Proceedings of the SPIE Conference on Performance and Control of Network Systems III, Vol. 3841, pp. 40-48, Boston (MA), 20-21 September 1999.

Delay bounds for low bit rate voice transport over IP networks

Danny De Vleeschauwer*, Jan Janssen, Guido H. Petit
Alcatel Bell, Corporate Research Center
Francis Wellesplein 1
B-2018 Antwerp, Belgium

ABSTRACT

The mouth-to-ear delay bounds that can be tolerated for undistorted voice are well known and standardized in the ITU-T Recommendations. In this paper, similar delay bounds are determined for voice that is transported in compressed form (over an IP network). More precisely, the dependency of these bounds on the low bit rate codecs used, the amount of echo control performed and the way the IP network is accessed, is investigated in detail.

Keywords: codec, echo control, E-model, mouth-to-ear delay, voice over IP

1. INTRODUCTION

Currently the real-time transport of (compressed) voice over an IP network (VoIP) [1, 3, 5] is an important center of attention. More precisely, whether Quality of Service (QoS) can be guaranteed for voice flows transported over a reasonably loaded IP network remains an open question. QoS in the context of VoIP is mainly determined by the Mouth-to-Ear (M2E) delay, i.e., the time that elapses between the moment the talker utters the words and the moment the listener hears them.

The M2E delay bounds that can be tolerated in traditional telephony are well known. They are standardized in ITU-T Recommendations G.114 [9] and G.131 [7] for undistorted voice, i.e., voice in analogue or the G.711 format. In this paper we extend these recommendations and determine the tolerable M2E delay bounds for voice that is transported in compressed form over an IP network. In particular, we investigate the dependency of these delay bounds on the low bit rate codecs employed, the level of echo control performed and the scenario used to access the IP backbone network. When designing a VoIP network, the voice packet sizes and dejittering delays then should be chosen such that the obtained delay bounds are met. Also from the analysis presented here, it may be concluded under which circumstances Echo Control (EC) is required.

The next section recalls the tolerable M2E delay bounds for traditional telephony. In Section 3 a method, based on the E-model, is described to obtain the tolerable M2E delay when the parameters of a voice call are known. Section 4 calculates the minimal delay associated with a certain codec. Section 5 assesses the tolerable M2E delays for two reference scenarios. Finally, in the last section some conclusions are drawn.

2. TRADITIONAL TOLERABLE MOUTH-TO-EAR DELAY

In the ITU-T Recommendations G.114 [9] and G.131 [7], which deal with tolerable M2E delays, the following rules are found for undistorted voice, i.e., voice in analogue or the G.711 format.

- 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 already necessary for M2E delays below 25 ms.
- When the echo is adequately controlled (i.e., the impairment of the echo is negligible compared to the impairment caused by the loss of interactivity)
  - 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 delay impact on the quality of the user applications, and
  - M2E delays above 400 ms are unacceptable.

The aim of this paper is to extend these M2E delay bounds to cases where voice is transported in compressed form, i.e., when a low bit rate codec is used (for bandwidth resource economy purposes).

* Correspondence: E-mail: danny.de_vleeschauwer@alcatel.be; Telephone: ++32-3-240 81 96; Fax: ++32-3-240 99 32

3. THE ETSI E-MODEL

The ETSI E-model[2, 4, 10] predicts 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 factor denoted as *R*. From this R-factor, which lies in the interval [0,100], subjective user reactions as for example the Mean Opinion Score (MOS) can be predicted.

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

$$R = R_0 - I_s - I_d - I_e + A \quad . \tag{1}$$

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, by too loud a side tone, etc. The third term I_d represents delayed impairments. Included are impairments caused by talker and listener echo and impairments caused 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. The fifth term A is the expectation factor. It 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 (i.e., GSM) equals 10. In this paper we take A=0, i.e., the quality of VoIP calls is compared with the quality of traditional wire-bound telephony.

From eq. (1) we notice that two calls having the same R-rating and therefore will be given the same MOS, can give a totally different subjective impression. One call might produce crystal clear, undistorted speech (e.g., I_e=0) but suffer from a relative large delay (e.g., I_d=10). The other call may distort the speech a little (e.g., I_e=10), while its delay is not noticeable (e.g., I_d=0).

ITU-T draft Recommendation G.govq[8] defines the range the R-factor has to fall in for the call to be rated of best ($90 \le R < 100$), high ($80 \le R < 90$), medium ($70 \le R < 80$), low ($60 \le R < 70$) or poor ($50 \le R < 60$) quality. Connections with R-values below 50 are not recommended. The same draft mentions that the term "toll quality" is an ill-used term. Here, we have taken the R-value of 72 (corresponding to a MOS of 3.7) as limit for *traditional quality*. Why this value was taken is explained in the last paragraph of Section 5.1. If the R-factor of a VoIP call is larger than 72, then the quality of the VoIP call is comparable to the quality of wired-bound telephony.

We consider the quality perceived by one party. Because this paper looks at VoIP, we only consider the impairment I_d caused by delay and the impairment I_e caused by the use of low bit rate codecs.

The impairment factor I_d is the sum of three contributions: the impairment caused by talker echo, the impairment caused by listener echo and the impairment caused by the loss of interactivity.

First, talker echo disturbs party 1 (see Figure 1), because he hears an attenuated and delayed echo of his own voice. This echo may be caused by a reflection close to party 2, e.g., in the hybrid of the terminating Public Switched Telephone Network (PSTN) node or in the terminal equipment of party 2. The Talker Echo Loudness Rating (TELR)[10] is the amount (expressed in dB) with which the echo signal is attenuated with respect to the original signal. The TELR is determined by the Echo Loss (EL) EL_2 close to party 2 (measured with respect to a certain reference point), by the attenuation of the signal from party 1 to the reference point (the Send Loudness Rating (SLR)) and by the attenuation of the signal from the reference point to party 1 back again (the Receive Loudness Rating (RLR)). That is,

$$TELR = SLR + RLR + EL_2 \quad . \tag{2}$$

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 caused by a reflection close to party 1 with attenuation EL_1, followed by a reflection close to party 2 with attenuation EL_2. The Weighted Echo Path Loss (WEPL)[10] is the amount (expressed in dB) the listener echo is attenuated with respect to the original signal heard by party 1, i.e.,

$$WEPL = EL_1 + EL_2 \quad . \tag{3}$$

Typical values for *SLR*, *RLR*, EL_1 and EL_2 depend on the access network and the terminal equipment and will be given in Section 5. The values for the ELs (EL_1 and EL_2) can be increased by the use of an Echo Controller (EC). It is advisible to deploy the EC close to the source of echo, because the less delay there is between the original and echo signal the easier it is

to attenuate the echo. A simple EC can easily increase the EL by 30 dB and a more elaborate EC can even get rid of the echo completely, in which case the EL is infinite.

The third delay-related factor that may disturb party 1 is the loss of interactivity. If the (one-way) delay is too large, an interactive conversation becomes impossible.

All those impairments (party 1 hears a too loud echo of his own voice when he is talking, party 1 hears a too loud echo of party 2 and party 1 has to wait too long for the response of party 2) influence the subjective quality observed by party 1. The E-model that predicts this subjective quality takes all those impairments into account.

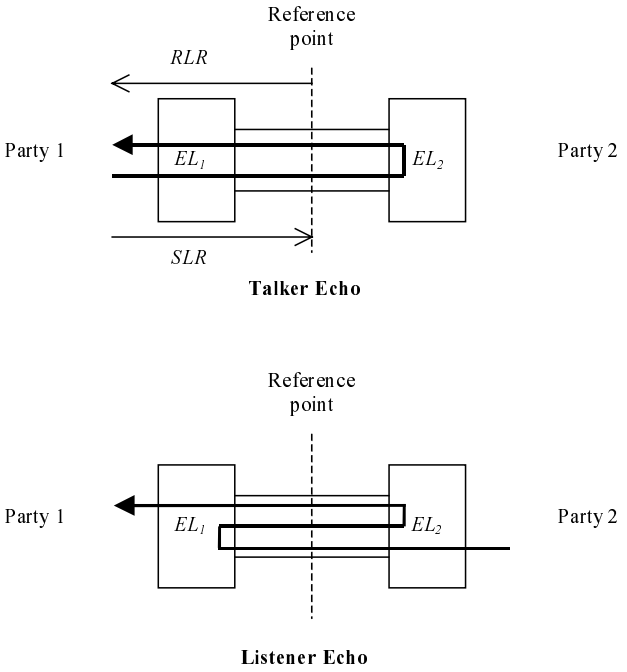

Figure 1: Talker and listener echo.

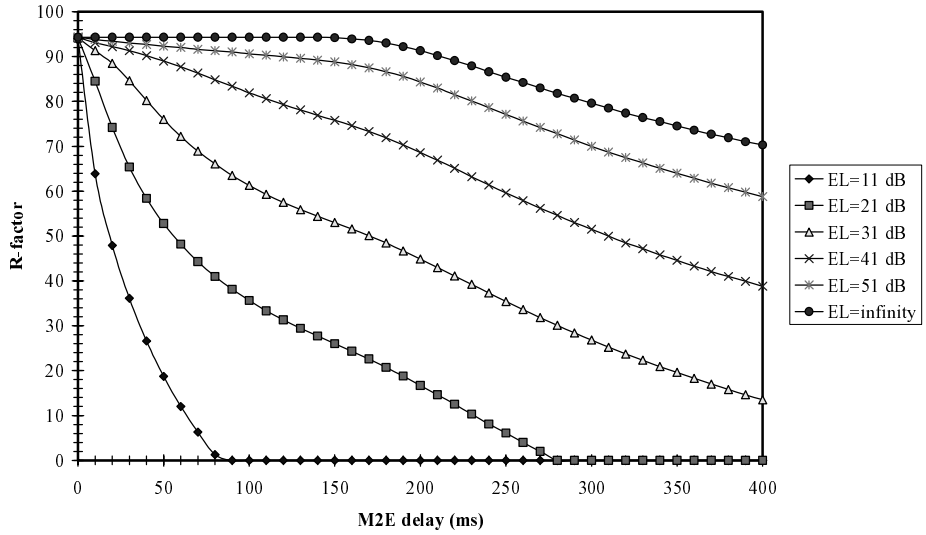

Figure 2: The R-factor as a function of the M2E delay for the G.711 codec and various levels of echo loss (EL).

Figure 2, calculated with the E-model, shows the influence of the M2E delay on the R-factor for calls transported in the G.711 format over the IP network. As expected from the reasoning above, the impairment associated with delay is strongly influenced by the ELs close to party 1 and 2. Figure 2 gives the result for a symmetric scenario, i.e., when EL_1=EL_2. Various levels of EL are considered. Observe that the R-factor is a non-increasing function of the M2E delay and that the maximal R-factor in Figure 2 equals 94.3 for zero M2E delay.

If the voice is transported in compressed form, the R-factor decreases by an amount I_e associated with that codec. These I_e-values, tabulated in Table 2, are averages of (lots of) subjective tests. Curves similar to the curves of Figure 2 can be made for every (standardized) codec. These new curves can be obtained from Figure 2 by a downward shift equal to the impairment term I_e associated with the codec. From Table 2 it is then clear that certain codecs, i.e., G.726 or G.727 at 16 and 24 kb/s and GSM-HR, never obtain traditional quality, because their intrinsic impairment factor I_e is too large, i.e., larger than 94.3-72=22.3. These codecs will not be considered further.

If transcoding occurs, i.e., if somewhere along the route the voice is translated from codec X to codec Y, the G.711 codec is used as an intermediate format. As such, the impairment terms associated with these three codecs (X, G.711 and Y) should be added to obtain the overall I_e, because impairments are additive on the R-scale. Yet, as I_e=0 for G.711, the latter operation is equivalent to adding only the impairment terms of codecs X and Y. The R-factors for all combinations of two codecs can be found in Table 1, the diagonal entries of which obviously correspond to the use of one particular codec with the G.711 format in between. As the slope of the perfect EC curve in Figure 2 is horizontal for M2E delay values between 0 and 150 ms, these R-factors are valid for M2E delays in the latter range under the assumption of perfect EC. We can conclude that transcoding can be very harmful to the quality of a call. In practice the order of tandeming the codecs has a small influence, which cannot be seen in the symmetric table of Table 1 as the E-model neglects this phenomenon. This effect is strongest for large impairment factors, and, thus, for poor R-factors, which is of low interest to this paper.

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	75.3	67.3	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	69.3	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

R-value range	90 - 100	80 - 90	70 - 80	60 - 70	0 - 60 *
Speech transmission quality category	best	high	medium	low	(very) poor

*** R-values below 50 are not recommended**

Table 1: Influence of transcoding on the R-factor for voice calls with perfect Echo Control (EC).

4. MINIMAL DELAY INHERENT TO A CODEC

All codecs work according to the same principle, illustrated in Figure 3. They first collect a few samples of speech of length T_F (in ms), referred to as the voice frame. Sometimes they also need some samples after the ones being encoded in order to better encode the samples of the current voice frame. The length of this block of samples (in ms) is referred to as the look-ahead T_{LA}. Then, they calculate a code word of length B_F (in bits). To calculate this code word the processor takes an encoding time T_{enc}. At the receiver side the decoder uses the code word to produce a close copy of the original voice frame. The time it takes to perform this operation is referred to as the decoding time T_{dec}. Remark that the bit rate of the codec is given by

$$R_{cod} = \frac{B_F}{T_F} \quad . \tag{4}$$

Code words need to be transported from the encoder to the decoder. This is done by packing code words in IP packets. Several consecutive code words may be collected in one IP packet. This introduces additional delay, called the packetization delay. We define the codec delay as

$$T_{cod} = T_{enc} + T_{dec} + T_{LA} \quad , \tag{5}$$

and the packetization delay as

$$T_{pack} = N_F T_F \quad , \tag{6}$$

where N_F is the number of voice frames per IP packet.

As the encoding and decoding process has to be performed in real-time, the time needed to encode T_{enc} and decode T_{dec} the voice signal are upper bounded by the voice frame length T_F. These values for T_{enc} and T_{dec} include, besides the time elapsed for the execution of the instructions required to process the code words, also the time that the encoding or decoding process needs to wait because the processor is occupied by other processes.

Along the route other delays may be introduced: a service delay and queueing delay in each IP node, propagation delay along the transmission lines, a dejittering delay at the destination, etc., but these are not directly related to the codec. Since an IP packet needs to wait for at least 1 code word, the minimal delay inherent to a certain codec is given by

$$T_{enc} + T_{dec} + T_{LA} + T_F \tag{7}$$

Table 2 gives the relevant parameters for the standardized codecs.

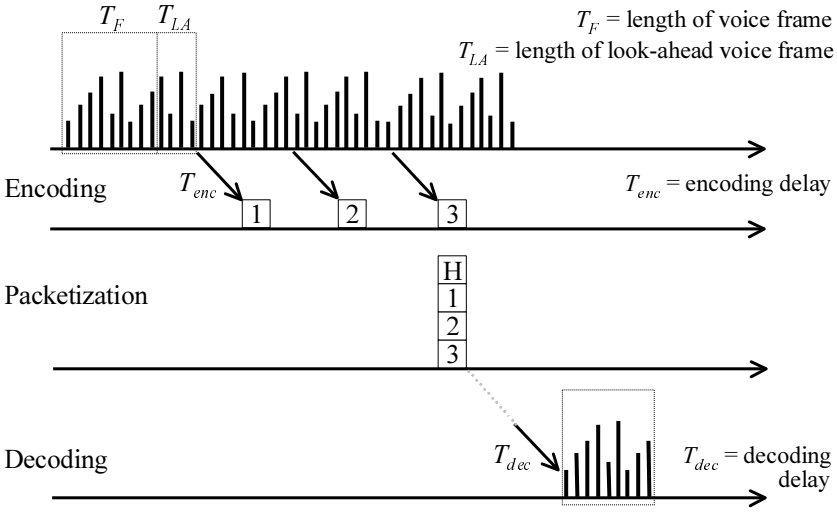

Figure 3: Codec and packetization delay.

Origin	Standard	Type	I_e	T_F (ms)	T_{LA} (ms)	B_F (bits)	R_{cod} (kb/s)
ITU-T	G.711	PCM	0	0.125	0	8	64
	G.726, G.727	ADPCM	50	0.125	0	2	16
			25			3	24
			7			4	32
			2			5	40
	G.728	LD-CELP	20	0.625	0	8	12.8
			7			10	16
	G.729(A)	CS-ACELP	10	10	5	80	8
	G.723.1	ACELP	19	30	7.5	158	5.3
		MP-MLQ	15			189	6.3
ETSI	GSM-FR	RPE-LTP	20	20	0	260	13
	GSM-HR	VSELP	23	20	0	112	5.6
	GSM-EFR	ACELP	5	20	0	224	12.2

Table 2: Standardized codecs with impairment factor I_e, voice frame length T_F (ms), look-ahead length T_{LA} (ms), code word length B_F (bits) and bit rate R_{cod} (kb/s).

5. RESULTS FOR SOME SYMMETRIC SCENARIOS

5.1. Tolerable mouth-to-ear delays in the phone-to-phone scenario

The first voice communication scenario, referred to as the phone-to-phone scenario, is illustrated in Figure 4. The VoIP calls use the traditional PSTN as access network from the source to the ingress GateWay (GW) and from the egress GW to the destination. In the ingress GW at the edge of the IP network the voice is encoded and packetized. The voice packets are then routed over the IP network from the ingress GW to the egress GW. At the egress GW the jitter introduced in the network is compensated and the voice signal is decoded.

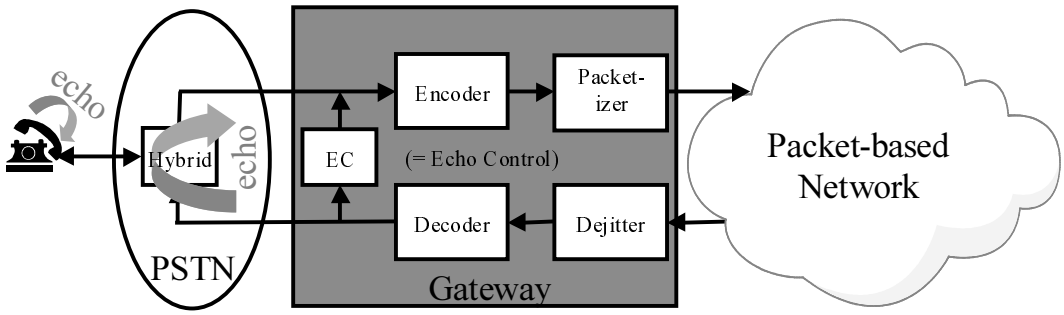

Figure 4: Phone-to-phone scenario

There are two sources of echo in this scenario: hybrid echo may be generated in the 4-to-2-wire hybrid and acoustic echo may be generated in a terminating phone in case it has a low TCL, i.e., in case it has a high coupling between the microphone and the speaker. If the terminal is a traditional phone, the EL is mainly determined by the hybrid and 21 dB is a typical value if no EC is performed [7]. For SLR and RLR, we take the standard values of respectively 7 and 3 dB [10]. Two EC are considered: one that reduces the echo by an additional 30 dB and a perfect EC, which gets rid of all echo.

All necessary processing, i.e., encoding, decoding and EC, is performed by dedicated Digital Signal Processors (DSPs) in the GWs. It is assumed that the DSPs in the GWs are chosen such that they are exploited to the fullest. This means that if a GW handles the maximum number of voice calls it is designed for, its DSPs are busy all of the time. Therefore, the

encoding time T_{enc} and decoding time T_{dec} associated with one call are likely to be equal to their maximal value T_F. Hence, in this scenario the delay inherent to the codec (see eq. (7)) is $3T_F+T_{LA}$.

Table 3 gives the tolerable M2E delay under which traditional quality is reached for various codecs in the phone-to-phone scenario and for various levels of EC. Also indicated is the fact whether voice calls, transported in a certain codec format, need EC under all circumstances or not. EC is always needed if the tolerable M2E delay without EC is smaller than the delay inherent to the codec. A "no" in Table 3 just states that if other delay components are negligible, no EC is required. If other delay components (e.g., propagation delay for long distance calls) considerably contribute to the M2E delay, EC may still be needed.

Origin	Recommendation/ Standard	$3T_F+T_{LA}$ (ms)	T_{M2E}(EC-0) (ms)	EC always required?	T_{M2E}(EC-30) (ms)	T_{M2E}(EC-p) (ms)
ITU-T	G.711	0.375	23	No	285	379
	G.726, G.727	0.375	15	No	237	305
		0.375	20	No	270	356
	G.728	1.875	2	No	60	192
		1.875	15	No	237	305
	G.729(A)	35	12	Yes	216	278
	G.723.1	97.5	3	Yes	90	203
		97.5	7	Yes	175	237
ETSI	GSM-FR	60	2	Yes	60	192
	GSM-EFR	60	17	Yes	250	324

(EC-0=no echo control, EC-30=echo control by 30 dB, EC-p=perfect echo control)

Table 3: The tolerable M2E delay under which traditional quality is reached for various codecs in the phone-to-phone scenario.

The results for the G.711 codec in Table 3 correspond to the traditional M2E delay bounds described in Recommendations G.114 and G.131 (see Section 2). The 25 ms bound above which EC is required and the 400 ms bound above which interactivity is impossible, are found (more or less) in Table 3. This is in fact a direct consequence of and the reason for the choice R=72 as definition for traditional quality. Moreover, it can be seen from Figure 2 that at about 150 ms, a value mentioned in Recommendation G.114, the quality of a call with perfect EC starts to drop. Hence, Table 3 can be interpreted as an extension of Recommendations G.114 and G.131 for other types of codecs.

Concerning this 150 ms bound, an additional comment can be made. From the reasoning in Section 3 on how the curves in Figure 2 should be adapted for other codecs (by a downward shift equal to the impairment factor of the codec), it follows immediately that this bound is codec independent. That is, the quality of a voice call with perfect EC remains more or less constant for M2E delays below150 ms, and this for any possible codec (or combination of codecs).

5.2. Tolerable mouth-to-ear delays in the PC-to-PC scenario

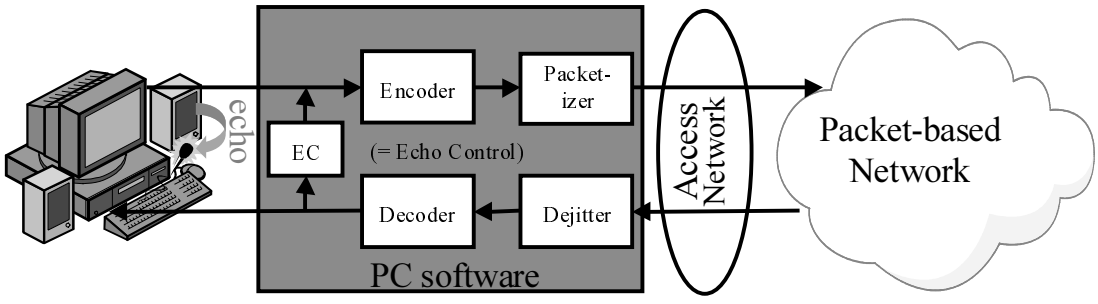

Figure 5: PC-to-PC scenario

In the second scenario, referred to as the PC-to-PC scenario, the functionality of the GWs of the previous scenario resides in the PCs (see Figure 5). The originating PC compresses and packetizes the voice signal and routes it via an access network (e.g., via an ADSL-modem) on the IP network. The terminating PC dejitters the incoming packet flow and decodes the voice signal.

In this scenario only acoustic echo occurs and the EL equals the TCL of the terminal. As a multimedia PC is not optimized to have a large TCL, i.e., as a lot of the energy of the speaker might be coupled into the microphone, this acoustic echo is likely to be larger than the acoustic echo in the phone-to-phone scenario. We assume the EL to be equal to 11 dB if no EC is performed, while we still use the default values SLR = 7 dB and RLR = 3 dB [10]. Again, we consider EC of 30 dB and perfect EC.

All necessary processing, i.e., encoding, decoding and echo control, is performed in the processor of the PC, which may be have to perform other user tasks as well. In the PC-to-PC scenario it is assumed that the processor in the PC is fast and gives absolute priority to encoding and decoding tasks. Therefore, the encoding time T_{enc} and decoding time T_{dec} are likely to be small and negligible with respect to other delay components. Hence, we take T_F+T_{LA} as the delay inherent to a codec in this scenario (see eq. (7)).

Table 4 gives the tolerable M2E delay under which traditional quality is reached for various codecs in the PC-to-PC scenario and for three levels of EC. Also indicated is the fact whether voice calls, transported in a certain codec format, need EC under all circumstances or not. As in the previous scenario, EC is always required if the tolerable M2E delay without EC is smaller than the delay inherent to the codec. Again, a "no" in Table 4 just states that if other delay components are negligible, no EC is required. Other delay components can considerably contribute to the M2E delay, so that EC may still be needed.

Observe that the M2E delay bounds for perfect EC are exactly the same as in the phone-to-phone scenario, which can be easily explained by the infinite value of the EL in both cases. The 150 ms delay bound above which calls with perfect EC start to degrade in quality also turns out to be a codec-independent fact in this scenario.

Origin	Recommendation/ Standard	T_F+T_{LA} (ms)	T_{M2E}(EC-0) (ms)	EC always required?	T_{M2E}(EC-30) (ms)	T_{M2E}(EC-p) (ms)
ITU-T	G.711	0.125	7	No	164	379
	G.726, G.727	0.125	4	No	110	305
		0.125	6	No	150	356
	G.728	0.625	1	No	20	192
		0.625	4	No	110	305
	G.729(A)	15	4	Yes	90	278
	G.723.1	37.5	1	Yes	30	203
		37.5	2	Yes	59	237
ETSI	GSM-FR	20	1	Yes	20	192
	GSM-EFR	20	5	Yes	124	324

(EC-0=no echo control, EC-30=echo control by 30 dB, EC-p=perfect echo control)

Table 4: The tolerable M2E delay under which traditional quality is reached for various codecs in the PC-to-PC scenario.

6. CONCLUSIONS

In this paper the tolerable mouth-to-ear delay bounds under which traditional quality is reached, were calculated for calls that are transported over an IP network (or any other packet-based network) in compressed form. These tolerable mouth-to-ear delays depend on the codec that is used to compress the voice, the communication scenario and the amount of echo control that is performed. In particular, they are in fact extensions of the delay bounds for undistorted voice, reported in ITU-T Recommendations G.114 and G.131, but now for low bit rate codecs.

On the other hand, staying below 150 ms mouth-to-ear delay and using perfect echo control guarantees the best possible quality obtainable. The quality then only depends on the distortion introduced by the codec(s).

Furthermore, it was demonstrated that each codec has an associated inherent delay. If this inherent delay is larger than the tolerable mouth-to-ear delay when no echo control is used, this means that voice calls using this codec require echo control under all circumstances to reach traditional quality.

ACKNOWLEDGMENTS

This work was carried out within the framework of the project LIMSON sponsored by the Flemish Institute for Science and Technology (IWT).

REFERENCES

1. A. Cray, "Voice over IP, Hear's how", *Data Communications*, pp. 44-58, Apr. 1998.
2. N. O. Johannesson, "The ETSI Computation Model: A Tool for Transmission Planning of Telephone Networks", *IEEE Communications Magazine*, pp. 70-79, Jan. 1997.
3. T.J. Kostas, M.S. Borella, I. Sidhu, G.M. Schuster, J. Grabiec, J. Mahler, "Real-Time Voice over Packet-Switched Networks", *IEEE Network*, pp. 18-27, Jan./Feb. 1998.
4. P. Meschkat, "TPE: Transmission Planning (end-to-end) using the E-model (supporting ETSI EG 201 050)", software tool for Windows, Alcatel Telecom, Dec. 1997.
5. K. Van Der Wal, M. Mandjes, H. Bastiaansen, "Delay Performance Analysis of the New Internet Services with Guaranteed QoS", *Proceedings of the IEEE*, Vol. 85, No. 12, pp. 1947-1957, Dec. 1997.
6. "Acoustic Echo Controllers", ITU-T Recommendation G.167, Mar. 1993.
7. "Control of Talker Echo", ITU-T Recommendation G.131, Aug. 1996.
8. "Definition of Categories of Speech Transmission Quality", ITU-T Recommendation draft G.govq, Dec. 1998.
9. "One-Way Transmission Time", ITU-T Recommendation G.114, Feb. 1996.
10. "Speech Processing, Transmission and Quality Aspects (STQ); Overall Transmission Plan Aspects for Telephony in a Private Network", ETSI Guide draft 201 050, Nov. 1998.