Impact of Saturation on Speech Quality in VoIP

Z. Becvar, J. Zelenka and M. Brada
Department of Telecommunication Engineering
Czech Technical University
Technicka 2, Prague, 166 27, Czech Republic
Phone: (+420) 2-2435 5994 Fax: (+420) 2-3333 9810 E-mail: becvaz1@fel.cvut.cz

Keywords: Speech Quality, VoIP, Saturation

Abstract – VoIP technology becomes more and more profitable and many people prefer this kind of connection instead of classic TDMA telecommunication network. Very helpful is a fact that with voice services becomes another related services like a video, a text or even a file transmission. These services are integrated both into simple software VoIP clients and hardware VoIP telephones. The usage of this type of the communication is very cost effective especially in the international connections. One single disadvantage of this technology is dependence on an internet connection quality. For example a delay of a mobile internet connection causes a huge decrease of the quality of service which is the most used parameter in VoIP networks. We can distinguish many factors affecting a quality of speech, such as packet delay, jitter, noises, etc. This paper deals with an impact of the saturation of a speech signal to the subjective voice quality.

1. INTRODUCTION

The Quality of Service (QoS) is the most observed parameter by telecommunication operators hence for vendors who have to deal with. The QoS parameter is very closely connected with user satisfaction about a final speech quality. The user satisfaction is expressed with a subjective listening score that is a result of the subjective test. Further, for quality assessment is used many types of objective tests such as PESQ (Perceptual Evaluation of Speech Quality) [1] and 3SQM (Speech Quality Measurement) [2]. Speech quality is measured in MOS (Mean Opinion Score) units [3]. Both subjective and objective methods have outputs in specific MOS units but it is possible to convert them so it is easy to compare these results and make some conclusions [4].

The quality of speech can be affected by many factors like packet loss, packet delay, jitter, echo, noise, harmonic and inharmonic distortion, etc. [5]. For example, the impact of an additional noise in speech utterances was discussed in [6]. In this paper will be described one of the inharmonic distortion called speech saturation. The sensitiveness of this distortion to human hearing system will be shown. Results of an investigation are obtained by subjective listening test and objective testing methods.

The saturation of the speech signal occurs generally in systems where analog to digital converter is used. This happens in classic telephone TDM networks, VoIP networks and many other systems. The important thing is that the saturation causes inconsiderable decrease of the quality which has to be measured and it is required to know how much this degradation influences the final speech quality.

The rest of the paper is organized subsequently. Next chapter describes the method for degradation of utterances by the saturation. The third section is focused on the test composition where both subjective and objective testing conditions and procedures are mentioned. Further section provides the results of the subjective test and compares them with results obtained by the objective tests. Last section presents our conclusions and future work plans.

2. UTTERANCES GENERATING

To test the effect of the saturation to the perception quality we take high quality studio recordings which were processed with a well defined distortion.

An example of the original (non-degradated) signal can be seen in Fig.1.

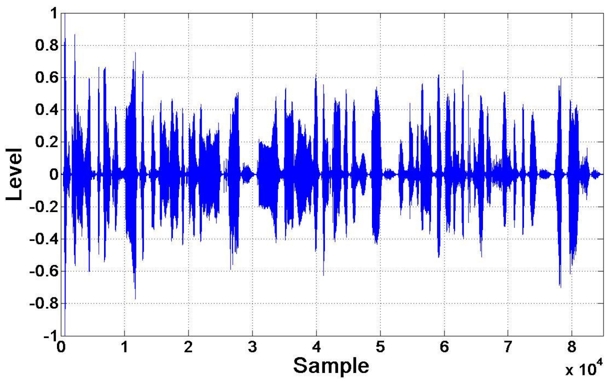

Fig. 1. Original speech signal

We have damaged the utterances with fourteen levels of the saturation distortion. The saturation level k was defined as a ratio between the maximum level of distorted signal y and the maximum level of the original speech signal x (1).

$$k = \frac{M(|y_i|)}{M(|x_i|)} \tag{1}$$

The levels of saturation vary from 1 to the 0.00005 level with step respecting the logarithmic scale.

In the next two figures is illustrated the process of saturation on the speech signal. At first, the speech signal was normalized, then degraded to the selected saturation level (Fig.2) and then normalized again (Fig.3). This should ensure that all the testing signals would have the same loudness level.

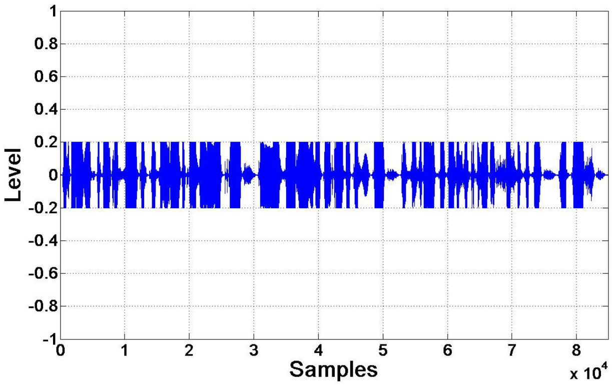

Fig. 2. Saturated speech signal – saturation level = 0.2

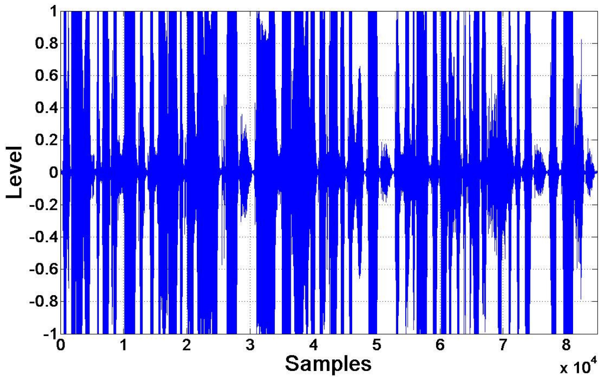

Fig. 3. Degraded and normalized speech signal.

For each level of the saturation we prepared six different utterances to improve the reliability of the measurement. In summary, this represents the total of 84 speech signals that were played to the listeners in random but for all subjects the same order.

3. TEST COMPOSITION

As mentioned above, two types of the tests were proceeded: the subjective test and the objective tests. The results were then compared. We evaluated the results in the MOS LQS (Listening Quality Subjective) score [7], which maps the subjective listening quality to the five point scale. The value "five" represents superior quality; the "one" denotes the inaudible and displeasing sensation.

The subjective test was proceeded according to the ITU-T P.800 [8] a ITU-T P.830 [9] recommendations, comprehending many factors.

Tested utterances were original digital studio recordings, mostly separated from dialogs of two people with neutral purport. Their content should not draw any emotional response at the listeners. All recordings were in Czech language and all speakers were natural born Czech without speech aberrations.

The length of utterances was selected to correspond to the requirements of both types of tests. The average length was between 8 an 12 seconds. Tested signals were 16-bit linear PCM (Pulse Code Modulation) sampled with 8 kHz sample rate (downsampled from the studio quality of 48 kHz).

Testing environment was represented by a quiet room, free of all accidental sounds or any other disturbing elements. Every listener passed a short training about the test conditions at the beginning of testing procedure. The test length was estimated to the approximately 25 minutes to prevent the testing subjects from being bored and then loosing their concentration.

The set of listeners was acquired from the university students and employees, chosen only those of them who did not pass similar test in the last six months (according to [8]).

The test itself consists of testing and measured part. The testing part was placed first to enable the listeners to adapt to the various levels of distortion. It helped them to create imaginary scale of distortion. The measured part followed immediately. The test itself was carried out using ourselves developed tool called "Tester". The testing tool was developed in Java.

The objective part of the test was performed using the standardized ITU-T P.563 (3SQM) [1] and ITU-T P.862 (PESQ) [2] methods. In these tests the same utterances as in the subjective tests were used.

The 3SQM method is a non–intrusive method evaluating the quality from the end point of the telecommunication chain. It does not need the referential signal and outputs a MOS LQS score. The PESQ method compares the original and distorted speech signal. It outputs a value in range from –0.5 to 4.5. The output must be adapted to the subjective MOS score according to the recommendation ITU-T P.862.1 [10] using the following formula (2),

$$y = 0.999 + \frac{4.999 - 0.999}{1 + e^{-1.4945 * x + 4.6607}} \qquad (2)$$

where x is the objective MOS PESQ score according to ITU-T P.862.1 and y is the matching MOS LQS value.

To determinate the conformity among the subjective and the objective tests results we used the Pearson's correlation coefficients (3) and residual errors (4).

$$r = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum (x_i - \bar{x})^2 \sum (y_i - \bar{y})^2}} \qquad (3)$$

$$e_i = x_i + y_i \qquad (4)$$

4. TESTS RESULTS

The results obtained by the objective methods (PESQ and 3SQM) and by the subjective tests are shown in Fig.4 (linear x axis) and Fig.5 (logarithmic x axis). The results are also summarized in Table 1.

Saturation level	MOS score		
	Subjective	PESQ P.862.1	3SQM P.563
0,00005	1,46	1,52	2,93
0,00010	1,50	1,38	2,49
0,00020	1,57	1,30	2,25
0,00050	1,69	1,35	1,61
0,00100	1,85	1,38	1,59
0,00200	1,82	1,39	1,68
0,00500	2,17	1,51	2,15
0,01000	2,52	1,54	2,92
0,02000	2,94	1,60	2,99
0,05000	3,33	1,75	3,14
0,10000	3,73	2,07	3,58
0,20000	4,15	2,83	3,79
0,50000	4,51	4,24	3,69
1,00000	4,82	4,55	3,41

Table 1. Subjective and objective MOS score for different saturation level

Fig.4 and Fig.5 show that the speech quality assessment acquired by the subjective tests has nearly logarithmic dependence on the saturation level and the impact of the saturation effect is insignificant for the saturation level higher than 0.2. If the speech is saturated to the higher levels, MOS score is increasing very slowly and MOS score is approximately 4.5 MOS. This behavior is caused by the distribution of the energy in the speech. The major part of the energy is allocated in the lower levels of signal amplitude. The saturation at the higher levels brings only very low energy decrease.

The maximum speech quality obtained by the subjective test is 4.8 (it corresponds to the speech without saturation; saturation level = 1). This value is very close to the theoretical maximal MOS score (5.0 MOS).

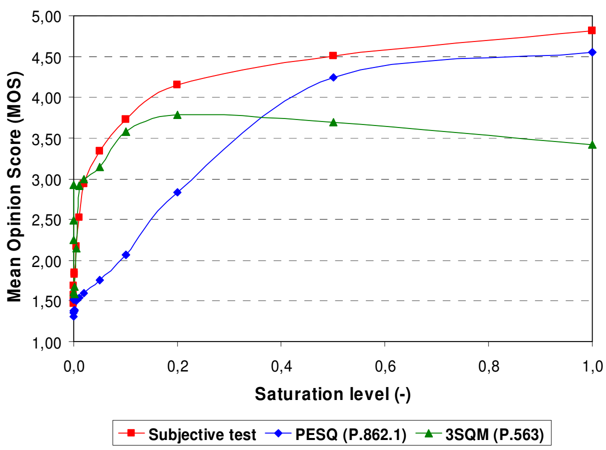

Fig. 4. Subjective and objective results of impact of saturation on speech quality

From Fig.5 can be observed that the results of subjective tests are nearly constant at the saturation level up to 0.002.

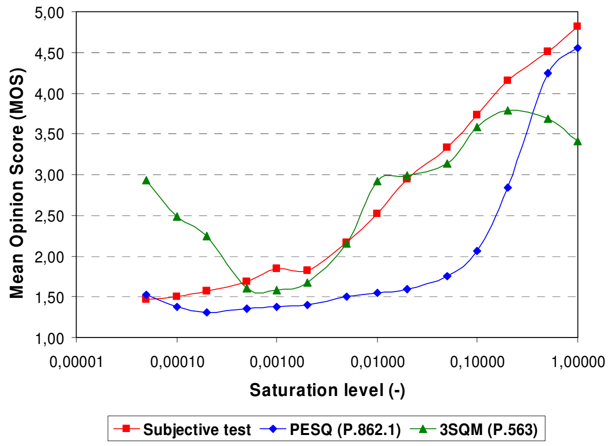

Fig. 5. Subjective and objective results of impact of saturation on speech quality in logarithmic scale

An impact of the saturation on the results obtained by the objective method PESQ is not logarithmic and the behavior is different from the subjective tests. PESQ results are nearly constant at the lower saturation levels (up to 0.05) and it is followed by the rapid increase at the higher saturation levels. PESQ results correspond to subjective results only at lower saturation levels (up to 0.002) and at higher saturation levels (grater then 0.5).

The results obtained by 3SQM are not much uniform. It points to the unsuitability of 3SQM method for measuring of the quality of speech affected by the saturation. Opposite to PESQ, 3SQM results correspond to the subjective tests in the middle part of figure (from 0.005 to 0.2).

Figure obtained by PESQ results is not logarithmic; the speech quality is increasing more slowly then in a case of the subjective results.

Pearson's correlation coefficients and average residual errors were used for the comparison of the results obtained by the subjective tests and the objective methods.

The Pearson's correlation coefficients between PESQ and subjective test and between 3SQM and subjective tests are presented in Table 2. The correlation coefficients show not excellent, but acceptable conformity of both objective methods and subjective test. Pearson's correlation coefficient as a comparison of PESQ and subjective tests is 0.88; the coefficient for 3SQM is 0.82.

Results acquired by 3SQM algorithm shows small correlation between subjective tests and 3SQM (Pearson's correlation coefficient is 0.82). Average residual error is lower then in PESQ case, but it is still too high (0.5).

The parameters presented in Tables 2 and Table 3 shows the differences between subjective and objective methods.

Pearson's correlation coefficients	
PESQ	0,88
3SQM	0,82

Table 2. Comparison of results obtained by subjective tests and objective methods by Pearson's correlation coefficients

Good conformity, but very different behavior pictured in the Fig.4 and Fig.5 are given by the not taking the direct component into account. This difference is expressed by residual error. Average residual errors for the same cases are summarized in Table 3. The average residual errors show wide difference among subjective and objective methods. The average residual error for 3SQM was smaller (0.5 MOS) opposite to PESQ (0.7 MOS).

Average residual errors (MOS)	
PESQ	0,70
3SQM	0,50

Table 3. Average residual errors between subjective tests and objective methods

The average residual errors and Pearson's correlation coefficients show very small conformity of both methods with the subjective results. It results to the poor capability of both objective methods to detect and interpret the distortion of the speech caused by the saturation effects.

5. CONCLUSIONS AND FUTURE WORK

The subjective tests for speeches affected by the saturation were done. The behavior of the subjective speech quality assessment is nearly logarithmic. The subjective test results show the major impact of the saturation on the speech quality for the saturation level set to approximately 20 percent of a maximum value. It is especially since the amplitude of the most of speech samples is lower the 20 percent of maximum amplitude. The samples with amplitude under saturation level are not influenced by the saturation effect so the energy eliminated from the speech signal by the saturation is very small in cases with higher saturation level. The small decreases in the energy are not distinguishable by the listeners.

Besides the subjective tests, the objective assessment of speech quality was done too. As the objective algorithms were used PESQ (recalculated according to ITU-T P.862.1) and 3SQM (according to ITU- P.563).

Results obtained by PESQ and 3SQM are not logarithmic. In the PESQ case the speech quality is increasing more slowly then subjective results. PESQ results correspond to subjective result only for lower saturation levels (up to 0.002) and for higher saturation levels (grater then 0.5). Pearson's correlation coefficient is 0.88, but the average residual error is 0.7.

Results acquired by 3SQM algorithm shows small correlation between subjective tests and 3SQM (Pearson's correlation coefficient is 0.82). Average residual error is lower then in PESQ case, but it is still too high (0.5). Opposite to PESQ, the 3SQM results correspond to subjective tests in the middle part of figure (from 0.005 to 0.2). The comparison of subjective and objective results shows poor ability of PESQ and 3SQM to detect and interpret the saturation effect.

In the future, we will investigate an impact of the other disturbing effects (e.g. harmonic distortion) on the speech quality in the VoIP networks.

ACKNOWLEDGEMENT

This work has been supported by the grant "Advance optimization of the communication systems designing by neural networks" No. GA 102/07/1503 funded by Grant Agency of Czech Republic.

REFERENCES

[1] ITU-T Recommendation P.563, Single Ended Method for Objective Speech Quality Assessment in Narrow-band Telephony Applications, 2004.

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

[3] ITU-T Recommendation G.107, The E-model, a computational model for use in transmission planning, 2003

[4] Z.Becvar, J. Zelenka, M. Brada, T. Valenta, "Comparison of Subjective and Objective Speech Quality Testing Methods in the VoIP Networks," *IWSSIP 2006*. Budapest, 2006.

[5] A. Raake, *Speech Quality of VOIP, Assessment and Prediction*, Wiley, Chichester, England, 2006.

[6] Z. Becvar, L. Novak, J. Zelenka, M. Brada, P. Slepicka, "Impact of Additional Noise on Subjective and Objective Quality Assessment in VoIP," *IEEE International Workshop on Multimedia Signal Processing*. Chania, Greece, 2007.

[7] ITU-T Recommendation P.800.1, Mean Opinion Score (MOS) terminology, 2003.

[8] ITU-T Recommendation P.800, Methods for Subjective Determination of Transmission Quality, 1996.

[9] ITU-T Recommendation P.830, Subjective Performance Assessment of Telephone-Band and Wideband Digital Codecs, 1996.

[10] ITU-T Recommendation P.862.1, Mapping function for transforming P.862 raw result scores to MOS-LQO, 2003.