Low Bit-rate Networks – A Challenge for Intrusive Speech Transmission Quality Measurements

J. Holub[1], M. D. Street[2]

[1]FEE CTU Prague, Dept. of Measurement K338, Technická 2, 166 27 Prague 6, Czech Republic

[2]NATO Consultation, Command and Control Agency, PO Box 174, 2501 CD The Hague, The Netherlands

holubjan@feld.cvut.cz

michael.street@nc3a.nato.int

Abstract

Some elementary questions and remarks concerning with inefficiencies of currently used speech quality measurement algorithms when applied to low bit-rate transmissions are shown. Basic hints and recommendations to improve their performance are also given. This kick-off article does not offer any significant measured results, however, it tries to give a direction for future research.

Keywords: Speech quality measurements, MOS, low bit-rate networks

1 Introduction

Intrusive speech quality measurements are defined for telecommunication networks (ETSI ETR 250, ITU-T P.861, P.862). The basic principle is as follows (see also Fig.1):

A speech sample (SS) fulfilling a set of conditions according to P.800 (among other noise- and distortion-free) is transmitted via a communication channel between sending and receiving stations. The received version of the SS is acquired and saved into a file. The original and received SS are then compared using an algorithm modelling the established psychoacoustic aspects of human hearing / listening. The final parameter is usually called MOS (Mean Opinion Score). The MOS scale covers interval from 1 (worst quality) to 5 (best quality). Good ISDN lines usually achieve 4.2-4.5, normal GSM transmission can achieve 3.2-3.9 for Full Rate (FR codecs). The algorithm is fitted internally to match the results of listening tests according to P.80.

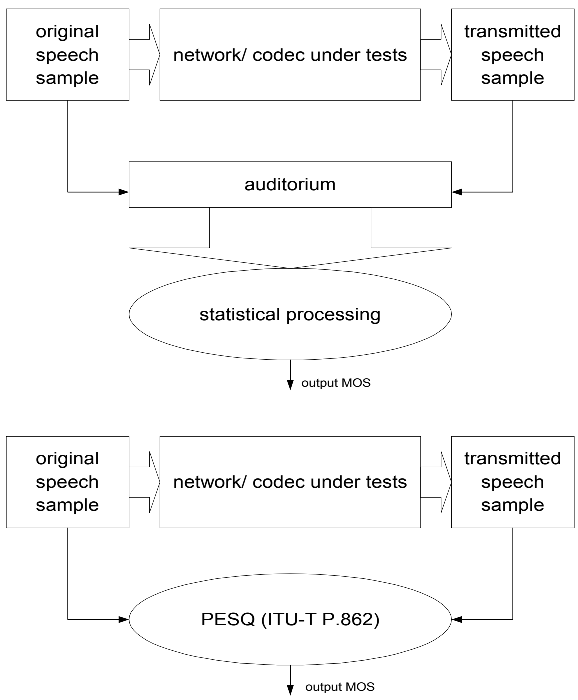

Fig. 1. Listening tests (top) and automatic measurement of speech quality (bottom)

The above described procedure allows correct evaluation of the following transmission impairments:

- Noise (arising during transmission)
- Extraneous signals in output sample (clicks, cracks, tones)
- Amplitude clipping
- Signal dropouts / temporal clipping
- Harmonic and non-harmonic speech distortion
- Multiple echo (even orders only - "listener echo")

The newest (Q1/2001) standard ITU-T P.862 PESQ (Perceptual Evaluation of Speech Quality) is capable of dealing correctly even with packet-oriented data streams affected by delay jitter. This standard is successfully applicable to both fixed (PSTN, ISDN) and mobile networks (NMT, GSM) in the following problem areas:

- codec comparison / evaluation / assessment
- transmission quality check (e.g. interconnect in GSM)
- automatic end-to-end functionality check (e.g. roaming tests in GSM networks)
- improvement verification after network modification / reconfiguration / upgrade
- verification of influence of different network loads (load tests) and interference levels to speech quality

There are obvious advantages of speech quality end-to-end testing in comparison with common parameter measurements (Bit Error Rate BER, Frame Erasure Ratio FER, Received Radio Signal Level RxLev) since it is the objective measurement very close to final end-user opinion. It can also serve as an automatic final functionality check (even excellent values of BER, FER and RxLev in parallel do not necessarily mean that the communication channel is really established and ready to use!).

As mentioned in [3], speech quality correlates well with speech intelligibility; this means that speech intelligibility tests can also, up to certain level of confidence, be substituted by speech quality measurements. The experimental level of confidence has been found as 0,92 for noisy samples and 0,73 for all the tested samples.

2 Motivation

In the environment of low bit-rate networks (LBRN) as used in satellite and military communications (1200 and 2400 bits per second) the above mentioned automated and repeatable approaches are not used at all yet. The reason is that there are pending questions and problems that do not allow direct application of those procedures to special communications in the low bit-rate environment:

- Low bit rate coders are delivering generally lower speech quality where the cognitive power of human listener considerably influences the speech perception.
- Environmental noise at the sending side is significantly higher.

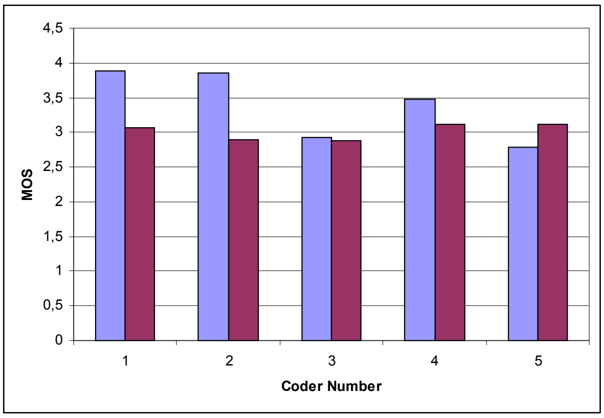

Fig. I – Comparison between listening test results (left/light/blue bars) and PESQ score (right/dark/red bars)

The comparison between listening test results and PESQ score is given in the Fig.1 for a set of 5 coders running at 1200 or 2400 bits per second. It is obvious that while human opinions vary significantly between 2.8 and 3.9, the PESQ scoring is almost stable for all the coders and remains on the MOS value of cca 3. The monotonicity of the results does not fit either.

3 Necessary Algorithm Changes

3.1 No Magic Inside

We do not consider the cognitive power of a human listener as a blocking problem for algorithm design. The common psychoacoustic aspects modelled by the speech quality measurement algorithms (human ear sensitivity, frequency and temporal masking etc.) are still valid.

3.2 Frequency Overlapping

The contemporary speech quality measurement algorithms such as PESQ [1,2] compare short time spectrum module content, having FFT bins grouped into so called Bark bands to simulate frequency resolution of the human ear.

In voice coders for LBRN, the frequency content of the signal is not maintained so precisely. Therefore a set of peaks of short time spectrum may be transformed into another set of peaks with slightly different positions of their maximas. If too precise frequency comparison is used or if this comparison is performed on non-overlapping bands (as in the case of PESQ and similar algorithms), it may happen that energy is missing in one band but it appears in the neighbouring band. Differences in both bands are then interpreted as faults and contribute to false indication of output degradation.

The solution is to introduce a frequency overlapping approach. This can be achieved either by real overlapping of suitably chosen bands of grouped FFT bins, or – in a more efficient way – by using e.g. wavelet transforms with a suitably designed mother wavelet allowing intrinsic "frequency" overlapping effect.

3.3 Noise Resistance and Cancellation

Silence/speech period separation has to be performed in a different way than it currently is (signal energy thresholding) and/or noise cancellation procedures have to be introduced to the algorithm. A cepstral approach seems to be appropriate for the VAD (voice activity detector) of the algorithm. The cepstral VADs are generally insensitive to non-speech signal components, thus offering a good opportunity for speech detection even in high levels of input noise.

Modern low bit-rate coders are often equipped with powerful noise cancellation algorithms. This fact has to be reflected in the speech quality measurement algorithm as an optional feature. It must be activated if noise cancellation procedure is active in the tested connection (otherwise the PESQ-like algorithm will consider missing noise power as missing signal and, paradoxically, the results may be worse with noise cancellation procedures activated).

4 Conclusion

Some elementary comments for the introduction of automatic intrusive measurements into low bit-rate networks were given.

Acknowledgements

This research has been supported by GACR project 102/01/1355 "Advanced Measurements in Mobile Networks".

REFERENCES

[1] ITU-T P.862, Perceptual Evaluation of Speech Quality, ITU-T, February 2001

[2] Rix, A., Beerends, J.G., Hollier, M.P., Hekstra, A. P.: Perceptual Evaluation of Speech Quality (PESQ) - a new method for speech quality assessment of telephone networks and codecs. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Salt Lake City, May 2001

[3] Street, M.D., Future NATO Narrow Band Voice Coder Selection: STANAG 4591, NC3A Technical Note, 2001

[4] Street, M.D., The NATO Post-2000 Narrow Band Voice Coder, Test and Selection of STANAG 4591, NC3A Technical Presentation, 2002

[5] Holub, J., Očenášek, J., Šmíd, R.: A Novel Intrusive Voice Transmission Quality Test System for Mobile Networks, IEEE 9th International Workshop on Systems, Signals and Image Processing (IWSSIP), November 2002, Manchester UK, p. 158-162

[6] Dresler, T., Holub, J., Šmíd, R.: Wavelet Transform in Voice Transmission Quality Measurements, accepted for ISCA Workshop, Germany, May 2003

[7] Holub, J., Šmíd, R: Methodology for Voice Transmission Quality Measurement Systems Verification, accepted for XVII IMEKO World Congress, June 2003, Dubrovnik, Croatia