

Child Listeners as the Test Subject – Comparison with Adults and P.862

J. Holub, R. Šmíd, M. Bachtík

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

holubjan@feld.cvut.cz

Abstract

This article reports some conclusions on comparison between listening tests performed on younger (age of 8-10 years) and adult (age of 18-25) listeners. Also a comparison between ITU-T Recommendation P.862 and subjective tests results is given for both groups.

Keywords

listening tests, PESQ, MOS

1 Introduction

Contemporary telecommunication services are hitting wider and wider portfolio of users. In particular, telecom operators are trying to attract children and younger people by offering new services and features. However, speech quality measurements are calibrated mostly on adult listeners. Our goal was to check whether the same algorithms could be used also for prediction of speech quality as perceived by children.

2 Experimental Setup

2.1 Speech Sample Preparation

43 speech samples have been prepared totally. Those were based on four (two male and two female) original samples (all talkers were adults) that were artificially affected by several types of distortion and impairments (amplitude clipping $-5/-10/-20$ dB, listener echo $16\text{dB}@50\text{ms}$, $10\text{dB}/200\text{ms}$, white noise $-15 \dots -35$ dB). Some samples were affected just by one impairment type, some of them carries more distortion types to achieve realistic acoustic appearance from common mobile networks. A special tool running under Matlab 6.0 has been used for distorted sample generation [3]. No pre-filtering of the samples has been applied. The original four samples featured studio quality.

2.2 Listener Types

Two groups of listeners have been identified according to the following pictures. Totally, 47 children and 71 adults were used for listening tests. None of them was

experienced in the field, the just obtained a short training on absolute category rating (ACR) listening quality (LQ) scale defined in ITU-T P.800 in the test beginning. The age structure of both groups is depicted on Fig. 1.

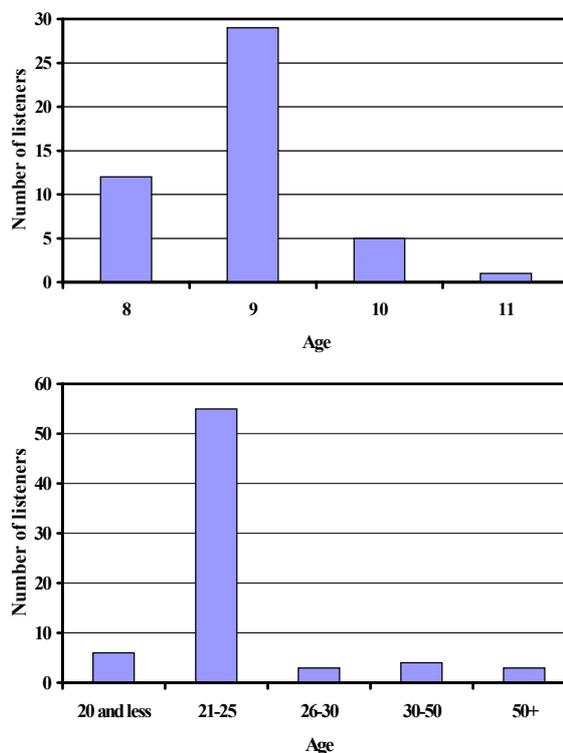


Fig. 1 Age structure of the listeners groups: children (top) and adults (bottom)

3 Test Results

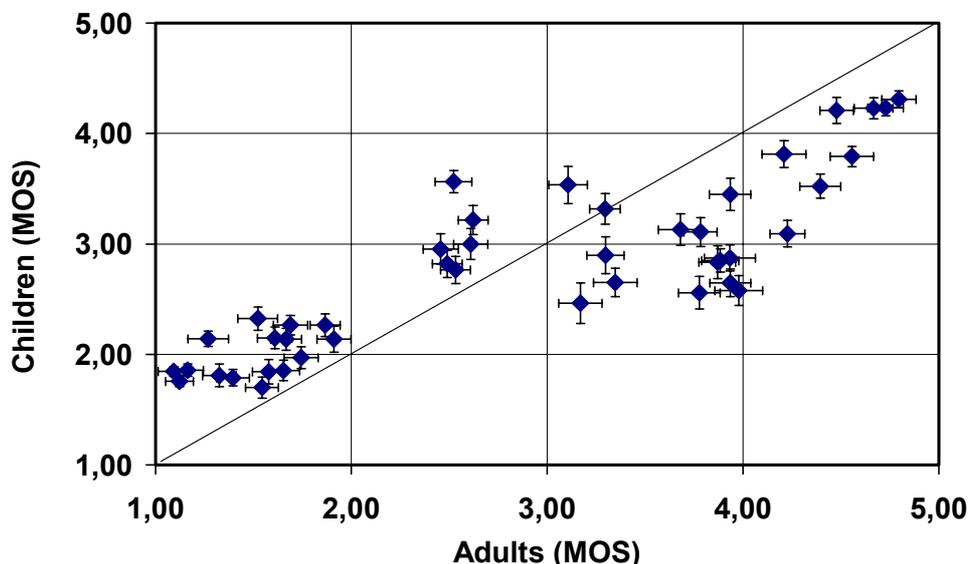


Fig. 2 Comparison between results of listening tests performed on adult (horizontal) and child (vertical) listeners (bars are showing “one-sigma” uncertainty of the average value, means 67% confidence level)

The comparison between adult and child listener opinions is given in the Fig. 2. There is an obvious dependency noticeable - children listeners overestimate all the samples with (adult) MOS lower than 3 and underestimate all the samples with (adult) MOS higher than 3. In both cases, the shift is cca 0.5 MOS.

3.1 Listening test results vs. PESQ and PSQM

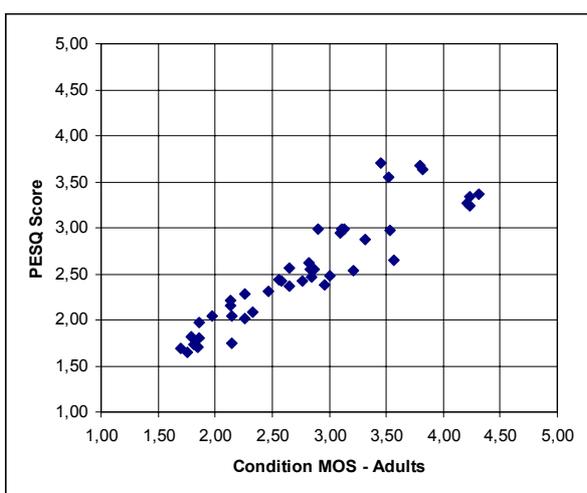


Fig. 3 PESQ vs. opinion of adults

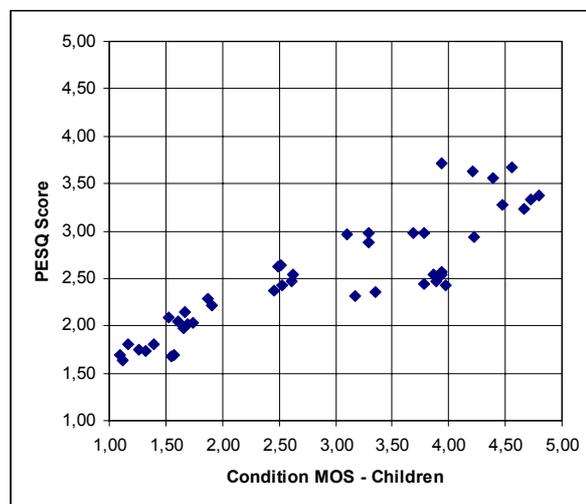


Fig. 4 PESQ vs. opinion of children

The comparison between PESQ scores and grouped results of listening tests is depicted on the Fig. 3 and 4. The set of samples where adult MOS is around 4.3 but PESQ Score gives 3.3 (Fig. 3) are samples affected by listener echo (16dB@50ms). The correlation for PESQ (P.862) and PSQM (P.861) are given in the Table 1 (The PSQM results are not represented graphically here due to their obsolete nature).

	Children	Adults	Combined
PSQM	0,62	0,74	0,72
PESQ	0,88	0,92	0,93

Tab. 1 Correlation between listening test results and normalised (ITU-T) algorithms

4 Conclusion

Even the numbers of speech samples and of listeners used are not exhausting, some evident conclusions can be made and some speculative explanations can be estimated:

1. Child listeners can easily sustain degraded signal. In fact, they evaluate all the significantly degraded samples 0,5 MOS higher than adults. This may be caused by their more adaptive perception and also by their generally still more optimistic views.

2. On the contrary, for less degraded signals the young listeners are more demanding than adults (0.5 MOS negative shift). This can be explained by CD-quality environment that is common today for all audio-recording devices that are commonly used.

3. In general, child listeners can be used as subjects of the listening tests in cases when it may be advantageous or more relevant.

REFERENCES

- [1] ITU-T P.862, Perceptual Evaluation of Speech Quality, ITU-T, February 2001
- [2] ITU-T P.861, Objective Quality Measurement of Telephone-band 300-3400 Hz speech codecs, 1996
- [3] Holub J., Šmíd, R., Horyna, J.: Intrusive Speech Quality Measurements in Czech Environment, MESAQIN2002, on-line at <http://wireless.feld.cvut.cz>, February 2002
- [4] Beerends, J.G., Stemerdink, J.A.: A Perceptual Speech Quality Measure based on a psychoacoustic sound representation, J. Audio Eng. Soc., vol. 42, pp. 115-123, (1994 March)
- [5] Rix, A.W., Hollier, M.P., Hekstra, A. P., Beerends, J.G.: PESQ, the new ITU standard for objective measurement of perceived speech quality, Part 1 - Time alignment, J. Audio Eng. Soc., vol. 50, pp. 755-764, 2002 Oct..
- [6] Beerends, J.G., Hekstra, A.P., Rix, a.W. and Hollier, M. P.: PESQ, the new ITU standard for objective measurement of perceived speech quality, Part II - Perceptual model, J. Audio Eng. Soc., vol. 50, pp. 765-778, 2002 Oct. (equivalent to KPN Research publication 00-32228).