Proceedings of WTC/ISS 2000, 7-12 May 2000, Birmingham, United Kingdom.

TOLERABLE DELAY BOUNDS FOR LOW BIT RATE VOICE TRANSPORT

Danny De Vleeschauwer, Jan Janssen,
Emmanuel Desmet, Guido H. Petit

Alcatel
Francis Wellesplein 1
B-2018 Antwerp
Belgium

Key Theme **T1**

Tel : (+32) 3 240 8196
Fax : (+32) 3 240 9932
E-mail : danny.de_vleeschauwer@alcatel.be

Tel : (+32) 3 240 8146
Fax : (+32) 3 240 9932
E-mail : jan.janssen@alcatel.be

Tel : (+32) 3 240 8613
Fax : (+32) 3 240 9932
E-mail : emmanuel.desmet@alcatel.be

Tel : (+32) 3 240 9869
Fax : (+32) 3 240 9932
E-mail : guido.h.petit@alcatel.be

TOLERABLE DELAY BOUNDS FOR LOW BIT RATE VOICE TRANSPORT

Abstract

Mouth-to-Ear delay significantly determines the perceived subjective quality of voice communications. In the case of the traditional voice service, ITU-T Recommendations G.114 and G.131 advise keeping the mouth-to-ear delay below 25 ms if no echo control is performed. However, when echo is adequately controlled, the quality of the conversation is solely determined by the perceived degree of interactivity. While a mouth-to-ear delay below 150 ms is hardly noticeable, a value above 400 ms is considered unacceptable.

For compressed voice signals, the determination of the mouth-to-ear delay bounds is more complex. The main objective of the paper is to use the E-model to assess the tolerable delay bounds for various low bit rate codecs. Our results show that echo control is essential for voice conversations transported in compressed form.

When perfect echo control is performed, the subjective quality remains constant up to a codec-independent delay bound of 150 ms. From this value onward, the voice quality steadily drops because of a loss of interactivity. The tolerable mouth-to-ear delay values above which the quality becomes unacceptable are codec-dependent and sometimes significantly smaller than the bound of 400 ms mentioned above.

1. Introduction

Currently the real-time transport of (compressed) Voice over IP (VoIP) is an important focus of attention. Whether or not quality can be guaranteed for voice flows transported over IP networks remains an open question. Quality, in this context, depends largely on the one-way Mouth-to-Ear (M2E) delay in combination with the level of the echo, and the distortion introduced. M2E delay is defined as the time that elapses between the moment the talker utters the words and the moment the listener hears them. Distortion may be introduced by the low bit rate codec or be caused by packet loss in the IP-based network or in the dejittering buffer.

The M2E delay bounds that can be tolerated in traditional telephone networks are standardized in ITU-T Recommendations G.114 and G.131. This paper determines the tolerable M2E delay bounds for voice transported in compressed form over an IP network introducing no packet loss. In particular, it investigates the dependency of these delay bounds on the low bit rate codecs employed and on whether or not Echo Control (EC) is employed.

The next section recalls the tolerable M2E delay bounds for traditional telephony. Section 3 introduces the E-model, which can be used to obtain the tolerable M2E delay bounds for any voice call in terms of its characterizing parameters. Section 4 assesses the tolerable M2E delays for voice calls (between traditional telephones) routed in compressed format over an IP backbone. General conclusions are drawn in the final section.

2. Traditional tolerable mouth-to-ear delay bounds

ITU-T Recommendations G.114 and G.131 specify the following delay bounds for traditional phone calls [1,2].

- Under normal circumstances, EC is needed if the M2E delay is larger than 25 ms. If the echo is exceptionally large (i.e. less than 33 dB attenuated with respect to the original signal), EC is necessary even for M2E delays below 25 ms.
- When the echo is adequately controlled (i.e. the impairment caused by echo is negligible compared to the impairment caused by the loss of interactivity),then:
  - M2E delays up to 150 ms are acceptable for most user applications;
  - M2E delays between 150 ms and 400 ms are acceptable, provided that one is aware of the impact of delay on the quality of the user applications; and
  - M2E delays above 400 ms are unacceptable.

This paper derives similar M2E delay bounds for voice that is compressed using a low bit rate codec.

3. ETSI E-model

The ETSI E-model predicts the subjective quality of a phone call as perceived by the calling or called party [3]. Therefore, the model uses the characterizing transmission parameters of the call and combines the impairments caused by these parameters into a rating factor R. Subjective user reactions, such as the Mean Opinion Score (MOS), can be predicted from this R-factor, which lies in the interval [0,100].

The R-scale was chosen such that impairments are approximately additive. This approximation (inherent to the E-model) is valid for the R-range of interest. The R-factor consists of the terms:

$$R = R_0 - I_s - I_d - I_e + A$$

The first term R_0 represents the basic signal-to-noise ratio. The second term I_s represents impairments occurring simultaneously with the voice signal, such as impairments caused by quantization, by too loud a connection or too loud a side tone. The third term I_d represents delayed impairments, including those caused by talker and listener echo and by the loss of interactivity. The fourth term I_e represents impairments caused by the use of special equipment. For example, each low bit rate codec has an associated impairment value I_e which increases as the packet loss experienced by the codec increases. The fifth term A is the expectation factor. This expresses the decrease in R-factor a user is willing to tolerate because of the advantage of access that certain systems have over traditional, wire-bound telephony. As an example, the expectation factor A for mobile telephony equals 10. In this paper we take $A = 0$ (i.e. the quality of VoIP calls is compared with the quality of traditional telephony).

We have taken $R = 72$ as the lower limit for "traditional" quality. The reason for this choice is explained in Section 4. Roughly speaking, the quality of a VoIP call is comparable to the quality of wire-bound telephony if its R-factor is larger than 72. Note that we deliberately avoid the use of the term "toll quality" as it is an ill-used term according to ITU-T Draft Recommendation G.GOVQ [4].

Because this paper puts the emphasis on voice transport over IP networks, the following only considers the delay impairment term I_d and the equipment impairment term I_e. As packet loss is assumed negligible, the equipment impairment term I_e is solely determined by the low bit rate codec. All other terms in the R-factor are set at their default values.

Figure 1, calculated using the E-model, shows the influence of the M2E delay on the R-factor for calls transported in the 64 kbit/s G.711 format. The impairment associated with delay is strongly influenced by whether or not (perfect) EC is provided. Observe that for both curves (without EC and with perfect EC), the R-factor is a non-increasing function of the M2E delay, which has a maximum (attained at zero M2E delay) of 94.3. The latter value is referred to as the intrinsic quality of the G.711 codec.

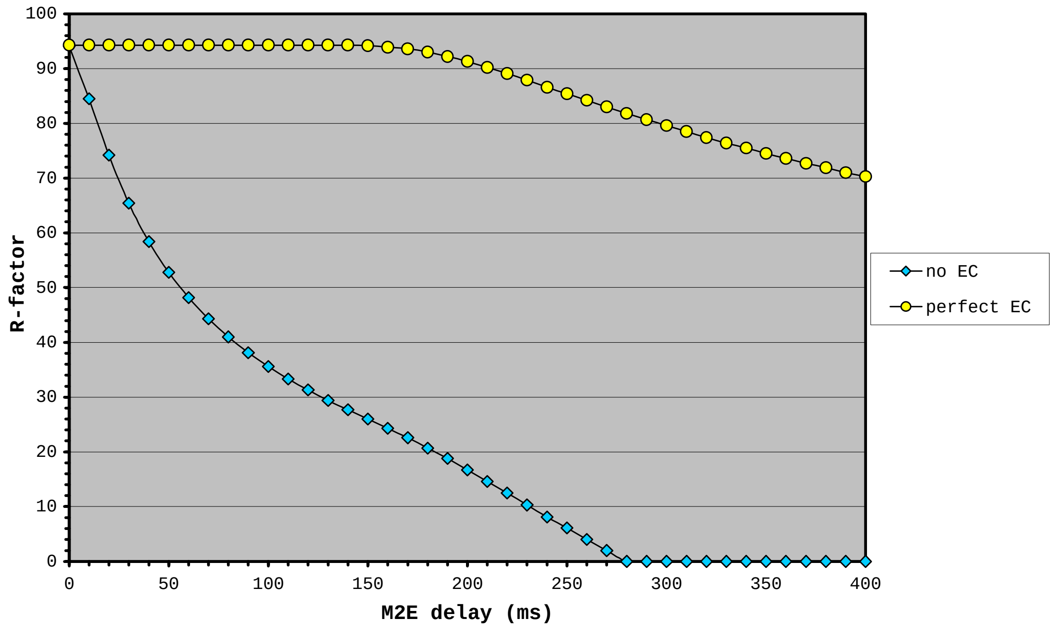

Figure 1 – The R-factor versus the M2E delay for the G.711 codec without and with (perfect) echo control

If the voice is transported in compressed form, the R-factor decreases by the amount I_e associated with that codec. These I_e-values, tabulated in Table 1, are averages of (lots of) subjective tests [5]. Curves similar to the curves of Figure 1 can be drawn for every standard codec. These new curves can be obtained from Figure 1 by a downward shift equal to the impairment factor I_e associated with the codec. From Table 1 it immediately follows that certain codecs (i.e. G.726 or G.727 at 16 and 24 kbit/s and GSM-HR) never achieve traditional telephone quality because their impairment factor I_e is too large (i.e. larger than 94.3 - 72 = 22.3). For this reason, these codecs are not considered here.

Origin	**Recommendation/ Standard**	**Type**	**Bit Rate (kbit/s)**	I_e
ITU-T	G.711	PCM	64	0
	G.726, G.727	ADPCM	16	50
			24	25
			32	7
			40	2
	G.728	LD-CELP	12.8	20
			16	7
	G.729(A)	CS-ACELP	8	10
	G.723.1	ACELP	5.3	19
		MP-MLQ	6.3	15
ETSI	GSM-FR	RPE-LTP	13	20
	GSM-HR	VSELP	5.6	23
	GSM-EFR	ACELP	12.2	5

Table 1 – Standardized codecs with their bit rates and impairment factors I_e

4. Results for (phone-to-phone) calls over an IP backbone

In the scenario considered here, voice calls use the traditional Public Switched Telephone Network (PSTN) as the access network from the source and destination telephones to gateways at the edge of the IP backbone network. The voice signal is encoded and packetized in the ingress gateway. Voice packets are then routed over the IP backbone from the ingress gateway to the egress gateway. The latter compensates the jitter introduced in the network and decodes the voice signal.

There are two sources of echo in this symmetric scenario. Hybrid echo may be generated in the 4-to-2-wire hybrids of the PSTN, and acoustic echo may be generated in the terminating phones as a result of the acoustic coupling between microphone and speaker. The Echo Loss (EL) is the amount by which the reflected echo is attenuated with respect to the incoming signal. Two ELs are important, one at the source and one at the destination side. However, since we only consider a symmetric scenario, they are assumed to have the same value. If no EC is performed and for traditional telephones, the EL is mainly determined by the hybrids. Figure 1 is based on an EL value of 21 dB. In the case of perfect EC, the EL becomes infinite.

Table 2 gives the tolerable M2E delay bounds below which traditional quality is achieved for various codecs in the phone-to-phone scenario. The cases without EC and with perfect EC are both considered.

Origin	Recommendation/ Standard	Bit Rate (kbit/s)	T_{M2E} (ms) no EC	T_{M2E} (ms) perfect EC
ITU-T	G.711	64	23	379
	G.726, G.727	32	15	305
		40	20	356
	G.728	12.8	2	192
		16	15	305
	G.729(A)	8	12	278
	G.723.1	5.3	3	203
		6.3	7	237
ETSI	GSM-FR	13	2	192
	GSM-EFR	12.2	17	324

Table 2 – Tolerable M2E delay bounds for a phone-to-phone call over an IP backbone without and with (perfect) echo control

The results in Table 2 for the G.711 codec correspond to the traditional M2E delay bounds described in ITU-T Recommendations G.114 and G.131 (see Section 2). More precisely, the 25 ms bound above which EC is required and the 400 ms bound above which interactivity is impossible, are (more or less) found in Table 2. This is in fact a direct consequence of, and the reason for, the choice of R = 72 as a definition of traditional quality.

The reasoning in Section 3 concerning how the curves in Figure 1 should be changed to cater for low bit rate codecs (by a downward shift equal to the impairment factor of the codec) immediately leads to the other bounds reported in Table 2. They can be interpreted as an extension of the ITU-T Recommendations G.114 and G.131 for compressed voice.

The former reasoning also implies that the 150 ms bound (ITU-T Recommendation G.114) is codec independent. That is, for any possible codec, the quality of a voice call with perfect EC remains more or less equal to the codec's intrinsic quality for M2E delays below 150 ms.

In comparison with calls transported over the PSTN, additional delays (codec, packetization, service, queuing, dejittering, etc) are introduced in an IP-based network. Packetization delay is unavoidable; it is the time taken to collect a number of voice code words in an IP packet. Hence, it scales with the number of code words. Since the M2E delay bounds are small when no EC is performed, the packetization delay must also remain small. This means that only a few (if any) code words can be put in an IP packet, which may lead to a large proportion of overhead bytes to be transported and an inefficiently exploited network. EC increases the M2E delay bound allowing the packetization delay to increase. This leads to larger packets, a smaller proportion of the overhead bytes to be transported and better efficiency. Thus, EC is strongly recommended. The influence of the level of EC on the delay bounds is studied in detail in [6].

A final remark deals with the fact that the M2E delay bounds in the last column of Table 2 are also valid for other, possibly non-symmetric, scenarios such as phone-to-PC and PC-to-PC. Indeed, as we apply perfect EC, the ELs at source and destination side are infinite and do not influence the delay bounds at all.

5. Conclusions

This paper has calculated the tolerable M2E delay bounds under which traditional quality is achieved for voice calls transported in compressed form over an IP network with no packet loss. As the bounds achieved without EC are very small, EC is strongly recommended. When perfect EC is applied, the best possible quality is achieved by staying below 150 ms M2E delay. In this case the quality only depends on the distortion introduced by the codec. The use of certain codecs (G.726/727 at 16 kbit/s, G.726/727 at 24 kbit/s, GSM-HR) should be avoided as they cannot attain traditional quality.

On the other hand, calls experiencing M2E delays between 150 ms and the bounds of Table 2 (for perfect EC), are all rated to be at least of traditional quality. Nevertheless, the larger the perceived M2E delay, the larger the loss of interactivity.

6. Acknowledgments

This work was carried out within the framework of the LIMSON project sponsored by the Flemish Institute for the Promotion of Scientific and Technological Research in the Industry (IWT).

7. References

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

[2] "One-Way Transmission Time", *ITU-T Recommendation G.114*, February 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] "Definition of Categories of Speech Transmission Quality", *ITU-T Recommendation (Draft) G. GOVQ*, December 1998.

[5] "Provisional Planning Values for the Equipment Impairment Factor I_e", *Appendix to ITU-T Recommendation G.113 (Draft)*, December 1998.

[6] D. De Vleeschauwer, J. Janssen, G.H. Petit: "Delay Bounds for Low Bit Rate Voice Transport over IP Networks", *Proceedings of Performance and Control of Network Systems III, SPIE Symposium on Voice, Video and Data Communications*, Boston, USA, 19-22 September 1999.