PARAMETRIC SIMULATION OF IMPAIRMENTS CAUSED BY TELEPHONE AND VOICE OVER IP NETWORK TRANSMISSION

PACS REFERENCE: 43.72 KB

Rehmann, Sebastian; Raake, Alexander; Möller, Sebastian
Institute of Communication Acoustics
Ruhr-University Bochum
Universitätsstraße 150
44801 Bochum
Germany
Tel: +49 234 32 23978
Fax: +49 234 32 14165
E-mail: {rehmann,raake,moeller}@ika.ruhr-uni-bochum.de

ABSTRACT
Modern telecommunication networks increasingly comprise trunks of packet-based transmission (e.g. VoIP). When used for speech communication, new types of degradations occur, which are of time-varying nature – in contrast to stationary or signal-correlated degradations in PSTN/ISDN networks. In order to investigate the quality impact of both types of degradations, it is necessary to carry out auditory conversation tests. An online simulation tool is presented, which allows all quality-relevant parameters of combined PSTN/ISDN/VoIP networks to be set in a defined way.

1. INTRODUCTION

Already today, packet-based transmission techniques such as voice over internet protocol (VoIP) or voice over ATM (Asynchronous Transfer Mode) are replacing or complementing existing circuit switched telephone networks with their synchronous transmission. Packet-based transmission is of great interest for service providers as well as for telephone network solutions in companies or small businesses. VoIP is especially attractive as it allows to use existing networks, e.g. in-company LANs, for speech communication, eliminating the need for additional telephone networks.

However, such transmission techniques still suffer from relatively poor speech quality when compared to traditional wireline networks. The low quality results from the nature of packet-based transmission: Speech packets can get lost during transmission, consecutive packets may arrive across different transmission paths and thus in the wrong order, or waiting for particular packets can lead to considerable delay, all depending on the availability of network resources. On the user's side, these time-varying network characteristics – packet-loss, jitter and delay – are perceived as speech quality degradations that are different from those present in wireline Public Switched Telephone Networks or Integrated Services Digital Networks (PSTN/ISDN). Although solutions such as speech data prioritisation may be applied in the long run to solve some of the current quality problems of packet-based systems, these will certainly remain an important issue for the acceptance of such systems in the years to come. A particular challenge for high quality will be the interconnection of different types of networks, where trunks of PSTN or ISDN are combined with trunks of packet-based transmission or mobile networks.

In order to predict the quality of networks in the planning phase, or to monitor quality of already existing networks, it is necessary to assess the quality as perceived by a user of the system. For this purpose, quality prediction models have been developed in the past, mapping network parameters on quality ratings (for an overview of quality models cf. Möller and Raake, 2002). Currently, the existing quality prediction models do not include the effect of time-varying distortions, and hence have to be expanded. However, only limited information is available in the literature on how network conditions typical for packet-based systems could be described in a parametric way (e.g. ITU-T Del. Contr. D.20, 2001), and what exact impact they have on quality (e.g. Gros and Chateau, 2001).

This information can only be reliably obtained from human subjects in auditory tests. In telecommunications, such tests are typically carried out as listening-only or as conversational tests. Especially in the case of time-varying distortions, conversation tests are advantageous, as they reflect the real-life conversation situation – loss of information due to packet-loss, for example, is assumed to be perceived differently in a real conversation than when listening to pre-recorded speech samples. With the aim of collecting subjective data for later modelling of speech quality under VoIP- or concatenated network-typical conditions in conversation tests, an integral online telephone-line simulation tool was developed at IKA. The test set-up as well as the simulation tool are described in the following.

2. SETUP FOR AUDITORY CONVERSATION TESTS

In order to carry out conversation tests, two test subjects are placed in separate rooms, each of which is equipped with an ordinary telephone set. They are then asked to carry out a number of dialogue tasks according to some previously composed scenarios (Wiegelmann et al., 1999). Invisibly for the test subjects, the seemingly "ordinary" telephone sets are connected via the multifunctional PSTN / ISDN / VoIP-simulation developed at IKA.

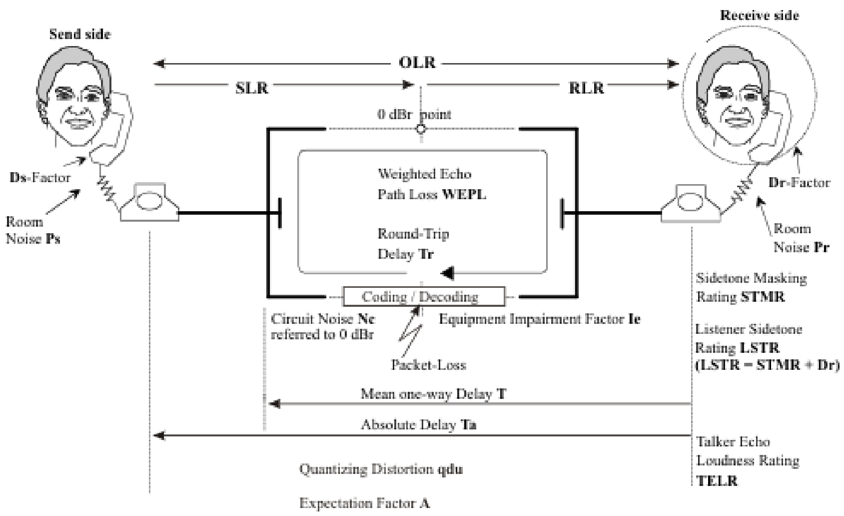

Fig. 1: *Parametric description and schematic of the transmission path for a combined PSTN / ISDN / VoIP-network (cf. ITU-T Rec. G.107, 2000).*

Such a simulation system is preferable to real network equipment, as it permits inexpensive tests under controlled laboratory conditions to be conducted. The real-time IKA system is based on a parametric description of a PSTN / ISDN / VoIP-transmission path, a schematic overview of which is given in Fig. 1. Examples for such parameters are the loudness of the transmission (i.e. send loudness rating *SLR*, receive loudness rating *RLR* and overall loudness rating *OLR*) or loudness and delay of echoes (*TELR* and *T* for talker echo, *WEPL* and *Ta* for listener echo). The simulation has been designed as to allow controlled variation of all parameters shown in the graph, which in their total determine the quality of a real-life speech transmission line. By letting the test subjects conduct several dialogues with different sets of transmission parameters and asking them for a quality judgement after each dialogue, the relationship between the absolute parameter settings and subjective speech quality perception can be established.

The parametric approach chosen here corresponds to the one taken in the currently recommended speech quality model for network planning, namely the *E-model* (ITU-T Rec. G.107, 2000). It assumes that the different degradations can be transformed onto a psychological scale, and are – on this scale – additive, yielding a transmission rating *R* for the particular transmission condition, see Equation 1.

$$R = Ro - Is - Id - Ie + A \qquad (1)$$

The additive features of the transmission are referred to as *impairment factors*. According to this principle, the transmission rating *R* is the rating for the basic signal-to-noise ratio (*Ro*) degraded by the impairment factors. Different parameters are combined to different impairment factors: The degradations simultaneous to the transmitted speech signal (e.g. non-optimum loudness or signal-correlated noise) form the Simultaneous Impairment Factor *Is*; the degradations delayed to the signal (e.g. pure delay or echoes) build the Delayed Impairment Factor *Id*; the degradations caused by specific signal processing equipment such as low-bitrate speech codecs are expressed with the Equipment Impairment Factor *Ie*. The *Ie*-value is often used to characterize degradation of quality due to a specific codec (see also Table 1). *A* is the Advantage Factor accounting for the effect of user expectation (i.e., users will be much more tolerant for bad quality, when additional service features such as mobile access are available).

3. OVERVIEW OF THE PSTN / ISDN SIMULATION

A block diagram of the technical layout of the PSTN / ISDN simulator is given in Fig. 2 (Möller, 2000; modified). It is implemented based on a signal-processing hardware which may be wired and programmed via software. The simulation addresses the perceptual characteristics of the transmission, excluding problems such as signaling. The ISDN / VoIP module may be interfaced by replacing the two “Codec” modules in the diagram by the set-up depicted in Fig. 3.

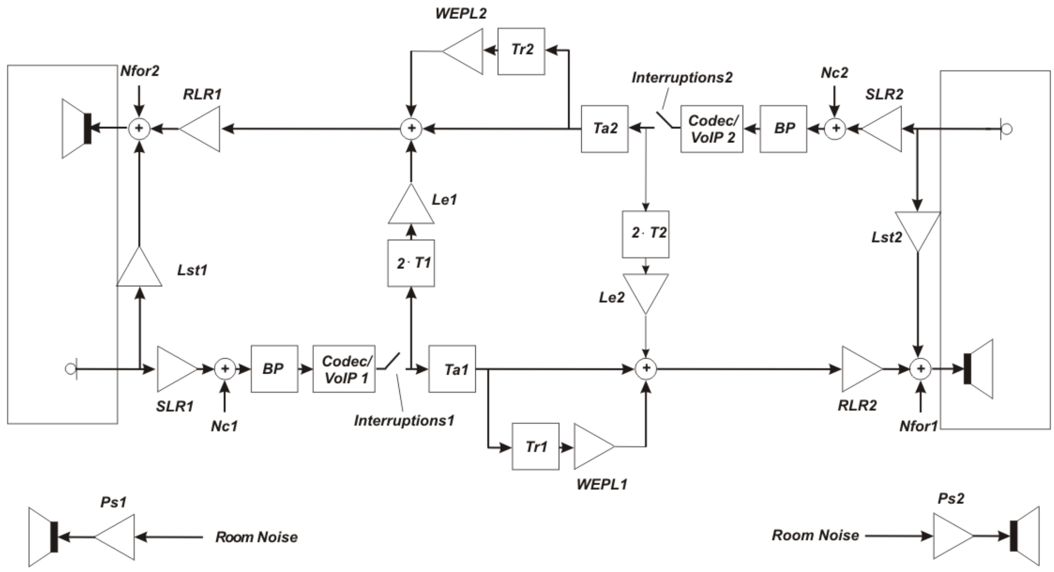

Fig. 2: *Schematic of the PSTN / ISDN line simulation tool including possible VoIP pathways. Other configurations are possible.*

The triangles represent filters or programmable attenuators. The rectangles stand for delay lines, external codecs (or a possible position of the VoIP-trunk), and the channel band-pass filter. Other positions of the VoIP trunk are possible and depend on the type of network to be simulated. The filter-characteristics can freely be predefined and uploaded according to the condition to be simulated. By using filters at the interfaces with the terminal equipment, their electro-acoustic characteristics may be shaped, and different types of equipment may be used (e.g. handset telephones, headsets, hands-free terminals). The parameters covered by the line simulation are those introduced in Section 2. In addition, the simulation allows interruptions of different lengths and distributions to be introduced into the speech path, which can serve as a reference for scalable time-varying distortions. The simulation may also be used in wide-band mode (50 - 7000 Hz), but currently, the wide-band version cannot be combined with the ISDN-based VoIP-trunk.

4. TRANSMISSION IMPAIRMENTS TO BE EXPECTED IN IP NETWORKS

Voice over IP is the ability to transmit normal telephone speech over IP-based networks. To transmit speech via the Internet Protocol poses problems different from those of PSTN or ISDN, because speech is carried as data-packets over a packet-switched data network, instead of as a synchronous stream of analogue or binary data over a circuit-switched voice network (Thomson and Jani, 2000). While PSTN and ISDN connections reserve a certain bandwidth for the duration of a call, this reservation of bandwidth is generally impossible to accomplish in the internet. Instead, a continuous stream of voice data is divided into packets which – roughly speaking – average 20 ms in size (this figure varies with the codec used and also depends on the number of speech frames per packet, see below). Each of these packets finds its way from source to destination "on its own" by being routed from one internet node to the next.

Speech transmission over IP networks is highly sensitive to transmission impairments. Impairments of VoIP connections qualitatively differ from impairments in circuit-switched telephone connections, resulting mainly from the lack of synchronicity of the Internet Protocol. This in turn arises from two characteristic attributes of the internet: (1) *Limited bandwidth:* when a connection between two routers operates at full capacity, the packets to be transmitted are stored in a queue *(store-and-forward*-principle), resulting in a transmission delay. If the capacity of the queue is exceeded, packets are dropped *(packet loss).* (2) *Individual routing of packets:* resulting from the fact that different packets may be routed differently through the network, packets may arrive at their destination in a different order than that in which they were sent.

These characteristics of the internet yield six different kinds of transmission impairments:

1. *coding distortion:* In order to reduce the bandwidth costs of voice transmissions, a variety of different coder/decoder algorithms (codecs) are in use which reduce the voice signal data rate. Each codec results in a perceivable degradation of speech quality. Speech signals may even be multiply coded (tandeming). Table 1 gives an overview of common codecs used for VoIP (Equipment Impairment Factors from ITU-T Rec. G.113, 2001).

Codec	Bit Rate/kbps	Ie	Codec	Bit Rate/kbps	Ie
G.711 a-law	64	0	G.726	32	7
G.711 ì-law	64	0	G.728	16	7
G.723.1	5.3	19	G.729	8	10
G.723.1	6.3	15	G.729b	8	11
G.726	16	50	GSMFR	13.2	20
G.726	24	25	GSMEFR	12.2	5

Table 1: *Overview of codecs for VoIP and corresponding Equipment Impairment factors Ie.*

2. *delay:* In wire-line networks, transmission delays are usually constant and below 30 ms. In the internet, delays are generally larger. Moreover, they are variable, resulting in *jitter* (see below). Delay in VoIP-transmission results from several factors: Firstly, the speech signal has to be recorded and coded. Secondly, it is packetized in IP packets. Thirdly, routing and transmission delays on the network have to be taken into account. Finally, the packets are buffered in the *jitter buffer* (see below) and have to be decoded again. These factors add up to a total between 150 and 400 ms according to Table 2.

Latency delay source	Typical value (ms)
Recording	10–40
Coder	10–20
Internet delivery	70–120
Jitter buffer	50–200
Decoder	10–20
Total	150–400

Table 2: *End-to-end VoIP packet latency delay (Thomson and Jani, 2000)*

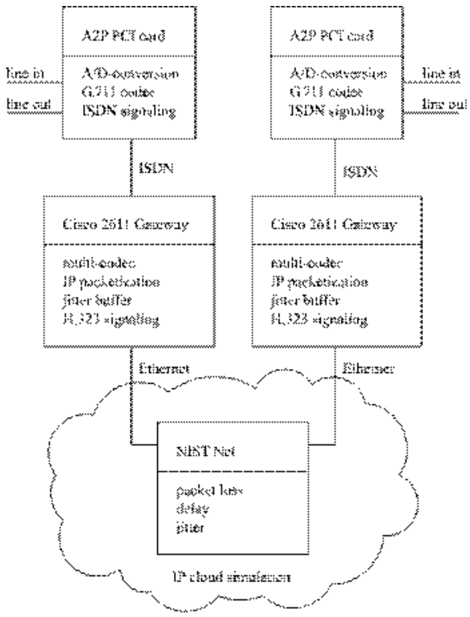

Fig. 3: *Schematic of the VoIP simulation*

3. *jitter:* Delay jitter represents the variation of the delay due to variable packet routing over the internet. VoIP calls may suffer from significant jitter. As it has been shown that even small jitter values lead to severe degradations in perceived voice quality, jitter buffers are essential for VoIP terminals. In a jitter buffer, packets are stored for a specific period of time and are sorted into the correct order before being decoded and played out. This in turn introduces an additional delay, which however can be more easily tolerated than jitter.

4. *packet loss:* When networks or parts of a network are used at their capacity limits, chances are that some packets do not arrive at their destination in time, or even never arrive at all. This is rather unproblematic for typical data transmission, as lost packets are simply sent again. However, speech transmissions have to be conducted in real time, so re-transmission of lost packets is useless. Depending on the applied coding scheme, packet loss may lead to severe perceptual degradations of the transmitted speech signal (e.g. Gros and Chateau, 2001).

5. *packet doubling:* This effect may occur due to faulty router configurations. It is rather rare, however, and does not have a direct impact on speech quality, as duplicate packets can simply be ignored by the receiver.

6. *echo:* As it has been shown that delays are high for VoIP transmissions, room echos at the far end particularily degrade ease of communication. For this reason, VoIP terminals need to implement efficient algorithms for echo cancellation.

5. DESCRIPTION OF THE VOIP SIMULATION

The VoIP transmission may be integrated into the simulation described in Fig. 2, by replacing the "Codec" modules of the tool with the ISDN/VoIP simulation module outlined in Fig. 3.

Transformation of the analogue speech signal into ISDN is conducted by two A2P modules constructed by T-Nova GmbH, Berlin. The A/D- and D/A-conversions use codec G.711, according to the ISDN definition. The A2P modules are furthermore responsible for ISDN signaling communication with the VoIP gateway, which is necessary for call build-up and release. These modules are implemented as two PCI cards to be inserted into a personal computer. They are connected to two VoIP gateways by two ISDN BRI lines.

The gateway components of the VoIP simulation are responsible for packetizing the ISDN speech signal and for building up an IP call according to the ITU-T H.323 protocol (ITU-T Rec. H.323, 2000). They are implemented using two Cisco 2611 routers, each equipped with an ISDN BRI voice port and an Ethernet port. They are connected via 10BaseT Ethernet to the *IP cloud simulation,* which is capable of simulating the effect of an arbitrary number of routers forming an internetwork.

IP typical network impairments are introduced by the NIST Network Emulation Tool (NIST Net), which is also available as public domain (NIST Net, 2001). This software operates as a kernel module for the Linux operating system and is installed on a PC acting as IP router. It allows controlled introduction of packet loss, packet doubling, delay and jitter into one or more of the IP routes configured on the router.

The VoIP simulator presented in this paper is capable of simulating any desired values of delay and packet loss. Furthermore the VoIP gateway supports all of the codecs shown in Table 1.

Jitter may be simulated, but as it is usually suppressed by a jitter buffer, jitter normally leads to a combination of increased delay and packet loss. Packet doubling may also be simulated, but has been shown to have no direct effect on voice quality. Echoes directly result from the selected delay, and echo cancellation should not be deployed as not to modify the speech signal in an uncontrolled way; however, as echo is a general problem of telephone transmissions, it is not regarded as a VoIP-specific parameter here but rather considered in the PSTN/ISDN simulation, see section 3.

6. CONCLUSIONS

An online simulation tool to perform subjective tests of voice transmission quality has been presented. It is suitable for both listening-only and real-time conversation tests, which will be carried out at IKA in the future. Network typical transmission characteristics may be simulated and adjusted during the course of the tests. This applies to both "traditional" telephony parameters, as present in analog or digital circuit-switched connections, and to parameters of packet-based voice transmission like VoIP, as well as to any combination thereof. Investigating such combinations is especially important in order to introduce time-varying distortions into quality prediction models like the E-model. The effect of the distribution of impairments over time will also be evaluated with this tool.

7. ACKNOWLEDGEMENT

The present work has been performed at IKA, Ruhr-University Bochum (Prof. J. Blauert, PD U. Jekosch). The authors would like to thank T-Nova GmbH, Berlin, Germany for providing the A2P PCI cards.

8. REFERENCES

WIEGELMANN, S., MÖLLER, S., JEKOSCH, U. (1999). Scenarios for Economic Conversation Tests in Telephone Speech Quality Assessment. Joint Meeting ASA/EAA/DEGA, Forum Acusticum 1999, D-Berlin, ACUSTICA/acta acustica 85 Suppl. 1, 48.

THOMSEN, G., JANI, Y. (2000). Internet telephony: going like crazy. IEEE Spectrum, 52–58.

GROS, L. and CHATEAU, N. (2001). Instantaneous and Overall Judgements for Time-Varying Speech Quality: Assessments and Relationships. Acta acust. Vol. 87 (2001), 367-377.

ITU-T SG12 DELAYED CONTRIBUTION D.20 (2001). The Burst Ratio: A Measure of Bursty Packet Loss. Source: Lucent Technologies, USA. International Telecommunication Union, CH-Geneva.

ITU-T REC. G.107 (2000). The E-Model, a Computational Model for Use in Transmission Planning. International Telecommunication Union, CH-Geneva.

ITU-T REC. G.113 (2001). Transmission Impairments due to Speech Processing. International Telecommunications Union, CH-Geneva.

ITU-T REC. H.323 (2000). Packet-based Multimedia Communications Systems. International Telecommunications Union, CH-Geneva.

NIST NET (2001). National Institute of Standards and Technology, USA. (www.nist.gov)

MÖLLER, S. (2000). Assessment and Prediction of Telephone Speech Quality. Kluwer Academic Publishers, Boston-USA.

MÖLLER, S. and RAAKE, A. (2002). Telephone Speech Quality Prediction: To-wards Network Planning and Monitoring Models for Modern Network Scenarios. Accepted in Speech Communication.