Electronic Communications Committee (ECC)
within the European Conference of Postal and Telecommunications Administrations (CEPT)

VOICE QUALITY OVER IP BASED NETWORKS

Gothenburg, July 2004

EXECUTIVE SUMMARY

This Report is intended as a tutorial and status report on voice quality on IP networks. This issue is receiving considerable attention in both the telco and Internet communities, and the quality achievable on the Internet will be an important factor in determining how quickly and to what extent voice communications over the Internet will grow.

The introduction of IP based communications changes quite significantly the technical issues that affect quality. With circuit switched communications the quality level achieved was essentially static and predictable from a knowledge of the networks plans. The unpredictability of traffic demand affected the probability of being able to establish a connection but not the quality of the connection. With packet based communications the situation changes and the quality varies continually with traffic loading.

In general terms, two different approaches are taken to the fundamental effects of packet technology. The telcos in their developments for fixed next generation networks in ETSI have been taking the approach of trying to simulate the effects of circuit switching and maintain voice quality by blocking attempted call set-up when the network resources are congested. The developments that are based on the Internet allow quality to vary and do not block calls, giving the user the opportunity to make their own decisions on whether or not to continue with a call.

This report summarises the current work in this area and so provides an understanding of developments in this important but complex area and also gives background information for policy discussions about the future of voice communications.

The report covers the following topics:

- Impairments in packet networks
- End-to-end quality classes
- Speech quality guarantees
- Quality of service signalling
- Modern designs for codecs and network equalisation.

INDEX TABLE

1 INTRODUCTION

This Report is intended as a tutorial and status report on voice quality on IP networks. This issue is receiving considerable attention in both the telco and Internet communities, and the quality achievable on the Internet will be an important factor in determining how quickly and to what extent voice communications over the Internet will grow. The report considers the quality achievable over IP but does not consider the interactions of the codecs used for IP and low bit rate codecs used in circuit switched networks when telephone calls are carried by a combination of IP and circuit switched technologies when the combination of different codecs can interact badly.

The introduction of IP based communications changes quite significantly the technical issues that affect quality. With circuit switched communications the quality level achieved was essentially static and predictable from a knowledge of the networks plans. The unpredictability of traffic demand affected the probability of being able to establish a connection but not the quality of the connection. With packet based communications the situation changes and the quality varies continually with traffic loading.

Until recently the view held by most people was that the quality of voice communications achievable over the Internet was very poor. This impression seems to have been caused by the use in early software packages of codecs that were designed for circuit switched applications with a low rate of bit errors and were especially sensitive to packet loss. Demonstrations with codecs designed to be tolerant to packet loss rather than bit errors and to adapt to the current state of the Internet have shown that the Internet is capable of providing good quality wideband communications. It is therefore wrong to assume that the development of packet technology and the use of the Internet will lead to a degradation in quality as in many cases it will lead to a growth in wideband communications that will be perceived as having better quality than the traditional PSTN.

In general the situation for voice quality had been turned inside out since the days of analogue networks. With analogue networks, the main sources of impairment were in the core networks and hence there was great emphasis on network planning and on the apportionment of impairments such as loss and delay. With new packet based networks including the Internet, the sources of impairment are mainly at the network edges where access capacity limitations introduce delays and there is less control by operators of the quality achievable by terminals. The performance of the core of packet networks is normally not a major factor. This fundamental change means that the approaches developed in the analogue world and carried over in to the digital world such as apportionment and guarantees cannot easily be applied or do not add value in the packet based world.

In general terms, two different approaches are taken to the fundamental effects of packet technology. The telcos in their developments for fixed next generation networks in ETSI are taking the approach of trying to simulate the effects of circuit switching and maintain voice quality by blocking attempted call set-up when the network resources are congested. The developments that are based on the Internet allow quality to vary and do not block calls, giving the user the opportunity to make their own decisions on whether or not to continue with a call.

This report summarises the current work in this area and so provides an understanding of developments in this important but complex area and also gives background information for policy discussions about the future of voice communications.

This report draws heavily on work in the Speech Transmission and Quality technical body of ETSI, and some parts of the text are copied from its work.

2 BASICS

The end-to-end speech quality of the connection is affected by various transmission impairments that depend on the transmission and switching technology used and affect the end-end performance in different ways.

The overall objective of QoS work is to achieve an end-end quality that is satisfactory for the user. This is not at all a simple issue because:

- User's requirements and views on what they consider to be satisfactory differ especially in terms of their past experience and what they are paying for their service.
- No one party has responsibility for the complete end-end connection (mouth to ear) nor control over the whole connection, since the terminals are normally owned by the user without control being exercised by the network operator. Furthermore whilst users may choose their own network operator, ie their access operator and

perhaps the nearest transit operator, calls that they make will terminate on networks chosen by the called user and the caller has no control over the termination arrangements for the call.

The historical approach for circuit switched networks was to take a single common approach to quality based on limited bandwidth 3.1kHz handset telephony with the aim of achieving a high degree of probability that the end to end quality would be adequate. This involved networks exercising control over the terminals of users through supplying these terminals themselves or controlling their quality through type approval. Restrictive rules based on apportionment were applied to network design to achieve a high probability that impairments that summed, such as loss in analogue networks or delay in both analogue and digital networks, would remain within an acceptable level in almost all practical configurations.

Circuit switched technology was conducive to this approach because its performance was constant and independent of the network's traffic level, provided a connection could be obtained. The nature of the technology also led to a single bit rate of 64kbit/s for the switches, and this meant that deviation from this single common approach was largely impracticable. Both loss and delay in analogue networks depended on distance and the greatest distances occurred in the public networks, giving the telcos control over the major sources of impairments even when terminals were liberalised. Digitalisation removed the problems of loss, leaving delay as the main impairment that was subject to network planning. Delay contributes to echo, but the reducing cost of echo cancellers has helped to keep echo to tolerable levels. Delay is also an impairment in its own right, and delay has generally increased with the increasing use of digital technology but in most cases delay has stayed within acceptable levels.

The introduction of packet technology has led to a wave of new interest in quality issues. The two main changes that packet technology introduces are:

- Replacement of the single rate of 64kbit/s for switches with a spectrum of rates increasing the scope for new codec designs and a wide range of levels of quality
- Replacement of the constant and repeatable level of quality for an active connection with a variable level of quality that depends on network traffic levels (unless the IP network is made to simulate circuit switching).

The dependence of quality on traffic levels is the most difficult issue and the area where there is least practical knowledge and experience. The traditional approach of apportionment can be used where the network topology is known and the impairments are constant and predictable. But in packet networks neither criterion is met.
In practice there are commonly two additional issues:

- The quality problems caused by traffic congestion may be greatest in the user's own networks and so are outside the control of the public network operators
- Packet based voice terminals are much more likely to be used from PCs whose acoustic characteristics may be intrinsically poor or suffer from being incorrectly configured.

These changes have raised many issues that have been studied by different groups. Unfortunately whilst there has been some coordination between the various studies not enough attention has been given to the overall system problem of quality in packet networks and different groups have worked with different assumptions.

The following is a list of the main issues that arise:

- Should services attempt to specify or guarantee a particular level of quality?
- How can different levels of quality be specified in a way that will be meaningful to users?
- How can interconnected networks be managed to achieve a particular level of quality?
- What coding techniques can be used to reduce the effect of network impairments on end-end quality
- What speech or audio processing techniques can improve quality for real-time speech by giving it priority over traffic that is insensitive to the impairments that affect speech?
- What is the relationship between network techniques to improve quality and the quality achieved with different levels of traffic?
- How can the quality of PC based terminals be maximised and measured?
- How effective and worthwhile are network-based techniques if in many cases the main problems are in terminals and access systems?

3 IMPAIRMENTS IN PACKET NETWORKS

At each node in a packet network, packets are held in a queue awaiting transmission. Congestion causes longer queues than normal and so increases the transmission delay for the packets. Nodes also have limited queuing capacity and queues may overflow resulting in packet loss.

The terminals at the ends of speech circuits provide jitter buffering to smooth the play-out of packets. The jitter buffer (also called "de-jitter buffer") is a store of packets awaiting processing. The store may have a limited size determined by the hardware or device configuration. The packets arrive at varying times and are extracted at regular intervals by the play-out algorithm. The effect of the jitter buffer is to convert variable delay into fixed delay. The fixed delay depends on how full the jitter buffer is. If the variable delay increases so that the buffer empties (called a jitter buffer under-run) then a packet is missed. When delays reduce, the jitter buffers will fill up and if the maximum capacity is exceeded then packets will have to be discarded, and this is called buffer over-run. The play-out algorithm may be quite sophisticated and adjust the fill of the jitter buffer so that it is filled only to the extent necessary so that the probability of an under-run is low, so that the additional delay is minimised. In order to make these adjustments, packets may have to be skipped or duplicated introducing some additional distortion. Some highly intelligent algorithms may observe the values in the packets and make adjustments only when there are gaps in the speech.

Thus end-to-end transmission delay affects the quality of interactive real time communications:

- Directly in terms of the average delay plus the settings of the play-out algorithms in the jitter buffers to account for delay variations (jitter)
- Indirectly through packet loss where:
  - Queue capacity in individual routers is exceeded and so packets are discarded (over-run)
  - Variable delay increases causing the jitter buffer to empty and resulting in missing packets (under-run)
  - Variable delay reduces causing the jitter buffer to overflow and resulting in lost packets (over-run).

The different effects are shown in the following diagram. The diagram distinguishes the network and terminal factors. The jitter buffer operation can be set to give lower delay and higher packet loss probability, or higher delay and lower packet loss probability. Codecs introduce delay that depends on the algorithm used and the processor power. There are also various different methods that execute packet loss concealment in conjunction with jitter buffering rather then keeping them as two separate processes.

Unlike digital circuit switched networks, packet networks are not normally synchronised to a common reference. Thus the clocks at the sending and receiving ends are likely to be running at slightly different rates causing the problem called "clock skew". Clock skew leads to overflow or underflow of playout buffers at the input to the decoder so that packets are lost in one direction and gaps occur in packet arrivals in the other direction. Both affect the decoding and contribute to distortion.

Since the connection is "4-Wire" in terms of echo, echo can arise only at the far end across the ear-mouth link, where the linkage is called terminal coupling. If the terminal coupling loss is high as with a conventional handset or high quality headset, then no echo control or cancelling may be needed. If the terminal coupling loss is low, and with a handsfree telephone it will be very low, then echo cancelling is needed and this is normally applied at the terminal itself. For example some PC operating systems offer loud speaking settings that include echo control. The echo control procedures may increase the distortion of the speech or clip the speech. Figure 1 shows the different impairments.

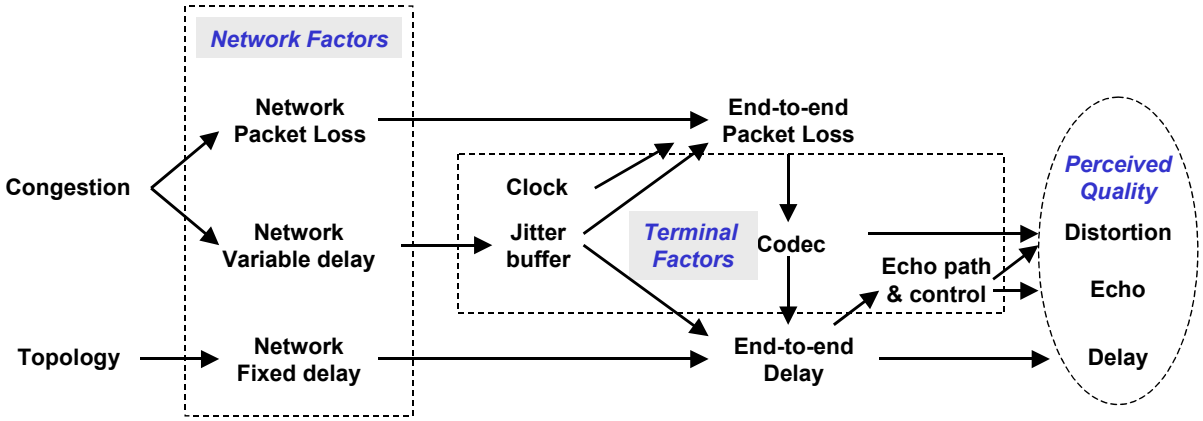

Figure 1: Impairment effects for packet based voice communications

The figure distinguishes between the effects that occur in the network and the mechanisms in the terminals that are affected and that can be used to correct for the effects in the network. The network fixed delay is sometimes called the transmission or propagation delay.

Many calls traverse multiple networks including end networks such as LANs, access networks and transit networks. End or customer networks are likely either to be wireline or wireless LANs. A wide variety of loadings is likely. Wireless LANs are especially likely to be overloaded at times and can in practice be the dominant source of quality problems for Internet access. Organisations where there is correlation between the activities of different LAN users are especially likely to experience high peaks of demand from simultaneous similar actions by users, eg all participants in a meeting downloading documents at the same time.

Figure 2 provides a summary of the different impairments and their sources. Those impairments that can typically degrade quality significantly are shown in bold:

	Circuit switched	Packet switched
Terminal equipment	Echo from terminal coupling Loudness rating **Delay and distortion from coding in mobile terminals**	**Echo from terminal coupling** **Distortion from packet loss** **Distortion from low bit rate coding** **Delay from jitter buffer** **Loudness rating** Delay from coding
End network eg LAN	None	**Congestion causing variable delay and packet loss**
Access network	None	**Congestion causing variable delay and packet loss**
Transit networks	**Delay** Distortion from transcoding	**Delay** and variable delay

Figure 2: Comparison of impairments

4 END-TO-END QOS CLASSES AT THE APPLICATION LEVEL

The main work on end-to-end QoS classes has been carried out in the ETSI TIPHON project and is described in ETSI TS 101 329-2 V.2.1.1. Three classes of end-to-end speech QoS are defined for TIPHON systems: WIDEBAND, NARROWBAND and BEST EFFORT. The TIPHON speech QoS classes WIDEBAND and NARROWBAND will provide performance guarantees for 95 % of all connections (ie a statistical guarantee). The BEST EFFORT class provides no speech performance guarantees. The classes are defined from mouth-to-ear and therefore include the transit and access networks, the end networks and the TIPHON terminal characteristics. Each of the classes defined is specified by three performance metrics: Overall Transmission Quality Rating (R), Listener Speech Quality (One-way non-interactive end-to-end Speech Quality) and End-to-end (mean one-way) Delay.

	Wideband	Narrowband			Best Effort
		High	Medium	Acceptable	
Description	IP telephony service using wideband codecs (codecs encoding analogue signals with bandwidth in excess of 3,1 kHz) and QoS-engineered IP networks	IP telephony service provided via QoS–engineered IP networks			voice communication service operated over non QoS-engineered IP networks such as the public Internet
	supposed to be better than the PSTN	supposed to be equivalent to recent ISDN services	supposed to be equivalent to recent wireless mobile telephony services in good radio conditions (EFR codec)	supposed to be equivalent to common wireless mobile telephony services (FR codec)	supposed to provide usable communications service but will not provide guarantees of performance (with periods of significantly impaired speech quality, and large end-to-end delays which are likely to impact the overall conversational interactivity)
Overall Transmission Quality Rating (R)	**not applicable**	**> 80**	**> 70**	**> 50**	**> 50** Target value !
Listener Speech Quality	**Better than G.711**	**Equivalent or better than ITU-T Recommendation G.726 at 32 kbit/s**	**Equivalent or better than GSM-FR**	**not defined**	**not defined**
End-to-end Delay	**< 100 ms**	**< 100 ms**	**< 150 ms**	**< 400 ms**	**< 100 ms** Target value !

The TIPHON classes are unusual in that they involve a combination of two independent parameters (delay and listener speech quality) and the overall transmission quality rating, which includes the delay and listener speech quality parameters as well as other parameters. Thus the three parameters are not orthogonal as might be expected.

If a designer has control over the whole end-to-end connection then he has some control over the delay by choosing the codec algorithm including the jitter buffer play-out, the routings and network loadings, and over listener speech quality (distortion) by choosing the codec, bit rate and network loadings. He can thus adapt his design to achieve a particular overall quality level. In many cases, however, the designer controls only the bearer service that works between the network termination points and TIPHON has not considered bearer services.

5 NETWORK DESIGN

5.1 Introduction

The approach to network design depends on the objective that the designer is trying to achieve. The common objective is to improve the quality for voice transmission, but telcos in particular are interested in being able to guarantee that a specified level of performance is achievable. Where they can guarantee a level of quality, some may be interested in offering different levels of quality at different tariffs.

In circuit switched networks, the transmission quality was normally constant and independent of traffic once a call was established. Traffic congestion meant that the call set-up might be blocked. Packet technology offers the possibility to trade quality against capacity and so, when there is a high level of congestion, there is the option to have communications of reduced quality or to wait until the congestion eases and then communicate at better quality. There is little knowledge of how the various choices may be presented to users and how the users will react to these options, as the choices have not been available with circuit switched technology. Intuitively the response is likely to depend on the circumstances of the caller and the party called. Callers with an urgent need to communicate will accept any quality that is intelligible whereas users whose communications are not urgent may prefer to wait for better quality especially if the main purpose is the pleasure of talking rather than a more functional requirement.

Network design, measurement and control issues can be handled at two levels:

- The service or application level
- The transport or network level.

The work in ETSI TIPHON has focussed on end-to-end QoS signalling as part of call set-up at the application level, whereas work in IETF is almost exclusively at the network level.

It is important to distinguish between the telco model and the Internet model when considering network design (see the concepts in the ECC Report on Next generation network developments and their implications for the new regulatory regime).

Techniques that apply to the application level or that make a network "application conscious" are relevant only where the telco model is being followed. Such techniques are likely to make interconnection more complex and this is a cost that needs to be taken into account in the overall design.

A further consideration at the service or application level is the substantial influence of the terminal design and especially the extent to which codecs can tolerate variable delay and packet loss in the networks. This means that the benefits of schemes that relate only to public networks and do not include the end networks and the terminals may be quite limited.

5.2 Guarantees

The word "guarantee" is used with two quite different meanings:

- an "absolute" guarantee
- a "statistical" guarantee.

In practice since the traffic demand is statistical and in effect unbounded, there can be no absolute guarantee of both quality and availability. Therefore an absolute guarantee means in practice that access may have to be denied to others and even also to those using the guaranteed service.

A statistical guarantee means that the network will be designed to achieve a certain quality level for x% of the time assuming a given level of demand.

Whereas with circuit switched technology call quality remains constant, with packet networks it may vary during the course of a call. With an absolute guarantee, the quality should not fall below the guaranteed level since the network loading will be controlled by denying access to those calls that would cause congestion.

The provision of an absolute quality guarantee in packet networks requires QoS signalling at the application level.

5.3 Quality of service signalling

There are two approaches to call-related QoS signalling:

- negotiating QoS during call set-up with the possibility of failing the call set-up attempt if the QoS is not adequate
- passing forward QoS information to guide routing or other decisions in networks that handle the call subsequently.

This report only considers the first of these approaches, since this was the approach that has been considered in the ETSI TIPHON project. This approach involves an end-to-end signalling interaction. The alternative approach of passing forward information, such as a call class, to guide routing decisions or decisions about the use of transport level techniques would not necessarily involve a signalling interaction or negotiation but only the passing forward of call class information. This information might be sent in the application layer signalling but could also be given by marker bits at the media packet level as in DiffServ.

5.3.1 Basic concept for call related signalling negotiation

The main work on QoS signalling negotiation has taken place within the ETSI TIPHON project. The following gives a brief description of the main concepts. The term "service provider" refers to the functions in the application plane, and the term "network operator" refers to the functions in the transport plane of the TIPHON architecture and the meanings of these terms may differ slightly from those normally used by regulators. Figure 3 shows an example of how the signalling would work:

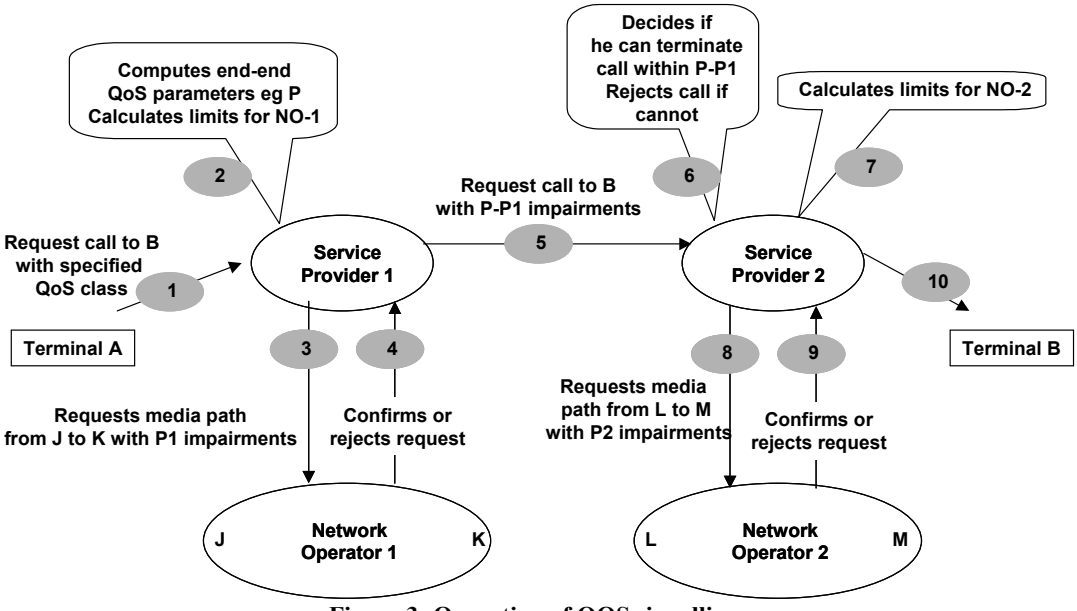

Figure 3: Operation of QOS signalling

The caller selects the desired QoS class and the main service provider transforms this class into transport impairments values. The service provider then starts a process of negotiation with network operators and other interconnected service providers.

The call/call QoS signalling between service providers is not a dialogue. In the forward going SETUP message, the sending service provider specifies the values of the impairment parameters within which it wishes the receiving service provider to complete the call (ie, the difference between the end-end impairment parameter limit (eg 200 ms) and the accumulated value for the impairment parameter).

The receiving service provider then decides whether or not to continue the set-up procedure and route the call onwards, or whether to fail the call on the basis of inability to meet the QoS objectives. This decision may use stored information on the recent performance, eg which destinations can be reached with what impairments.

If the call is failed, then the RELEASE message passes back to the sending service provider. The sending service provider may decide to pass the release back further or to recommence set-up with another service provider.

Thus the QoS signalling does not involve any additional message flows, just the inclusion of specific information in the messages that have already been defined for call control.

5.3.2 Discussion of the usefulness of call related QoS signalling negotiation

The precise role of QoS signalling is still under discussion and there are slightly different ways in which it could be used. The ETSI TIPHON project has now ended and the work is big continued as TISPAN with the objective of adopting the specifications developed by 3GPP for the IP Multimedia Service, whose approach is rather different with more emphasis on the treatment of different classes of traffic at the transport level and less on end-to-end performance and application level signalling.

The main function of QoS signalling in TIHON is to implement the absolute quality guarantee and to fail call set-up where the guaranteed quality cannot be met (an absolute quality guarantee would be a service option offered to the user). It also could be used to enable the service provider to adjust the charging where the guaranteed quality is not met. Different views are held as to whether this is helpful to the caller. One school of thought says that it is always better to connect the call and let the user assess whether he wants to continue with the call since this policy gives the user the maximum choice and lets him take into account the urgency of the communications.

There has also been discussion in TIPHON on the effectiveness of QoS signalling in general purpose networks where delay sensitive real-time speech can be given priority in the router queues. Assume first that less than say 20% of the traffic is delay sensitive. This means that prioritisation should have a substantial effect in reducing variable delay. The issue is then whether the potential QoS problem is caused by:

- the fixed delay that is a consequence of the network design, topology and the physical distances involved, or
- the variable delay caused by the queues in the routers.

If the main problem is the fixed delay, then the networks may be fundamentally unsuitable for the traffic being considered. This could be determined a priori from knowledge of the design and does not require call-related signalling, although call related signalling could help with route selection.

If the main problem is the variable delay, then prioritisation may solve the problem in which case QoS signalling does not add much real value as almost all calls will then be satisfactory. If the problem cannot be solved, then QoS signalling may help by screening the calls and denying access when the networks are too congested, but fundamentally the networks concerned cannot be relied on as a suitable means of carrying the traffic.

These arguments are summarised in the flow chart in figure 4.

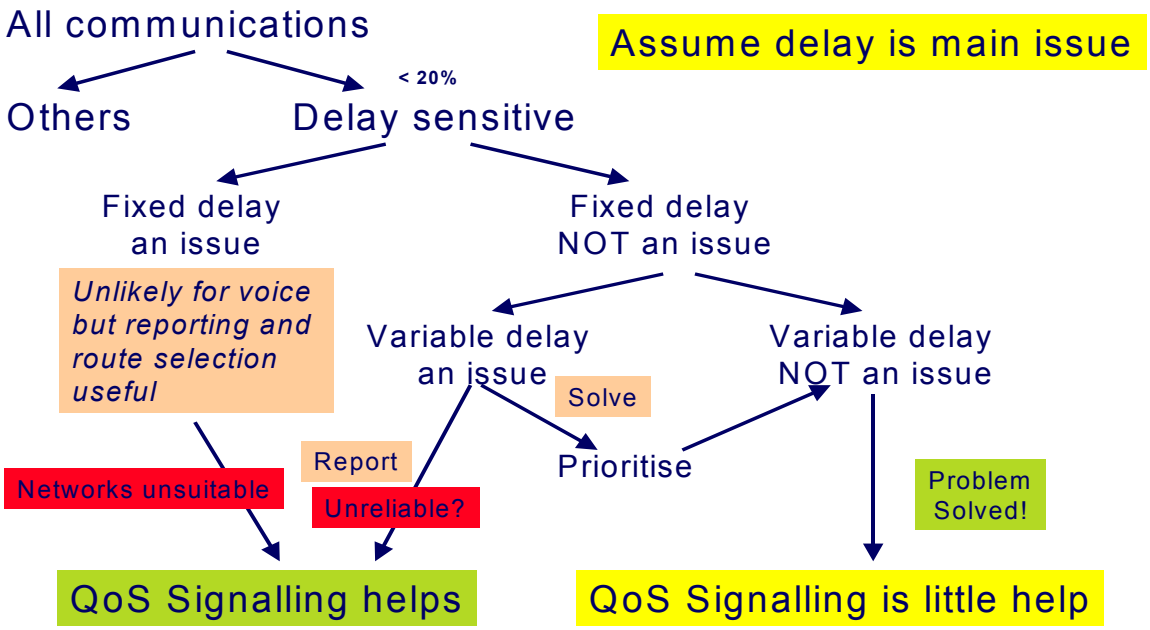

Figure 4: Effectiveness of call related QoS signalling

A further important point is that call related QoS signalling needs to be supported end-to-end to be of any use. Since calls are likely to traverse more than one network and many calls will traverse at least 5 networks, two end networks, two access networks and one transit network, the introduction of QoS signalling faces the problem of achieving critical mass ie of becoming sufficiently widely supported so that there is a high probability of QoS signalling being available end-to-end.

5.3.3 Conclusion on call-related QoS signalling negotiation

Call-related QoS signalling for IP based services is to date only a proposal that has been standardised within TIPHON. The concept is not yet mature and there is no practical experience of the extent to which users would value the functionality that it provides. Thus conclusions about it can be only provisional.

QoS signalling is designed to support the provision of an absolutely guaranteed QoS for a service, ie if the service can be provided the quality will exceed a specified level, but this means that access may be denied at times. It is not at all clear that users will prefer this approach to one where they always have service access but quality may vary.

The analysis of the quality factors suggests that QoS signalling negotiation is only likely to be useful if the network performance is marginal. Provided that the proportion of delay sensitive traffic to the total traffic is low, the use of prioritisation should be effective in ensuring adequate performance for delay sensitive traffic in networks that are designed to be fundamentally suitable for such traffic, ie where the fixed delays are within an acceptable level. (Prioritisation would require some indication of the call class but this might be given in the media packets and would not necessarily involve application layer signalling.)

Thus in summary the basic concept is open to doubt from the perspective of the user and the proportion of cases where it might add value is likely to be low. Thus overall it does not appear to be a very attractive approach.

5.4 Reservation, Segregation and prioritisation of traffic type

The sources of the impairments generated in the networks are the queues that delay the packets at the routers. Queues may be a problem in networks of all sizes, but at points where the transmission capacity is low they are especially critical because delay sensitive packets may be held up for significant lengths of time while large data packets are transmitted. These problems can occur especially in access networks, and some access systems will use techniques for subdividing long data packets to prevent long delays to smaller speech packets.

Three quite different solutions have been developed in IETF:

- RSVP, which reserves capacity on a per-call basis.
- DiffServ, which applies different behaviours at the routers to different classes of traffic.
- MPLS (Multi-Protocol Labelling System), which segregates traffic by adding labels to packets so that the internal routers can route on the labels. This has the effect of segregating a network into several separate virtual networks that can each be dimensioned differently to that for example the virtual network for speech can be dimensioned generously for low router queues.

These techniques are not widely implemented at present. The implementation in the Internet of DiffServ with a simple standardised set of behaviours would greatly improve the Internet. The practical difficulty is creating an incentive for users to classify traffic correctly.

6 NETWORK PERFORMANCE

The main standard that defines network performance for international connections from UNI to UNI is ITU-T Recommendation Y.1541. This Recommendation defines six different network QoS classes that are unrelated to the TIPHON classes. For each class the Recommendation specifies performance values for each of the IP-related performance parameters defined in ITU-T Recommendation Y.1540. The network QoS classes defined here are intended to be the basis of agreements between end-users and network service providers, and between service providers. The limited number of QoS classes defined in Y.1541 support a wide range of applications, including the following: real time telephony, multimedia conferencing, and interactive data transfer.

Y.1541 defines limits for the following network performance parameters:

- IPTD – IP packet transfer delay
- IPDV – IP packet delay variation
- IPLR – IP packet loss ratio
- IPER – IP packet error ratio.

The IP network QoS classes that have been defined provisionally based on these parameters are the following:

		QoS Classes					
Network Performance Parameter	**Nature of Network Performance Objective**	**Class 0**	**Class 1**	**Class 2**	**Class 3**	**Class 4**	**Class 5 Un-specified**
IPTD	Upper bound on the mean IPTD (Note 1)	100ms	400ms	100ms	400ms	1 s	U
IPDV	Upper bound on the 1-10^{-3} quantile of IPTD minus the minimum IPTD (Note 2)	50ms (Note 3)	50ms (Note 3)	U	U	U	U
IPLR	Upper bound on the packet loss probability	$1*10^{-3}$ (Note 4)	$1*10^{-3}$ (Note 4)	$1*10^{-3}$	$1*10^{-3}$	$1*10^{-3}$	U
IPER	Upper bound	$1*10^{-4}$ (Note 5)					U

General Notes:

The objectives apply to public IP Networks. The objectives are believed to be achievable on common IP network implementations. The network providers' commitment to the user is to attempt to deliver packets in a way that achieves each of the applicable objectives. The vast majority of IP paths advertising conformance with Recommendation Y.1541 should meet those objectives. For some parameters, performance on shorter and/or less complex paths may be significantly better.

An evaluation interval of 1 minute is provisionally suggested for IPTD, IPDV, and IPLR, and in all cases the interval must be reported.

Individual network providers may choose to offer performance commitments better than these objectives.

"U" means "unspecified" or "unbounded". When the performance relative to a particular parameter is identified as being "U" the ITU-T establishes no objective for this parameter and any default Y.1541 objective can be ignored. When the objective for a parameter is set to "U", performance with respect to that parameter may, at times, be arbitrarily poor.

Figure 5 gives guidance for the applicability and engineering of the network QoS classes in order to support different applications:

QoS Class	Applications (Examples)	Node Mechanisms	Network Techniques
0	Real-Time, Jitter sensitive, high interaction(VoIP, VTC)	Separate Queue with preferential servicing, Traffic grooming	Constrained Routing and Distance
1	Real-Time, Jitter sensitive, interactive (VoIP, VTC).		Less constrained Routing and Distances
2	Transaction Data, Highly Interactive, (Signalling)	Separate Queue, Drop priority	Constrained Routing and Distance
3	Transaction Data, Interactive		Less constrained Routing and Distances
4	Low Loss Only (Short Transactions, Bulk Data, Video Streaming)	Long Queue, Drop priority	Any route/path
5	Traditional Applications of Default IP Networks	Separate Queue (lowest priority)	Any route/path

Figure 5: QoS classes in Y.1541

The QoS classes in Y.1541 are not levels of quality like the TIPHON classes, but rather QoS requirements profiles for different types of traffic. The general quality level of the requirements is high. For example the packet loss levels are very low, and for speech would allow use of codecs such as G.711 that are designed for circuit switched applications.

7 TERMINAL AND CODEC ISSUES

7.1 Introduction

Terminal and codec design is very important because of the scope for compensating for impairments that arise in the networks. Terminals are especially important where the public Internet is being used and the general level of impairments may be higher than in generously dimensioned networks run by the telcos specifically for carrying voice traffic.

7.2 Speech coding basics

Speech codecs are intended to compress the voice in such a way as to minimally affect voice quality and in turn the quality of service (QoS) delivered by IP telephony systems.

The speech encoder converts the digitized speech signal (after A/D conversion) to a bit-stream, which is packetized and sent over the IP network. The speech decoder then reconstructs the speech signal from the packets received. The reconstructed speech signal is, therefore, an approximation of the original signal. Speech codecs are deployed at end points and so, determine the achievable end-to-end quality.

A speech codec has several important features, including speech quality, bit or compression rate, robustness, delay, sampling frequency, and complexity.

The quality of speech produced by the speech codec will define the upper limit for achievable end-to-end quality. This will determine sound quality for perfect network conditions - no packet loss, delay, jitter, echo, or other quality-degrading factors. Other factors affecting the overall sound quality include the handling of different voices as well as the effect of non-speech signals such as background noise.

Historically, a number of speech codecs have been designed to address bit errors in the communication channel. However, in packet networks the speech codec must be able to deal with lost packets. This ability determines the sound quality in congested situations where packet loss is likely to happen.

The delay introduced by the speech coder can be divided into algorithmic and processing delay. The algorithmic delay occurs because of framing for block processing, since the encoder produces a set of bits representing a block of speech samples. Furthermore, many coders using block processing also have a look-ahead function that requires a buffering of future speech samples before a block is encoded. This adds to the algorithmic delay. Processing delay is the time it takes to encode and decode a block of speech samples.

The complexity of a speech-coding algorithm dictates the computational effort required and the memory requirements. Complexity is an important cost factor for implementing a codec and generally increases with decreasing bit rate.

Increasing the sampling frequency from the 8 kHz used for telephony band products to the 16 kHz used for wide-band speech coding produces distinctly more natural, comfortable, and intelligible speech. To date, wide-band speech coding has found limited use in applications such as videoconferencing because speech coders mostly interact with the public switched telephone network (PSTN) and so have been limited to compatibility with codecs used in the network. There is no such limitation in calls carried wholly over an IP network. Therefore, because of the significant quality improvement attainable, the next generation of speech codecs for VoIP will be wide-band.

7.3 Use of traditional circuit switched codecs for voice over IP

Figure 6 lists the most common codecs and their characteristics.

Type		Standard	Bit rate (Kbit/s)	Delay (ms)	Ie
PCM	Narrow	G.711 G.712	64	0.75	0
ADPCM	Narrow	G.726/G.727	40 32 24 16	0.25	2 7 25 50
LD-CELP	Narrow	G.728	16 12.8	2	7 20
CS-ACELP	Narrow	G.729	8	35	12
VSELP	Narrow	GSM 06.20 Half rate	5.6	60/95	23
RPE-LTP	Narrow	GSM 06.20 Full rate	13	60/95	20
ACELP	Narrow	GSM 06.60 E-FR	12.2	60/95	3
ACELP	Narrow	G.723.1	5.3	97.5	19
MP-MLQ	Narrow	G.723.1	6.3	97.5	15
AMR Narrowband	Narrow	EN 301 703 GSM 06.71	11.4/22.8	Not known	Not known
AMR Wideband	Wide	TS 26.171	6.6-23.8	Not known	Not known
7 kHz	Wide	G.722	64		

Figure 6: Standardised codec types

The most commonly used codecs for IP telephony today are G.711, G.729A, and G.723.1 (at 6.3 kbps). These were designed for (or based on technology designed for) circuit switched telephony.

G.711 was designed for use in circuit switched telephony, and as such it does not include any means to counter packet loss. The common remedy of inserting 'zeros' (zero stuffing) whenever packet loss occurs leads to voice break-up and a steep degradation of quality. Error concealment can be introduced by extrapolating/interpolating received speech segments. An example is the method described in Appendix I to G.711 [4], which provides some improvement but does not guarantee robust operation.

G.729 and G.723.1 belong to the CELP coder class, which is also based on a coding model that was designed for circuit switched mobile telephony. Basic speech quality is worse than PSTN quality.

Mobile telephony has been the major driver for development of speech coding technology in recent years. All the codecs used in mobile telephony, as well as G.729A and G.723.1, are based on the code excited linear prediction (CELP) paradigm. Due to their design for use in circuit switched networks, these codecs were intended to handle bit errors rather than packet loss.

The CELP coding process uses inter-frame dependencies that lead to inter-packet dependencies. Error propagation resulting from such dependencies leads to poor performance when packets are lost or delayed and so speech quality degrades rapidly with increasing packet loss.

Figure 7 compares the convergence times for resynchronizing encoder and decoder states after packet loss for the traditional G.729 and G.723 codecs and the iLBC codec, which has been specified through the IETF AVT (audio/video transport) Working Group. The iLBC codec, which is designed not to have any inter-packet dependencies, recovers much more quickly giving much better performance.

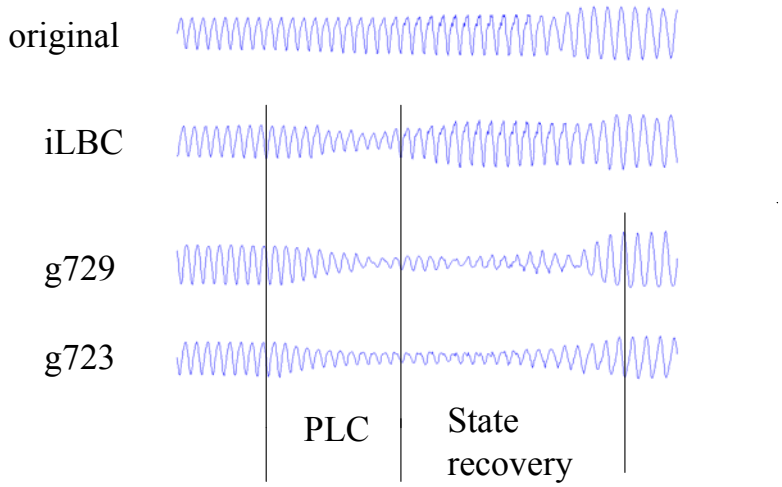

Figure 7: Recovery from packet loss

Subjective evaluations have confirmed that it is more important to restore the state of the decoder after a frame erasure than to attempt to restore the speech that was lost during the frame erasures. Comparisons have also shown that it is always better to use the information in an errored packet than to discard a packet when the error level exceeds a certain threshold.

With the development of voice communications over packet, a new generation of codec tolerant to packet loss has been developed by GlobalIPSound in Sweden (see http://www.globalipsound.com). They offer the codecs listed in figure 8.

Type	Description	Quality
Enhanced G.711	Narrowband	Comparable to G.711
GIPS iLBC	Narrowband	Better than G.729/G.723.1
GIPS iPCM	Wideband	Better than G.722
GIPS iSAC	Wideband adaptive	

Figure 8: GlobalIPSound codec types (tolerant to packet loss)

7.4 Speech processing designed for speech over packet networks

7.4.1 Introduction

Speech processing software designed for real time communications over IP networks aims to provide an edge-device QoS solution with high voice quality, even under severe network degradations due to jitter and packet loss. Its design is characterized by the following principles:

- Speech quality in IP telephony should generally be equal-to or better-than PSTN.
- Speech quality should degrade gracefully with increasing packet loss and delay. Moderate packet loss should be inaudible.

7.4.2 Codec enhancements

The main approach achieving good speech quality in the presence of packet loss is to avoid inter-packet dependencies and so to prevent error propagation. The new speech processing algorithms support diversity, which results in minimal loss of speech from packet loss but they do so not by adding redundancy as with forward error correction (FEC) methods, but by generating multiple descriptions of the source signal of equal importance, called Multiple Description Coding (MDC).

These descriptions can be decoded independently at the receiver. If all descriptions are received, the source signal can be faithfully reconstructed. If only a subset of the descriptions is received, the quality of the reconstruction is lower

These new techniques achieve a significant increase in performance and can achieve adequate quality at packet loss levels as high as 20-30%, whereas traditional codecs degrade significantly as packet loss levels rise above 3-5%. Some of this improvement comes from using a higher bit rate but the narrowband codecs are still capable of working over dial-up access.

7.4.3 Playout buffer control

Delay variation is the fundamental main problem in real-time voice applications over IP, consequently one of the important functions to be implemented at the receiver is the Playout Controller, which buffers the variable delay (jitter) in the network to give a constant stream of packets for the codec. The playout controller controls the size of the buffer at the receiving end and therefore trades off packet loss and overall delay or latency. The design and algorithms for playout controllers have become much more sophisticated in their ability to minimise additional delay and also conceal packet loss. Furthermore the playout controller can reduce the effect of clock skew, which can occur due to the sender and receiver not being correctly synchronised.

7.4.4 "Traditional" Playout Buffer

A traditional playout buffer removes the jitter in the arrival times of the packets by adding delay so that the total delay in the network and the buffer is constant. The objective of a playout buffer algorithm is hence to keep the buffering delay as short as possible while minimizing the number of packets that arrive too late to be used. A larger playout buffer causes increase in the delay and decreases the packet loss. A smaller playout buffer decreases the delay but increases the packet loss.

The traditional approach is to store the incoming packets in a buffer (packet buffer) before sending them to the decoder. The most straightforward approach is to have a buffer with a fixed number of packets. This results in a constant system delay (if there is no clock drift) and requires no computations and therefore gives a minimal complexity. The drawback with this approach is that the length of the buffer has to be made sufficiently large to accommodate the maximum jitter (which in practice occur quite seldom).

In order to keep the delay as short as possible it is important that the jitter-buffer algorithm adapts rapidly to changing network conditions. Therefore, playout buffers with dynamic size allocation, so called adaptive playout buffers, are most common nowadays.

The adjustment of delay is achieved by inserting packets in the buffer, when the delay needs to be increased, and removing packets when the delay can be decreased. The insertion of packets usually consists of repeating the previous packet. Unfortunately, this will almost always result in audible distortion and hence most adaptive playout buffer algorithms are very cautious when it comes to delay adaptations in order to avoid such effects. To avoid audible distortion, the removal of packets can only be done during periods of silence. Hence, delay builds up during a period of speech and it can take several seconds before a reduction in the delay can be achieved. Also, high delay at the end of a period of speech will have a severe effect on the conversation since it increases the probability of double talk.

This traditional packet buffer approach is limited in its adaptation granularity by the packet size since it can only change the buffer length by adding or discarding integral numbers of packets.

Some of the current implementations of adaptive playout buffers have been shown to experience problems when there are packet losses in the network. For example, studies for TIPHON show that the playout buffer delay can increase significantly in cases where packet losses are present.

7.4.5 Packet Loss Concealment

Until recently, two simple (codec independent) approaches for packet loss concealment have been used.

The first method, referred to as zero stuffing (ZS), is obtained by simply replacing a lost packet with a period of silence of the same duration as the lost packet.

The second method, referred to as packet repetition (PR), assumes that the difference between two consecutive speech frames is quite small. Hence, the lost packet is replaced by simply repeating the previous packet. In practice, though, even a minor change in, for example, the pitch frequency is easily detected by the human ear. In addition, it is virtually impossible to achieve smooth transitions between the packets with this approach. However, this approach performs fairly well for very

small probabilities (less than 3 %) of packet loss. Packet repetition outperforms zero stuffing but both methods are very sensitive to packet loss compared to the more advanced methods.

Recently, the ITU standardized a method for packet loss concealment in G.711 Appendix I (usually referred to as G.711 PLC). This is a more sophisticated method that tries to estimate the lost packet from previously decoded speech and hence cannot be implemented in the packet buffer.

Some codecs, such as those based on CELP, have their own built-in packet loss concealment algorithm. In many cases this gives a reasonable concealment during the loss. Unfortunately, as previously noted, many of these codecs suffer from their packet inter-dependencies instead.

7.4.6 Clock drift (skew)

Clock drift is the difference in the rates of the clocks at the sending and receiving end. The traditional approach (in a TDM network) is to deploy a clock synchronization mechanism at the receiver to correct for clock drift by comparing the number of samples received with the local clock. In an IP network, however, it is hard to do reliable clock drift estimation. The reason is that the estimates of drift only can be based on averaging packet arrivals at a rate of typically 30 - 50 per second instead of averaging on a per sample basis at a rate of 8000 per second (as done in TDM networks). In addition, because of the jitter present in IP networks it is almost impossible to obtain an accurate estimate of the clock drift and hence many algorithms designed to mitigate this effect fail.

7.4.7 Advanced Algorithms

Advanced algorithms, eg those designed by GlobalIPSound, are coming onto the market that include both delay adaptation and error concealment in one unit and adapts quickly to changing network conditions to ensure high speech quality with minimal buffer latency. The following is a generic discussion of the techniques that can be used. These algorithms work on both the input and the output of the decoder, whereas traditional algorithms worked only on the inputs. By working on the outputs, these algorithms are able to smooth the speech when a packet is lost rather than just replacing the lost packet.

Figure 9 shows a simple block diagram of an advanced algorithm and contrasts it with traditional packet loss concealment. .

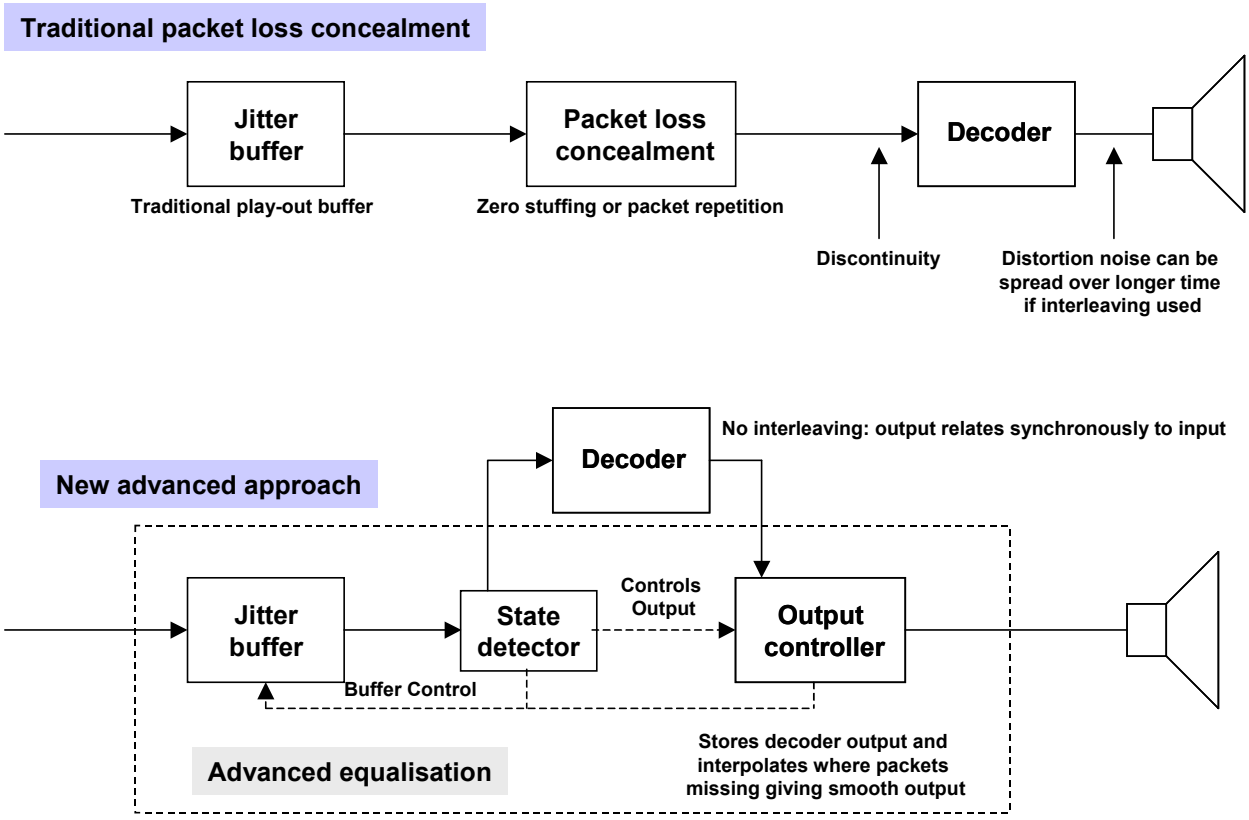

Figure 9: Simple block diagram depicting how an advanced algorithm can interacts with the speech decoder and the jitter buffer

The algorithm works in the following way. When packets are available from the jitter buffer, the packets are sent to the decoder and the decoder output is used but with some additional fixed delay in the output controller. When packets are not available (because the jitter buffer has emptied as a result of increasing network delay) the packet loss detector stops the output controller from using the output of the decoder and instead the output controller inserts a new short synthetic speech segment instead of the output of the decoder to interpolate across the gap caused by the packet loss. The purpose of the additional delay added in the packet controller is to allow for this interpolation.

The effectiveness of this approach depends on the coder not using spreading so that the effects of lost packets on the decoder output are not spread over a significantly greater time period than the duration of the lost packets. The main advantage over traditional systems is that the concealment takes place on the decoder output rather than the input and includes interpolation rather than extrapolation allowing for the output to be smoothed.

The loss detector and the output controller are able to measure the extent and jitter at the network output and they are used to control the jitter buffer so that the size of the buffer can be reduced when jitter is low. Changes to the jitter buffer require corresponding changes to the speech output as described in 7.4.4.

The superior performance in terms of clock drift is achieved because the algorithm does not have to estimate the actual clock drift but is able to mitigate its effect automatically at the speech output rather than the decode input. It is therefore a very attractive alternative to a standard clock synchronization mechanism.

Studies were carried out to compare the delay of the NetEQ algorithm used by GlobalIPSound in their codecs to handle jitter variation with alternative solutions. The graph below illustrates the delay performance of different algorithms on a channel with quite a lot of jitter. NetEQ manages to keep the delay much lower than both the adaptive and the fixed playout buffer. In general a delay improvement of 30-80 ms can be expected with the NetEQ algorithm compared to traditional approaches. Figure 10 also shows that NetEQ adapts to the envelope of the jitter very efficiently, since the delay is reduced in less then a second after a jitter peak.

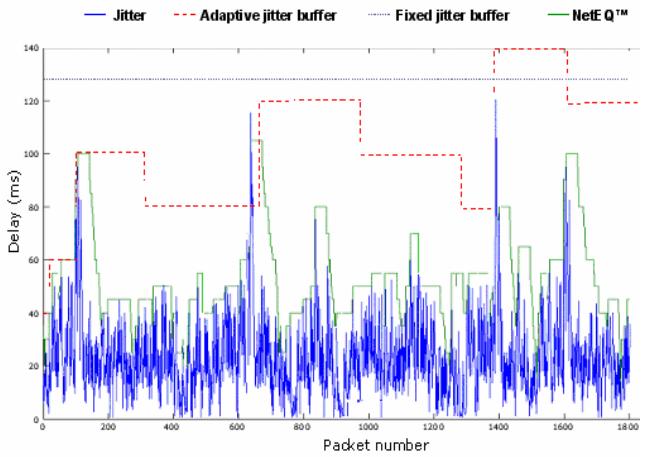

Figure 10: Delay performance for different playout buffers on a system with 20 ms packets.
(Note that the constant system delays of about 80ms have been removed from all curves.)

Figure 11 shows the performance improvements that can be achieved for narrowband codecs. MOS=5 denotes excellent quality, MOS=1 denotes bad quality.

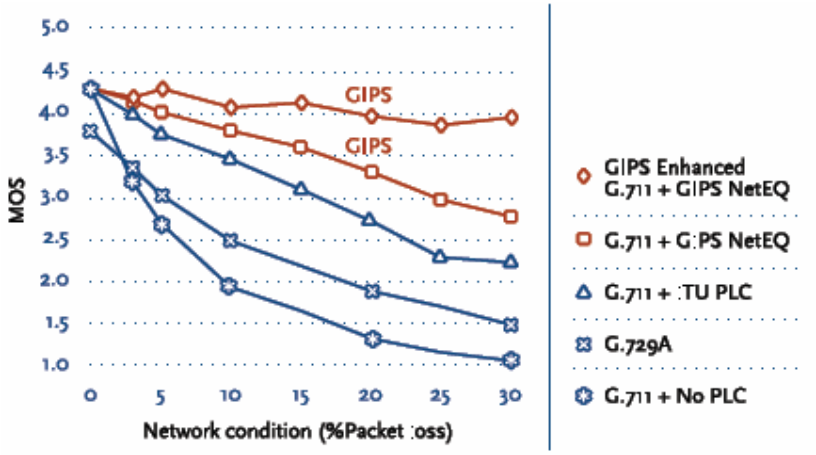

Figure 11: Narrowband performance improvements in presence of packet loss
(Source: Lockheed Martin - COMSAT)

Figure 12 shows the performance improvements that can be achieved for wideband codecs.

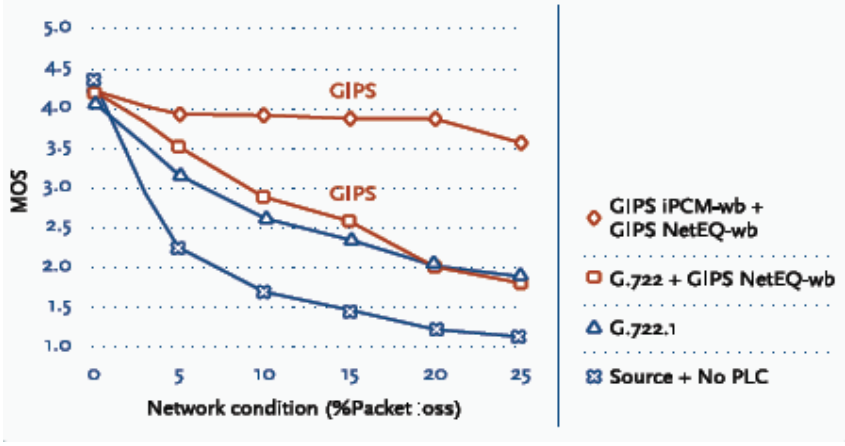

Figure 12: Wideband performance improvements in presence of packet loss
(Source: Lockheed Martin - COMSAT)

7.4.8 Other approaches

Other approaches to deal with packet loss are forward error correction schemes and interleaving schemes.

Forward error correction schemes are designed to correct bit errors that are well distributed in time, whereas packet loss results in errors in many consecutive bits (a burst), which significantly decreases the efficiency of FEC schemes. In order to combat a burst or errors, redundant information has to be added and spread over several packets, which introduces greatly increased delay. Hence, the repair capability of forward error correction is limited by the delay budget. Furthermore the use of additional bits increases the network loading and so may aggravate the problem whose effects the scheme is trying to reduce.

Another technique for reducing the effect of packet loss is interleaving where the coder output frames are interleaved to that a burst of errors is spread over more than one frame. Whilst interleaving does not increase the data rate of transmission

it does increase delay significantly. The efficiency of loss recovery increases if the source packet is interleaved and spread over more packets but the more packets that are used the higher the delay.

Both forward error correction and interleaving are used in GSM.

7.4.9 Terminals

With many voice over IP applications, voice communications will be provided at personal computers and users will tend to use headsets or speakers rather than the traditional handsets, although handsets that can clip onto the video display unit are available. Where loudspeakers are used to give hand free operation, echo control is needed in the computer. This is provided in the latest versions of Microsoft Windows™. In general the use of PCs rather than traditional handsets creates more scope for terminals to be configured incorrectly and correspondingly degrade quality, although it does provide better support for wideband speech.