen et al. [9]. Here, the relationship between media users' perceived sense of interactivity and the scope of the audience they address is studied. As a result, email messages which address large audiences indicate low levels of perceived interactivity while interpersonal messages result in high perceived interactivity.
Wu [10] has focused on perceived interactivity in the context of users' attitude towards a website. He claims that

> *"Perceived interactivity can be defined as a two-component construct consisting of navigation and responsiveness".*

Recently, McMillan and Hwang [11] identified measures of perceived interactivity (MPI) in the context of "interactive advertising" in the Internet. The 18 items included in the MPI (e.g. "loads fast" and "keeps my attention") provide a tool for measuring perceived interactivity for this specific application, but do not allow for a straightforward transfer to our VoIP environment .

Sligthly more appropriate for our purposes, a recent study on the performance of default speech codecs carried out by 3GPP [12] reports on the use of interactivity in subjective quality tests. The test subjects were asked to judge the conversation when interacting with the conversation partner based on the occurrence of double talk (test persons talking simultaneously) and interruptions. As an example, "fair" interactivity was described as "sometimes, you were talking simultaneously, and you had to interrupt yourself".

Summarizing, despite there are many definitions of interactivity in the literature , they are not really applicable to our problem as we neither want to take the context (by means of third-order dependency) nor the users' perception of interactivity into account. Instead, we aim at an instrumental metric which at least somehow is reflected by the aspect of speed, or rate of information flow, given by Steuer. However, we conclude that no formalized approach is available at the time which is able to describe the interactivity of conversations.

3 Approaching an Instrumental Metric for Conversational Interactivity

This section presents our approach to defining a metric, which meets the requirements stated in Section 2.

3.1 Conversational Model

We model the structure of a two-way conversation in accordance with ITU-T Recommendation P.59 [13] by distinguishing four states as presented in Figure 2. States A and B refer to the situations in which either person A or person B is speaking alone. State M (mutual silence) represents the case of both persons being silent and state D (double talk) refers to both persons talking simultaneously.

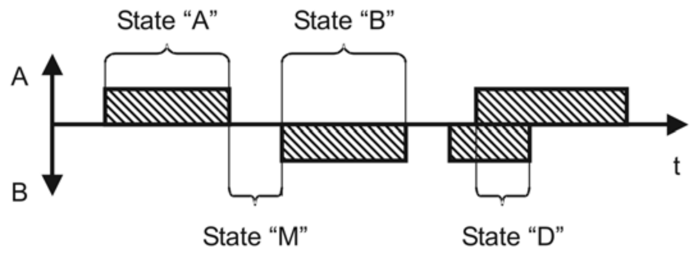

Figure 2. States of a two-way conversation.

In a further step, we include the transitions between these four states I = {A, B, M, D} and end up with a stochastic process as presented in Figure 3. Note that direct transitions between states A and B as well as between states M and D are omitted due to the fact that these transitions represent extremely rare events.

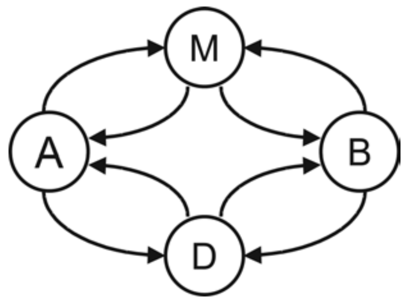

Figure 3. 4-state Conversation Model.

This stochastic process can be described by the mean probabilities that the conversation actually stays within one of the states and the corresponding sojourn times which represents the average period of time that the conversation stays in that given state. Since we focus on two-way conversations, we avoid the use of alternative models The model's simplicity makes it a reasonable candidate for our purposes.

3.2 Instrumental Metric for Conversational Interactivity

As we did not find an explicit definition for conversational interactivity in the related work, we propose an implicit approach [14] which is based on three properties as described as follows.

Limiting Behavior
First, we state that a conversation is *not* interactive if either A or B are talking all the time (monologue), both speakers are silent (no activity), or both speakers are talking simultaneously all the time (lack of information exchange). This property describes the basic relation between sojourn times and the interactivity. If (at least) one of the sojourn times is large, the interactivity tends to zero, whereas in the case of small sojourn times, the interactivity increases.

Normalization

As a second property, we want our interactivity metric to scale alongside a “norm conversation”. This scaling is based on the mean sojourn times of a large set of conversations which can be derived from ITU-T Recommendation P.59 [13].

Monotonicity

Finally, we describe the first-order behavior of our metric. An increasing state sojourn time in one of the states (assuming the other sojourn times to be constant) leads to a decrease in interactivity and vice versa.

From these properties, in [14] we have derived the following expression:

$$\lambda_I = \alpha \exp(\frac{-\bar{\tau}}{\tau}), \qquad (1)$$

where λ_I represents the total transition rate out of state *I*, τ equals the interactivity of the actual conversation, $\bar{\tau}$ equals the average interactivity of the norm conversation (reference interactivity) and α represents a constant.

The function corresponds to a well-known result from statistical thermodynamics. Consider a particle moving within a quantum well which is divided into several regions of different energy. Equation (2) describes the behavior of the particle with regard to its transition rate λ_I out of state *I*:

$$\lambda_I = \nu \exp(\frac{-\Delta E}{kT}) \qquad (2)$$

where ν represents the oscillation frequency of the particle, ΔE equals the height of the potential walls, *k* equals Boltzmann’s constant, and *T* represents the temperature of the system. From this analogy, we conclude that our interactivity metric can be interpreted as a “conversational temperature”. Using least-squares estimation, we can calculate the conversational temperature given the sojourn times of the conversation under investigation. For further information on the mathematical background of this metric, we refer to [14].

The conversational temperature provides a monotonic function following our limiting criteria and can be scaled to a norm conversation by setting the corresponding reference temperature to room temperature, e.g. 21.5° with regard to the Celsius scale (corresponding to 294.65 K or 70.5° F).

4. Measurements and Evaluations

Our results as presented in Section 5 are mainly based on conversations that have been recorded during speech quality tests carried out at the Institute of Communication Acoustics (IKA) at Ruhr-University Bochum. The laboratory setup for these tests is described in [15]. The conversational tests consisted of VoIP connections using the ITU-T G.729 codec with different bursty packet loss rates (0%, 3% 5% and 15%) combined with transmission delay of 60 ms, 360 ms, 660 ms and 960 ms. In this study, we restrict ourselves to scenarios without packet losses in order to explore the pure delay effect on both interactivity and subjective quality. The quality ratings comprise mean opinion scores (MOS) based on the 5-point absolute category rating scale recommended by the ITU-T [16], which reflects the overall quality perceived by the user and ranges from 5.0 (“excellent”) to 1.0 (“bad”). The CR-10 category rating scale [15, 17] indicates the perceived impairment of the connection: here, a

value of “0” denotes that the user has not perceived any impairment at all, whereas, e.g. “2”, “5” and “10” correspond to “weak”, “strong”, and “extremely strong” impairment, respectively.
The test subjects were asked to accomplish interactive Short Conversation Tests (iSCT, [18]) which represent telephone scenarios like the rapid exchange of information about new employees, e.g. e-mail addresses and telephone numbers, leading to comparable and balanced conversations of higher interactivity compared to the standard Short Conversation Test (SCT) scenarios [15]. The iSCTs were designed to create a conversation situation in which the participants would be more sensible for increased transmission delay.

For our investigations, we have recorded the microphone signals of both speakers and manually coded the talkspurts in order to reach high accuracy in the derived parameters. The talk spurts were shifted in time as to obtain the conversational patterns as perceived by the individual participants. The speakers have been re-ordered according to their average speech activity (i.e. ratio between total sum of talk-spurts and total duration of the conversation). Therefore, generally speaker A is the one talking more. We have evaluated both the conversational parameters and the conversational temperature and put them into relation with the subjective quality ratings. We mainly focus on the evaluation of the parameters for increasing transmission delay times.

5. Results

In the following, we present the analysis of conversations performed by 7 pairs of test persons (8 female, 6 male) who knew each other. We consider the average call duration of 143 sec as long enough for the test persons to obtain a useful impression of the properties of the connections.

5.1. Conversational Parameters

Our first analysis addresses the evolution of the conversational parameters for increasing delay conditions as presented in Figure 4. As a first observation, the mean state probabilities and sojourn times of states *A* and *B* are decreasing with increasing delay, while
the values for mutual silence (*M*) show a significant increase between 360 ms and 660 ms. The latter behaviour is not unexpected, because mutual silence increases with delay due to the time lag of the responses. However, the present data suggests a change in behaviour of the test subjects when the delay is increased from 360 ms to 660 ms. In contrast to the variation of these parameters, double-talk (*D*) rises only slightly with delay.

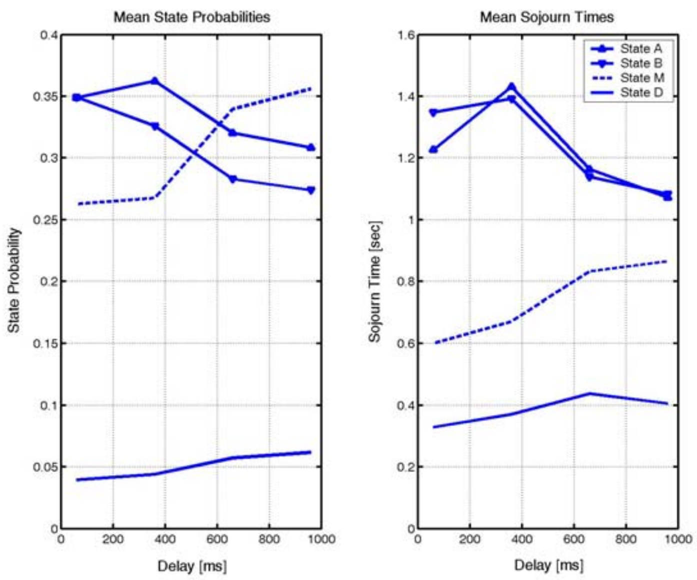

Figure 4. Conversation Model Parameters versus Transmission Delay.

5.2. Conversational Temperature and Perceived Speech Quality

As described in Section 3.2, we derive our Temperature Metric from the sojourn times of the conversations. The limiting property states that the interactivity increases with decreasing sojourn times, thus, as a consequence of the results regarding the sojourn times, we would expect the interactivity to increase with increasing delay.
Figure 5 presents the mean conversational temperature (TEMP). We indeed observe an increase in conversational temperature between 360 ms and 960 ms. While at 360 ms the temperature minimum of 19.9° is attained, the subsequent rise towards 23.7° at 960 ms is significant also w.r.t. the depicted standard deviations.

As a final result of our investigations, we explore the correlation of the perceived speech quality with the conversational temperature at different delay times. Figure 5 illustrates the evolution of MOS and CR-10 value, and Conversational Temperature with respect to increasing delay, where each parameter has been averaged over all conversations.
Both the MOS and CR-10 ratings indicate only a slight decrease in perceptual quality at very high delay. Starting from about 3.6 and 3.8 at a one-way delay of 60 ms and 360 ms, respectively, the MOS decreases to about 3.4 and 3.1 at delay values of 660 ms and 960 ms, respectively. The perceived impairment as reflected by the CR-10 scale monotonically increases with delay, starting from 1.6 at 60 ms and ending up at 2.6 at 960 ms.
As shown above, the conversational temperature increases with delay from 360 ms. Thus, for high amount of delay, i.e. greater than 360 ms, the interactivity correlates with the perceptual speech quality.

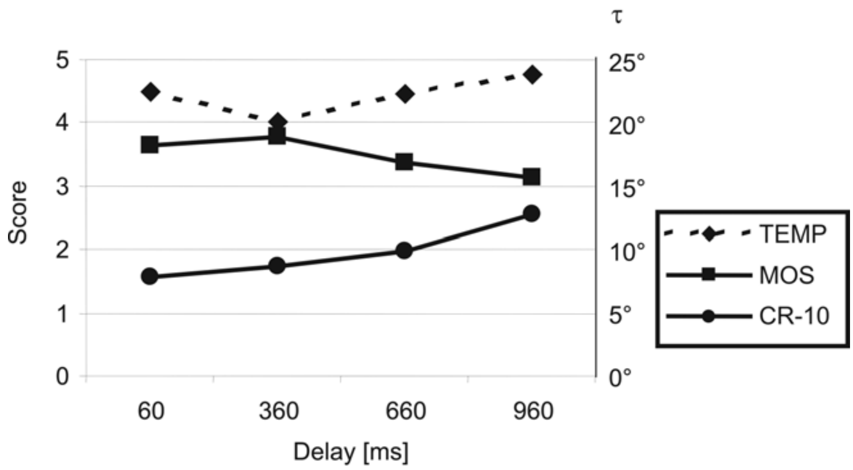

Figure 5. Comparison of Average Temperature (TEMP), Average MOS (MOS) and Average CR-10 vs. Delay. Note that a one-way delay of 360 ms does not significantly degrade the perceptual quality. Moreover, a further increase of the delay only slightly affects the quality.

6. Conclusions

This paper has presented a comprehensive overview on the variety of interactivity definitions, before introducing an approach for formally describing telephone conversations, i.e. a temperature metric for conversational interactivity. To this end, we have analysed a total of 28 conversations carried out over connections with different amounts of delay.
Our results show that for the type of task in use the conversational interactivity increases with increasing delay. Moreover, high delay has only a slight impact on the quality perceived by the users which is reflected both in MOS and CR-10 scores.
Future work focuses on testing and comparing the interactivity of different tasks used for conversational speech quality tests. Furthermore, we plan the development a measurement methodology for ``perceived interactivity". Thus, we will be able to obtain a more comprehensive picture about telephone users, their conversational habits and their perception of transmission delay.

Acknowledgements

This work has been funded under the Austrian government's Kplus Competence Center Program. The authors like to thank Alexander Raake for the collaboration and the fruitful discussions.

References

[1] International Telecommunication Union, “The E-model, a computational model for use in transmission planning”, ITU-T Recommendation G.107, March 2005.

[2] Kiousis, S., “Interactivity: a concept explication”, New Media Society, Vol. 4, pp. 355-383, Sept. 2002.

[3] Steuer, J.S., “Defining Virtual Reality: Dimensions Determining Telepresence”,

Journal of Communication, Vol. 42, No. 4, pp. 73–93, 1992.

[4] Novak, T. P., Hoffman, D. L., and Yung Y.-F., “Measuring the Customer Experience in Online Environments: A Structural Modeling Approach”, Marketing Science, Vol. 19, No. 1, pp. 22-42, 2000.

[5] S. J. McMillan and E. J. Downes, Defining Interactivity: A qualitative identification of Key Dimensions”, New Media & Society, Vol. 2, No. 2, pp. 157-179, 2000.

[6] Rafaeli, S., “Interactivity: From New Media to Communication”, in R.P. Hawkins, J.M. Wieman and S. Pingree (eds) *Advancing Communication Science: Merging Mass and Interpersonal Processes*, pp. 110–34. Newbury Park, CA: Sage, 1988.

[7] Williams, F., Rice, R. E. and Rogers, E., *Research Methods and the New Media*. New York: Free Press, 1988.

[8] Zack, M. H., “Interactivity and Communication Mode Choice in Ongoing Management Groups”, Information Systems Research, Vol. 4, No. 3, pp. 207-239, 1993.

[9] Newhagen, J. E., Cordes, J. W., and Levy, M. R., “Nightly@nbc.com: Audience Scope and the Perception of Interactivity in Viewer Mail on the Internet’, *Journal of Communication* Vol. 45, No. 3, pp. 164–75, 1995.

[10] Wu, G., “Perceived Interactivity and Attitude Toward Websites”, paper presented at the Annual Conference of the American Academy of Advertising in Albuquerque, New Mexico, March 1999.

[11] McMillan, S. J. and Hwang, J.-S., “Measures of Perceived Interactivity: An Exploration of the Role of Direction of Communication, User Control, and Time in Shaping Perceptions of Interactivity”, Journal of Advertising, Vol. XXXI, No. 3, Fall 2002.

[12] 3GPP, “3rd Generation Partnership Project; Technical Specification Group Services and System Aspects; Packet Switched Conversational Multimedia Applications; Performance Characterisation of Default Codecs (Release 6)”, 3GPP TR 26.935 v6.0.0, June 2004.

[13] International Telecommunication Union, “Artificial conversational speech,” *ITU-T Recommendation P.59*, Mar. 1993.

[14] Reichl, P. and Hammer, F., “Hot Discussion or Frosty Dialogue? Towards a Temperature Metric for Conversational Interactivity”, Int. Conf. Speech and Language Processing, Vol. 1, pp. 317-320, Jeju Island, Korea, Oct. 2004.

[15] Möller, S., *Assessment and Prediction of Speech Quality in Telecommunications*. Boston: Kluwer Academic Publishers, 2000.

[16] International Telecommunication Union, “Methods for subjective determination of transmission quality,” *ITU-T Recommendation P.800*, Aug. 1996.

[17] G. Borg, *Borg’s Perceived Exertion and Pain Scales*. Champaign, IL: Human Kinetics, 1998.

[18] International Telecommunication Union, “E-model: Additivity of burst loss impairment with other impairment types,” *Source: Ruhr-University Bochum (A. Raake), ITU-T Delayed Contribution 221*, 2004.