PERFORMANCE ISSUES FOR VOICE OVER ADSL

Annelies Van Moffaert, Jan Janssen, Danny De Vleeschauwer,
Maarten Büchli, Guido H. Petit
Alcatel Bell, Network Strategy Group
Francis Wellesplein 1
B-2018 Antwerp, Belgium
e-mail: {jan.janssen, danny.de_vleeschauwer, annelies.van_moffaert,
maarten.büchli, guido.h.petit }@alcatel.be
and

Marc Verhoeyen
Alcatel Bell, Carrier Internetworking Division
De Villermontstraat 38
B-2550 Kontich, Belgium
E-mail: marc.verhoeyen@alcatel.be

ABSTRACT

In this paper, we study the performance of (multiple) voice calls (multiplexed) over an Asymmetric Digital Subscriber Line (ADSL). More precisely, we determine the effective bit rate and the subjective quality of voice-over-ADSL calls. The influence of several parameters (codec, packetization interval, number of calls multiplexed onto one ATM connection) is investigated. It turns out that if these parameters are chosen intelligently and the echo is perfectly controlled, the subjective quality of such voice-over-ADSL calls is comparable to the quality attained in traditional, circuit-switched networks.

Keywords: ADSL, ATM, performance, quality, voice

1. Introduction

Currently the real-time transport of voice over packet-based networks is an important center of attention. One of the most important topics to be tackled in this context, is whether such packet-based voice calls can be transported at quality levels that are comparable to the quality in traditional, circuit-switched networks.

In this paper, we consider the performance of multiple voice calls multiplexed over an Asymmetric Digital Subscriber Line (ADSL). More precisely, we determine the effective bit rate and the subjective quality (i.e. the rating R) of Voice-over-ADSL (VoADSL) calls. We consider voice calls between a VoADSL phone and a Public Switched Telephone Network (PSTN) phone as well as voice calls between two VoADSL phones. Other parameters considered are the choice of the codec and packetization delay, the amount of calls multiplexed onto one ATM connection and the level of echo control. With optimally tuned parameters, the subjective quality of such VoADSL calls is shown to be similar to the quality attained in traditional PSTN networks.

The paper is structured as follows. In the next section, the reference network under consideration is described, both from an architectural and a functional point of view. Section 3 discusses the E-model, with emphasis on the influence of the Mouth-to-Ear (M2E) delay and distortion. In Section 4, the capacity requirements for VoADSL calls are calculated for different codecs, packetization intervals, and a varying number of voice sources to be multiplexed onto one ATM connection. After a description and estimation of the encountered delay components, Section 5 evaluates the performance for several VoADSL scenarios. Finally, some conclusions are drawn and some future study topics are identified.

2. The considered reference network

2.1. Architectural description

In Figure 1, the ADSL access network considered in this paper is depicted.

The Customer Premises Equipment (CPE) gets access to the voice or data network (PSTN or Internet) via a copper pair. The spectrum of this copper pair is partitioned into two parts: the lower frequency band is used for the Plain Old Telephone Service (POTS) or Integrated Services Digital Network (ISDN) telephony services and the rest of the spectrum is used for the ADSL service. Although, most of the considerations are valid for POTS and ISDN, we mainly concentrate on POTS. The Splitter (S) at the customer premises and the splitter at the network separate both frequency bands.

The phone directly attached to the splitter (denoted as “lifeline” in Figure 1) is connected via the base band of the copper pair to the PSTN, over which it is switched to its destination. Except from the splitters the signals from or to this phone do not traverse any other network components than in the case that the copper pair does not carry an ADSL service.

The user can use the available ADSL capacity for gaining access to the Internet or for serving additional telephone lines, called “derived lines”, both via an Integrated Access Device (IAD). A phone connected to this device is referred to as a VoADSL phone. An IAD groups several components in one box. The main components of interest to this paper are the Subscriber Line Interface Circuit (SLIC), the Voice over ATM (VoATM) GateWay (GW) and the ADSL Network Terminator (ANT).

The data terminal(s) and VoATM GW use a different ATM connection to transport their ATM cells.

First, the ATM connection between the data terminal(s) (e.g. a PC) and the Network Access Server (NAS) can consume part of the ADSL capacity. The user terminal (e.g. a PC) segments its IP packets via ATM Adaptation Layer (AAL) 5 into ATM cells. These ATM cells are transported via the ANT and the Digital Subscriber Line Access Multiplexer (DSLAM) over an ATM network to a NAS. There, the IP packets are reassembled and routed over the Internet. IP packets destined for the data terminal follow the reverse route. Remark that the data terminal might also run a Voice over IP (VoIP) application, e.g., over the Internet to a VoIP GW somewhere in the Internet. The performance of this VoIP scenario is not considered in this report.

Second, the ATM connection between the VoATM GW in the IAD and a VoATM GW in the network consumes part of the ADSL capacity. An IAD at the customer premises can typically serve 2 to 32 VoADSL phones. The voice streams generated by the VoADSL phones (connected to the SLIC) are encoded and packed into ATM cells in the VoATM GW in the IAD using AAL2. These ATM cells are transported via the ANT and the DSLAM over an ATM network to the VoATM GW of the network operator. There, the stream of ATM cells with voice payloads is dejittered and decoded, and switched over the PSTN to its destination. The voice streams destined for the VoADSL phone follow the reverse path. The performance of both the VoADSL-phone-to-phone scenario and the VoADSL-phone-to-VoADSL-phone scenario will be considered in this paper. In the latter case the call is still partly switched over a PSTN, the voice is only transported over an ADSL network in the access part.

A variant of the configuration discussed so far is the one where the DSLAM and VoATM GW (of the network operator) are integrated into one box. Yet, as the resulting configuration does not have a large impact on the performance issues discussed in this paper, it is not discussed explicitly in the following.

As mentioned before, data and voice cells are transported over different ATM connections. In this paper, we assume that AAL2 is used in the VoATM GWs. As such, there are three possibilities: all voice packets can be multiplexed onto a single Virtual Circuit (VC) / Virtual Path (VP), each voice stream can use its separate VC within a VP or some intermediate solution can be used.

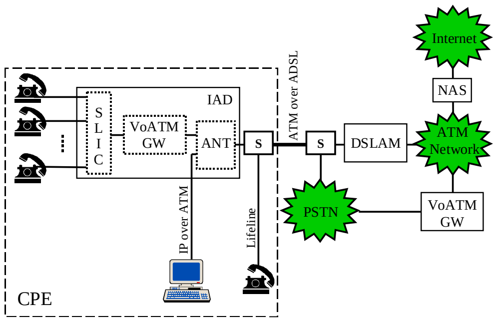

Figure 1: Reference network under consideration.

2.2. Functional description

In this section we describe the functionality of each block in Figure 1. We only consider the functions that can have an impact on the voice quality.

The SLIC serves as an interface between the VoADSL phones and the VoATM GW in the IAD. From a performance point of view, its important functions are digitizing the voice signal and conversion from four to two wires in a (4-to-2-wire) hybrid. The digitizing function consists of sampling and quantizing the analog signal (according to the A- or μ-law), i.e., the voice signal is transformed from the analog into the G.711 format. The 4-to-2-wire hybrid function of the SLIC impacts the voice quality because it generates (hybrid) echo. We assume that the echo generated in this hybrid is attenuated by 21 dB with respect to the incoming signal, which is a typical value for a PSTN [6]. This attenuation is referred to as the Echo Loss (EL).

At least two VoATM GWs are involved in a VoADSL conversation. In the VoADSL-phone-to-phone scenario there is one in the IAD and one in the network between the ATM network and a PSTN. In the VoADSL-phone-to-VoADSL-phone scenario there are four, i.e. one in each IAD and two in the network. In one direction the VoATM GW encodes the G.711 voice signal (if a low bit rate codec is used) and packs the code words into ATM cells using AAL2. Also Voice Activity Detection (VAD) can be used to avoid sending ATM cells when there is no voice signal. In the other direction the VoATM GW dejitters the received ATM cell stream and decodes the received payloads. The VoATM GWs also perform echo control. We assume that the GW in the IAD controls the echo generated in the hybrid of the connected SLIC and that the GW between the ATM network and the PSTN controls the echo generated in the hybrids of the PSTN in the VoADSL-phone-to-phone scenario. Remark that in the

VoADSL-phone-to-VoADSL-phone scenario the echo controllers of the VoATM GWs in the network have no echo to compensate, because there is no 4-to-2-wire hybrid conversion in the PSTN. The same holds for the echo controllers in the IADs when ISDN phones are attached to them.

The ANT and DSLAM (de)multiplex the ATM connection(s) carrying data (between PC and NAS) and the ATM connection(s) between the VoATM GWs. We assume that the ATM cells carrying voice are transported over the Constant Bit Rate (CBR) service category. Even in the case that VAD is used no attempt is made to exploit statistical multiplexing. For the CBR connection carrying the voice ATM cells the Peak Cell Rate (PCR) is allocated in the network and a low delay and jitter is guaranteed. The ANT, DSLAM and any other traversed ATM switches give priority to ATM cells containing voice over cells containing data (based e.g. on the VP/VC Identifier of the ATM cells).

The capacity of the ADSL link depends on the quality of the copper pair. In ADSL the upstream capacity is much smaller than the downstream capacity. These (upstream and downstream) capacities of the ADSL link are determined at configuration time. The upstream and downstream bit rates are configured such that (in the absence of impulsive noise) a bit error rate of about 10^{-7} results. For our purposes this bit error rate of 10^{-7} is negligible, since the loss of ATM cells over the ADSL link due to this bit error rate is extremely small. Impulsive noise can introduce additional cell loss on the ADSL link. To protect the ADSL bit pipe from impulsive noise the bit stream is interleaved and protected with a Forward Error Correction (FEC) method (i.e. a Reed Solomon code). This introduces additional delay. We refer to the delay to transport one bit over the ADSL bit pipe as ADSL bit pipe delay. The larger the interleaving depth is, the longer the impulses that can be corrected, can be. The maximum interleaving depth introduces an ADSL bit pipe delay of 20 ms and can correct noise pulses up to 250 μs. The fast mode, where the minimal interleaving depth is used, introduces the smallest possible ADSL bit pipe delay of 2 ms, but can correct no noise impulses. Interleaving is required for data applications where low bit error rates are mandatory. Hence, the ADSL interleaved mode has to be used for data applications. For voice a larger cell loss can be tolerated [1]. Hence, for voice services the ADSL interleaved mode is not strictly required. The influence of uncorrectable noise pulses on the cell loss (and, hence, on the voice quality) is not considered in this paper.

3. The E-model

As far as quality is concerned the difference between a VoADSL call and a PSTN call, lies in the M2E delay and distortion that is introduced.

Assuming optimally tuned telephone sets, traditional (wire-bound) PSTN calls do not suffer from distortion. As such, the key factor that determines the quality is the M2E delay, defined as the delay incurred from the moment the talker utters the words until the instant the listener hears them.

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 the case of VoADSL calls, distortion may be introduced by the codec that compresses the voice signal or by the loss of voice payloads (in the network or in the dejittering buffer).

The E-model [2] can be used 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 can be used to predict subjective user reactions, such as the Mean Opinion Score (MOS). The *R* scale was defined such that impairments are approximately additive in the *R* range of interest. The rating *R* is given by:

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

The first term R_0 groups the effects of noise, such as background noise 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 is also used to take into account the influence of packet loss. 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 (e.g. GSM) is 10.

Comparing a VoADSL call with a traditional call, we notice (from Figure 1) that the only difference lies in the way the PSTN is accessed. This has an impact on the quality of the call: VoADSL access introduces more delay and distortion than traditional circuit-switched access does. First, the delay for VoADSL access (where the most important contributions are encoding, packetization, queuing, service, dejittering and decoding delay) is larger than for traditional circuit-switched access (where the M2E delay is mainly made up of switching delay). Second, in contrast with circuit-switched access, as a result of voice compression and cell loss during transport or payload loss in the dejittering buffer, the distortion of packetized voice calls is not negligible. Notice that also on the PSTN itself delay (mainly propagation and switching delay) is introduced. We assume that no distortion is introduced in the PSTN[1].

[1] On transcontinental links compression (and VAD) may be used to exploit the capacity of this link optimally.

We have studied the impact of the one-way M2E delay (via I_d) and the distortion (via I_e) on the quality of a call between two parties, referred to as party 1 and party 2 (see Figure 2).

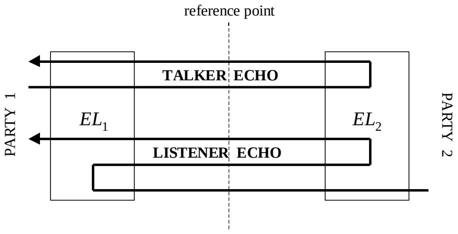

Figure 2: Talker and listener echo.

Other factors, like background noise and a connection that is too loud, also impair the quality (via R_0 and I_s) of a VoADSL call, but as these factors are not fundamentally different from a traditional PSTN call they were not considered. Furthermore, as the objective was to make a fair comparison between the quality of VoADSL calls and traditional wire-bound PSTN calls, the expectation factor A was set to 0.

3.1. 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 M2E delay. This impairment is the sum of three contributing impairments: talker echo, listener echo and loss of interactivity.

First, talker echo disturbs party 1, who hears an attenuated and delayed echo of his/her own voice. This echo is caused by a reflection close to party 2. The level of this echo is determined by the echo loss EL_2 of party 2. The Echo Loss (EL) is defined as ratio of the level of the original signal and the level of the echoed signal expressed in dB.

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 hybrids of the local PSTN or in the SLIC at the customer premises. If the call is terminated over a PSTN to a traditional handset, the echo is mainly caused by the hybrid in the PSTN. A typical value for the EL is 21 dB [6]. If the call is terminated over a VoATM GW (in an IAD) to a VoADSL phone, we take the same value for the EL, since a SLIC contains a hybrid similar to the one in the PSTN.

The echo losses EL_1 and EL_2 can be increased by using an echo controller. An echo controller usually consists of a linear part (which estimates the echo and subtracts it from the disturbed signal, a process referred to as echo cancellation) and a non-linear part that smothers the remaining part of the echo. The linear part does not affect the original signal and should suppress the echo by 30 dB in order to be standard-compliant [8]. The non-linear part of the echo controller can get rid of the remainder of the echo (in which case the echo losses EL_1 and EL_2 increase to infinity), but affects the original signal as well. This may introduce some distortion in the voice signal. Since this distortion is similar to the distortion introduced by VAD and the distortion introduced by a good VAD scheme is small [3], we assume that the distortion the echo controller introduces, is negligible as well. Perfect echo control can be achieved at moderate computational cost.

Since it gets harder to remove the echo when it is delayed more with respect to the original signal, the echo controller should be installed as close to the source of the echo as possible. This means that echo generated in the SLIC should be compensated in the VoATM GW in the IAD and echo generated in the hybrid of a local PSTN should be compensated in the VoATM GW between the ATM network and the PSTN. Remark that in the VoADSL-phone-to-VoADSL-phone scenario the echo controllers of the VoATM GWs in the network have no echo to compensate, because no 4-to-2-wire hybrid is traversed in the PSTN. In addition, we assume that the tandeming of echo controllers (in the VoATM GWs) does not contribute noticeably to the distortion.

The third delay-related factor that may disturb party 1 is the loss of interactivity. If the M2E delay is too large, an interactive conversation becomes impossible.

We have used the E-model, which takes all these delay-related impairments into account, to calculate the rating R given by party 1 in the case of undistorted voice. In the case of VoADSL calls, undistorted calls are calls transported without payload loss in the G.711 format. Figure 3 shows the influence of the M2E delay on the rating R for different values of the EL when the ELs at both end points are equal ($EL_1 = EL_2$). The impairment associated with delay is strongly influenced by this EL value.

Observe that the rating R is a non-increasing function of the M2E delay. The intrinsic quality of a voice call is defined as the rating R associated with a zero M2E delay. The intrinsic quality of a VoADSL call transported in the G.711 format over the ATM network without packet loss using optimally tuned telephone sets corresponds to $R = 94.3$. Figure 3 shows that if echo is perfectly controlled ($EL_1 = EL_2 = \infty$), this voice call retains its intrinsic quality up to a M2E delay of 150 ms.

ITU-T Recommendations G.114 [4] and G.131 [5] specify the following tolerable M2E 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 M2E delay is larger than 25 ms.
- When the echo is adequately controlled:
  - a M2E delay of up to 150 ms is acceptable for most user applications;

This compression may introduce some distortion. We assume this distortion to be negligible.

– a M2E 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
– a M2E delay above 400 ms is unacceptable.

It can be seen from Figure 3 that for an echo loss of 21 dB, the rating R drops below 70 at a M2E delay of 25 ms. For calls with perfect echo control, the rating R drops below 70 at a M2E 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. Hence, for a call to be of "traditional quality" its rating should be above 70. Remark that the interactivity bound of 150 ms can also be observed in Figure 3 for infinite EL.

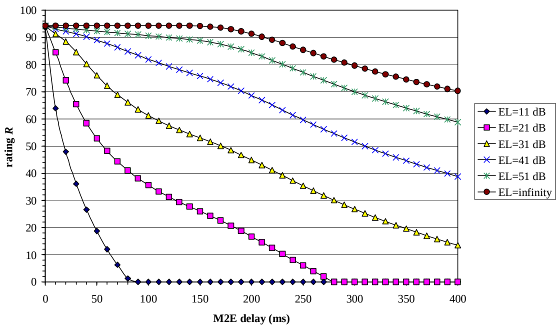

Figure 3: The rating R versus the M2E delay for undistorted voice and various EL values.

3.2. 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 payload loss during the transport of voice payloads from party 2 to party 1.

Table 1 summarizes the distortion impairment and intrinsic quality associated with the codecs considered in this paper, i.e., the G.711 codec at 64 kb/s, the G.726 codec at 32 kb/s and the G.729 codec at 8 kb/s [3]. The distortion impairment I_e associated with a codec increases as the payload loss ratio increases. In [3], this effect is investigated for four codecs (G.711, G.723.1, G.729, GSM-EFR), assuming that voice payloads are lost at random. Results are not yet known for other codecs and payload loss patterns.

In both scenarios considered in this paper, the voice signal is not transported in the same format end-to-end. More precisely, in the VoATM GWs in the network the voice is transcoded from the codec format used in the VoATM GW in the IAD into the G.711 format (or vice versa). Consequently, in the VoADSL-phone-to-VoADSL-phone scenario, the impairment terms associated with the codecs used in the (VoATM GWs in the) IADs have to be added to obtain the overall distortion impairment I_e, because the rating scale of the E-model was defined such that impairments are approximately additive on the R scale. As such, it can be seen that transcoding can be very harmful to the quality of a call. Additional results on transcoding are reported in [1].

Name	R_{cod} (kb/s)	B_F (bits)	T_F (ms)	T_{LA} (ms)	I_e	intrinsic quality
G.711	64	8	0.125	0	0	94.3
G.726	32	4	0.125	0	7	87.3
G.729	8	80	10	5	10	84.3

Table 1: Major parameters of considered codecs.

If the M2E delay, echo loss and distortion impairment are known, the quality of a VoADSL call (i.e. its rating R) can be derived from Figure 3 as follows. First, identify the curve on Figure 3 that corresponds to the given EL. Then, using this curve, read the rating R corresponding to the given M2E delay. Finally, subtract the (total) distortion impairment I_e from this rating R.

4. Capacity requirements

In this section, we calculate the required capacity to transport a specified number of voice calls over a VP between two VoATM GWs using AAL2. We assume that no VAD scheme is deployed and that all voice sources use either the G.711, the G.726 or the G.729 codec mentioned above. Due to the CBR nature of the voice sources, the CBR ATM service category will be used for the VP, whose PCR equals the corresponding required capacity.

We start by calculating the effective bit rate R_{eff} (in kb/s) that one voice source (to be transported over a VP between two VoATM GWs) requires. The effective bit rate of a voice source is defined as the rate at which this source puts (payload and overhead) bits onto the network. It can be expressed as

$$R_{eff} = \frac{R_{cod}}{\phi}, \qquad (2)$$

with ϕ the filling factor, i.e., the ratio of the number of payload bits and the total amount of bits transported per payload.

Remark that most low bit rate codecs are frame-based, that is, they produce a voice code word of B_F bits every T_F ms. For the codecs of interest, T_F and B_F are tabulated in Table 1. Besides the fact that the packetization delay T_{pack} must be chosen to be a multiple of T_F, the total length of the corresponding voice code words should be byte-aligned too. For the G.726 codec, the latter requirement implies that T_{pack} needs to be a multiple of $2\,T_F = 0.25$ ms.

In AAL2 several voice sources can be multiplexed onto one ATM VC. As shown below, this has a beneficial effect on the filling factor, because on average less overhead bits need to be transported. For the configuration we study (see Figure 1) it is reasonable to assume that the sources are synchronized[2]. We calculate the filling factor for one of the N_{VC} voice sources that are multiplexed onto the same ATM VC. We assume that all of the N_{VC} voice sources use the same codec. Because of the assumption above, the AAL2 payloads of all these sources are simultaneously delivered to the ATM layer at particular moments in time (at a rate of one delivery each T_{pack}). Remark that the difference between two such particular time instants needs to be larger than the time needed to put the sequence of consecutive ATM cells transporting one payload of each voice source onto the link, otherwise the voice queue in the VoATM GW is unstable (see Section 5.1). The last ATM cell of the latter sequence needs to be stuffed with padding bytes. Hence, the filling factor is given by

$$\phi_{AAL2} = \frac{N_{VC}\ P_v}{\left\lceil \frac{N_{VC}\ (P_v + 3)}{47} \right\rceil 53}, \tag{3}$$

with $\lceil x \rceil$ the smallest integer larger than x and P_v the AAL2 payload size (in bytes), which cannot be larger than 64 (or the default value of 45) bytes. Notice that a Service Specific Convergence Sublayer (SSCS) of AAL2 has to be used if AAL2 payloads larger than 64 bytes or packetization delays T_{pack} larger than $(64 \times 8)/R_{cod}$ are allowed. We have

$$P_v = \left\lceil \frac{T_{pack} R_{cod}}{8} \right\rceil \text{ for } T_{pack} \le \frac{64 \times 8}{R_{cod}}, \tag{4}$$

with R_{cod} the codec bit rate (in kb/s) and T_{pack} the packetization delay (in ms). The resulting effective bit rates R_{eff}, which depend on N_{VC}, are shown in Table 2.

The first table (a) contains the results for uncompressed (G.711) voice. The largest possible AAL2 packet payload was chosen equal to 64 bytes, such that eq. (4) implies that the maximal packetization delay equals $(64 \times 8)/R_{cod}$ = 8 ms. The larger the number of voice sources N_{VC} multiplexed onto 1 VC, the smaller the effective bit rate is observed to be (for a fixed packetization interval). For an AAL2 payload of 44 bytes, or, equivalently, $T_{pack} = (44 \times 8)/R_{cod}$ = 5.5 ms, the effective bit rates turn out to be independent of N_{VC}. Indeed, in this case every ATM cell is completely filled with one AAL2 packet, resulting in a filling factor of 44/53.

Similar conclusions can be drawn for the other tables corresponding to the G.726 codec, (b) and the G.729 codec, (c), respectively. However, for the latter codec, a packetization delay value corresponding to 44 bytes payload cannot be attained due to the voice frame granularity of this codec (see Table 1).

T_{pack} (ms)	N_{VC}=1	N_{VC}=4	N_{VC}=8	N_{VC}=16	N_{VC}=32
0.5	848	212	212	159	132.5
1	424	106	106	106	106
1.5	282.67	141.33	106	106	97.17
2	212	106	106	92.75	86.125
2.5	169.6	84.8	84.8	84.8	84.8
3	141.33	106	88.33	88.33	83.92
3.5	121.14	90.86	90.86	83.29	83.29
4	106	79.5	79.5	79.5	79.5
4.5	94.22	94.22	82.44	82.44	79.5
5	84.8	84.8	84.8	79.5	79.5
5.5	77.09	77.09	77.09	77.09	77.09
6	141.33	88.33	79.5	79.5	77.29
6.5	130.46	81.54	81.54	77.46	77.46
7	121.14	90.86	83.29	79.5	77.61
7.5	113.07	84.8	77.73	77.73	75.97
8	106	79.5	79.5	76.19	76.19

(a) G.711@64 kb/s

T_{pack} (ms)	N_{VC}=1	N_{VC}=4	N_{VC}=8	N_{VC}=16	N_{VC}=32
1	424	106	106	79.5	66.25
2	212	53	53	53	53
3	141.33	70.67	53	53	48.58
4	106	53	53	46.38	43.06
5	84.8	42.4	42.4	42.4	42.4
6	70.67	53	44.17	44.17	41.96
7	60.57	45.43	45.43	41.64	41.64
8	53	39.75	39.75	39.75	39.75
9	47.11	47.11	47.11	47.22	47.22
10	42.4	42.4	42.4	39.75	39.75
11	38.55	38.55	38.55	38.55	38.55
12	70.67	44.17	39.75	39.75	38.65
13	65.23	40.77	40.77	38.73	38.73
14	60.57	45.43	41.64	39.75	38.80
15	56.53	42.4	38.87	38.87	37.98
16	53	39.75	39.75	38.09	38.09

(b) G.726@32 kb/s

T_{pack} (ms)	N_{VC}=1	N_{VC}=4	N_{VC}=8	N_{VC}=16	N_{VC}=32
10	42.4	21.2	15.9	13.25	11.93
20	21.2	10.6	10.6	10.6	10.6
30	14.13	10.6	10.6	10.6	10.16
40	10.6	10.6	10.6	9.94	9.94
50	16.96	10.6	10.6	10.07	9.81
60	14.132	10.6	9.72	9.72	9.5

(c) G.729@8 kb/s

Table 2: Effective bit rates of 1 voice source for G.711@64 kb/s, G.726@32 kb/s and G.729@8 kb/s codecs. The results are presented as a function of the number of voice calls N_{VC} multiplexed onto the same VC.

In order to determine the total effective bit rate of N voice sources transported over a VP between 2 VoATM GWs, the effective bit rates given in Table 2 should be multiplied by N. Observe that we can choose how many

[2] Although this assumption will not be satisfied in general, it can be easily imposed by the ingress VoATM GW at the expense of loosing some (maximally T_{pack} ms) voice information at the beginning of the voice calls. One could also impose "synchronization" by collecting AAL2 payloads/packets of the N_{VC} different sources during a specific period of time (TIMER_CU) before they are presented to the ATM layer. Obviously, the latter approach introduces some additional delay.

voice sources are multiplexed onto 1 VC, and we have to use the corresponding values of Table 2. As the effective bit rate of 1 source is a non-increasing function of the number N_{VC} of voice sources multiplexed onto 1 VC, the optimal effective bit rate is attained when the multiplexing gain is maximal, i.e., when all N sources are transported onto the same VC.

In summary, the current best-practice choices for the packetization delay T_{pack} are 5.5 ms, 11 ms and 10 ms for the G.711, G.726 and G.729 codecs, respectively. In case of the G.711 and G.726 codecs, these choices are often referred to as AAL2-lite: one ATM cell is completely filled with one AAL2 packet such that the optimal filling factor of 44/53 is reached, independent of the number of voice sources N_{VC} multiplexed onto the same VC.

5. Performance evaluation

5.1. General

In this section we calculate the M2E delay and distortion introduced in the VoADSL-phone-to-phone and VoADSL-phone-to-VoADSL-phone scenario. As explained in Section 3, these parameters, together with the ELs at both ends, determine the subjective quality, i.e. the rating R, of the call.

The M2E delay consists of several contributions. We describe them and estimate their value next.

The packetization delay T_{pack} was already introduced in the previous section. It is defined as the time needed in the ingress GW to collect the voice information that is transported in one AAL2 packet. The packetization delay is also equal to the inter-packet time between packets of the same source (during periods of active voice). As explained in the previous section (see Table 2), there is a trade-off between efficiency and delay.

The encoding delay T_{enc} is the time needed by the Digital Signal Processor (DSP) in the VGW or in the IAD to possibly digitize the voice signal and then encode it from the G.711 format into the low bit rate codec format. Most low bit rate codecs are frame-based, i.e., they encode a voice interval of duration T_F ms, referred to as the voice frame, in a single encoding operation. Some codecs even need to collect the voice signal of an interval (referred to as the look-ahead of length T_{LA}) after the voice frame to be encoded. The lengths of these intervals are given in Table 1 for the codecs considered here. Analogously, the decoding delay T_{dec} is the time needed by the DSP in the VGW to decode the voice from the low bit rate codec format into the G.711 format and possibly convert it back to an analog signal.

The encoding and decoding delays depend on the DSPs used in the VGWs in the network and in the IAD and the number and priority of the other processes (e.g. echo control) running on the DSPs. For the sum of both, which we call T_{DSP}, a value of 13 ms was measured in the case that the G.711 codec was used. This value should be increased with the codec look ahead T_{LA} in case a low bit rate codec is used. This rather high value is due to other processes running on the DSP such as tone detection. In that case the signal has to be delayed over a certain time, referred to as the look ahead, to enable the tone detection algorithms to distinguish between in-band signaling or fax tones and voice.

In the upstream direction (i.e. from the IAD to the network) the voice payloads compete in the ANT for the ADSL capacity, in the downstream direction (i.e. from the network to the IAD) they compete in the DSLAM for the ADSL capacity. The voice load on the ANT, respectively DSLAM, is given by

$$\rho_{up} = \frac{N\, R_{eff}}{R_{up}} \quad , \tag{5}$$

respectively,

$$\rho_{down} = \frac{N\, R_{eff}}{R_{down}} \quad , \tag{6}$$

where R_{up} is the upstream capacity of the ADSL link, R_{down} is the downstream capacity of the ADSL link and R_{eff} is the effective bit rate of one voice source (see Table 2) and N is the number of active voice sources. The upstream and downstream capacity are determined during the configuration of the ADSL link. The upstream capacity is much smaller than the downstream capacity. An ADSL link typically has link rates of 512 kb/s in the upstream and 2 Mb/s in the downstream direction. The loads ρ_{up} and ρ_{down} are upper bounded by 1 in order to have finite queuing delays in the ANT and DSLAM. This restricts the number N of VoADSL phones that can be supported per ADSL link.

During the set-up of the CBR connection between the VoATM GW in the IAD and the VoATM GW in the network, capacity is reserved in each of the traversed ATM switches such that an end-to-end minimal delay and an end-to-end maximal jitter is guaranteed [7]. We assume that in this set-up procedure the end-to-end minimal delay and end-to-end maximal jitter are specified to values just larger than the ones introduced in the bottleneck node (i.e. the ANT, respectively the DSLAM, for the upstream, respectively downstream, direction). In that way, if the connection set-up succeeds, the queuing delay in all but the bottleneck nodes is negligible. The connection set-up fails if one of the traversed ATM switches has not enough capacity left to guarantee the required quality of service. We assume that the ATM network is of high capacity (155 Mb/s or more), that it carries voice and data with data constituting the bulk part and that it gives priority to voice over data. In such a network the connection set-up with the parameters specified above is not likely to fail.

Hence, the ANT and DSLAM are the main points where queuing delay is incurred in upstream and downstream direction respectively. We assume that ATM cells transporting voice get non-pre-emptive Head-Of-Line (HOL) priority over ATM cells carrying data in the ANT and DSLAM. This implies that except for the data cell in service, voice cells experience only competition from other voice cells. Since VAD is not studied in this paper all competing voice sources are CBR streams with the same characteristics: each VoADSL phone produces a voice payload every T_{pack} ms. If a multiplexing point is loaded up to 1 Erlang with CBR streams (with identical characteristics), the maximum time a certain payload has to wait before being served, equals T_{pack} ms. Hence, we take as the upper bound for the maximum queuing delay

$$T_{que} = T_{pack} \quad . \tag{7}$$

If the multiplexing point (i.e., the ANT or DSLAM) is not loaded up to 1 Erlang, but up to a load ρ, the maximum queuing can be relaxed to

$$T_{que} = \rho T_{pack} \quad . \tag{8}$$

The minimal queuing delay occurring for ATM cells that do not need to queue, is of course 0 ms.

Over the ADSL link there is the bit pipe delay T_{ADSL}. This is the delay incurred when transporting a bit over the ADSL link. This delay depends on the interleaving depth that is used. The larger the interleaving depth is chosen, the larger the bit pipe delay, but the more robust the transport is against impulsive noise. Several interleaving modes are possible in ADSL. The ADSL interleaved channel typically introduces a ADSL bit pipe delay of 20 ms, the ADSL fast channel typically 2 ms and ADSL-lite 3 ms. The influence of impulsive noise on the cell loss, and hence, the voice quality, is beyond the scope of this paper. Hence, in this report we do not quantitatively discuss the trade-offs involved in the choice of the interleaving depth.

The dejittering delay is needed in the egress GW to compensate for the difference in queuing delay of voice payloads. From the previous discussion we know the maximum queuing delay T_{que} and the minimal queuing delay (0 ms). If we delay the first payload of a certain VoADSL call over the maximum queuing delay and then read the dejittering buffer at constant rate (one payload every T_{pack} ms), it is guaranteed that every payload arrives in time in the dejittering buffer. Hence, if we take the dejittering delay equal to the maximum queuing delay

$$T_{jit} = T_{pack} \quad , \tag{9}$$

no payloads are lost in the dejittering buffer.

Hence, the worst case delay introduced in the VoADSL access network sums up to

$$T_{DSP} + 3T_{pack} + T_{ADSL} \quad . \tag{10}$$

This delay is worst case in the sense that

- the bottleneck (upstream ADSL) link is fully loaded with CBR voice traffic,
- the first arriving payload is the slowest possible, and
- the dejittering delay is chosen equal to its maximum value so that no payloads are lost in the dejittering buffer.

There are several other constant delays T_{oth} that can be important. Of them propagation delay is the most important if the call is long distance. We take 5 µs per km for the propagation speed of light in optical fibers and multiply the distance between the 2 calling parties with an engineering factor 1.5 to account for the fact that the cables do not follow a straight line. The echo control also needs some time. We include the time needed for echo control in T_{oth}.

Next, we consider distortion. We assume that there is no (or a negligible amount of) cell loss in the ATM backbone. Because we use the interleaved mode, there is no cell loss on the ADSL link due to impulsive noise. Studying the trade-off involved when choosing a smaller interleaving depth at the expense of possibly more packet loss on the ADSL link, is beyond the scope of this paper. There are no payloads lost in the dejittering buffer, because we have chosen both the dejittering delay and the buffer capacity large enough. Hence, the distortion is solely introduced by the compression of the voice signal.

5.2. Examples

In this section we apply the theory of the previous section to the VoADSL-phone-to-phone and the VoADSL-phone-to-VoADSL-phone scenario. In all scenarios all calls traversing an IAD use the same codec, packetization delay and dejittering delay. We assume perfect echo control. The VoADSL-phone-to-VoADSL-phone scenarios are chosen to be symmetric.

We assume that the quality of the copper pair is such that at configuration time the upstream and downstream capacity are configured to R_{up} = 512 kb/s and R_{up} = 2 Mb/s respectively. The upstream capacity R_{up} impacts the number of voice sources that can be supported. As ADSL bit pipe delay we take the (worst-case) delay of the interleaved mode, i.e. T_{ADSL} = 20 ms. We take for T_{oth} 5 ms for a local call and 50 ms for a long distance call (covering about 6000 km, e.g. a call from the US to Europe).

We calculate the maximum number of VoADSL phones that can be supported for each scenario. This number depends on the upstream capacity, the type of codec, the packetization interval and the number N_{VC} of calls multiplexed onto one VC. We also calculate the quality of the call. For the VoADSL-phone-to-phone scenario, the delay in the upstream direction (from VoADSL phone to PSTN) is always higher than the delay in the downstream direction. As such, we only calculate the quality in the upstream direction. For the VoADSL-phone-to-VoADSL-phone scenario, the quality is equal in both directions.

First, we consider the G.711 codec. For this codec we conclude from Table 2 that for a packetization interval T_{pack} = 5.5 ms, an effective bit rate of 77.09 kb/s results. A slightly smaller effective bit rate can be reached at the expense of higher packetization delays if enough sources can be multiplexed onto one VC. With a packetization interval of T_{pack} = 5.5 ms, 6 VoADSL phones can be supported by the IAD.

With this value for T_{pack}, Table 3 (a) gives the M2E delays and R ratings for different scenarios. It can be seen that in every case the 25 ms, which is the bound above which echo control is needed, is exceeded. Therefore, we assumed echo control from the beginning[3]. For all calls considered (local and long distance; VoADSL-phone-to-phone and VoADSL-phone-to-VoADSL-phone), the mouth-to-ear delay remains below 150 ms and, as such, the intrinsic quality (R=94.3) of the G.711 codec is reached.

Next, we consider the G.726 codec. From Table 2 we conclude that for a packetization delay of T_{pack} = 11 ms we have an effective bit rate of 38.55 kb/s. Increasing the packetization interval again leads to a slightly smaller effective bit rate if more calls are multiplexed onto one VC. With a packetization interval of T_{pack} = 11 ms, 13 VoADSL phones can be supported by the IAD.

	VoDSL-phone-to-phone		VoDSL-phone-to-VoDSL-phone	
	M2E delay (ms)	rating R	M2E delay (ms)	rating R
local call	54.5	94.3	104	94.3
long distance call	99.5	94.3	149	94.3

(a) G.711@64 kb/s

	VoDSL-phone-to-phone		VoDSL-phone-to-VoDSL-phone	
	M2E delay (ms)	rating R	M2E delay (ms)	rating R
local call	71	87.3	137	80.3
long distance call	116	87.3	182	78.8

(b) G.726@32 kb/s

	VoDSL-phone-to-phone		VoDSL-phone-to-VoDSL-phone	
	M2E delay (ms)	rating R	M2E delay (ms)	rating R
local call	73	84.3	141	74.3
long distance call	118	84.3	186	72.5

(c) G.729@8 kb/s

Table 3 : The M2E delay and R rating for different VoADSL-phone-to-phone and VoADSL-phone-to-VoADSL-phone scenarios with perfect echo control.

For this value for T_{pack}, all but one of the resulting M2E delays given in Table 3 (b) remain below the interactivity bound of 150 ms. As such, (local and long distance) VoADSL-phone-to-phone calls reach a quality of R = 94.3 - 7 = 87.3. In the VoADSL-phone-to-VoADSL-phone scenario, transcoding (from the G.726 via G.711 to G.726) is performed, leading to a quality of R = 94.3 - 7 - 7 = 80.3 for the local scenario. For long distance calls the quality suffers a small extra interactivity impairment since there the M2E delay exceeds the 150 ms. This leads to $R = 80.3 – 1.5 = 78.8$.

Finally, we consider the low bit rate codec G.729. From Table 2 we conclude that at the expense of a large packetization delay the effective bit rate can be quite low. However, we choose the minimal packetization delay possible, i.e. T_{pack} = 10 ms, in order to keep the M2E delay small enough, because for this codec the encoding and decoding delay also contribute considerably to the M2E delay. With this packetization delay it is clear from Table 2 that in principle 40 VoADSL phones can be supported, e.g., 32 multiplexed onto one VC and 8 multiplexed onto another, leading to a total effective bit rate of ($32 \times 11.93 + 8 \times 15.9 =$) 508.96 kb/s, a value just below R_{up} = 512 kb/s.

With this value for T_{pack}, Table 3 (c) gives the M2E delays for different scenarios. In the VoADSL-phone-to-phone scenario the interactivity bound of 150 ms is met, leading to a rating R of 94.3 - 10 = 84.3. Again remark that in the VoADSL-phone-to-VoADSL-phone scenario transcoding is performed, leading to an intrinsic quality of 74.3 which is reached for local calls. However, as the M2E delay is larger than 150 ms for long distance calls, this intrinsic quality is not reached in this case. More precisely, the ratings R of long distance VoADSL-phone-to-VoADSL-phone calls drops to 72.5.

One could think of a future scenario in which the voice cell stream bypasses the PSTN and remains in the ATM network for VoADSL-phone-to-VoADSL-phone calls. Such a configuration would have a beneficial effect on the performance of such voice calls as the transcoding step could be avoided. In particular, the R ratings derived above for VoADSL-phone-to-VoADSL-phone calls need to be increased with the distortion impairment I_e of the codec used in the IADs, if no PSTN is traversed. In addition, also the decrease in delay (no dejittering, decoding, encoding and packetization in the VoATM GWs between the ATM network and the PSTN) may lead to an additional increase in quality. In particular, the latter quality increase is observed when bypassing the PSTN reduces the M2E delay from a value above 150 ms to a value below 150 ms (see Figure 3 and Table 3). This is e.g. the case for the VoDSL-phone-toVoDSL-phone scenario with the G.729 codec. In this case bypassing the PSTN would decrease the delay upper bound with 28 ms (18 ms less DSP delay and 10 ms by avoiding one packetization stage). This would lead to a M2E delay of 113 ms and the intrinsic quality R = 84.3 for local calls and a M2E delay of 158 ms and R = 84.0 for long distance calls.

Exactly the same exercise can be done for non-perfect echo control. In this case the intrinsic quality can never be reached since every delay causes impairment as can be seen from Figure 3. Only the curve corresponding to perfect echo control has a flat part (between 0 and 150 ms) while all other curves immediately have a negative slope. The extra impairment caused by a certain delay can be read from the curves and should be substracted from the rating R. The results for an echo loss of 51 dB, i.e. an echo controller that is compliant with ITU-T Rec. G.168 [8], are given in below.

[3] If the ADSL fast mode instead of the interleaved mode is used, (20 – 2 =) 18 ms can be gained for each scenario. In that case the local VoADSL-phone-to-phone scenario does not strictly require echo control.

	VoDSL-phone-to-phone		VoDSL-phone-to-VoDSL-phone	
	M2E delay (ms)	rating R	M2E delay (ms)	rating R
local call	54.5	92.2	104	90.5
long distance call	99.5	90.6	149	88.8

(a) G.711@64 kb/s

	VoDSL-phone-to-phone		VoDSL-phone-to-VoDSL-phone	
	M2E delay (ms)	rating R	M2E delay (ms)	rating R
local call	71	84.6	137	75.3
long distance call	116	83.1	182	72.4

(b) G.726@32 kb/s

	VoDSL-phone-to-phone		VoDSL-phone-to-VoDSL-phone	
	M2E delay (ms)	rating R	M2E delay (ms)	rating R
local call	73	81.5	141	69.2
long distance call	118	80.0	186	66.0

(c) G.729@8 kb/s

Table 4: The M2E delay and R rating for different VoADSL-phone-to-phone and VoADSL-phone-to-VoADSL-phone scenarios for an *EL* of 51 dB.

6. Conclusions

In this paper the performance of VoADSL calls (calls switched over an ADSL access network to the PSTN) was studied.

First, the quality of a (VoADSL) call was studied with the E-model. With respect to quality, ADSL access introduces more delay and distortion than traditional PSTN access. As we assume no (or a negligible amount of) cell loss to occur in the ATM network, distortion only depends on the codec used.

Next, the effective bit rate (i.e., the codec bit rate plus the overhead bit rate) produced by each standard codec was calculated for various values of the packetization interval. An important parameter to be considered here is the number of voice sources to be multiplexed onto 1 ATM connection.

Finally, the mouth-to-ear delay and distortion introduced for various reference scenarios were calculated and compared with the bounds mentioned above. We also calculated the number of VoADSL phones that could be supported by one IAD if a certain standard codec was used.

The main conclusions were the following.

- (Perfect) Echo control is required in the gateway in the IAD and in the gateway in the network. The former echo controller controls the echo generated in the SLIC; the latter compensates the echo introduced in the 4-to-2-wire hybrid in the PSTN.
- If perfect echo control is deployed, all scenarios studied (local and long distance; VoADSL-phone-to-phone and VoADSL-phone-to-VoADSL-phone; G.711, G.726 and G.729 codecs) attain traditional quality, i.e., a rating R considerably larger than 70.

Acknowledgement

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] D. De Vleeschauwer, J. Janssen, G. H. Petit, F. Poppe, “Quality Bounds for Packetized Voice Transport”, Alcatel Telecom Review, First quarter 2000, pp. 19-23, 2000.

[2] "The E-model, a computational model for use in transmission planning", *ITU-T Recommendation G.107*, 1998.

[3] "Provisional planning values for the equipment impairment factor I_e", *Appendix to ITU-T Recommendation G.113*, 1999.

[4] "One-way transmission time", *ITU-T Recommendation G.114*, 1996.

[5] "Control of talker echo", *ITU-T Recommendation G.131*, 1996.

[6] "Speech processing, Transmission and Quality aspects (STQ): Overall transmission plan aspects for telephony in a private network", *ETSI EG 201 050*, 1999.

[7] "ATM User-Network Interface (UNI) Signalling Specification", *af-sig-0061.000 (Version 4.0)*, 1996.

[8] "Digital network echo cancellers", *ITU-T Recommendation G.168*, 1997.