HISTOGRAM BASED APPROACH FOR NON-INTRUSIVE SPEECH QUALITY MEASUREMENT IN NETWORKS

Neintrusivní měření kvality hlasových přenosů pomocí histogramů

Jan Křenek *, Jan Holub *

Abstract

This article describes the usage of histograms for speech audio quality assessment in GSM and other networks.

Abstrakt

Článek popisuje způsob použití histogramů pro neintrusivní hodnocení kvality hlasových přenosů v GSM a jiných sítích.

Introduction

The networks, such as GSM, UMTS, Tetra or Local Area Network (with proper VoIP software) represent technology used for speech transmission. Transmitted voice signal experiences a set of distortions on its way through the communication channel from a receiver to transmitter. Each distortion type (e.g. attenuation, noise, delay, echo, packet loss, clipping, jitter etc.) could cause considerable voice quality degradation. Along with the need for quality improvement goes the need for a methodology for its measurement. Two different methods are described below followed up with a deeper focus on histogram based non-intrusive voice quality measurement.

Voice quality measurement

There are two different methods for voice quality measurement: Intrusive and Non-intrusive.

Intrusive method

Based on a comparison of both, original and transmitted signal using proper algorithm, e.g. ITU-T P862 (PESQ). The usage of intrusive method gives more accurate results in comparison with the quality assessed by the average listener that is acquired from listening tests. High cost and time consumption forms a space for the second, Non-intrusive method.

Non-intrusive method

The non-intrusive method estimates the quality just from the transmitted sample! It is easy to see the difficulty of such an algorithm to give reliable results. The main advantage of the non-intrusive method is cost efficiency (“unlimited” number of speech samples can be

* Ing. Jan Krenék, Department of Measurement, CTU in Prague, FEL, Technická 2, 162 27 Prague
tel.: +420 2 2435 2187, fax.: +420 2 3119 929, e-mail: krenek@fel.cvut.cz
* Doc. Ing. Jan Holub, Department of Measurement, CTU in Prague, FEL, Technická 2, 162 27 Prague
tel.: +420 2 2435 2131, fax.: +420 2 3119 929, e-mail: holubjan@fel.cvut.cz
assessed for the quality for the total accuracy improvement), the measurement is conducted within real network data and states (as the call establishment is not necessary unlike in the intrusive method). Nevertheless, some distortion types, like the harmonic distortion of some codec’s (e.g. ADPCM) cannot be detected, due to the lack of the original sample. It is therefore recommended to use a combination of both, intrusive and non-intrusive methods.

**MOS scale**

When discussing quality of voice transmission, clarification of the term „QUALITY“ is expedient. Surprisingly, the term is not defined unambiguously. Considering the signal transmission, the quality is treated as the level of similarity of both the transmitted and received sample. In a view of human perception, the quality indicators could be: clarity, delay, noise, level, drop-outs, etc.

For the purpose of the quality assessment, the MOS (Mean Opinion Score) scale is widely used, which corresponds to the meaning of the average listener. It is, therefore, a subjective assessment.

<table>
<thead>
<tr>
<th>MOS scale</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
</tr>
<tr>
<td>4</td>
</tr>
<tr>
<td>3</td>
</tr>
<tr>
<td>2</td>
</tr>
<tr>
<td>1</td>
</tr>
</tbody>
</table>

Tab. 1 MOS scale

Unlike the standard custom, the best “excellent” grade is a 5, whereas the worst “bad” grade is a 1.

**Histogram based non-intrusive algorithm for voice transmission quality measurement**

For the algorithm development, a proper library of speech samples is essential. Such a library was at our disposal. It consists of four original speech samples made by two male and female professional speakers. These samples were distorted by clipping (simulating the distortion caused by a Voice Activity Detector used in mobile phones), jitter, noise and filtering. In total, the library contains 40 samples (4 original and 36 distorted). Quality of those distorted samples was assessed during listening tests, which are considered as an etalon of value that is basically desirable by every quality appraising algorithm to achieve.

The first step is to obtain histograms of all samples. Such an algorithm can work as the one in Fig. 1. The examined sample is at first normalized in amplitude and re-sampled to an 8kHz sample rate. After that, the whole sample is divided into 16ms packets. Using FFT, amplitude spectrum is computed for every packet. Depending on the type of each packet (active speech-voice or pause-noise) determined by a Voice Activity Detector, the sum of the amplitude spectra related to the packet type is computed. Sum of voice packets – “HIST_voice”, sum of noise packets – “HIST_noise” and sum of both – “HIST” were examined for invention of parameters with relation to quality.
If speech sample rate ≠ 8kHz, resample

Voice Activity Detection

LOOP START
Until end of sample

Take first or next packet of 16ms from the sample

HIST = HIST + \text{ABS}(\text{FFT}(\text{packet} \ast \text{hann}))

HIST_voice = HIST_voice + \text{ABS}(\text{FFT}(\text{packet} \ast \text{hann}))

HIST_noise = HIST_noise + \text{ABS}(\text{FFT}(\text{packet} \ast \text{hann}))

LOOP END

Sample noise power computation

plot HIST_voice
plot HIST_noise

Fig. 1 *Simplified flowchart for Histograms creation*

The simplified flowchart in Fig. 1 was more complex in real in order to provide other histograms creation. The expanded flowchart gives 4 more histograms – “HIST_voice_delta”, “HIST_voice_delta_delta”, “HIST_noise_delta” and “HIST_voice_delta_delta”. Delta (Δ) represents that difference between the amplitude spectrum of the actual packet and the next packet (with 50% overlap) was used for the histogram computation. In the double delta variant, there is a difference from differences made in the previous step, also with a 50% overlap. The reason for making such differences with a 50% overlap is that they can increase resolution and highlight possible parameters.

Fig. 2 *Histogram of pure sample of MOS=5*
There are frequency bins on the horizontal axis (with increasing position towards the right, the frequency increases). The vertical axis is the axis of amplitudes. Counts are represented by the color scale, from white to black. The presented histograms are actually three dimensional figures presented in two dimensions.

In Figures 2 – 4, sample histograms can be seen. The first histogram in Fig.2 is a histogram of the sample with the best quality (MOS=5). In Fig. 3 – the histogram of the
sample is just a bit distorted by clipping with MOS=4.88. Fig. 4 shows a histogram significantly distorted by clipping (MOS=2.76).

It can be seen even at first glance that the number of counts in first row (from bottom) of histograms increases with decreasing quality. On further review, it can be seen that the ratio between for example five frequency bins from the middle of the second row and the middle five from the first row is bigger than 1 for relatively high quality. This ratio decreases with the quality. The procedure of searching for parameters applied on all speech samples and all (=seven) histograms leads to spotting many desirable parameters. Such parameters were spotted and could be used as an input for a neural network in order to find out which of them are suitable to use and how each should affect the result of the algorithm. However, high computational requirements for all histogram creation made usage of such an algorithm unsuitable for real-time quality assessments. The primary advantage of the non-intrusive method is its speed. Further investigations to speed up the computations are in progress at present. Consideration is being given whether the usage of a Voice Activity Detector and the computation of all seven histograms for all samples are necessary.

Conclusion

In this paper, a novel approach for voice transmission quality assessment in a non-intrusive way was presented.

Starting with brief introduction of VTQoS (Voice Transmission Quality of Service) measurement, principle of proposed histogram based algorithm was described. Even though this approach has shown its functionality, further research is necessary for the performance improvement.

References

