# A systematic study of PESQ's behavior

Martín Varela and Ian Marsh and Björn Grönvall
Swedish Institute of Computer Science (SICS)
Kista, Sweden
{mvarela,ianm,bg}@sics.se

February 27, 2006

PESQ [1], the ITU-T's Perceptual Evaluation of Speech Quality is among the most widely used objective assessment tools in telecommunications and IP networks. Several commercial offerings incorporate it as a central component for voice over IP quality assessment. In terms of accuracy, i.e. correlation with subjective assessments, it has an advantage over all the other purely objective quality metrics. In this paper we take a systematic, black–box approach to analyzing the performance of PESQ.

We consider G.711 streams with and without packet loss concealment (PLC). We present results on how PESQ determines the speech coding quality under various network conditions found both in wired and wireless networks. We focus on the influence that the packet loss process found in the Internet has on PESQ results. We use both uniform and Gilbert loss models, covering a very wide range of possible network conditions which can occur both in wired and wireless networks. We also show results on language and gender dependencies.

We study how the loss rate, and the temporal distribution of losses affect PESQ results. Figure 1, for instance, shows median PESQ scores over a big loss space. Each of the points of the surface corresponds to the median of 200 PESQ assessments. These results, combined with validation with subjective assessment, allow us to have a detailed view of PESQ's performance and reliability. It has been noted [3, 4] that the PESQ performance is degraded when used for VoIP, especially when losses are bursty. We systematically study the consistency of the results over a large loss space. We study how slight variations in the way losses occur impact PESQ results, and how this varies for different streams.

We have already performed over half a million PESQ assessments in several experiments covering different scenarios, and we are currently preparing two more objective assessment campaigns, plus a subjective one, to be analyzed and included in the full version of this paper. We will describe the variability of the PESQ scores with respect to the parameters mentioned above, and will compare them to subjective scores in order to determine their correlation with the latter. We will also present a comparison of the PESQ results obtained with those of

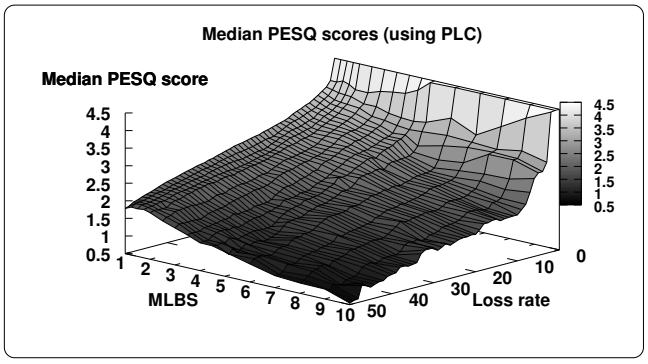


Figure 1: Preliminary results: median PESQ scores over a wide range of loss rates and mean loss burst sizes, when packet loss concealment (PLC) is being used. These results are based on the PESQ scores of 163200 degraded samples.

the ITU's P.563 [2] single–sided quality assessment algorithm.

We anticipate the results of our full paper to be useful in a number of different ways. The first is the possibility of developing a single–sided quality metric that mimics PESQ's performance closely. The fact that signal–independent one–sided metrics such as PSQA [5] provide very good correlations with subjective scores indicates that this should be feasible, and it would be useful for cases in which the cost of subjective tests is undesirable. We perform a statistical analysis of the stability of the results obtained, in order to determine their suitability for the intended applications.

# References

[1] ITU-T Recommendation P.862. Perceptual Evaluation of Speech Quality (Pesq), an Objective Method for End-To-End Speech Quality Assessment of Narrowband Telephone Networks and Speech Codecs, 2001.

[2] ITU-T Recommendation P.563. Single ended method for objective speech quality assessment in narrow-band telephony applications, 2004.

[3] S. Pennock. Accuracy of the perceptual evaluation of speech quality (PESQ) algorithm. In *Measurement of Speech and Audio Quality in Networks Line Workshop, MESAQIN '02*, January 2002.

[4] Psytechnics Ltd. PESQ: an Introduction. `http://www.psytechnics.com`, September 2001.

[5] G. Rubino, M. Varela, and S. Mohamed. Performance evaluation of real-time speech through a packet network: a random neural networks-based approach. *Performance Evaluation*, 57(2):141–162, May 2004.