Proprietary Low Bit-rate Radio-communication Network - Objective and Subjective Speech Transmission Quality Assessment

Jan Holub, Andrea Drozdová

FEE CTU Prague, Technická 2, Prague 6, Czech Republic, holubjan@fel.cvut.cz

University of Defence, Kounicova 65, 61200 Brno, Czech Republic

Abstract

The article describes test methodology and result of both subjective and objective testing of speech transmission quality in proprietary military radio network, operating in SW (30-88 MHz) band. The subjective tests has been based on ITU-T P.800 Recommendation. The objective tests has been performed using ITU-T P.862 (PESQ) and P.563 (3SQM) algorithms. The results confirm applicability of both algorithm for the given transmission technology, however, some systematic differences and consequent potential risks for routine objective testing are highlighted.

Keywords

Speech Transmission Quality, Low Bit-rate Speech Coding

1 Introduction

Speech transmission quality measurements are widely and widely used to compare different coding and transmission technologies, or to monitor the network performance. The traditionally proven but expensive subjective methods [1] have been partially replaced by objective digital signal processing algorithm based measurements that either compare the original undistorted signal to the transmitted one [2] (so called intrusive or double-sided algorithms) or process only the transmitted version [4]. All these methods have been designed and tested on past & contemporary telecommunication transmission standards that are widely used in common mobile and fixed telecommunications. Application possibility to any other area, like special proprietary radio communication network that deploys low bit-rate speech coding, must be carefully verified by proper testing and result comparison with subjective assessment. Due to generally worse narrow-band coding of female voices, a special attention has been paid to male/female voice coding quality comparison and also to comparison of their assessment by different objective algorithm.

2 Work Performed

2.1 Speech Database Design and Recording

Speech database fulfilling P.800 requirements and containing various transmission settings (open/encrypted) and impairments (jamming) has been recorded on network model deploying two portable radio stations that are used in the radio network and set of splitters and generators. The original sample contained 4 short sentences in Czech language, spoken by four (two male and two female) talkers, and recorded in studio environment. Each transmitted version has been recorded 5 times per one transmission/impairment setting (See Table 1). Totally, 160 speech samples have been recorded and processed.

Table 1 Transmission setting/impairment combinations (Open/Encrypted Mode, Co-channel/ Adjacent Channel/ Wideband Jamming)

Combination	Combination number
Co-channel Light Jamming, Open Mode	1
Co-channel Heavy Jammig, Open Mode	2
Encrypted, No Jamming	3
Encrypted, Adj-Channel Jamming	4
Encrypted, Wide-band Jamming	5
Co-channel Heavy Jamming, Encrypted	6
Open mode, No Jamming	7
Original (IRS filtered)	8

2.2 Subjective Tests

The ACR-MOS (Absolute Category Rating – Mean Opinion Score) [1] have been performed on the recorded samples, using 20 naive listeners. Single (always the first) multi-sentence sample record of each transmission setting/impairment has been used since no audible difference between consequent 5 recordings has been identified in any of transmission setting/impairment combination. Except of omitted MNRU test, all requirements of P.800 have been followed. The listening level has been equalized to 79 dB SPL, wide-band headphones have been used. The results are shown aggregated per transmission /impairment case in Fig. 1, and drilled down by speaker sex in Fig. 2 and 3 for male and female voices, respectively.

2.3 Objective Testing

Two up-to-date measurement methods have been applied to the recorded samples: PESQ (ITU-T P.862 [2]) as widely accepted [3] example of intrusive method and 3SQM (ITU-T P.563 [4]) as an advanced non-intrusive measurement algorithm. Each of 5 recordings (of each sentence for given transmission setting/impairment combination) have been evaluated separately and the average score has been calculated. The results are again shown aggregated per transmission /impairment case in Fig. 1, and drilled down by speaker sex in Fig. 2 and 3 for male and female voices, respectively.

3 Conclusions

Even though the number of tested samples as well as number of transmission setting / impairment combinations is not exhausting, several evident conclusions can be made:

-Female voices are transmitted worse than male voices, the difference is app. 0.5 MOS in average (1 MOS maximum) for all tested cases.

-For encrypted transmission, lightly jammed cases (case 4) are both subjectively and objectively assessed better than respective case without jamming (3).

-Both PESQ and 3SQM perform satisfactorily even on the tested transmission technology for which the methods have not been obviously designed except of the following systematic differences:

-For MALE voices, the 3SQM is generally “too optimistic” and gives scores of up to about 0.5 MOS higher. However, this effect can be suppressed by additional polynomial fit as recommended [2], supposing the a-priori knowledge about the speaker sex (as this effect does not appear significantly for FEMALE voices).

-For MALE voices, PESQ sensitivity is lower than that of subjective tests (see cases 5-6 in Fig. 2). Also this can be probably decreased by the polynomial fit as above.

-Very distorted recordings are not evaluated by objective methods properly (the objective score is mostly significantly higher that the subjective one), means the decreased quality can not be reliably detected by means of objective methods, if the threshold is set too low (e.g. under MOS=2).

4 References

[1] ITU-T Rec. P. 800 “Methods for subjective determination of transmission quality”, International Telecommunication Union, Geneva, 1996.

[2] ITU-T Rec. P. 862 “Perceptual Evaluation of Speech Quality”, International Telecommunication Union, Geneva, 2001.

[3] Pennock, S.: Accuracy of the Perceptual Evaluation of Speech Quality (PESQ) Algorithm, MESAQIN 2002, Praha, CTU.

[4] ITU-T Rec. P. 563, “Single-ended method for objective speech quality assessment in narrow-band telephony applications”, International Telecommu-nication Union, Geneva, 2004.

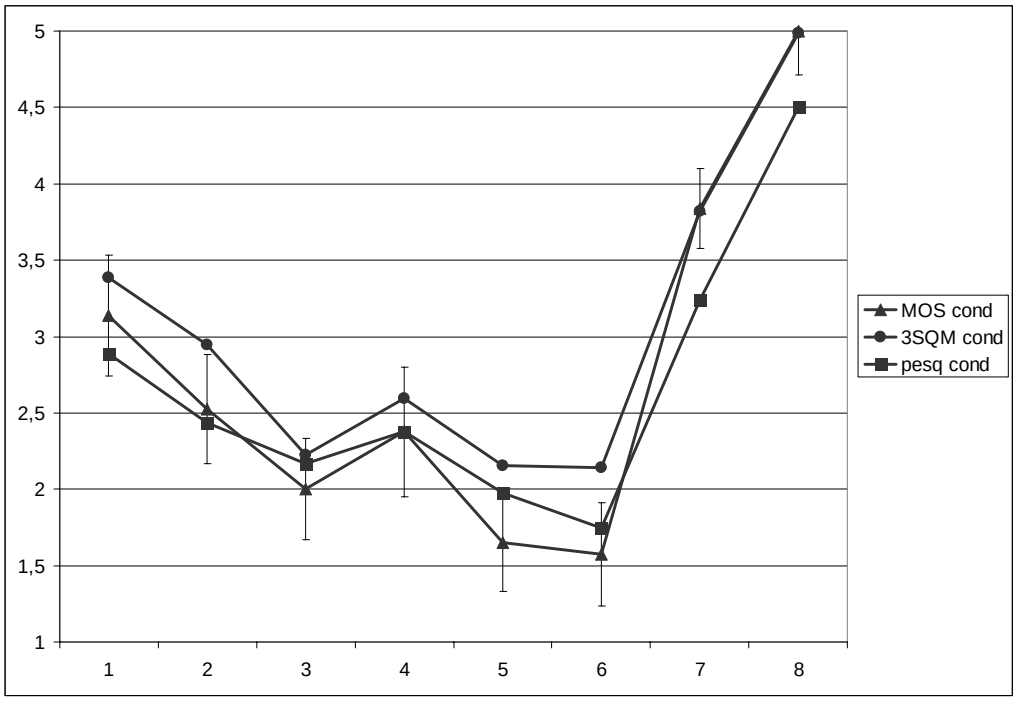

Fig. 1 – Total experimental results ("MOS cond" – subjective testing results, "3SQM cond" – 3SQM results, "pesq cond" – PESQ results). The vertical bars show the worst (among the 4 speakers) 95% CI for each transmission setting / impairment combination. The sample numbers correspond to Table 1. The connecting lines between neighboring points are to highlight the points of the same evaluation method type and have no interpolating meaning.

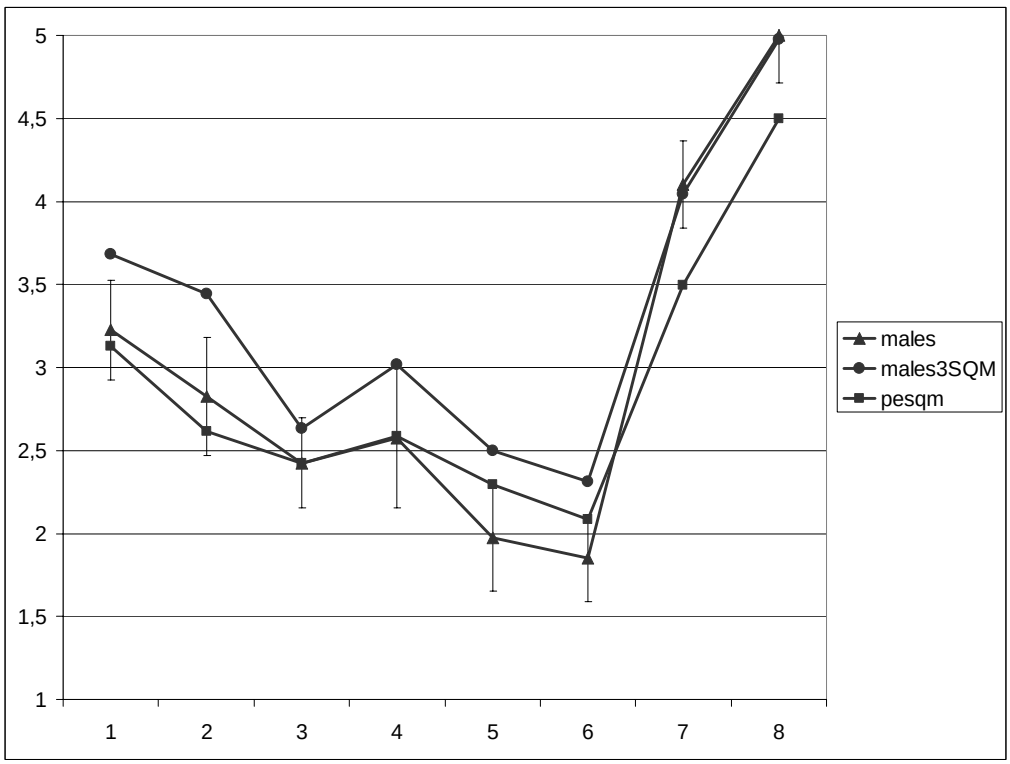

Fig. 2 – "Male speakers" experimental results. Other detailed descriptions of Fig. 1 apply appropriately.

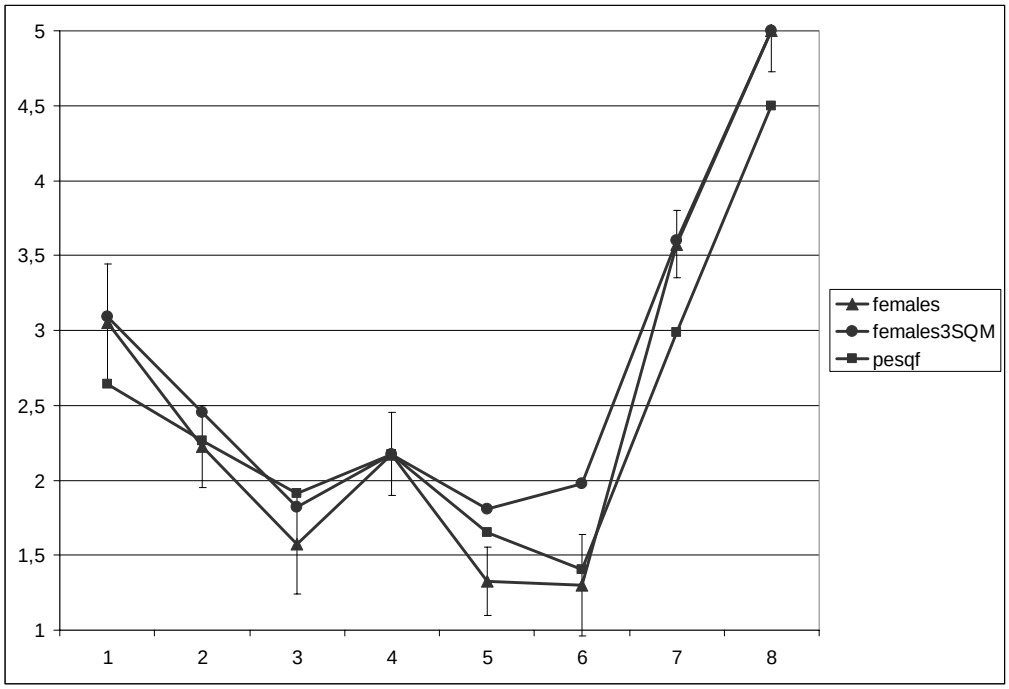

Fig. 3 – "Female speakers" experimental results. Other detailed descriptions of Fig. 1 apply appropriately.