Echo Assessment in Narrowband and Wideband Scenarios

Frank Kettler, Marc Lepage

HEAD acoustics GmbH, 52134 Herzogenrath, Germany,

Introduction

A new approach to quantify echo disturbances was introduced during the last NAG/DAGA 2009 [1]. It is based on a hearing adequate echo analysis and considers the masking by the user's own voice. Furthermore it is applicable for narrowband and wideband scenarios, which makes it suitable for upcoming telecommunication networks and the new generation of terminals. Results of subjective tests and latest modifications on the echo assessment model are given in this paper. The correlation to the objective results is discussed.

Subjective Test Procedures

The basis for a new echo model -like for all other objective analyses- must be the subjective impression of test subjects. Subjective echo assessment tests were therefore carried out. In principle these tests can be conducted as so called Talking-and-Listening Tests or as Third-Party-Listening Tests based on artificial head recordings. The principle of the recording procedure is shown in **figure 1**. The listening tests provide the advantage that the same audio files as assessed in the subjective test can be used for the objective analyses.

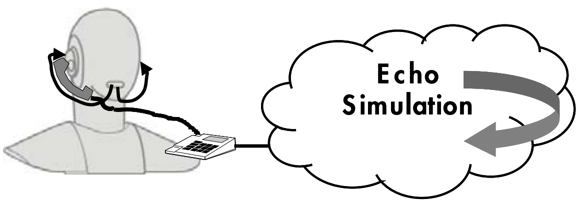

Figure 1: Principle of binaural recordings for Third-Party-Listening Tests

It is important to consider the masking due to the audible sidetone (self masking) during the LOT (see **figure 1**). The results of TALT and Third Party LOT correlate very good for narrowband scenarios [2]. In order to verify the correlation between TALT and LOT results these tests were carried out in a wideband test scenario. Seven experts (six male, one female) participated in a wideband TALT. Fourteen test conditions which were also assessed in the LOT are tested.

The LOT was conducted using two speech files (male and female voices) and the simulation of wideband and narrowband phones. In total 146 conditions plus 6 training conditions were assessed by 21 test persons in the narrowband LOT. The wideband LOT covered 150 conditions plus 6 training conditions assessed by 19 test persons. A condition is represented by a combination of echo delay (varied between 100 ms and 500 ms), the echo attenuation (variation between 25 dB and 55 dB), echo shaping filters (unfiltered plus two transfer functions) and non-linear effects realized by a threshold operation. The narrowband and the wideband LOT were conducted in separate sessions.

WB TALT vs. WB LOT

The comparison between TALT and LOT results in wideband mode is given in **figure 2**. The results from the TALT are indicated by the magenta bars, the blue bars represent the LOT results both based on the male voice.

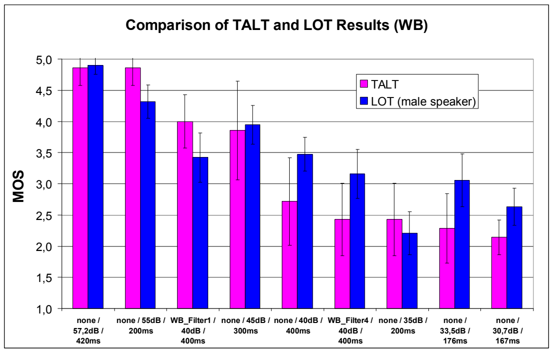

Figure 2: TALT MOS (magenta) vs LOT MOS (blue)

The confidence intervals overlap for all test conditions indicating that the differences between both test methods are not significant. However, a significant difference could be observed in the LOT results for the male and female voices. A reason can be found in the combination of spectral content of male and female voices used for the speech recordings and the sending frequency response of the wideband handset used for the recordings.

Figure 3 shows that the frequency response attenuates the power density spectrum below approximately 200 Hz which affects more the male voice than the female voice.

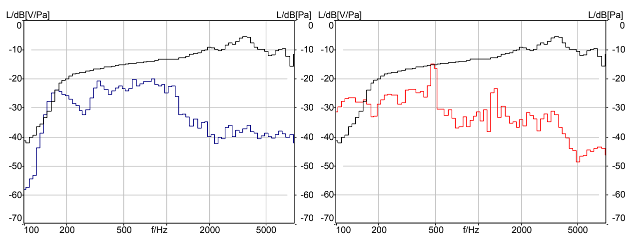

Figure 3: Sending frequency response (black) and power densitiy spectrum (left (blue): female, right (red): male)

Both signals were applied during the recordings with an average speech level of -4.7 dB_{Pa} at the mouth reference point of the artificial head. However, the frequency response of the wideband phone leads to level differences of approximately 4 dB in the network. The female voice is 4 dB higher in level compared to the male voice. This leads to a systematical offset in the LOT results. The MOS for the female voice is between 0 and 1.5 MOS lower compared to the quality ratings for the male voice.

Objective Echo Assessment Model

The block diagram of the echo assessment model is shown in **figure 4**. The electrically recorded echo signal $e_{el}(k)$ is converted via the terminal simulation to the echo signal e(k). The right ear sidetone signal $s_R(k)$ and echo signals e(k) are used for analysis in the model.

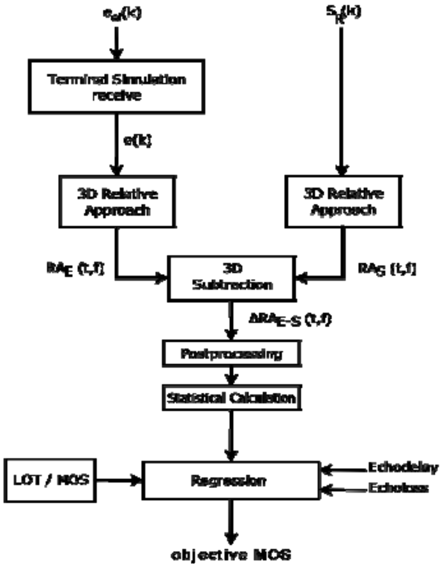

Figure 4: Block diagram of echo assessment model

The 3D Relative Approach (RA [3], [1]) analysis is calculated for e(k) and $s_R(k)$. The combination of both RA pattern (ΔRA) are postprocessed and statistically analyzed. The regression analysis further relies on the LOT MOS and the two "traditional" parameters echo delay and echo loss.

Correlation Analysis

The first correlation results between the wideband auditory MOS (x-axis) and objective MOS (y-axis) calculated with the echo assessment model from **figure 4** are given in the two analyses in **figure 5**. The left hand correlation plot is based on the two parameters echo loss and echo delay for the regression analysis without the hearing model based input.

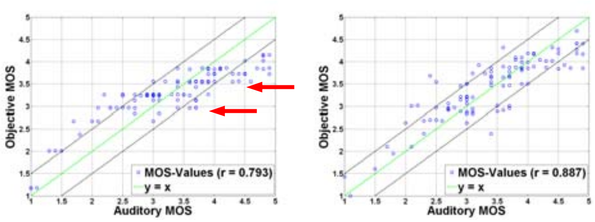

Figure 5: Objective vs. auditory MOS (wideband)
left: input echo loss and echo delay; right: input $m\Delta RA_{e\text{-}r}$

The correlation is r = 0.79 and a known systematical error can be observed in this data (see red arrows): listening examples generated with the same echo delay and echo loss but different spectral characteristics (echo shaping filters) and non-linearities lead to identical objective scores even the auditory MOS spread over a wide range. This represents the known shortcoming of analyses based only on standard parameters: relevant echo characteristics like non-linearities or the spectral content of echo signals can not be adequately considered. Again these results demonstrate the motivation for the new echo assessment model. The correlation analysis based on the ΔRA input values is shown in the right hand plot. The correlation coefficient r increases up to 0.89 even without using the postprocessing block in **figure 4**. Furthermore the systematical error between the objective and auditory MOS as indicated in the left hand plot does not occur anymore. **Figure 6** shows the correlation plot for the combination of ΔRA, echo loss and echo delay vs. auditory MOS. The residual error distribution is shown in the right hand plot.

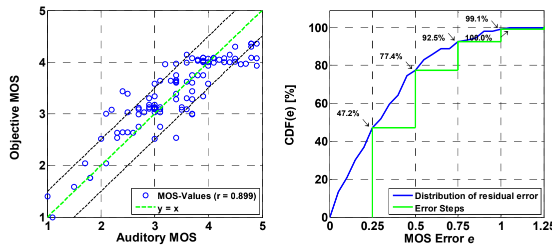

Figure 6: Objective vs. auditory MOS (wideband) and residual error distribution; input parameter $\Delta RA_{E\text{-}S}$, echo loss and echo delay

One tuning step to be discussed is the use of individual coefficients for male and female test signals for the linear regression. This can be motivated by the fact that the echo model is designed for the same speech samples used in the subjective tests. As indicated above the two voices show different characteristics relevant for echo assessment like the level difference and lead to slight systematical differences (see above). The resulting correlation factor r and the CDF (e=0.5) are indicated in **table 1** using the three input parameters ΔRA, echo loss and echo delay.

	Male voice	Female voice
Narrowband	0.93 / 84.9%	0.93 / 83.0%
Wideband	0.91 / 92.6%	0.92 / 88.9%

Table 1: Correlation factor r and CDF (e=0.5) for signal individual regression coefficients

Conclusion

The subjective test results provide the data basis for the development and fine tuning of the objective echo assessment model. The first correlation results between objective and auditory MOS results are -even without any further optimization steps- very promising.

References

[1] Evaluation of Aurally-aedequate Analyses for Echo Assessment, F. Kettler, M. Lepage, M. Pawig, NAG/DAGA 2009, Rotterdam, Netherlands

[2] Echobeurteilung beim Abhören von Kunstkopfaufnahmen im Vergleich zum aktiven Sprechen, F. Kettler, H.W. Gierlich, E. Diedrich, J. Berger, DAGA 2001, Hamburg, Germany

[3] K. Genuit, Objective evaluation of acoustic quality based on a Relative Approach, Inter-Noise'96, Liverpool, England