Proceedings of the 1st IP-Telephony Workshop (IPTel 2000), GMD Report 95, pp. 105-110, Berlin, Germany, 12-13 April 2000.

DELAY AND DISTORTION BOUNDS FOR PACKETIZED VOICE CALLS OF TRADITIONAL PSTN QUALITY

Jan Janssen, Danny De Vleeschauwer and Guido H. Petit
Alcatel, Network Strategy Group
Francis Wellesplein 1, B-2018 Antwerp, Belgium
{jan.janssen, danny.de_vleeschauwer, guido.h.petit}@alcatel.be

***Abstract* --When voice is to be transported over packet-based networks with particular quality guarantees, bounds on the mouth-to-ear delay and distortion should be adhered to. In this paper, the E-model is used to calculate such (delay and distortion) bounds when traditional (circuit-switched) telephone quality is aimed for.**

***Index Terms* -- delay, distortion, E-model, quality, voice**

A. INTRODUCTION

For traditional Public Switched Telephone Network (PSTN) calls, which do not suffer from distortion, the key factor that determines the quality is the mouth-to-ear delay, defined as the delay incurred from the moment the talker utters the words until the instant the listener hears them. The mouth-to-ear delays that can be tolerated depend on the level of echo disturbing the voice call [1, 2].

Voice calls can also tolerate some distortion, that is, the voice signal heard by the listener does not need to be an exact copy of the voice signal produced by the talker. In contrast to circuit-switched calls (which do not suffer from distortion), distortion is likely to be introduced in packetized voice calls by the codec that compresses the voice signal or by the loss of voice packets.

As such, controlling both the mouth-to-ear delay and distortion is key to offering high-quality packetized voice calls. In this paper, we will use the E-model to calculate the mouth-to-ear delay and distortion bounds that should be respected for packetized voice calls of traditional PSTN quality.

B. E-MODEL

The E-model [3, 4, 6, 7] is a computational tool to predict the subjective quality of a telephone call based on its characterizing transmission parameters. It combines the impairments caused by these transmission parameters into a rating *R*, which ranges between 0 and 100 and can be used to predict subjective user reactions such as e.g. the Mean Opinion Score (MOS). The model was developed such that its results are in accordance with the results of extensive subjective laboratory tests.

The *R*-scale was defined so that impairments are approximately additive in the *R*-range of interest, i.e., $R = R_0 - I_s - I_d - I_e + A$. The first term R_0 groups the effects of (background and circuit) noise. The second term I_s includes impairments that occur simultaneously with the voice signal, such as those caused by quantization, by too loud a connection and by too loud a side tone. The third term I_d encompasses delayed impairments, including impairments caused by talker and listener echo or by a loss of interactivity. The fourth term I_e covers impairments caused by the use of special equipment. For example, each low bit rate codec has an associated impairment value. This impairment term can also be used to take the influence of packet loss into account. The fifth term *A* is the expectation factor, which expresses the decrease in the rating *R* that a user is willing to tolerate because of the "access advantage" that certain systems have over traditional wire-bound telephony. As an example, the expectation factor *A* for mobile telephony is 10.

ITU-T draft Recommendation G.109 [5] states that a rating *R* in the ranges [90,100], [80,90], [70,80], [60,70] and [50,60] corresponds to best, high, medium, low and poor quality, respectively. A rating below 50 indicates unacceptable quality. Throughout this paper, these quality classes are color coded according to Table 1.

As far as quality is concerned, a packetized voice call introduces more delay and distortion than a traditional PSTN call.

First, the delay for packetized voice calls, where the most important contributions are encoding, packetization, propagation, queuing, service, dejittering and decoding delay, is larger than for a traditional circuit-switched voice call, where the mouth-to-ear delay is mainly made up of the propagation delay and switching delay.

Second, in contrast to circuit-switched voice calls, as a result of voice compression and packet loss during transport or in the dejittering buffer, the distortion of packetized voice calls is not negligible.

We have studied the impact of the one-way mouth-to-

Table 1: Speech quality classes

R-value range	100 - 90	90 - 80	80 - 70	70 - 60	60 - 0
speech transmission quality category	best	high	medium	low	(very) poor
	← PSTN quality →				

ear delay (via I_d) and the distortion (via I_e) on the quality of a packetized voice call. Other factors may also impair the quality of a packetized voice call (via R_0 and I_s), but as these factors are not fundamentally different from a traditional PSTN call, they were not considered in this paper. Furthermore, as the objective was to make a fair comparison between the quality of packetized voice calls and traditional PSTN calls, the expectation factor *A* was set to 0.

Consider a packetized voice call between two parties, referred to as party 1 and party 2 (see Fig. 1). Using the E-model, we calculated how party 1 will judge the call, that is, what rating *R* will be assigned to it. The influence of delay was studied first, followed by the influence of distortion.

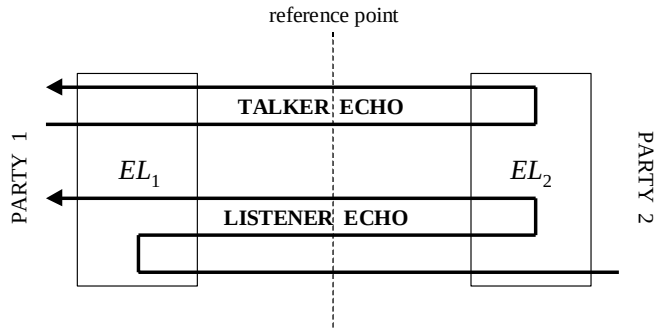
1st. Influence of mouth-to-ear delay

If the voice signal party 1 hears is delayed, the rating *R* decreases by an amount equal to the impairment I_d associated with the mouth-to-ear delay. This impairment is the sum of 3 contributing impairments, caused by talker echo, listener echo and loss of interactivity.

First, talker echo disturbs party 1, who hears an attenuated and delayed echo of his own voice. This echo is caused by a reflection close to party 2 with attenuation or echo loss EL_2 (measured with respect to a certain reference point) [7].

Second, listener echo also disturbs party 1, who hears the original signal from party 2 followed by an attenuated echo of this signal. This echo is determined by a reflection close to party 1 with attenuation EL_1, followed by a reflection close to party 2 with attenuation EL_2.

Echo may occur in the 4-to-2-wire hybrid (if the packetized voice call is terminated over a local PSTN) or in the callers' terminal equipment. For PSTN calls from traditional handsets, where echo is mainly caused by the hybrids, a typical value for the echo loss is 21 dB [7]. The same value is valid for packetized voice calls terminated over a local PSTN to traditional handsets. Handsfree phones are likely to have a lower echo loss value due to acoustic echo. When there is no hybrid (e.g. in packet-based access networks), the echo loss only depends on the acoustic echo introduced in the used terminals. Multimedia PCs are expected to have rather low echo loss values, while well-designed IP-phones will have high echo loss values. The echo losses EL_1 and EL_2 can be increased by using an echo controller, which should be deployed as close to the echo source as possible. A simple echo controller can increase the echo loss by 30 dB. Perfect echo control, in which the echo losses EL_1 and EL_2 increase to infinity, can be achieved at moderate computational cost.



EL – Echo Loss

Figure 1: Talker and listener echo

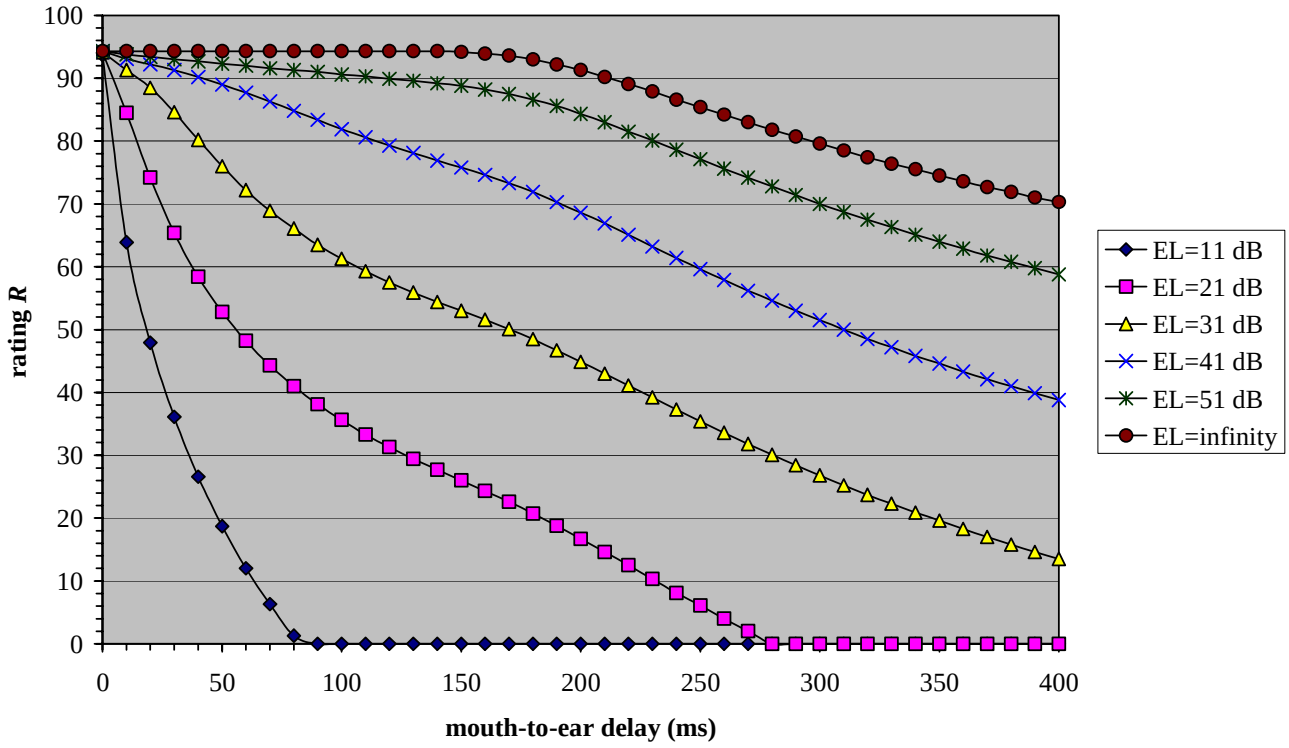
The third delay-related factor that may disturb party 1 is the loss of interactivity. If the mouth-to-ear delay is too large, an interactive conversation becomes impossible.

We have used the E-model, which takes all these impairments into account, to calculate the rating *R* given by party 1 to undistorted voice calls, i.e., calls transported without packet loss in the G.711 format. Fig. 2 shows the influence of the mouth-to-ear delay on the rating *R* for different echo loss values. The latter are assumed to be equal at both end points ($EL_1 = EL_2$). The impairment associated with delay is strongly influenced by this echo loss value.

Observe that the rating *R* is a non-increasing function of the mouth-to-ear delay. The intrinsic quality of a voice call is defined as the rating *R* associated with a mouth-to-ear delay of 0 ms. The intrinsic quality of a packetized voice call transported without packet loss in the G.711 format corresponds to $R = 94.3$. Fig. 2 shows that if echo is perfectly controlled ($EL_1 = EL_2 = \infty$), this voice call retains its intrinsic quality up to a mouth-to-ear delay of 150 ms.

ITU-T Recommendations G.114 [1] and G.131 [2] specify the following tolerable mouth-to-ear delays for traditional PSTN calls:

- Under normal circumstances (i.e. if the echo loss is at least 21 dB), echo control is needed if the mouth-to-ear delay is larger than 25 ms.
- When the echo is adequately controlled:
  - a mouth-to-ear delay of up to 150 ms is acceptable for most user applications,
  - a mouth-to-ear delay between 150 ms and 400 ms is acceptable, provided that one is aware of the impact of delay on the quality of the user applications, and



EL – Echo Loss

Figure 2: The rating *R* as a function of the mouth-to-ear delay for undistorted voice and for various echo loss values

- a mouth-to-ear delay above 400 ms is unacceptable.

It can be seen from Fig. 2 that for an echo loss of 21 dB, the rating *R* drops below 70 at a mouth-to-ear delay of 25 ms. For calls with perfect echo control, the rating *R* drops below 70 at a mouth-to-ear delay of 400 ms. Hence, ITU-T Recommendations G.114 and G.131 ensure that traditional PSTN calls have a rating *R* of at least 70. Also, the interactivity bound of 150 ms can be observed in Fig. 2 for infinite echo loss.

2nd. Influence of distortion

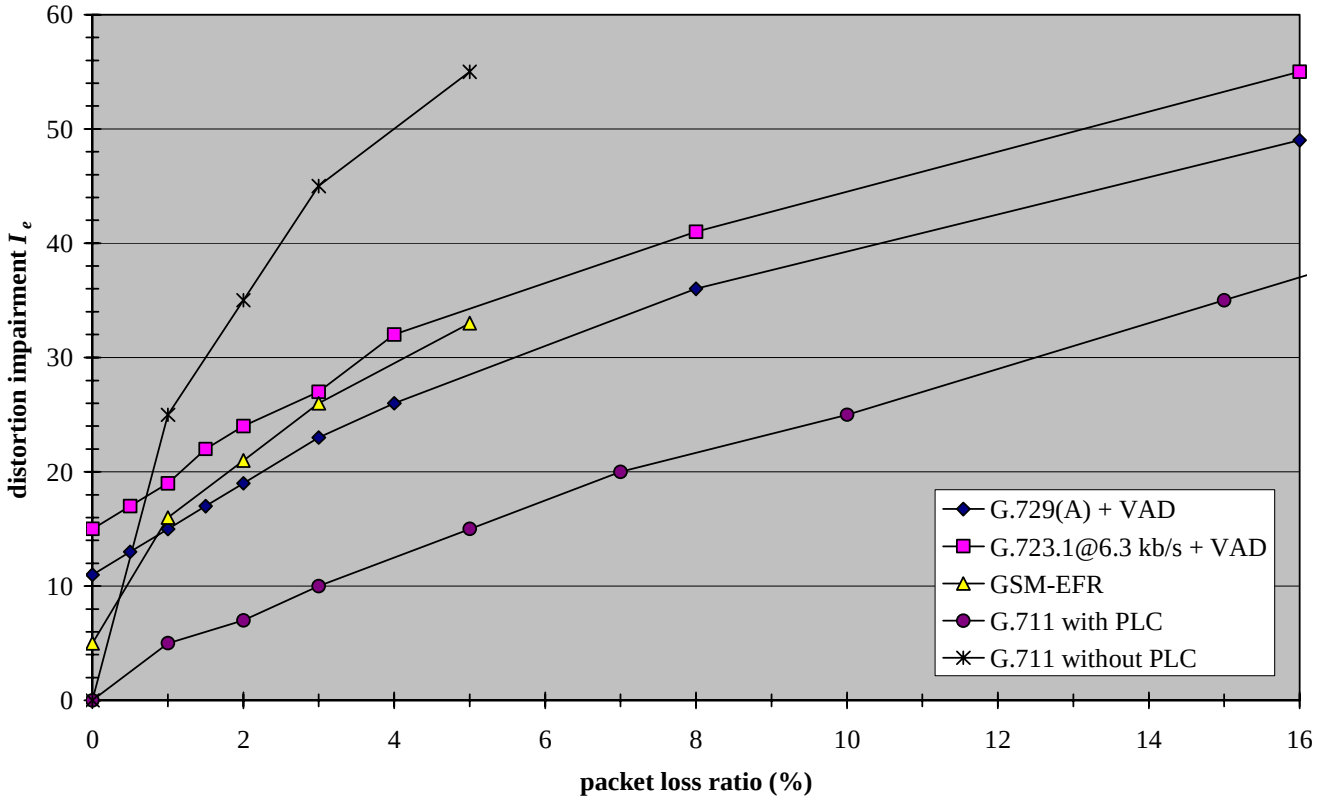
If the voice signal party 1 hears is distorted, the rating *R* decreases by an amount equal to the distortion impairment I_e. This impairment has two sources: encoding of the voice signal from party 2 and packet loss during the transport of voice packets from party 2 to party 1.

Table 2 summarizes the distortion impairment I_e and intrinsic quality (expressed in terms of *R*) associated with each standard codec [8]. It follows immediately that traditional PSTN quality ($R \geq 70$) cannot be attained with the G.726/G.727 codecs at 16 and 24 kb/s, while the intrinsic quality of the GSM-HR codec is very close to the acceptable limit. These particular codecs should be avoided.

The distortion impairment I_e associated with a codec increases as the packet loss ratio increases. Fig. 3, based on [8], shows this effect for 4 codecs, assuming that voice packets are lost at random. This figure deals only with one specific packetization interval per codec (10 ms for G.711, 20 ms for G.729 and GSM-EFR, 30 ms for G.723.1). Similar results are not yet known for other packetization intervals. The sensitivity to packet loss depends on whether a Packet Loss Concealment

Table 2: Distortion impairment and intrinsic quality of standard codecs

origin	standard	type	codec bit rate (kb/s)	I_e	intrinsic quality *R*
ITU-T	G.711	PCM	64	0	94.3
	G.726, G.727	ADPCM	16	50	44.3
			24	25	69.3
			32	7	87.3
			40	2	92.3
	G.728	LD-CELP	12.8	20	74.3
			16	7	87.3
	G.729(A)	CS-ACELP	8	10	84.3
	G.723.1	ACELP	5.3	19	75.3
		MP-MLQ	6.3	15	79.3
ETSI	GSM-FR	RPE-LTP	13	20	74.3
	GSM-HR	VSELP	5.6	23	71.3
	GSM-EFR	ACELP	12.2	5	89.3



VAD – Voice Activity Detection
PLC – Packet Loss Concealment

Figure 3: Distortion impairment as a function of the packet loss

(PLC) technique is used by the codec. More precisely, for codecs with PLC the impairments increase slower than for codecs without PLC. In contrast to the G.711 codec, most low bit rate codecs (i.e. G.729, G.723.1 and GSM-EFR) have a built-in PLC scheme. However, a PLC scheme can be implemented on top of the G.711 codec.

The voice signal does not need to be transported in the same format end-to-end. Somewhere along the route, voice might be transcoded from one codec format into another. Since all (considered) standard codecs need an 8 kHz stream of uniformly quantized voice samples at the input, the code words of the first codec need to be decoded before the signals can be encoded into another codec format. Consequently, the distortion impairment terms associated with the two codecs should be added to obtain the overall impairment I_e, because, in the E-model, impairments are additive on the *R*-scale. The intrinsic quality associated with all combinations of two codecs can be found in Table 3 (using the color code of Table 1). It turns out that transcoding can be very harmful to the quality of a call and should be avoided.

C. QUALITY BOUNDS

If the mouth-to-ear delay, echo loss and distortion impairment are known, the quality of a packetized voice call (i.e. its rating *R*) can be derived from Fig. 2 as follows. First, identify the curve on Fig. 2 that corresponds to the given echo loss. Then, using this curve, read the rating *R* corresponding to the given mouth-to-ear delay. Finally, subtract the distortion impairment I_e from this rating *R*.

As stated before, if there is no echo control, the echo loss is likely to be (smaller than) 21 dB for packetized voice transport. For that value of the echo loss, the rating *R* drops rapidly as the mouth-to-ear delay increases. Hence, if there is no echo control, there is only a very small delay budget for which traditional PSTN quality ($R \geq 70$) can be guaranteed.

From now on, we assume that perfect echo control is performed [9], in which case the intrinsic quality of the call is attained if the mouth-to-ear delay is kept below 150 ms. This intrinsic quality depends solely on the distortion impairment I_e, which in turn is determined by the codec(s) used and the overall packet loss experienced.

Since the intrinsic quality of an undistorted call is 94.3 and the bound for traditional quality is 70, there is an impairment budget of 24.3, part of which is

Table 3: Transcoding matrix

CODEC	G.711 (64kb/s)	G.726 (40kb/s)	G.726 (32kb/s)	G.726 (24kb/s)	G.726 (16kb/s)	G.728 (16kb/s)	GSM-FR (13kb/s)	G.728 (12.8kb/s)	GSM-EFR (12.2kb/s)	G.729 (8kb/s)	G.723.1 (6.3kb/s)	GSM-HR (5.6kb/s)	G.723.1 (5.3kb/s)
G.711 (64kb/s)	94.3	92.3	87.3	69.3	44.3	87.3	74.3	74.3	89.3	84.3	79.3	71.3	75.3
G.726 (40kb/s)	92.3	90.3	85.3	67.3	42.3	85.3	72.3	72.3	87.3	82.3	77.3	69.3	71.3
G.726 (32kb/s)	87.3	85.3	80.3	62.3	37.3	80.3	67.3	67.3	82.3	77.3	72.3	64.3	68.3
G.726 (24kb/s)	69.3	67.3	62.3	44.3	19.3	62.3	49.3	49.3	64.3	59.3	54.3	46.3	50.3
G.726 (16kb/s)	44.3	42.3	37.3	19.3	0	37.3	24.3	24.3	39.3	34.3	29.3	21.3	25.3
G.728 (16kb/s)	87.3	85.3	80.3	62.3	37.3	80.3	67.3	67.3	82.3	77.3	72.3	64.3	68.3
GSM-FR (13kb/s)	74.3	72.3	67.3	49.3	24.3	67.3	54.3	54.3	69.3	64.3	59.3	51.3	55.3
G.728 (12.8kb/s)	74.3	72.3	67.3	49.3	24.3	67.3	54.3	54.3	69.3	64.3	59.3	51.3	55.3
GSM-EFR (12.2kb/s)	89.3	87.3	82.3	64.3	39.3	82.3	69.3	69.3	84.3	79.3	74.3	66.3	70.3
G.729 (8kb/s)	84.3	82.3	77.3	59.3	34.3	77.3	64.3	64.3	79.3	74.3	69.3	61.3	65.3
G.723.1 (6.3kb/s)	79.3	77.3	72.3	54.3	29.3	72.3	59.3	59.3	74.3	69.3	64.3	56.3	60.3
GSM-HR (5.6kb/s)	71.3	69.3	64.3	46.3	21.3	64.3	51.3	51.3	66.3	61.3	56.3	48.3	52.3
G.723.1 (5.3kb/s)	75.3	73.3	68.3	50.3	25.3	68.3	55.3	55.3	70.3	65.3	60.3	52.3	56.3

consumed by the codec (see Table 2). Once the codec has been chosen, the remainder of the margin can be consumed either by allowing the mouth-to-ear delay to exceed 150 ms or by allowing some packet loss. Tables 4 and 5 give the codec-dependent bounds on the packet loss and mouth-to-ear delay, respectively, assuming only one of these phenomena is allowed to occur. Note that packet loss could be traded off against mouth-to-ear delay (e.g. by varying the dejittering delay), as long as the impairment budget is not exceeded.

D. CONCLUSIONS

The E-model has been used to study the quality of packetized voice calls. With regard to quality, for packetized voice calls more delay and distortion is introduced than for traditional PSTN calls.

Since the tolerable mouth-to-ear delay budget is small for compressed voice, echo control is recommended.

If the echo is perfectly controlled, the quality remains equal to the intrinsic quality up to a mouth-to-ear delay

Table 4: Tolerable packet loss bounds for a mouth-to-ear delay below 150 ms
PLC – Packet Loss Concealment
VAD – Voice Activity Detection

origin	standard	codec bit rate (kb/s)	packet loss bound (%)
ITU-T	G.711 without PLC	64	1
	G.711 with PLC	64	10
	G.729(A) + VAD	8	3.4
	G.723.1@6.3 kb/s + VAD	6.3	2.1
ETSI	GSM-EFR	12.2	2.7

Table 5: Tolerable mouth-to-ear delay bounds when there is no packet loss
NA – traditional quality is Not Attainable

origin	standard	codec bit rate (kb/s)	mouth-to-ear delay bound (ms)
ITU-T	G.711	64	400
	G.726, G.727	16	NA
		24	NA
		32	324
		40	379
	G.728	12.8	212
		16	324
	G.729(A)	8	296
	G.723.1	5.3	221
		6.3	253
ETSI	GSM-FR	13	212
	GSM-HR	5.6	180
	GSM-EFR	12.2	345

of 150 ms. The intrinsic quality depends on the amount of distortion that is introduced.

The intrinsic quality associated with some low bit rate codecs is lower than the traditional PSTN quality. Therefore, these codecs should not be used. For the same reason, transcoding should be avoided if possible.

The margin between the intrinsic quality of a codec and the bound for traditional quality can either be consumed by allowing a mouth-to-ear delay above 150 ms or by allowing some packet loss. The mouth-to-ear delay and packet loss bounds are reported here for the most common codecs. These bounds should be respected by any packetized voice call if traditional quality is to be maintained.

ACKNOWLEDGMENT

This work was carried out within the framework of the project LIMSON, sponsored by the Flemish institute for the promotion of scientific and technological research in the industry (IWT).

REFERENCES

[1] "One-Way Transmission Time", *ITU-T Recommendation G.114*, February 1996.

[2] "Control of Talker Echo", *ITU-T Recommendation G.131*, August 1996.

[3] N.O. Johannesson: "The ETSI Computation Model: A Tool for Transmission Planning of Telephone Networks", *IEEE Communications Magazine*, pp. 70–79, January 1997.

[4] P. Meschkat: "TPE: Transmission Planning (End-to-End) using the E-model (Supporting ETSI Guide 201 050)", Windows Software Tool, Alcatel Telecom, December 1997.

[5] "Definition of Categories of Speech Transmission Quality", *ITU-T Recommendation G.109*, September 1998.

[6] "The E-model, a Computational Model for Use in Transmission Planning", *ITU-T Recommendation G.107*, December 1998.

[7] "Speech Processing, Transmission and Quality Aspects (STQ); Overall Transmission Plan Aspects for Telephony in a Private Network", *ETSI Guide 201 050*, February 1999.

[8] "Provisional Planning Values for the Equipment Impairment Factor I_e", *Appendix to ITU-T Recommendation G.113 (Draft)*, September 1999.

[9] D. De Vleeschauwer, J. Janssen, G.H. Petit: "Delay Bounds for Low Bit Rate Voice Transport over IP Networks", *Proceedings of the SPIE Conference on Performance and Control of Network Systems III*, volume 3841, pp. 40–48, Boston (MA), USA, 20-21 September 1999.