Evaluation of Aurally-adequate Analyses for Echo Assessment

Frank Kettler[1], Marc Lepage[1], Matthias Pawig[2]

[1]HEAD acoustics GmbH, 52134 Herzogenrath, Germany,
[2]Institute of Communication Systems and Data Processing, 52074 Aachen, Germany

Introduction

The migration towards NGN (Next Generation Networks) is expected to introduce higher propagation delay in telecommunication. This emphasizes the need for reliable echo control. Wideband telephony will further change speech perception - and echo perception. New investigations on echo perception demonstrate the necessity to renew tolerances for wideband echo attenuation. These trends also motivate new echo assessment methods. Current analysis methods typically determine the echo attenuation of terminals as a one-dimensional dB value. Requirements for the echo attenuation are sometimes delay dependent, however, these parameters are inaccurate, neither perception oriented nor aurally adequate. They do not consider wideband specific aspects. A new approach based on a hearing model analysis suitable to extend current echo analysis methods is introduced and discussed in this paper.

Current Status of Echo Analyses

Echo disturbances are mainly classified by the combination of two parameters: round trip delay and echo attenuation. The attenuation is typically expressed by the terminal coupling loss for terminals or the talker echo loudness rating (TELR) for networks. Some guidelines can be found e.g. in ITU-T Recommendation G.114 (One way Transmission Time [1]) and in ITU-T Recommendation G.131 (Talker Echo and its Control [2]). **Figure 1** shows the E-model rating -a network planning tool- characterizing the user's satisfaction based on the mouth-to-ear delay (x-axis) and different TELR values. The upper curve represents a TELR of 65 dB, the lowest curve the 25 dB TELR condition. Quality degrades as a function of delay and talker echo loudness rating.

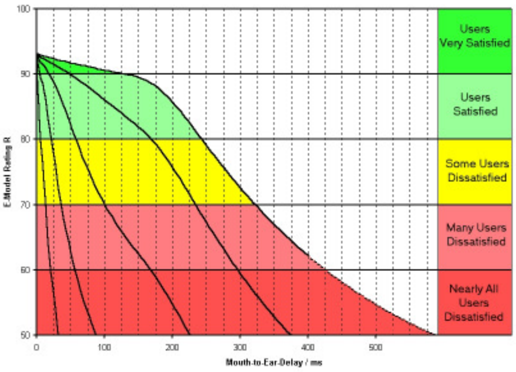

Figure 1: E-Model rating as function of one-way delay and TELR [1]

A similar relationship can be derived from **figure 2** taken from ITU-T Recommendation G.131 [2]. An "acceptable" and "limiting case" curve indicates user's satisfaction as a function of mean one way delay (x-axis) and talker echo loudness rating (y-axis).

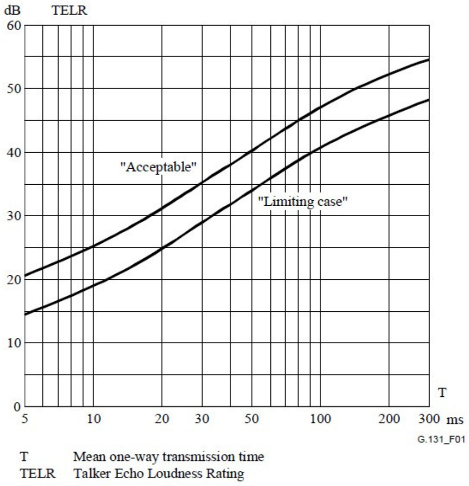

Figure 2: Recommended delay dependent TELR [2]

The echo performance of terminals is typically verified by determining the weighted terminal coupling loss TLC_W. The requirements are verified on the basis of this one-dimensional "dB" value. Other test specifications (e.g. ITU-T Recommendation P.1100 for Mobile Hands-free Implementations in Vehicles [3]) additionally measure the echo attenuation vs. time and spectral echo characteristics. These parameters are very important for practical tests because terminals sometimes lead to residual non-linear echo disturbances, e.g. caused by nonlinearities in the echo path due to high level peaks in transmitted speech. These components "pass" the signal processing of echo cancellers.

New investigations on wideband echo perception further point out that the spectral echo content in the frequency range between 3.1 and 5.6 kHz is especially crucial for echo disturbance [4]. New tolerances for the spectral echo attenuation have been introduced in [4].

An example for an aurally adequate analysis is shown in **figure 3**. The echo signal recorded for a mobile phone is analyzed as level vs. time in the left hand picture. The test signal (real speech) was applied at 2.5 s on the time axis. The echo level is low between approximately -60 and -70 dB_V. Some distinct temporal components can be detected. Further analyses like spectrograms can be used to provide additional spectral information. However, these are all analytical analyses and do not represent the aurally adequate assessment of the echo components.

A very promising method is the Relative Approach [6]. This method is especially sensitive to detect unexpected temporal and spectral components and can therefore be used as an aurally adequate analysis to assess temporal echo disturbances. An example is shown in the right hand picture in **figure 3** [5]. The peaks which can already be detected from the level vs. time analysis are clearly marked as unexpected, disturbing components especially in the high frequency range. This analysis provides important hints for tuning.

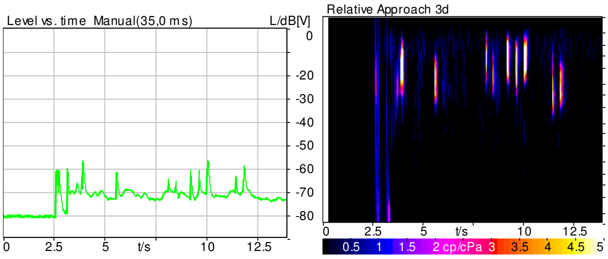

Figure 3: Echo analysis [5] (left: level vs. time; right: Relative Approach)

In summary, the current echo analyses combine various single measurements and verify the compliance to requirements and tolerances. The next step is the combination of these parameters to an objective model providing one-dimensional values with a high correlation to the MOS results from subjective tests. Models providing good correlations for echo assessment have already been evaluated for narrowband telephony, distorted sidetone and room reverberations [7]. The approach introduced here shall be applicable for narrowband and wideband telephony and shall deliver hints for improvement of devices under test such as acoustic or network echo cancellers.

Subjective Tests on Echo Perception

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 first under wideband conditions. In principle these tests can be conducted as so called Talking-and-Listening Tests acc. to ITU-T P.831 [8] or as Third-Party-Listening Tests based on artificial head recordings (ITU-T P.831, Test A [8], [9]). The principle of the recording procedure is shown in **figure 4**. Beside the more efficient test conduction -a group of test subjects can perform the tests at the same time- the listening tests provide the advantage that the same audio files as assessed in the subjective test can be used for the objective analyses.

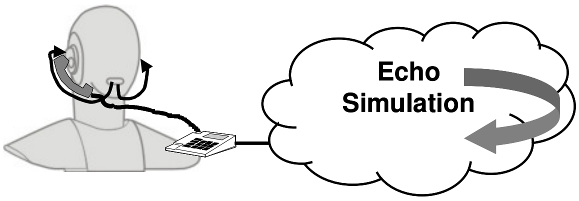

Figure 4: Principle of binaural recordings for Third-Party-Listening Tests (Type A [8], [9])

Figure 5 shows the simulation environment. A wideband capable handset was simulated at the right ear of the HATS [13].

The Third-Party-Listening Tests were carried out with twenty subjects in total, fourteen naïve and six expert listeners. The speech material consists of male and female voices.

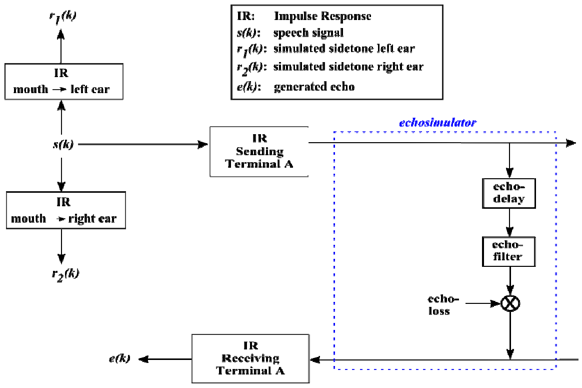

Figure 5: Block diagram of simulation environment

A total number of 33 test conditions including the reference scenarios (infinite echo attenuation) and different combinations of delay, echo attenuation and spectral shaping were included:

- round trip delays between 100 ms up to 500 ms
- echo attenuation between 35 dB up to 55 dB
- non-linear residual echoes.

The spectral echo content was realized by the following filter characteristics (subset of test conditions):

- NB: narrowband filter, 300 Hz to 3.4 kHz
- HF1: 3.1 kHz to 5.6 kHz
- HF2: 5.2 kHz to 8 Hz
- 1/3 oct.1: 900 Hz to 1120 Hz
- 1/3 oct.5: 2.24 kHz to 2.8 kHz
- 1/3 oct.7: 3.55 kHz to 4.5 kHz
- 1/3 oct.8: 4.5 kHz to 5.6 kHz

The 1/3 octave filter characteristics are shown in **figure 6** together with the hearing and speech perception threshold. The intention of these filters is a more detailed analysis of the critical frequency range between 1 kHz up to 5 kHz providing the highest sensitivity for sound and speech perception.

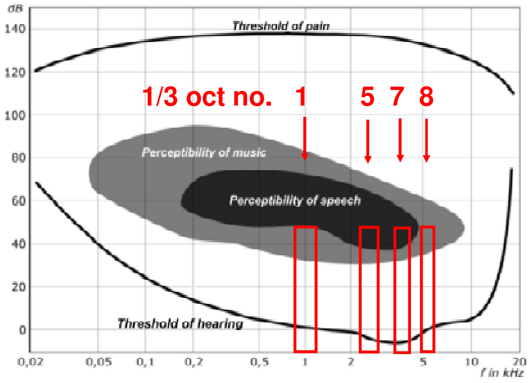

Figure 6: Filter characteristics (subset of test conditions)

A 5-point annoyance scale was used (5 points: Echo is inaudible, …, 1 point: echo is very annoying, [10]). The stimuli were presented without pair comparison. The results are analyzed on a MOS basis together with confidence intervals based on a 95 % level. A first analysis pointed out that the quality rating of both groups (naïve, expert listeners) were very similar. The results were therefore combined.

A small subset of results from the listening only test is shown in **figure 7**. The blue bar indicates the echo-free test condition. The rating of 4.8 MOS must be expected under this condition.

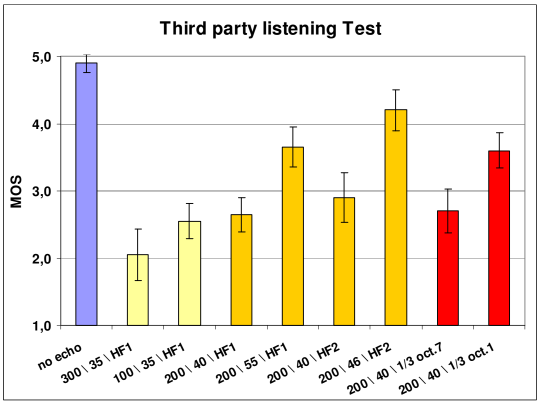

Figure 7: Subset of test results

The two light yellow bars represent the test condition with an echo attenuation of 35 dB (corresponding to a TELR of 45 dB) and a spectral echo content between 3.1 kHz and 5.6 kHz (echo filter “HF1”). This is the critical frequency range for human speech perception. The two bars represent the rating for the 300 ms and 100 ms round trip delay. For a narrowband scenario a TELR of 45 dB in conjunction with a one-way delay of 50 ms (round trip delay 100 ms) would lead to an E-model rating of approximately 82 in **figure 1** (“users satisfied”). In contrary **figure 7** indicates a MOS score of only 2.6 for this condition (“100\35\HF1”, echo characterization between “slightly annoying” and “annoying”). This clearly points out that the spectral echo content plays a crucial role. Similar results can also be found in [4].

The test conditions “200\40\HF1” and “200\55\HF1” differ only in the echo attenuation (40 dB vs. 55 dB). The same can be analyzed for the two conditions named “200\40\HF2” and 200\46\HF2” in **figure 7** (orange bars). As expected the MOS results increase for the higher attenuation. The “HF1” filter (3.1 to 5.6 kHz) seems to be more critical than the “HF2” filter (5.2 to 8 kHz). The results are lower for the same echo attenuation of 40 dB and even for the higher echo attenuation of 55 dB for “HF1” compared to 46 dB for “HF2”.

An example for a more detailed spectral echo analysis is given by the red bars in **figure 7**. Both conditions represent a 200 ms round trip delay in combination with a 40 dB echo attenuation. The two different filter characteristics “1/3 oct.1” and “1/3 oct.7” are introduced in **figure 6**. The results differ by approximately 1 MOS and point out the strong influence of spectral echo shaping on subjective assessment.

Objective Echo Assessment Model

The block diagram of the objective model is shown in **figure 8** [11]. The speech signal *s(k)* is processed through the terminal simulation in sending direction and sidetone simulation for the right and left ear. The terminal send signal is further processed through echo simulation (echo loss, echo delay and spectral shaping filter, see **figure 5**) and the terminals’ receiving direction (handset applied with 8 N application force to the type 3.4 artificial ear at the HATS [12]). The dashed lines in **figure 8** indicate the signal flow used for test material generation for the Third-Party-Listening Test.

Two 3D Relative Approach representations are calculated for the sidetone signal *r(k)* and the echo signal *e(k)*. These analyses are shown in **figure 9**. The 3D subtraction of both analyses considers the masking effect due to the sidetone. The resulting $\Delta RA_{e-r}(t,f)$ representation (**figure 10**) is then used for further processing and statistical analyses.

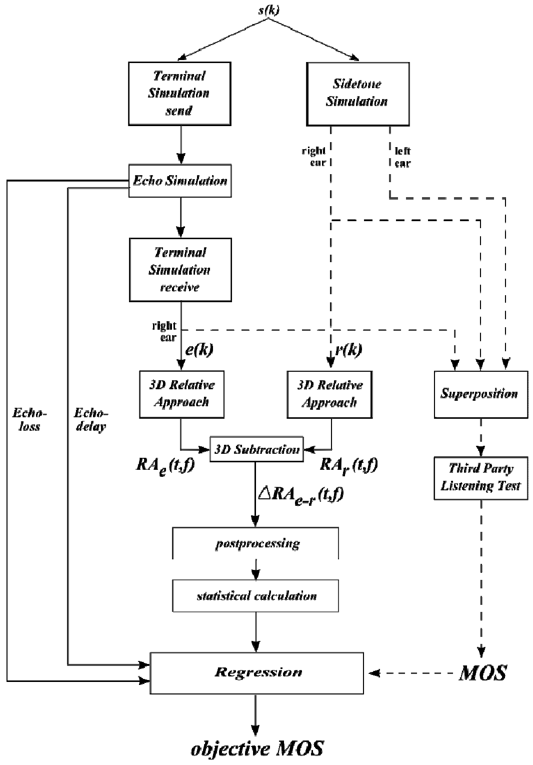

Figure 8: Block diagram of objective echo model [11]

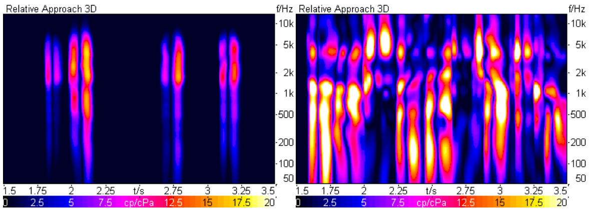

Figure 9: 3D Relative Approach $RA_e(t,f)$ for echo signal *e(k)* (left) and $RA_r(t,f)$ for sidetone simulation *r(k)* (right)

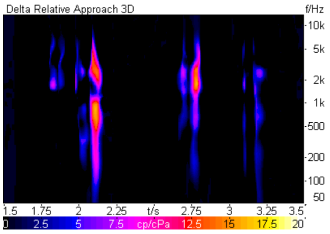

Figure 10: Δ 3D Relative Approach $\Delta RA_{e-r}(t,f)$

In a first approach the two dimensional mean value *m∆RA*$_{e-r}$ *is* calculated according to formula

$$m\Delta RA_{e-r} = \frac{1}{KM}\sum_{k=1}^{K}\sum_{m=1}^{M}\Delta RA_{e-r}(k,m)$$

K= no. of freq. bands, M= no. of samples per band

The parameters echo loss, echo delay and *m∆RA*$_{e-r}$ are used as input signal for a linear regression in order to correlate the objective results to the subjective MOS. In a first step only the two parameters echo loss and echo delay were used in the regression. The result is shown in the left hand scatterplot in **figure 11**. A correlation of r = 0.80 is achieved but the comparison of auditory MOS and objective MOS shows systematical errors: clusters of identical objective MOS occur in **figure 11** (see red arrows) spread over a wide range of auditory MOS (between approximately 1.7 and 3.7 MOS). This can be explained by the different spectral content of these echo signals leading to significant different echo ratings in subjective tests -although the objective parameters (echo delay, echo attenuation) are identical.

The plot on the right hand side in **figure 11** shows the correlation between the auditory MOS and the objective results based only on the two-dimensional mean value *m∆RA*$_{e-r}$. The correlation factor increases to r = 0.84. The systematical error is implicitly solved using the Relative Approach based analysis. In principal this could be expected because the Relative Approach considers the sensitivity of human hearing especially for different frequency characteristics of transmitted sounds.

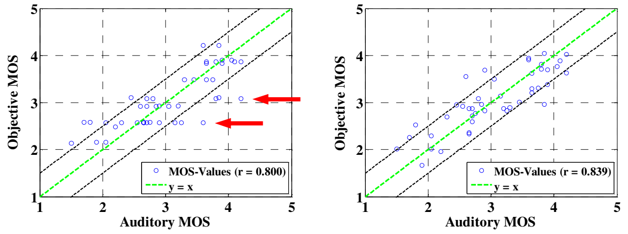

Figure 11: Objective vs. auditory MOS; left: input echo loss and echo delay; right: input *m∆RA*$_{e-r}$

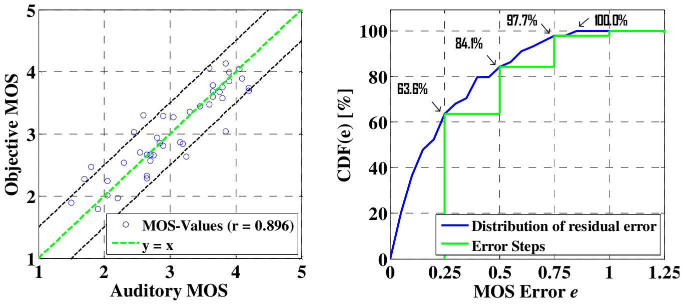

Figure 12: Objective vs. auditory MOS and residual error distribution; input parameter *m∆RA*$_{e-r}$, echo loss and echo delay

The combination of the three parameters *m∆RAe-r*, echo loss and echo delay to the objective MOS further increases the correlation (r = 0.90). The scatterplot is shown in **figure 12** (left hand side) together with the error distribution in the right hand picture. The residual error between objective and auditory MOS is below 0.5 MOS in 84 % of test conditions.

Conclusion

The extension of the standard parameters echo attenuation and echo delay by the hearing model based Relative Approach seems to bear a high potential for echo analyses. The results are particularly promising since so far they are achieved without any postprocessing. Comments from subjects' interviews after the listening tests about most disturbing echo characteristics need to be reviewed. The further adaptation of the Relative Approach analysis on speech characteristics and the masking effect -currently realized by relatively simple subtraction operation- bear further potential.

Acknowledgment

The authors wish to thank Deutsche Telekom Laboratories, Berlin, for making available the wideband echo simulator for these tests.

References

[1] ITU-T Recommendation G.114, One-way transmission delay, May 2003

[2] ITU-T Recommendation G.131, Talker echo and its control, Nov. 2003

[3] P.1100, Narrowband hands-free communication in motor vehicles, August 2008 (pre-published)

[4] S. Poschen, F. Kettler, A. Raake, S. Spors, Wideband Echo Perception, IWAENC, Seattle, USA, Sept. 2008

[5] F. Kettler et al., New developments in mobile phone testing, DAGA2008, Dresden, September 2006

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

[7] R. Appel, J. Beerendts, On the Quality of hearing one's own voice, JAES, April 2002

[8] ITU-T Recommendation P.831, Subjective performance evaluation of network echo cancellers, Dec. 1998

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

[10] ITU-T Recommendation P.800, Methods for subjective Determination of Transmission Quality, Aug. 1996

[11] Evaluation of Aurally-adequate Analyses for Assessment of Interactive Disturbances, M. Lepage, Diploma Thesis, IND Aachen

[12] ITU-T Recommendation P.57, Artificial Ears, Nov. 2005

[13] ITU-T Recommendation P.58, Head and torso simulator for telephonometry, Aug. 1996