**Source:** Telchemy Incorporated

**Title:** Comparison of TS101 329-5 Annex E with PSQM and PAMS

**Notice:** The author of this document declares that ETSI **may** make the document publicly available.

**Document for:**

| | |
|---|---|
| Decision: | **X** |
| Discussion: | **X** |
| Meeting Report: | |
| Liaison: | |
| Information: | X |

**Contact details:**

| | | |
|---|---|---|
| *First Name, Last Name* | Alan Clark | Renyi Liu |
| *e-mail:* | alan@telchemy.com | rliu@telchemy.com |

## 1. Decision/Action Requested

This contribution provides additional support for TS 101 329-5 Annex E. It is proposed that Annex E is moved from informative to normative status.

## 2. References

[1] ETSI TS 101 329-5 QoS Measurement Methodologies

[2] ITU-T Recommendation P.861 Objective Measurement of Telephone Band (300-3400Hz) Speech Codecs (PSQM)

[3] ITU-T Recommendation G.107 E Model

[4] *The perceptual analysis measurement system for robust end-to-end speech quality assessment (PAMS)* A. W. Rix and M. P. Hollier. Presented at the 2000 IEEE International Conference on Acoustics, Speech, and Signal Processing, Istanbul, 8 June 2000

## 3. Introduction

This contribution provides a comparison of the TS101 329-5 Annex E (VQmon) voice quality monitoring technology with the PSQM and PAMS objective testing methods. It is intended to provide test results and technical data to support evaluation and comparison of VQmon with other test methods.

PSQM [2] and PAMS [4] are widely known voice quality measurement techniques that are implemented in test equipment and PC based software. They both require that a speech file be transmitted through a telephony network in order that the received/ impaired file can be compared with the transmitted/ original file. This requires a test call to be made through a network in order that access to both sent and received signals is possible. The comparison process is computationally complex and generally involves operations such as FFTs.

VQmon (Annex E to [1]) is a lightweight non-intrusive monitoring technology that can be integrated into Voice over IP Gateways or IP Phones, providing a per-call quality metric. VQmon operates on the received voice packet stream and does not require access to the transmitted stream. This means that no additional test calls need to be made and hence no additional network traffic is generated. VQmon is able to monitor the quality of every call made through a Voice over IP network.

## *4. Methodology*

A set of impaired speech files was created from an original audio file by packetizing the file and then passing the packet stream through a random packet loss simulator. Last packet replay was used to simulate the packet loss concealment algorithm implemented by typical CODECs. The source file comprised an 8 bit 22kHz audio file 27 seconds in length with one male speaker.

One set of 40 impaired files was created using a 10mS frame size and a second set of 40 files using a 30mS frame size. Packet loss ranged from 0 to 20%, as shown in Figure 1, packet loss burst length from 0 to 2 seconds and burst density from 20 to 70% as shown in Figure 2. Each set of files was ranked according to average packet loss density and given randomized (non-descriptive) file names.

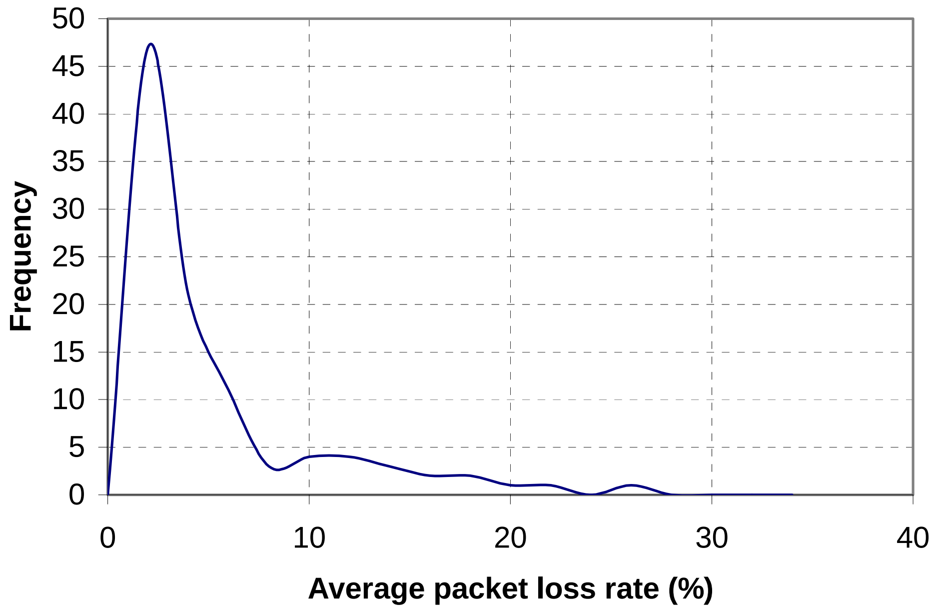


**Figure 1 Histogram showing distribution of packet loss rates amongst files in test set**

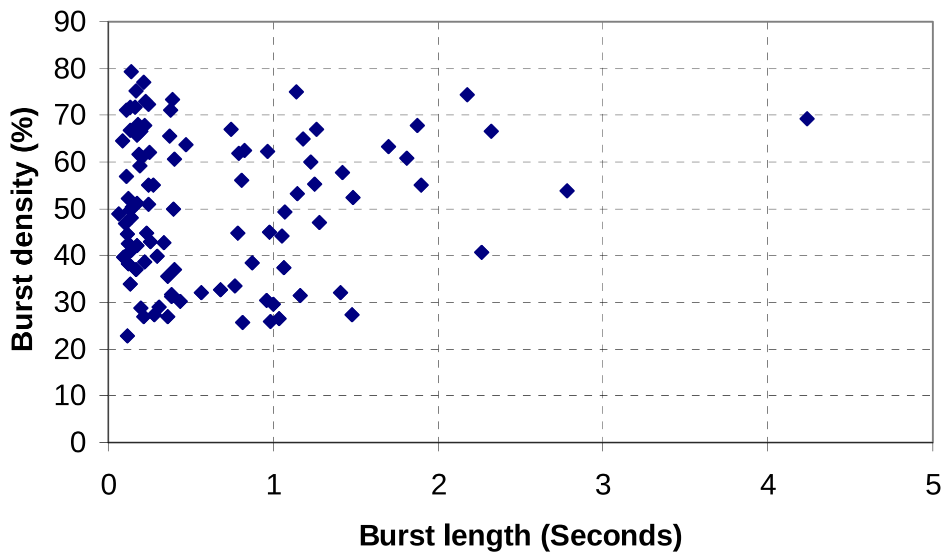


**Figure 2 Scatter diagram showing the burst length/ burst density of impairments in the test set**

Three non-overlapping groups of five sample files were selected from each of the two sets. The files were selected to be approximately equally spaced and ranged from the lowest packet loss to the highest (see Appendix). This file selection method produced six groups of files, which were numbered 1 to 6, with the 10mS frame size files numbered 1-3 and the 30mS frame size files numbered 4-6.

The files were then ranked by a number of listeners. Each person was asked to listen to each group of files and to rank the files from best to worst. The three objective test methods – VQmon, PSQM and PAMS – were then used to obtain metrics and a rank obtained for each set of files using each method.

The listening tests were conducted under controlled conditions. The audio files were replayed from a PC using studio quality headphones. Subjects were asked to listen to each set of 5 audio files and to rank them on the computer screen. They were allowed to play files multiple times if needed in order to "perfect" their ordering, and no time constraints were placed on them. The same test system and same methodology was used for each subject.

The results from PSQM and PAMS were obtained using a GL Communications test set that incorporated software licensed from Malden. Impaired files were compared with the original unimpaired audio file.

## *5. Results*

The *mean rank distance* is defined as the average of the absolute distance between the rank obtained by two different approaches. For example, if method A obtained a rank of 1 2 3 4 and method B a rank of 1 3 2 4 then the mean rank distance would be

$$(\text{abs}(1-1) + \text{abs}(2-3) + \text{abs}(3-2) + \text{abs}(4-4)) / 4 = 0.5$$

Therefore a low mean rank distance implies that two rankings were similar and a distance of 0 that they were identical. This approach was used to compare the ranking given objective test methods with the rank obtained using subjective testing.

The *average subjective rank* obtained using subjective comparison was determined by taking the average of the individual ranking given to each file by listeners. The *mean rank distance* for each of the objective test methods was then calculated by comparing the ranking given by that objective test method to the average subjective rank. This provides an estimate of how well each objective test method predicted relative subjective quality.

Figure 3 shows the mean rank distance obtained using this approach.

For set 1 the results for VQmon are similar to those of PAMS and PSQM. VQmon scored 0.27 compared to PSQM and PAMS' 0.58, i.e. slightly better.

For set 2 VQmon scored significantly better than PAMS or PSQM

For set 3 VQmon also scored significantly better than PAMS or PSQM

For set 4 VQmon obtained a worse score than PAMS and PSQM.

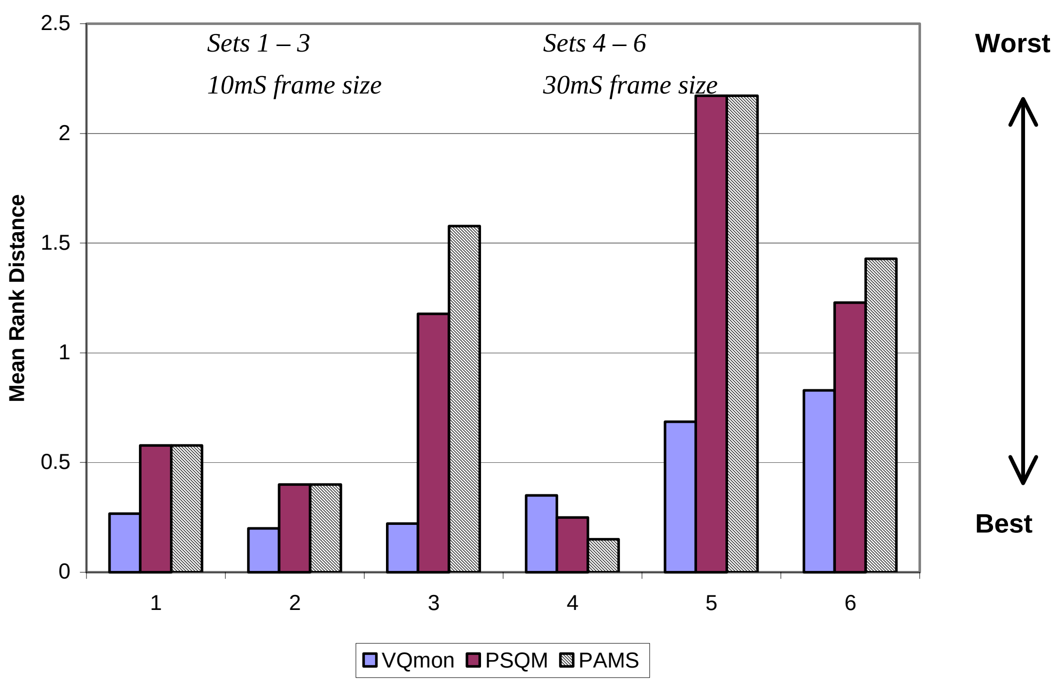


**Figure 3 Mean rank distance for VQmon, PSQM and PAMS**

For set 5 VQmon obtained a significantly better score than PAMS or PSQM. This was a very large difference (as was set 2) and appeared to be due to the impact of time varying effects, which VQmon does model and PAMS/PSQM do not.

For set 6 VQmon obtained a better score than PAMS or PSQM.

Figure 4 shows the correlation coefficients illustrating the same results. Vqmon obtained a correlation coefficient of greater than 0.9 for four of the five sets.

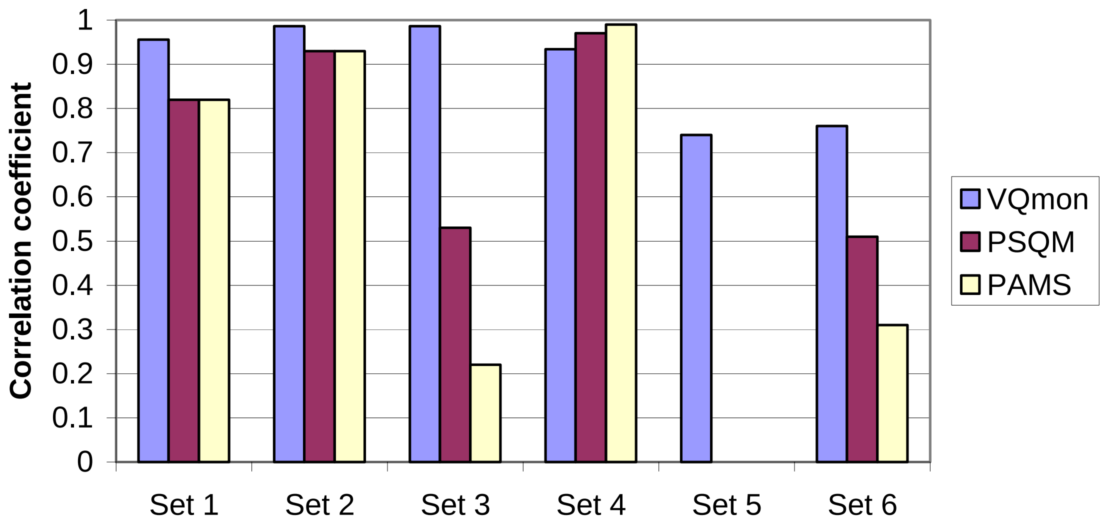


**Figure 4 Correlation Coefficients for Vqmon, PSQM and PAMS**

## *6. Conclusion*

VQmon outperformed PSQM and PAMS on 5 file sets out of 6 in predicting the relative subjective ranking of impaired voice files. The probable reason for this is that the files were impaired using time varying impairments – PSQM and PAMS assume that quality is relatively constant during a call whereas VQmon is able to model the effects of time varying impairments. PSQM and PAMS are generally used with short (6-10 second) audio files – although they are used in VoIP network testing with longer files they were not intended for this purpose.

In one case (set 4) VQmon performed worse than PSQM and PAMS. This may be due to some packet loss bursts occurring during repetitive sound patterns (e.g. aaaah) in which case packet loss concealment was extremely effective.

Under the test conditions used VQmon did provide a more accurate prediction of subjective voice quality in the presence of time varying network impairments than PSQM and PAMS.