Speech Quality Measurement Tools for Dynamic Network Management

Simon Broom, Mike Hollier

Psytechnics, 23 Museum Street, Ipswich, Suffolk, UK IP1 1HN
Phone +44 (0)1473 261800, Fax +44 (0)1473 261880

simon.broom@psytechnics.com

Summary

There is a clear trend in the telecommunications industry, away from conventional circuit-switched networks and towards flexible, dynamically managed packet-based "Next-Generation Network" (NGN) models. The success of NGNs will hinge on whether they can deliver services of acceptable quality at lower overall cost than the networks they replace. Experience from the development of voice over IP (VoIP) over the last five years has shown that the efficient delivery of the required quality is likely to remain a key challenge. Effective network management requires performance monitoring of live traffic, and so real-time non-intrusive monitoring techniques are especially relevant. Further since NGNs will be packet based, methods that can accurately predict network performance from packet statistics are desirable – we have developed such a method, psyVoIP, which has already been adopted by leading industry players.

This paper introduces objective speech quality measurements and examines how non-intrusive monitoring can be applied to dynamic network management. The paper begins by reviewing VoIP and techniques for measurement of perceived quality. PsyVoIP, a non-intrusive measurement model specifically designed for call-by-call monitoring of VoIP, is introduced. Finally, the application of psyVoIP to provide accurate voice quality data for SLA/SLG and network management is discussed.

Keywords

Quality, VoIP, Network Management, Real-time Measurement

Voice over IP (VoIP) Review

VoIP Network Architecture

Voice over IP (VoIP) can be divided into two broad categories; systems that transport voice over the Internet and systems that carry voice across a managed IP network. However, there are many VoIP solutions in design and development making it difficult to define a generic architecture.

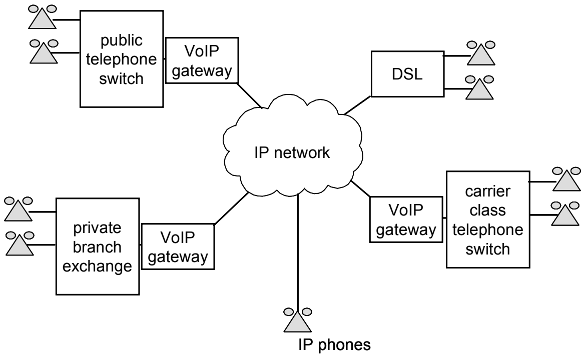

Figure 1 Voice over IP Network Architecture

Fortunately, from a monitoring perspective, the VoIP stream itself is generally well defined. Figure 1 provides a summary of how VoIP systems are being developed to interconnect the telecommunications networks of the world as well as provide direct services.

VoIP Fundamentals

When measuring the speech quality of a VoIP system from within the IP network, it is highly beneficial to account for the specific gateway or IP phone. This device is typically highly non-linear and can have a substantial impact on the quality perceived by the end-user from a given packet stream. The receive path in a VoIP gateway, or IP phone, can be simply described by the functional blocks shown in Figure 2. A jitter buffer processes the IP packet stream. This removes jitter and re-orders any mis-sequenced packets. The packets are then presented to the codec in the appropriate time sequence where they are decoded to audio. An attempt may be made to compensate for any lost packets using error concealment techniques.

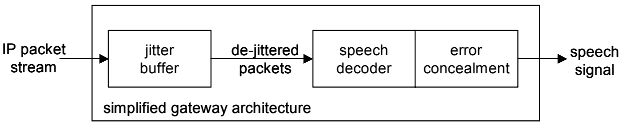

Figure 2 Simple VoIP gateway architecture – receiving side

We have found considerable diversity in the operation of VoIP gateways and IP phones. Each manufacturer often uses a variety of different jitter buffer and error concealment algorithms. This is demonstrated in Figure 3 where speech quality is plotted as a function of mean absolute packet delay variation for three different VoIP Gateways.

The impact on speech quality of the jitter buffer is heavily dependent on the effectiveness of a specific algorithm and implementation. The codecs used are generally standardised, while the effects of additional error concealment are less well known. Both jitter buffer and error concealment algorithms tend to be proprietary and can vary widely between manufacturers.

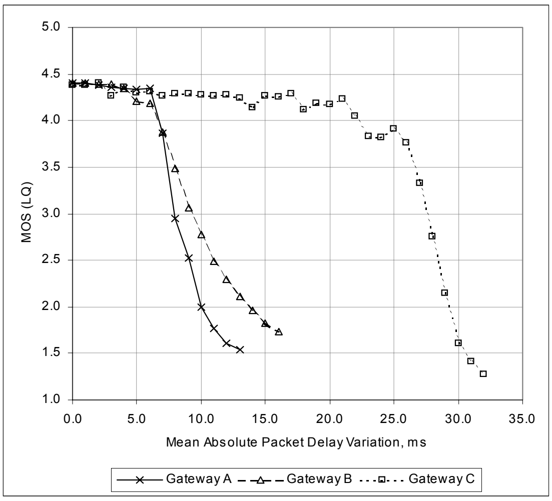

Figure 3 Speech quality (MOS) vs mean absolute packet delay variation for three VoIP Gateways

Significant performance improvements can be achieved by optimising jitter buffer design and error concealment techniques; and this is a key differentiator for manufacturers when similar products compete in the marketplace. This means that proprietary algorithms are widespread and can be upgraded to improve performance.

In consequence different gateways, or IP phones, can provide different speech quality from an identical IP stream. The probe location is not restricted and can be at any convenient network interconnect point. However, to accurately predict speech quality from an IP packet stream (or even a post jitter-buffer packet stream) non-intrusive probes, like psyVoIP, must take account of the specific gateway in use.

Whilst a unified voice and data network undoubtedly brings advantages, it also introduces new problems; a packet based network that was designed for data introduces a host of potential new impairments to real-time services such as voice.

Impairments Affecting Speech Quality

Packet network degradations have been extensively studied over the last 40 years, but most of this work has been focused on data networks. Only in the last few years, with the emergence of VoIP has this work turned to the study of real-time audio streams.

At the most basic level, network impairments can be classified into two broad categories:

- packet loss within the network
- packet delay variation (jitter) within the network

Ultimately both of these lead to missing packets at the output of the jitter buffer.

Many different models have been developed to simulate/measure packet loss distributions within the network. The most well known of these was devised in the 1960's to model bit errors on telephone channels by Gilbert [1]. This was later extended by Elliott to produce the well known Gilbert-Elliott model [2,3]. Since then many other more complex models have been devised. However these models are designed to operate over long periods of times, typically minutes or hours, whereas VoIP users can detect, and are affected by, impairments over a much shorter time period in the order of five to ten seconds.

There is however evidence to suggest that jitter is often a more significant problem than packet loss. IP routers tend to have large queues and good collision detection methods, so it is rare for packets to be lost in normal operating conditions. However switching delay can be highly variable, in particular with networks that are heavily loaded with data traffic or are close to maximum capacity. This can lead to significant jitter.

Jitter is much harder to model though, and studies have typically focused on measuring the distribution of jitter over a period of time. Essentially jitter is caused by variations in the delays incurred by a packet during its journey through the network. Modelling jitter relies heavily on queuing theory and models of queues within the network. These models need to be updated every time new queuing mechanisms are developed to help reduce the delay variation of VoIP packets.

Review of Speech Quality Measurement

Speech Quality Measurement

There is no absolute physical definition of speech quality with the "baseline" provided by the subjective perception of human listeners.

However, systems are designed and tested against what an "average" person thinks of quality. This is often summarised by the term MOS (Mean Opinion Score).

There are essentially two complementary approaches to testing speech quality:

– subjective tests, which seek the average opinion of users

– objective tests

Subjective testing

Subjective tests aim to find the average user's perception of a system's speech quality by asking a panel of users a directed question and providing a limited response choice. For example, to determine listening quality users are asked to rate "the quality of the speech" on a five-point scale: Excellent, Good, Fair, Poor, Bad. A mean opinion score, MOS, is calculated for a particular condition by averaging the votes of all subjects [4]. Subjective tests are time consuming to perform, must be very carefully designed and executed and are not a practical solution for most real-world measurement needs.

Objective testing

Objective testing techniques measure physical properties of a system in order to predict perceived performance. This is typically achieved for telephony systems by: the injection of a test signal for an intrusive (active) measurement, or the monitoring of live traffic for a non-intrusive (passive) measurement. Such measurements are repeatable, efficient and fast.

Significant work has lead to objective measurement techniques that replace the need for a large proportion of subjective testing to provide an automated prediction of speech quality.

Intrusive (active)

Intrusive techniques inject test signals into a system so they can be captured and assessed at a further point (Figure 4). To do this the system under test is taken out of service, although in terms of a telephone network this can be limited to a single telephone channel.

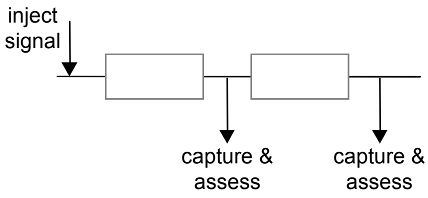

Figure 4 Illustration of an intrusive test

Assessment involves a comparison between the injected and captured signals. This method:

- enables isolation of the system under test,
- is capable of high accuracy
- can be used during development, commissioning and routine monitoring
- may incur a real cost, e.g. call charges when testing third-party networks
- is possible when no customers are on a system (during development before live traffic is present)
- allows control over external factors

In some instances this type of testing may incur a real cost, such as when testing over third-party networks. However a benefit of intrusive testing is that it can be done when no customers are on a system (during development before live traffic is present) and allows control over external factors.

The world standard for intrusive assessment of end-to-end speech quality is the perceptual evaluation of speech quality (PESQ) model, ITU-T Recommendation P.862 [8,9,10]. PESQ has been tested on a wide range of network conditions, and gives accurate predictions of subjective quality with most currently available technologies for fixed, mobile and VoIP networks.

Non-intrusive (passive)

Non-intrusive techniques monitor live customer traffic to determine the quality perceived by the customer (Figure 5). By performing measurements on live customer-traffic, network capacity is not lost to testing and the service provider can know what quality their network is actually delivering to customers.

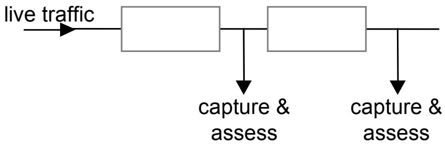

Figure 5 Illustration of a non-intrusive test

Non-intrusive techniques allow for a larger number of tests, at a much-reduced operational cost compared to intrusive monitoring. However, these measurement techniques are more difficult to develop and are typically less accurate than intrusive techniques.

More than a single MOS

For most telecommunications testing Listening Quality is the preferred measure of performance and is generally referred to as MOS. However, a MOS score can be created by asking subjects any question which has a discrete number of responses. ITU-T Recommendation P.800 [4] defines a number of mean opinion scales including listening quality. In addition, subjective tests can be conversation based and hence will result in a conversational MOS. " MOS" is typically used to refer to Listening Quality, but clarification should be sought if in any doubt.

PsyVoIP Overview

PsyVoIP is a set of software components for the assessment and management of VoIP speech quality. PsyVoIP operates by monitoring live VoIP customer calls to determine the speech quality delivered by a VoIP network.

At the core of psyVoIP is the Probe. The probe combines the software components that capture packets, extract call-streams, parameterise VoIP degradations and predict speech quality. The Probe is designed for integration into network test equipment or VoIP network elements.

Figure 6 illustrates one example of where the probe may be located in a VoIP connection. In this location, psyVoIP can monitor passing traffic streams to predict the quality of phone calls as they leave the VoIP network and enter the PSTN.

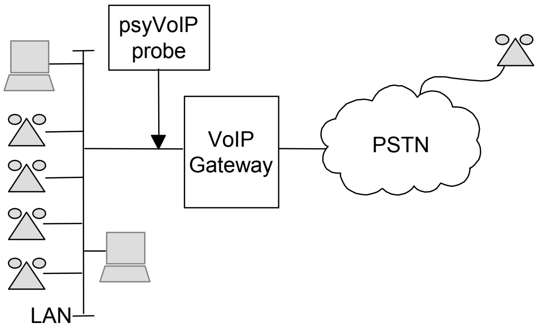

Figure 6 Probe Location Example

Probe Architecture

The architecture of the Probe is shown below in Figure 7. Packets are captured from the network and passed to the call identification module which maps the packet to a specific call. The pre-process block extracts the information required by the rest of the Probe so that the rest of the packet can be discarded. Out of sequence packets are accounted for by a re-sequence buffer which enables packets to be processed in their original order by the following blocks. The voice activity detection (VAD) block enables packets to be marked as either speech or non-speech wherever possible. This allows more accurate speech quality predictions to be made – for example, packets lost in speech have a greater impact on speech quality than those lost in silence. Finally relevant statistical descriptors are extracted from the packet stream and the internal parameter states are updated.

The current speech quality prediction for a call can then be queried whenever required and is calculated from the internal parameter states.

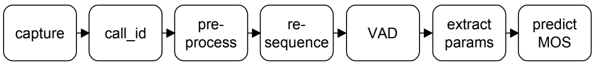

Figure 7 Probe Architecture

Characterisation of VoIP Devices

PsyVoIP handles the differences between gateways and IP phones by characterising each device. This process involves characterising a gateway's speech quality performance over a wide range of network conditions. A practical characterisation tool has been developed to allow this one-time off-line procedure to be easily performed. Once a gateway or IP phone has been characterised this information is stored in a light-weight file, which can be loaded into the Probe and used to achieve accurate quality monitoring of the chosen device. The average performance of the Probe over ten VoIP devices for a range of network conditions is presented in Table 1 below. The table shows the average condition (5 tests per condition) difference between psyVoIP measurements and ITU-T P.862 listening quality MOS scores.

Packet loss	**Jitter (ms)**					
(%)	**0**	**25**	**50**	**100**	**150**	**300**
0	0.08	0.08	0.08	0.14	0.20	0.20
5	-0.07	0.10	0.07	0.11	0.20	0.21
10	-0.16	0.12	0.15	0.15	0.20	0.17
20	0.04	0.09	0.06	0.00	0.12	0.07
30	0.19	0.21	0.25	0.15	0.17	0.12
OOP 5%	-0.11	0.07	0.12	0.05	0.01	0.05
OOP 10%	-0.11	0.07	0.12	0.05	0.01	0.05
OOP 20%	0.07	0.22	0.21	0.16	0.03	0.08

OOP - out of order packets

Table 1 psyVoIP Probe prediction performance – average condition error

PsyVoIP has been submitted to the ITU-T Study Group 12 P.VTQ competition.

Network Management

The existing economic model for telecommunications networks is not expected to be commercially viable beyond 5 to 7 years. This view is based on the rationale that cheaper more flexible infrastructure will facilitate the cost effective delivery of a wide range of new products and services - rendering existing network/service paradigms obsolete. Flexible, dynamically managed, packet-based and ubiquitous networks are at the core of the NGN model.

Quality assurance of performance is essential in NGNs since they are intended to provide high quality services at any time and anywhere. In particular the tolerance users have exhibited with regard to 2G mobile services cannot be expected to persist for NGN and 3G since the higher utility of mobile is now taken for granted and in any case will not apply when broadband users are unaware of traffic origin and demand ubiquitous service and quality.

Efficient real-time management must be based around best use of valuable network resources while ensuring that adequate quality, typically defined by SLAs (Service Level Agreements) and SLGs (Service Level Guarantees), is reliably delivered. A key component of such efficient delivery will be non-intrusive, real-time, perceptually relevant metrics.

Existing Network Management

Current network management methods are far more reactive than pro-active. Long term trends in network traffic patterns are used to plan extra network capacity and upgrade network equipment. There are two main problems with this approach if the NGN model is to succeed. Firstly, if problems do occur on the network that affect a VoIP stream then the network manager will only know about it well after it has happened, and after many customers have been affected. Secondly, even if the network manager could monitor the VoIP speech quality in real time and detect a drop in quality there is no way of immediately adapting the network to improve the quality before the user ended the call because of the problems.

Dynamic Network management

Tools such as PsyVoIP can provide a component of the solution to these problems. They give a real-time speech quality indicator on a per-call basis that can then be used to help manage the network dynamically. This dynamic network management function is the key to successful NGN operation and real-time perceptual measures are the missing link for its implementation.

Future networks will need to be able to continuously re-configure themselves to adapt to the changing demands of the users – per session service delivery. To be able to achieve this for real-time services is a huge challenge, and still some way off. In the meantime network managers can employ tools such as PsyVoIP to tell them how their networks are performing for real-time services such as VoIP.

Conclusions

Quality assurance of performance is essential to provide high quality services - at any time and anywhere. The final arbiter of service performance is the end user - whose opinion of quality is based on his perception. Core 3G networks and NGNs will be packet based, and hence methods that can accurately predict network performance from packet statistics are highly desirable. The psyVoIP algorithm is a non-intrusive measurement algorithm that predicts the listening quality of a live VoIP call. The PsyVoIP algorithm can use knowledge of the downstream device, i.e. the gateway or IP-phone, to increase the accuracy of measurements.

IP networks and NGN paradigms for "per session service delivery" offer greater potential network flexibility and efficiency than existing mixed technology networks. However, if this efficiency is to be delivered, services must make best use of network resources without compromising the quality experienced by the customers. Perceptually relevant metrics, such as those described in this paper, will play a vital role by enabling management of networks against the most appropriate criteria – human perception.

References

[1] Gilbert E. N. "Capacity of a Burst Noise Channel", BSTJ September 1960

[2] Elliott E. O. "Estimates of Error Rates for Codes on Burst Noise Channels", BSTJ, September 1963

[3] Elliott E.O. "A Model of the Switched Telephone Network for Data Communications", BSTJ, January 1965

[4] *Methods for subjective determination of transmission quality,* ITU-T recommendation P.800, August 1996.

[5] R. J. B. Reynolds and A. W. Rix. "Quality VoIP — an engineering challenge", BT Technology Journal, 19 (2), pp 23-32, April 2001.

[6] R. J. B. Reynolds and A. W. Rix. "Achieving VoIP Voice Quality". In R. P. Swale (ed.), *Voice over IP: systems and solutions,* IEE/BT Exact Communications Technology Series, Ch.2, 29-49, December 2001.

[7] A. W. Rix, S. R. Broom and R. J. B. Reynolds. "Non-intrusive monitoring of speech quality in voice over IP networks", ITU-T study group XII delayed contribution COM12-D049, October 2001.

[8] *Perceptual evaluation of speech quality (PESQ), an objective method for end-to-end speech quality assessment of narrow-band telephone networks and speech codecs,* ITU-T Recommendation P.862, February 2001.

[9] A. W. Rix, M. P. Hollier, A. P. Hekstra and J. G. Beerends. "Perceptual Evaluation of Speech Quality (PESQ), the new ITU standard for end-to-end speech quality assessment. Part I – Time-delay compensation", Journal of the Audio Engineering Society, 50 (10), 755-764, October 2002.

[10] J. G. Beerends, A. P. Hekstra, A. W. Rix and M. P. Hollier. "Perceptual Evaluation of Speech Quality (PESQ), the new ITU standard for end-to-end speech quality assessment. Part II – Psychoacoustic model", Journal of the Audio Engineering Society, 50 (10), 765-778, October 2002.