Intrusive Speech Quality Measurements in Czech Environment

Jan Holub, Radislav Šmíd, Jan Horyna

K338 – Department of Measurement, Faculty of electrical Engineering,
Czech Technical University, Technicka 2, 16627 Prague 6, Czech Republic,
Phone: +420 2 2435 2131, Fax: +420 2 3333 9929, e-mail: holubjan@feld.cvut.cz

Summary: The comparison between listening test results and voice quality measurement algoritm PAMS 3.0 is given for the set of Czech speech samples affected by several kinds of impairment.

Keywords: speech quality, PAMS

1. INTRODUCTION

Objective measurement of speech quality as described in P.86x of ITU-T or in other standards like PAMS and TOSQA belongs to the most actual topics in contemporary telecom issues. It enables to compare and benchmark different transmission technologies or codecs from the point of view that is rather close to that of end-user. Moreover, some transmission technologies and their chains that appear in today's converging telecom networks are difficult to compare by means of any other method (e.g. circuit-based and packet-based transmissions).

The basic scheme is almost identical for all the above mentioned standards: special dedicated test call is established and suitable speech sample is transmitted between calling and called station over the tested network. Received version of speech sample is digitally recorded and compared with original speech sample (after level and time correlation / adjustments). This comparison is performed by suitable algorithm that should precisely comprise all known features of human's ear and brain related to speech listening.

2. TASK DEFINITION

Our goal was to verify PAMS 3.0 (Perceptual Analysis Measurement System) applicability for Czech environment. PAMS algorithm has been used as a representative of a new generation voice transmission quality measurement standards (capable to correctly analyse packed-based transmissions affected by packed jitter delay).

3. SPEECH SAMPLES PREPARATION

28 speech samples have been used both for listening tests and for PAMS processing. The speech samples have been structured as follows:

1. Noisy transmissions (12 samples)
2. Jittered transmissions (8 samples)
3. Clipped transmissions (8 samples)

Speech samples have been prepared artificially from studio quality recordings. All of them were in Czech language, using two male and two female speakers.
Recordings have been distorted by adding well-defined amount of noise (corresponding approximately to Mean Opinion Scores 2,3 and 4), simulated packet jitter (having mean value of delay 30 and 70 ms, uniformly distributed jitter) and front-end clipping (30ms after each silent period longer than 100 and 300 ms). A special toolbox (for Matlab 6.0) has been developed to enable distortion of speech samples. Moreover, this toolbox can introduce amplitude clipping, multiple echo and simple frequency filtering, see Fig. 1.

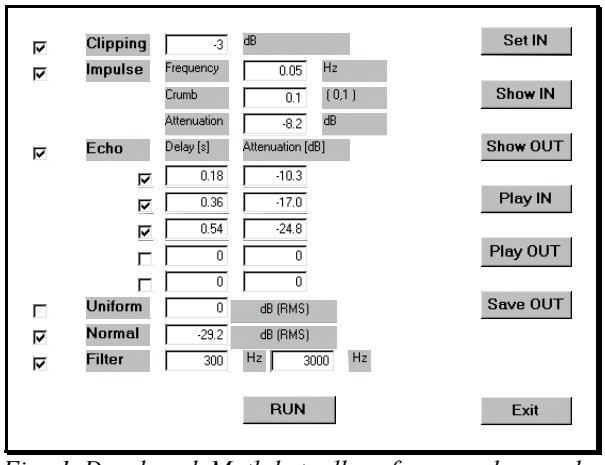

Fig. 1 Developed Matlab toolbox for speech sample distortion. Clipping, impulse noise, multiple echoes, uniform and/or normal noise and BP filtering can be introduced to the input speech sample at well-defined levels

4. LISTENING TESTS

Listening tests have been performed using approximately 80 listeners. Those were not trained in detail, they were just asked to evaluate speech samples using the scoring 1 to 5 (Mean Opinion Score, MOS) and the verbal description of MOS levels in accordance with P.830 has been given to them. First, 4 original (non-distorted) speech samples were played and then 28 distorted versions in random order were presented without repeating the originals (Absolute Category Rating). Wide-band aparature without any additional filtering has been used (implicit LP filter with cut-off frequency 4 kHz has been applied since 8-kSa/s speech samples have been used). The authors are aware that no IRS filtering may be considered as important simplification. Nevertheless, the final results (see conclusions) are not affected.

5. PAMS RESULTS

Perceptual Analysis Measurement System (PAMS) version 3.0 has been used. It was chosen because neither its 3.1 implementation neither final version of Perceptual Evaluation of Speech Quality (PESQ) algorithm has been available in the time of experiments.

Detailed measured results are given in the Table 1. Also PSQM values according to P.861 for some samples are given (most of jittered samples was not correlated by PSQM correctly).

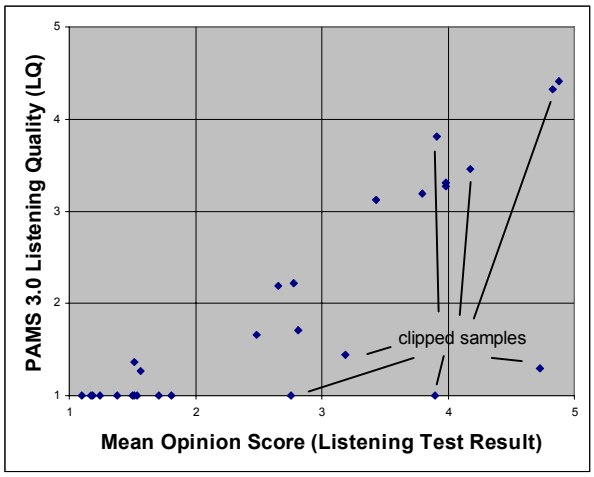

Fig. 2 Comparison between Listening Quality (LQ) and Mean Opinion Score (MOS) for 28 used speech samples

Tab. 1 Detailed results of PAMS application. Speech sample library structure: 1 to 4 – original non-distorted samples, 5-16 noisy samples, 17-24 jittered samples, 25-32 clipped samples.

Speech Sample Number	Listening Test Results (MOS)	PAMS 3.0 (LQ)	PAMS 3.0 (LE)	PSQM
1	5.00	5	5	0
2	5.00	5	5	0
3	5.00	5	5	0
4	5.00	5	5	0
5	3.43	3.12	4.04	3.34
6	2.48	1.66	2.76	6.3
7	1.53	1	1.42	9.1
8	3.79	3.19	4.06	3.18
9	2.81	1.71	2.79	6.1
10	1.52	1	1.33	8.88
11	3.98	3.31	4.08	3.23
12	2.78	2.22	3.44	6.58
13	1.50	1	1.44	9.4
14	3.98	3.27	4.06	3.69
15	2.66	2.19	3.33	7.2
16	1.38	1	1.33	10.13
17	1.57	1.27	2.24	no result
18	1.19	1	1.33	9.09
19	1.52	1.36	2.29	no result
20	1.17	1	1.48	no result
21	1.71	1	1.45	6.64
22	1.24	1	1.33	no result
23	1.81	1	1.41	no result
24	1.10	1	1.33	no result
25	4.72	1.29	1.33	0.53
26	4.88	4.41	3.92	0.13
27	3.90	1	1.33	0.87
28	4.17	3.46	2.97	0.25
29	2.76	1	1.33	1.02
30	4.83	4.32	3.7	0.08
31	3.19	1.44	1.33	0.53
32	3.91	3.81	3.39	0.16

PAMS provides two output parameters: Listening Quality (LQ) and Listening Effort (LE). The graphical comparison between LE and M.O.S. evaluated during listening tests is given in Fig. 2. Similarly, LE vs. MOS is depicted in Fig. 3. Since important differences between listening test results and PAMS outputs have been found for temporarily clipped samples, statistical analysis (cross-corellation) has been performed both for complete set of measurements and for restricted set of measurements (skipped temporarily clipped samples). The results are given in the Tab. 2.

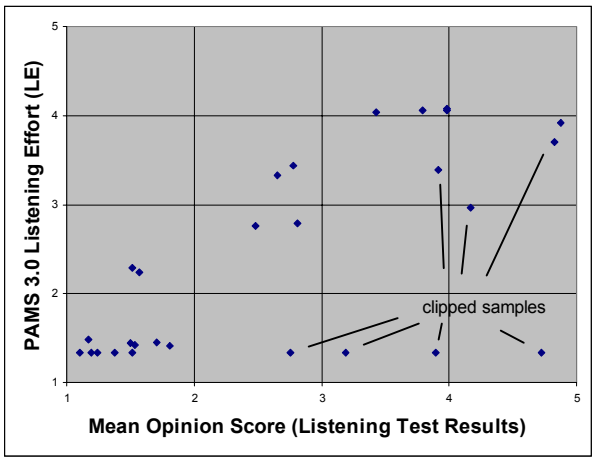

Fig. 3 Comparison between Listening Effort (LE) and Mean Opinion Score (MOS) for 28 used speech samples

Tab.2. Correlation between listening test results and PAMS 3.0 outputs (LQ, LE) for restricted (with skipped clipped samplest) and for complete set of speech samples

noise+jitter LQ vs MOS	0.967
noise+jitter LE vs MOS	0.954
noise+jitter+clipped LQ vs MOS	0.790
noise+jitter+clipped LE vs MOS	0.650

6. CONCLUSIONS

It was confirmed that the results using Czech speech samples are comparable with any other European language. It was also confirmed that PSQM as given by ITU-T P.861 can not be applied for packet transmissions. As follows from the results, PAMS 3.0 is well applicable for speech quality measurements in the case that the transmission distortion is mainly caused by noise or packet jitter (correlation 0.97). However, the results differ for temporarily clipped speech samples. The same set of experiments for new speech quality evaluating algorithms (e.g. PESQ) is currently being performed.

ACKNOWLEDGEMENT

This project is supported by Grant Agency of Czech Republic under the number GACR 102/01/1355, "Advanced Measurements in Mobile Networks". The final goal of this project is to build an open-architecture voice quality measurement system for PSTN, GSM and UMTS networks allowing easy implementation of any new standard. The current group of standards that is being implemented includes ETR250 by ETSI, P.861, P.862 by ITU-T and other selected recommendations like German TOSQA or British PAMS. The goal of the measuring system is not only to have the standards implemented but also to determine, assure and compare the measurement accuracy and uncertainty of measured parameters and results for each standard.

REFERENCES

[1] ITU-T P.830, 1996: Subjective performance assessment of telephone-band and wideband digital codecs

[2] ITU-T P.861, 1996: Objective Quality Measurement of Telephone-band (300-3400 Hz) speech codecs

[3] ITU-T P.862, 2001: Perceptual Evaluation of Speech Quality (PESQ), an Objective Method for End-to-end Speech Qauality Assessment of Narrow-band Telephone Networks and Speech Codecs

[4] Overview of PAMS, Agilent Technologies, 2000

[5] ITU Supplement 3 to P-series recommendations

[6] ETSI Technical Report ETR 250 - E-model, ETSI, June 1996

[7] Holub, J: *Intrusive Measurements in GSM Networks*, Sdelovaci technika 10/1999, October 1999, pp. 14-15