EQUEST: Echo Quality Evaluation of Speech in Telecommunications

HEAD acoustics
Application Note

HEAD acoustics GmbH

Ebertstraße 30a
D-52134 Herzogenrath
Tel: +49 (0)2407-577-0
Fax: +49 (0)2407-577-99
E-mail: telecom@head-acoustics.de
WEB: www.head-acoustics.de

EQUEST:
Echo Quality of Speech in Telecommunications

Contents

1. 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.

This application note describes a new approach based on a hearing model analysis suitable to extend current echo analysis methods: EQUEST, the Echo Quality Evaluation of Speech in Telecommunications.

2. Background and Motivation

2.1. 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 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.

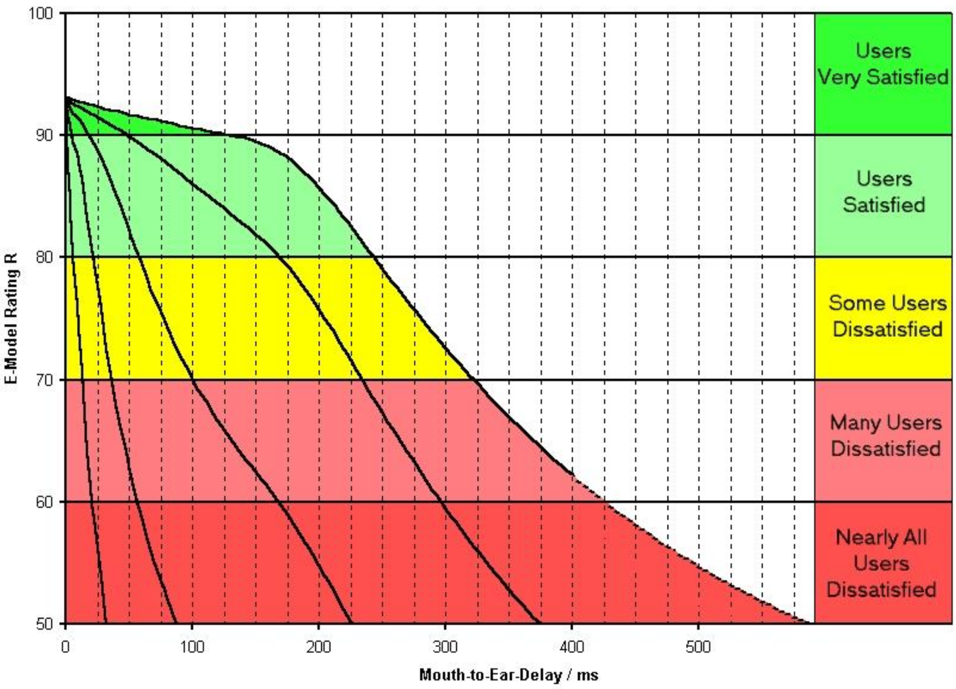

Figure 1: E-Model rating as function of one-way delay and TELR [1]

A similar relationship can be derived from ITU-T Recommendation G.131, see figure 2 [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). As noted in G.131 *"The 'limiting case'…should only be used in exceptional circumstances, as it corresponds to a 10% probability of encountering objectionable echo."*

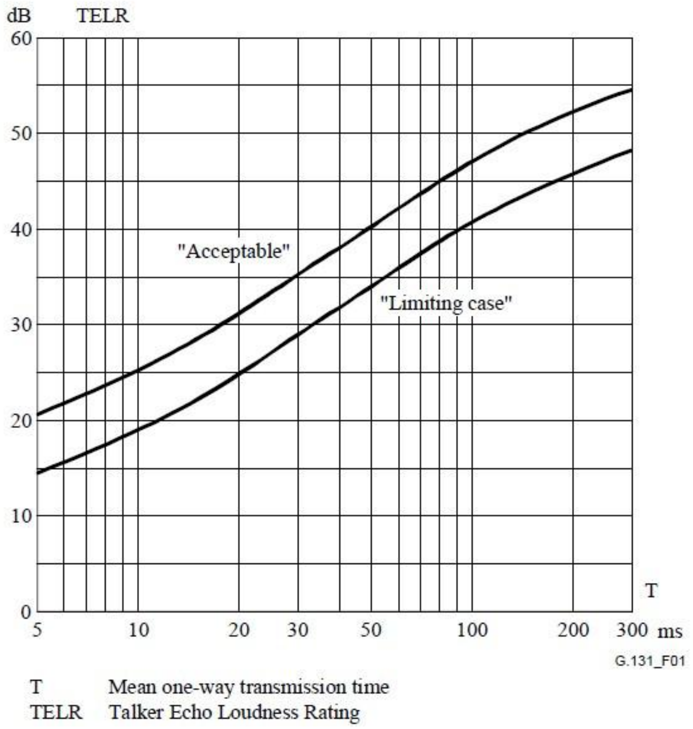

Figure 2: Recommended delay dependent TELR [2]

The echo performance of terminals is typically verified by determining the weighted terminal coupling loss TLCw. 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.

2.2. Wideband and Aurally Adequate Requirements

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 dBV. 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.

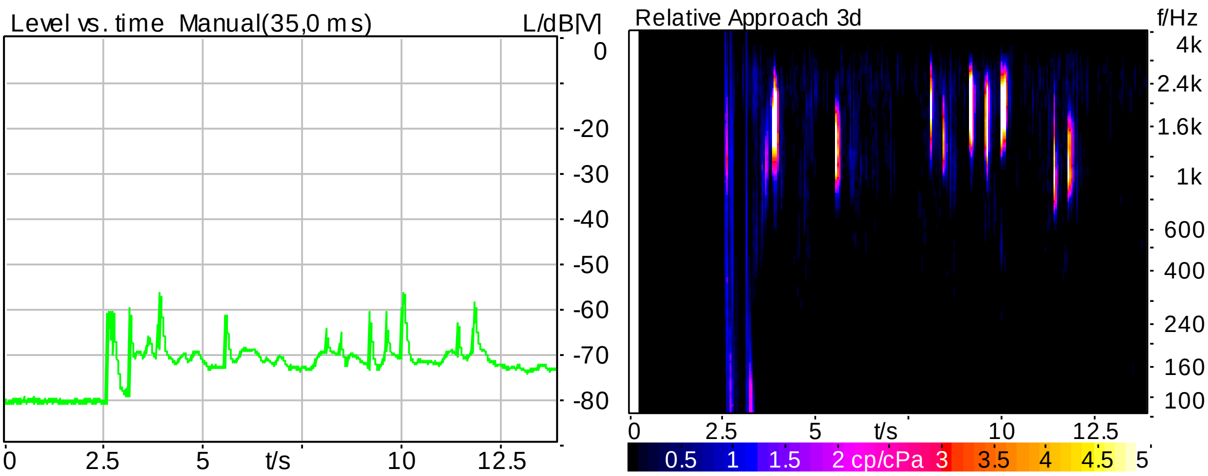

Figure 3: Echo analysis [5] (left: level vs. time; right: Relative Approach)

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.

3. Design of EQUEST

3.1. 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.

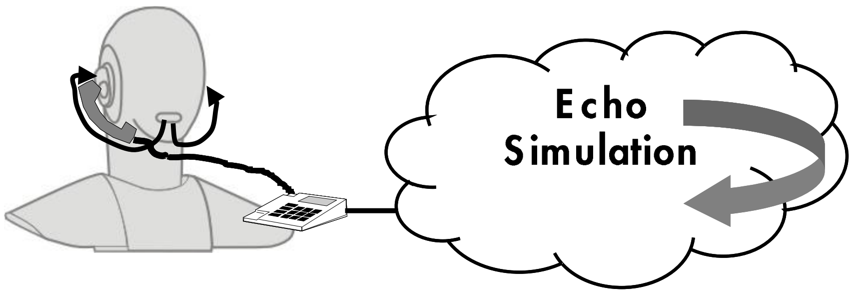

Figure 4: Principle of binaural recordings for Third-Party-Listening Tests (Type A [8], [9])

Figure 5 shows the simulation environment. A wideband capable handset applied to the right ear of the HATS [14, 15] was simulated.

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.

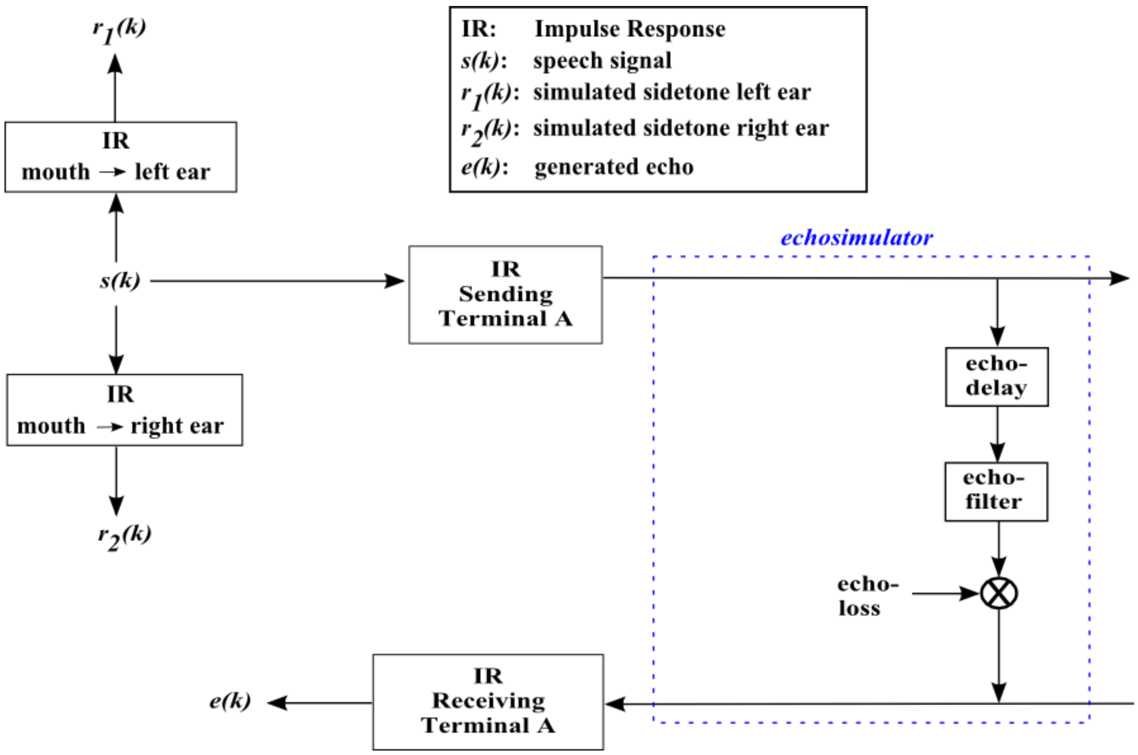

Figure 5: Block diagram of simulation environment

A total number of 146 test conditions for narrow band and 150 test conditions for wide band 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.
- Male & female speakers

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 [10] 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.

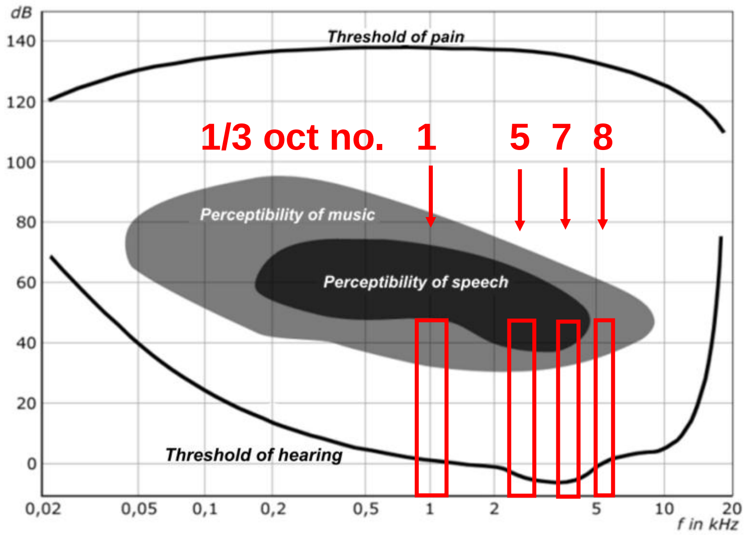

Figure 6: Filter characteristics (subset of test conditions) [10]

A 5-point annoyance scale was used (5 points: Echo is inaudible, …, 1 point: echo is very annoying, [11]). 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 leading to an MOS score of 4.8.

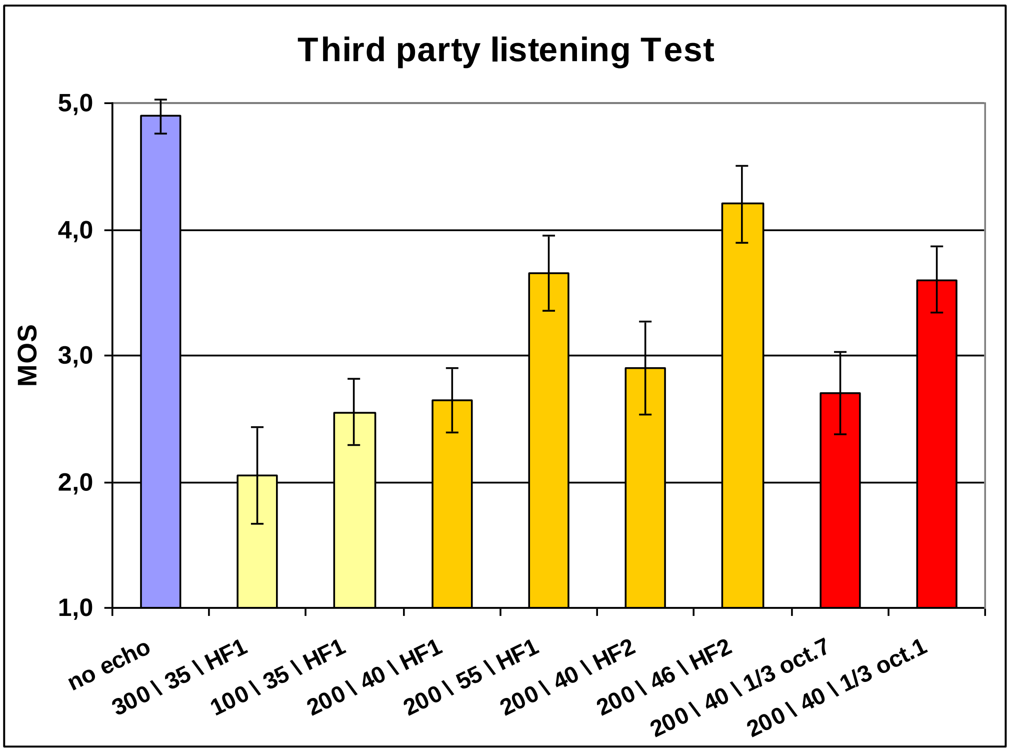

Figure 7: Subset of test results

The two light yellow bars represent the test condition with an echo attenuation of 35 dB (which is equivalent to a TELR of 45 dB, as TELR incorporates SLR and RLR) and spectral echo content between 3.1 kHz and 5.6 kHz; using echo filter “HF1”. This filter represents the most 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 (echo attenuation 35 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. Furthermore, the “HF1” filter (3.1 to 5.6 kHz) appears 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.

3.2. Objective Echo Assessment Model

The concept of a new echo assessment model is shown in figure 8. The analysis requires two input signals:

- The echo signal $e_{el}(k)$ captured electrically from line, which is filtered with a terminal simulation in receiving direction and results in the acoustical echo e(k).
- The acoustical sidetone $s_R(k)$ of the right ear as described in chapter 3.1.

These signals are transformed to their 3D Relative Approach representation. An example for the two analyses is shown exemplarily in figure 9.

In order to take into account the masking effect of the sidetone, the two resulting 3D Relative Approaches are combined by the Δ-Relative Approach and a zero-threshold (masking effect) is applied as a post-processing according to the following equation.

$$\Delta RA'_{E-S}(t_i, f_j) = \max\left(0, \Delta RA_{E-S}(t_i, f_j)\right) \quad \forall t_i, f_j$$

The two 3D Relative Approach representations calculated for the sidetone signal $S_R(k)$ and the echo signal $e(k)$ are shown in figure 9. The 3D subtraction of both analyses considers the masking effect due to the sidetone. The resulting $\Delta RA_{E\text{-}S}(m,n)$ representation (figure 10) is then used for further processing and statistical analyses.

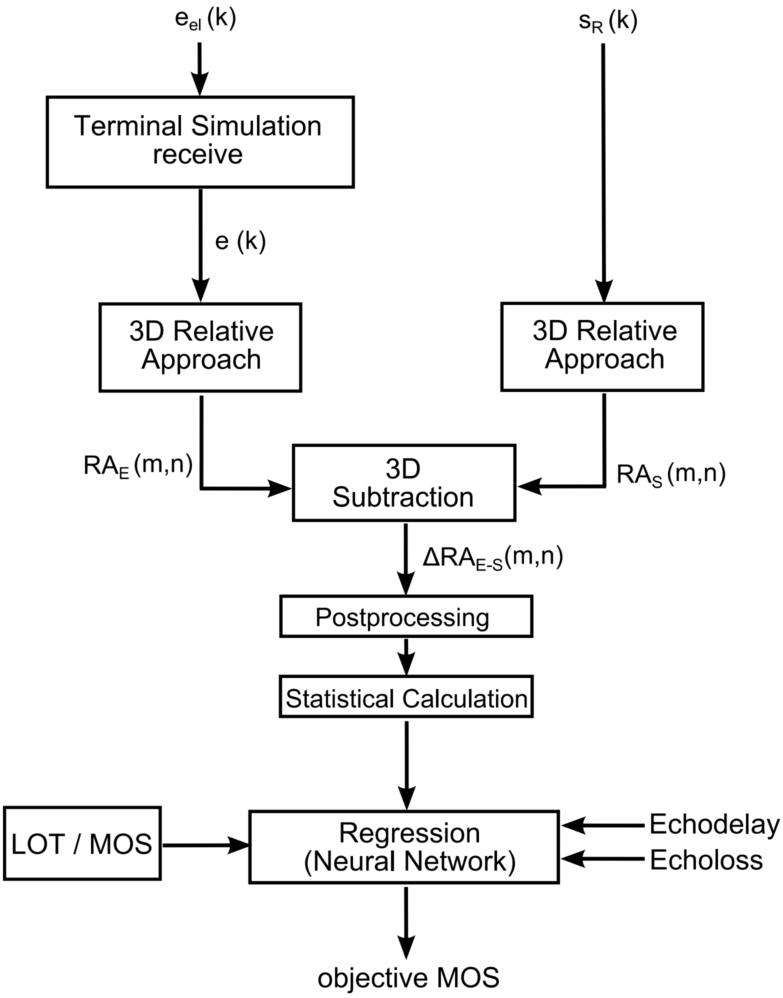

Figure 8: Block diagram of objective echo model [12]

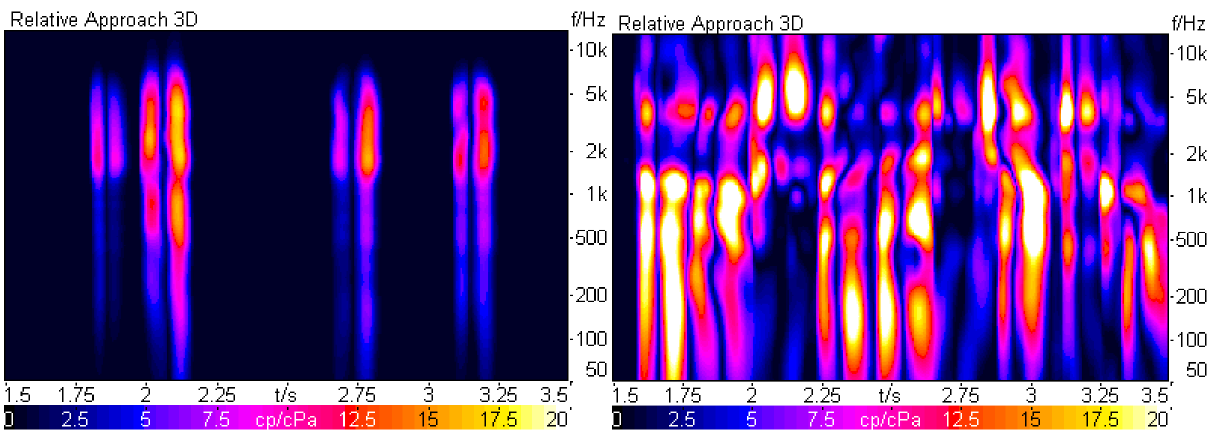

Figure 9: 3D Relative Approach RA_E(m,n) for echo signal e(k) (left) and RA_S(m,n) for sidetone simulation S_R(k) (right)

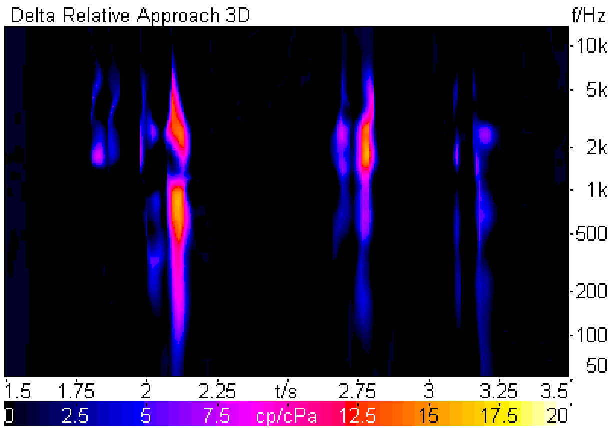

Figure 10: Δ 3D Relative Approach $\Delta RA_{E\text{-}S}(m,n)$

In a first approach the two dimensional mean value $m\Delta RA_{E\text{-}S}$ is calculated according to formula

$$m\Delta RA_{E-S} = \frac{1}{MN}\sum_{m=1}^{M}\sum_{n=1}^{N}\Delta RA_{E-S}(m,n)$$

M = no. of frames containing speech activity, N = no. of frequency bins

The parameters echo level, echo delay and $m\Delta RA_{E\text{-}S}$ 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.

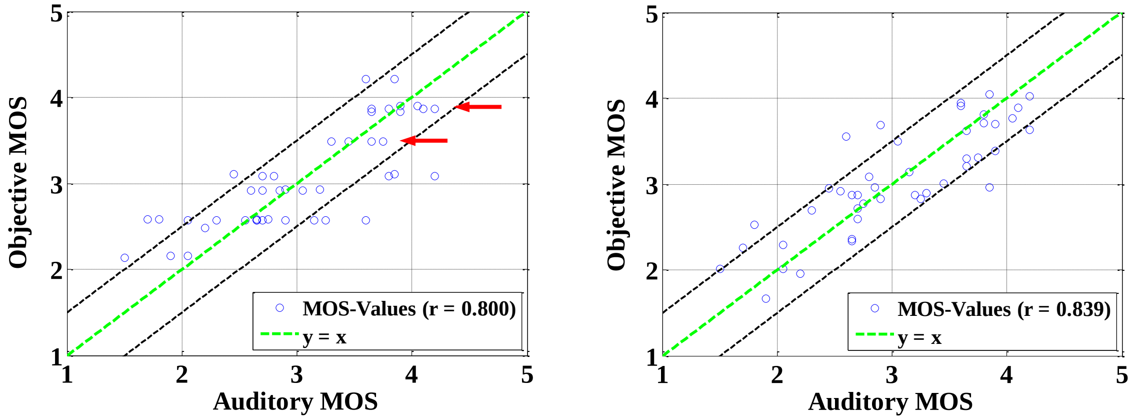

Figure 11: Objective vs. auditory MOS; left: input echo loss and echo delay; right: input mΔRA$_{E\text{-}S}$

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 *mΔRA$_{E\text{-}S}$*. 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.

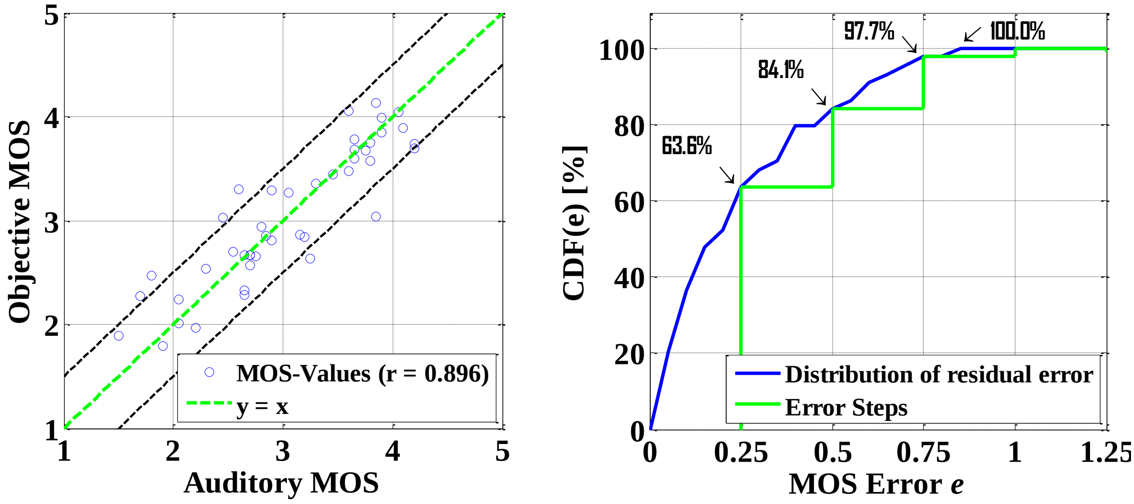

Figure 12: Objective vs. auditory MOS and residual error distribution;
input parameter mΔRA$_{E\text{-}S}$, echo loss and echo delay

The combination of the three parameters *mΔRA$_{E\text{-}S}$*, 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.

3.3. EQUEST as ACQUA Application

For testing EQUEST with ACQUA, the Advanced Communication QUality Analysis system of HEAD acoustics, the following three prerequisites have to be fulfilled:

- The ACQUA Software needs to be version 3.1 or higher
- The separate ACOPT 29 (Option EQUEST, code 6856) needs to be available
- All tests shall be conducted in a quiet environment, ambient noise < 30 dB_{SPL} (A)

A typical setup for EQUEST tests is shown in figure 13. The system represents a standard setup for testing echo on handsets. It should be noted that in general, all different setups for echo testing (mounting on HATS ear; suspending handset in free air; placing handset transducers on plain hard surface; generating variable echo path) are feasible, however validation of EQUEST has been accomplished by mounting handsets on a HATS ear with an application force of 2 newton. This is a common set up of a handset for echo testing.

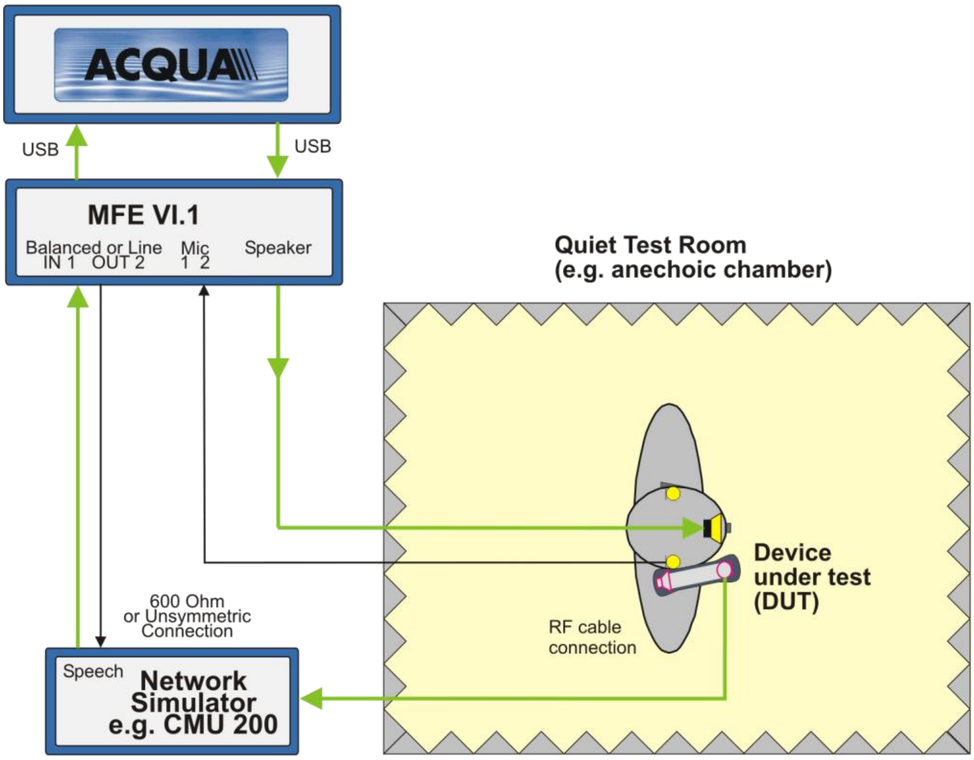

Figure 13: ACQUA setup for EQUEST tests on mobile handsets

Figure 14 demonstrates a sample SMD (Single Measurement Descriptor) of EQUEST. The content of the main SMD entries is as follows.

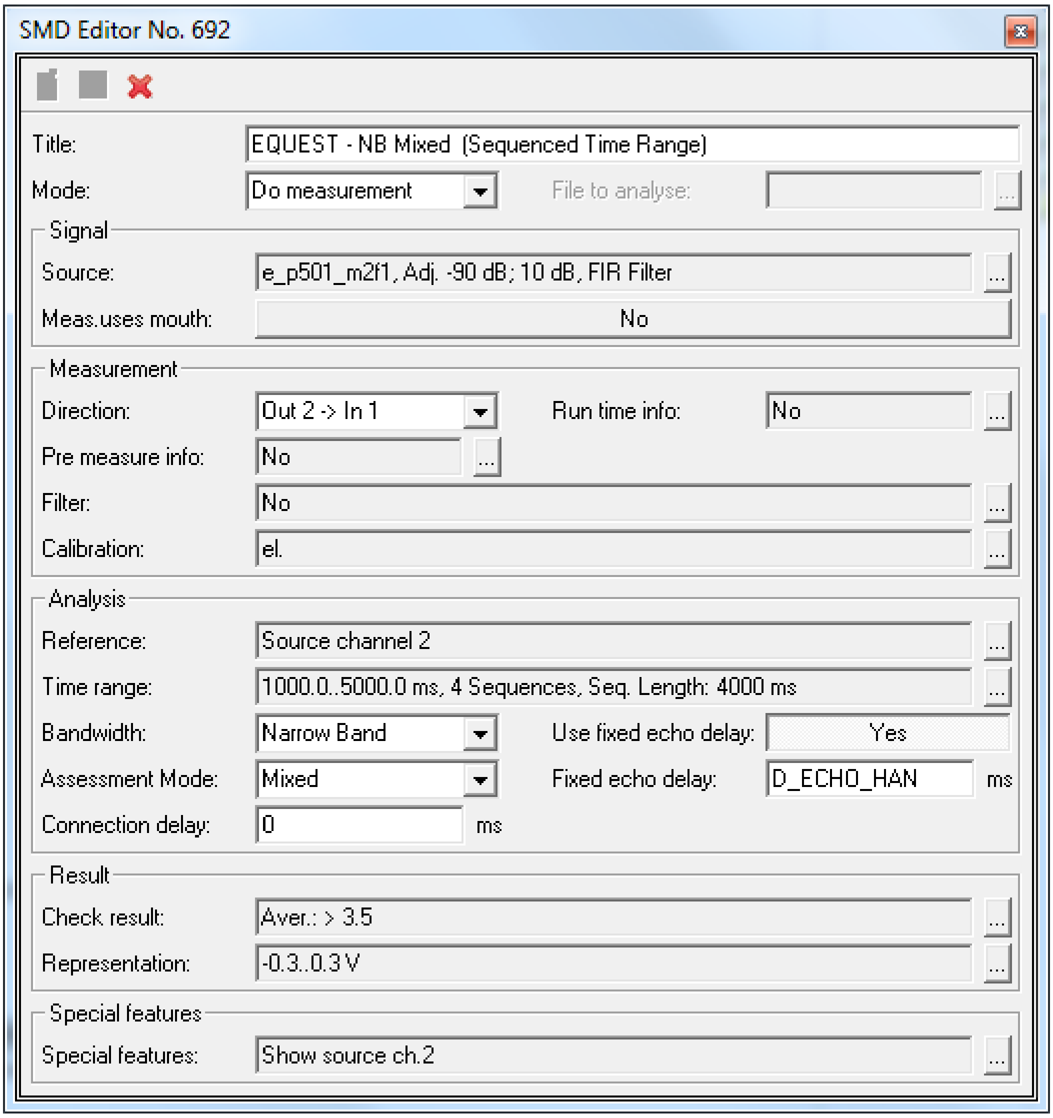

Figure 14: EQUEST SMD

- *Source*: The used stimulus needs to be real speech. Speech samples are made available with the framework of ACOPT 29, option EQUEST. Bandwidth of the source signal needs to match the bandwidth of the DUT (narrow band or wide band). As EQUEST has been validated with specific source signals of ITU-T P.501 [13] it is strongly recommended to solely use these signals. Suitable source files will be provided by HEAD acoustics within the framework of ACOPT 29, Option EQUEST.

- *Reference* is the time signal which reflects the clean speech that is used for calculation of EQUEST. Bandwidth of reference has to be full band (no narrow band nor wide band).
  The source signal can be derived from the reference signal by applying a filter (simulating narrow or wide bandwidths of reference handsets) that has been used for the validation of EQUEST. As the source signals require the corresponding band limitation, these reference filters need to be set. Figure 15 illustrates the two filters (NB / WB), the use in the SMD is shown by figure 14.

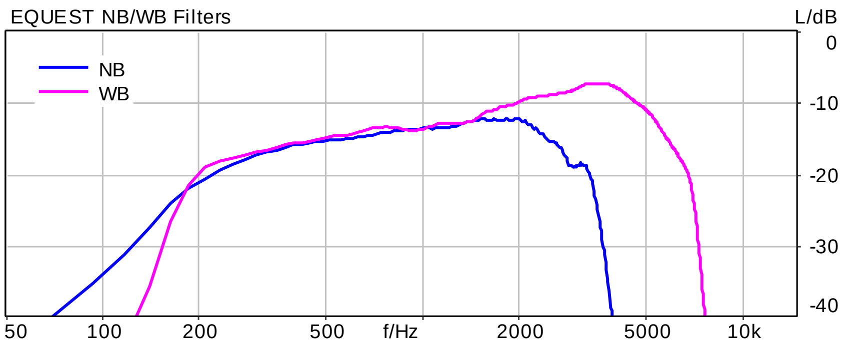

Figure 15: Reference NB/WB filters of EQUEST SMD

- *Time range* needs to be set according to the source signal (i.e. used adaptation sequences need to be taken into account, see SMD entry *Source*).
  Special attention has to be drawn to the *Repeated Sequence* setup of the analysis: The measured time signal can either be analyzed as entire sequence or sentence by sentence of the source file. The final EQUEST MOS score is then calculated as arithmetic mean of these single speech sequences. The SMD settings are shown by figure 16.

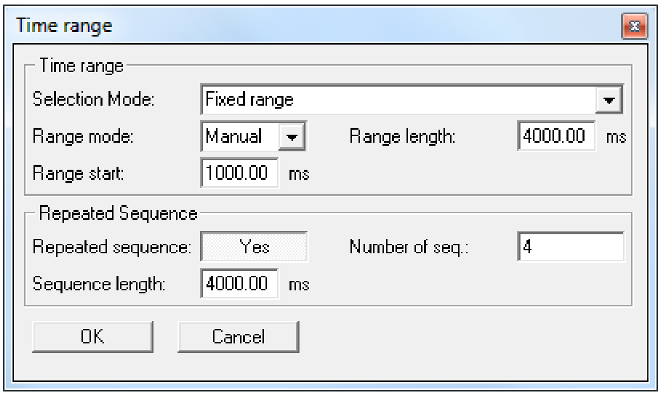

Figure 16: Submenu *Time range* of EQUEST SMD

When analyzing via *Repeated Sequence* it needs to be ensured that the *Time range* entries fit to the source signal: Figure 17 shows the used EQUEST source signal and its *Range length*, *Range start* and *Sequence length*. In this specific case, *Range length* is identical to *Sequence length*, however in general, *Range length* could be shorter than *Sequence length*, depending on the used source signal.

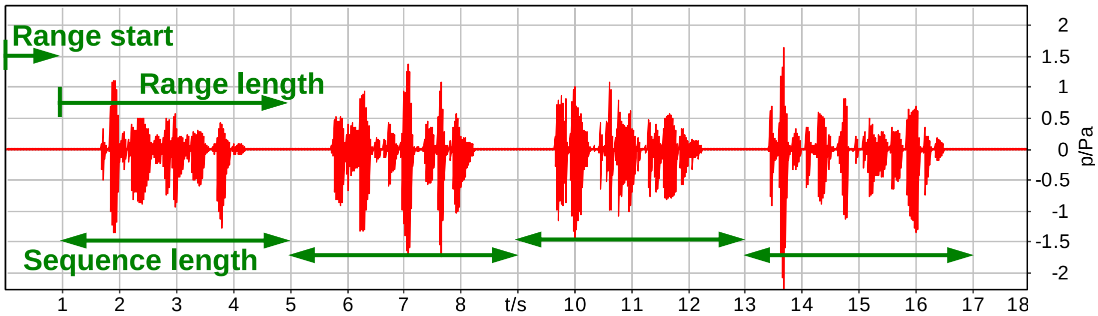

Figure 17: *Repeated Sequence* settings of EQUEST source file

- ***Bandwidth*** is switchable between *Narrow Band* and *Wide Band*, according to the DUT. The Bandwidth needs to correlate with the used source stimulus, see SMD entry *Source*.
- ***Assessment Mode*** depends on the used source signal: The options *Male*, *Female* and *Mixed* shall be chosen according to the speech content of the source signal. A typical use will be *Mixed* where the source signal includes male and female speech.
- ***Connection delay*** represents a simulated connection delay which can be used to insert additional latency to the calculation: When testing terminals that are directly connected to ACQUA (e.g. cordless handsets via MFE X) the measured echo (round trip) delay does not represent the delay that will occur during standard usage in a network. Thus, the delay of the entire setup / network can be simulated by entering the corresponding (additional) *Connection delay*.
- ***Fixed echo delay***: EQUEST offers two methods for delay synchronization:
  - Automatic delay synchronization determined by the EQUEST algorithm: In this case the *Use fixed echo delay* switch is set to *No*. This method has the benefit of calculating the delay automatically. However, it also has the risk of leading to wrong results due to miscorrelation. This can happen when the echo signal is poor and the algorithm may have not sufficient signal energy for a reliable correlation.
  - Manual delay synchronization: In this case the echo delay is measured as a preparation SMD prior to measuring EQUEST. If *Use fixed echo delay* is set to *Yes*, the *Fixed echo delay* can either be entered as a numerical value or as variable (in this case, *D_ECHO_HAN*).

For typical application cases, *Use fixed echo delay* is recommended.

- *Check result*: A requirement for the lower limit EQUEST MOS score can be entered (in the example of figure 14 a minimum MOS of 3.5 has been arbitrarily chosen).
- *Representation*: It is recommended to choose a comparatively low level voltage range so that any residual echo artifacts of the measured echo time signal can be made visible.

The use of the *Assessment Mode* needs some more detailed explanation: Subjective tests on echo perception revealed that echoes of female speech have been perceived as more annoying than male speech when being transmitted under similar conditions. The root cause of this effect is the different spectrum of female speech which has higher high frequency content. This results to a higher echo level when being transmitted via typical handsets due to the handset sending and receiving characteristics.

Thus, the EQUEST algorithm includes three different modes which weight the difference metrics according to a perceived male, female or mixed echo signal. Each weighting can be addressed with the *Assessment Mode* switch within the SMD.

When using the correct combination of source signal and assessment mode (both male or both female), the predicted scores yield the highest correlation according to the subjective listening tests which have been used for the training of the algorithm. In case of using the mixed mode for a pure male or female source signal, test results may slightly vary compared to the ones determined with the "correct" *Assessment Mode*.

Different *Assessment Modes* can be used to evaluate die sensitivity of a DUT with special regard to male or female talkers. A balanced DUT should give similar results for male, female and mixed source signals.

Note: When benchmarking different DUTs, only E-MOS scores with same source signal type and *Assessment Mode* should be compared. Valid combinations for reliable scores (source signal type / *Assessment Mode*):

- Male / Male
- Male / Mixed
- Female / Female
- Female / Mixed
- Mixed / Mixed

4. EQUEST Tests on Phone Handsets

4.1. Measuring EQUEST with ACQUA

For the following tests, handsets of different telephone transmission technology (mobile phone, VoIP & NG-DECT) as well as of different operation modes (handset, hands-free) have been selected. The EQUEST algorithm has been trained using round trip delays of 150 ms or higher, therefore no test scenarios with less than 150 ms are presented.

4.2. Sample Test Results – VoIP / NG-DECT Setup

Figure 18 illustrates the time signal of an EQUEST test result with a VoIP / DECT handset in narrow band handheld hands-free mode (grey: Source, 4 speakers, 2 male & 2 female speakers, green: Measured echo signal, SMD design as per figure 14). A full set of EQUEST result values is shown in figure 19.

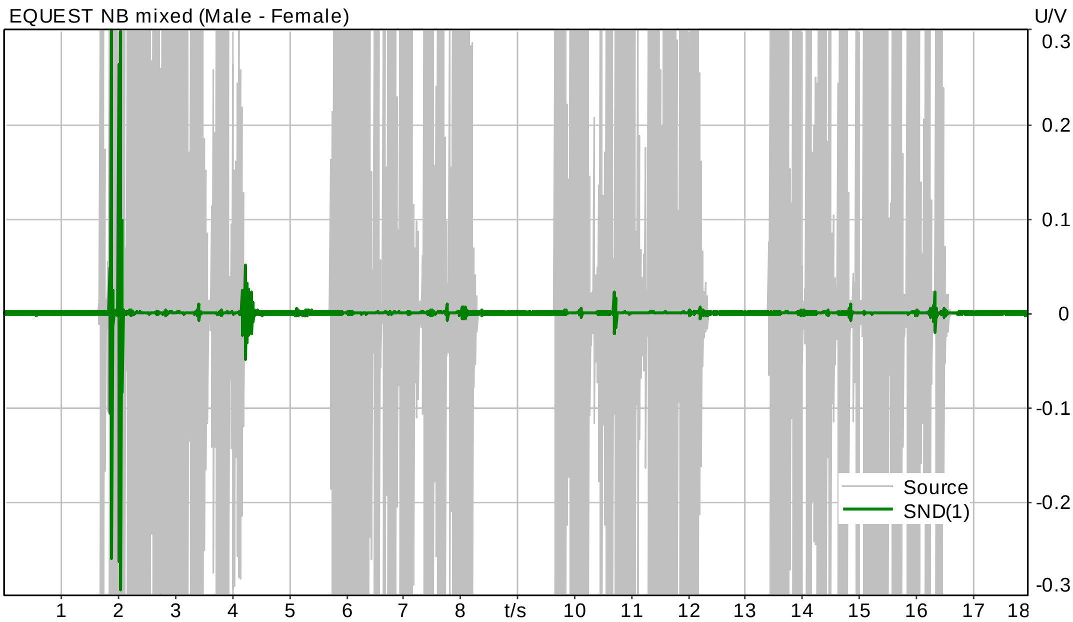

Figure 18: Time signal of EQUEST test result (NB, hands-free mode)

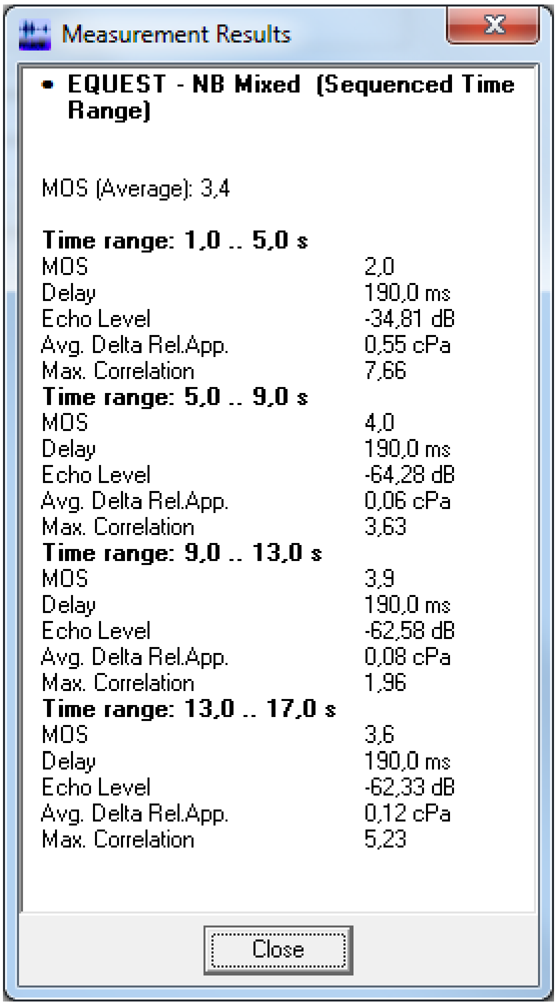

Figure 19: EQUEST result values

The five result values of EQUEST are:

- *MOS*: EQUEST result value (based on Degradation Category Rating, or DCR, acc. to ITU-T P.800 [11]);
- *Delay*: Average echo delay between measured signal and source signal (round trip delay);
- *Echo Level*: Active speech level of echo of measured signal (unweighted echo level, in contrast to TCL_w that is weighted acc. to ITU-T G.122);
- *Avg. Delta Rel. App.*: Average Delta Relative Approach (base for EQUEST result value, high values indicate poor EQUEST scores);
- *Max. Correlation*: Percentage of correlation between source signal and measured echo signal, which can be used as indication of the correlation quality: Low echo levels result to insufficient correlation, for such cases the use of fixed echo delays is recommended.

Figure 18 displays the four test sentences (2 male, 2 female) of the EQUEST *Mixed* source signal in grey color. As the gain of the source is an (approx.) standard electrical signal level it is higher in magnitude as compared to the measured signal. The illustration of the Y-axis has been magnified to a range of -0.3 V – 0.3 V, which is several times larger than usual scaling for signals in sending direction. Note that there is no delay compensation in the display of the signals, so the time shift of the measured signal and source signal is the round trip delay (in this example, 190 ms).

Test 1 – Result Interpretation

The measured signal (green) shows residual echo, at the beginning of the first test sentence and the end. It can also be observed that the measured echo signal bears a strong resemblance to the source. This indicates a low residual echo that is present during the entire signal.

The resulting EQUEST score of MOS 3.4 indicates that the overall echo quality is average up to good. On a DCR a MOS of 3.0 represents that disturbances are “slightly annoying”, DCR MOS 4.0 stands for “audible, but not annoying”.

Figure 20 lists the EQUEST result scores for the overall signal as well as for the four separate single test sentences. The threshold of minimum EQUEST has been arbitrarily chosen value as 3.5 (see figure 14).

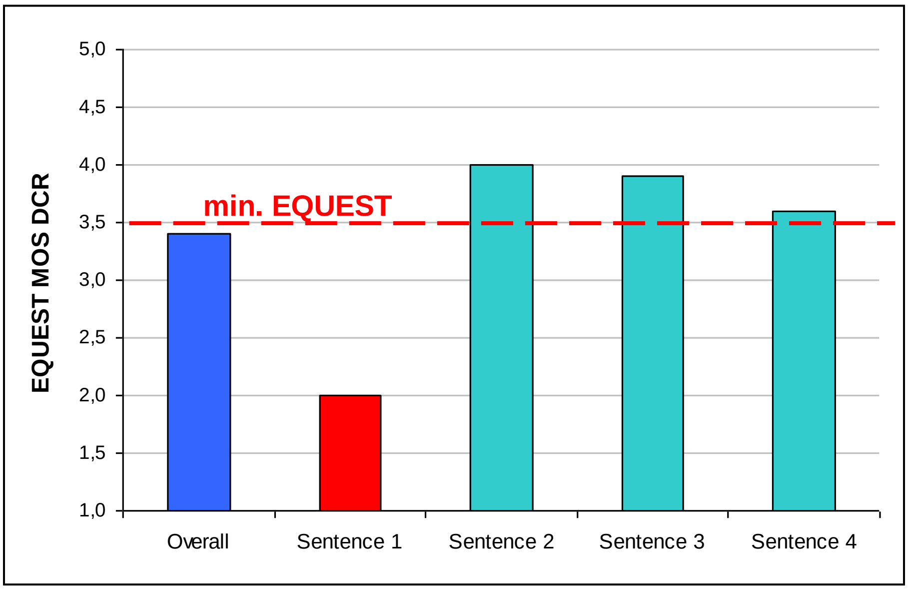

Figure 20: EQUEST result scores for single test sentences

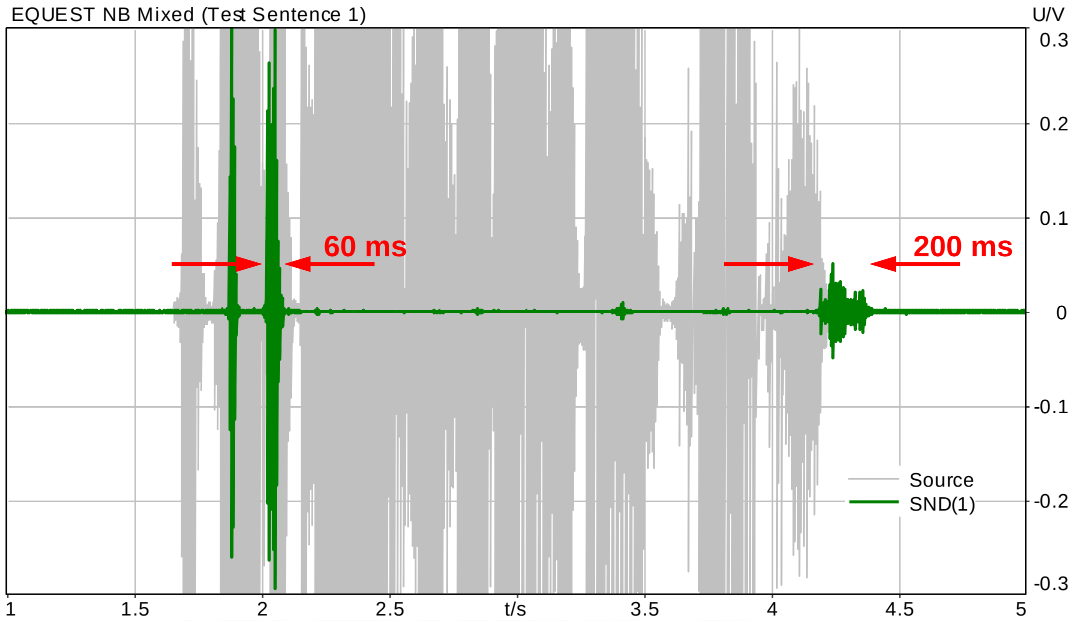

Figure 21: Time signal of test sentence 1

The detailed analysis reveals that the echo level peaks of the first test sentence have a significant impact on the overall EQUEST result. Test sentence 1 only reaches an EQUEST score of MOS 2.0 whereas the other three test sentences reach scores higher than MOS 3.5. When interpreting the time signal of test sentence 1 (see figure 21) it can be assumed that the last echo peak is influencing the score degradation: The first two peaks appear during active speech, while the last peak occurs after finishing the test sentence. Since the masking effect of the talker sidetone is no longer present at the end of the active speech, the echo will be more audible – even though its level is comparatively lower. Moreover, although the first two echo peaks reach considerably higher echo levels, their durations are very short (60 ms of peak 2 in contrast to 200 ms of the last peak, see figure 21).

The masking effect of the talker sidetone can be interpreted as having a significant influence for the comparatively high EQUEST overall result score. All echo disturbances of this sample test result are short, thus they are only slightly annoying (-> MOS DCR 3).

Test 2 – Wideband

In a second test the same DUT is evaluated in wide band handheld hands-free mode. Therefore, the SMD has been modified by two steps:

- Choosing a corresponding FIR filter for wide band limitation of the source signal;
- Switching the *Bandwidth* to *Wide Band*.

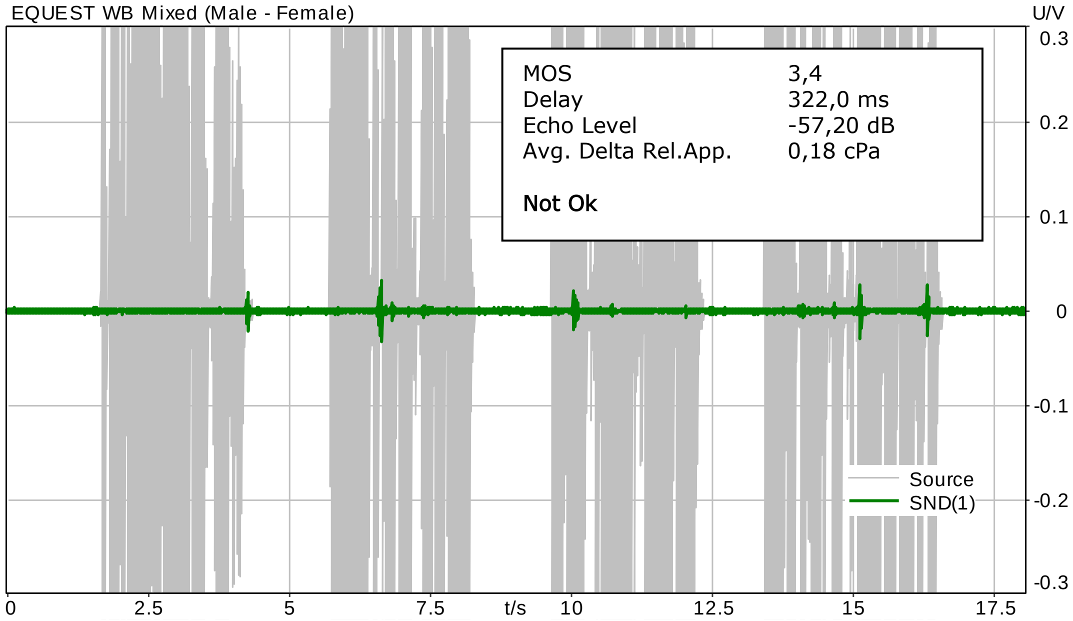

Figure 22: Time signal and result values of EQUEST test (WB, hands-free mode)

Figure 22 presents time signal and result values of the overall EQUEST measurement. Short and low level residual echo peaks are visible, thus the overall EQUEST score attests that echo disturbances are slightly annoying (-> MOS DCR 3) with tendencies for being audible but not annoying (-> MOS DCR 4).

Further objective analysis and subjectively listening to the measured signal provide insight in to the EQUEST result. In figure 23 the frequency analysis of the source and measured signal are presented. A high frequency sound is clearly seen in the FFT spectrum. Thus, the echo signal does not specifically sound annoying, as the sound cannot be easily identified as echo since the subscriber probably cannot clearly correlate the sound to the source. In addition, the residual echo peaks can again be classified as short enough to not be significantly annoying.

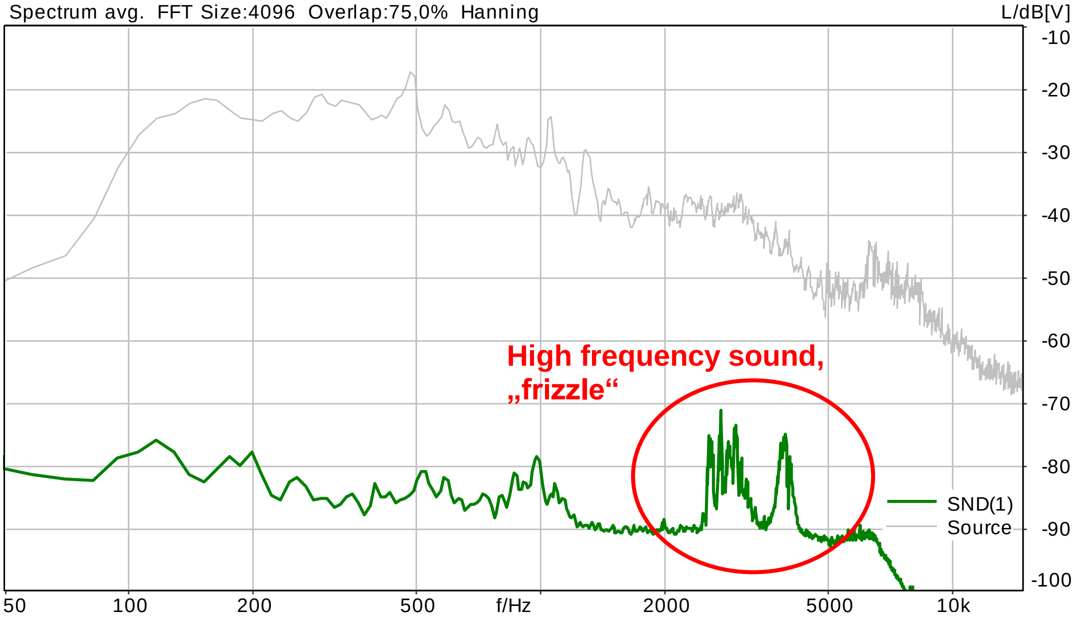

Figure 23: FFT analysis of EQUEST test result (WB, hands-free)

4.3. Sample Test Results – Mobile Phone Setup

In this test series, an up-to-date wide band mobile phone was examined. Figure 24 shows the time signal of handset operation in 3G wide band mode (AMR wide band 12.65 kbps), figure 25 presents the same phone in handheld hands-free operation.

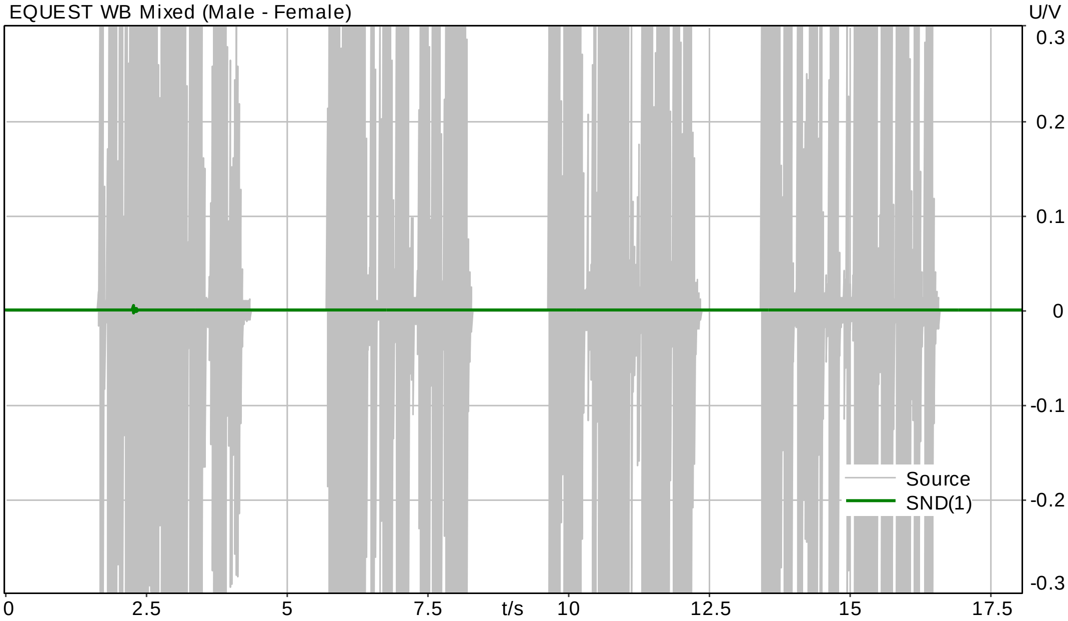

Figure 24: Time signal of EQUEST test (WB, handset mode)

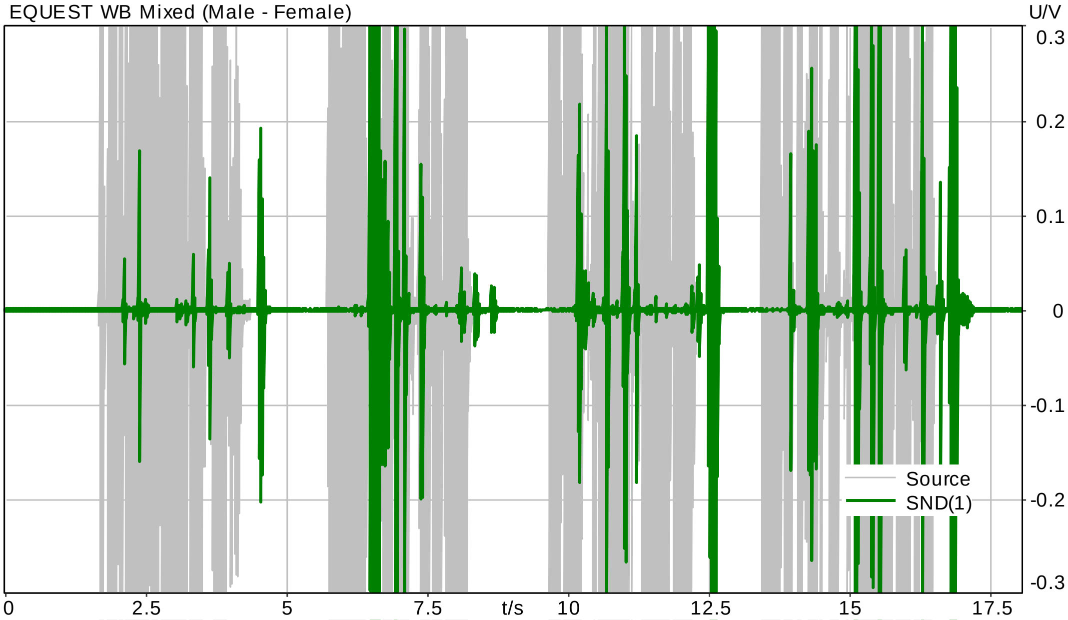

Figure 25: Time signal of EQUEST test (WB, hands-free mode)

Figure 26 lists the result values of the two measurements: As the tests have been carried out in *Repeated Sequence* time range (see figure 16) the MOS scores of all four speech sentences are determined and displayed as well. The *MOS (Average)* value is then calculated as arithmetic mean of the four single MOS scores.

Measurement Results

- **EQUEST - WB Mixed (Sequenced Time range)**

MOS (Average): 1,1

Time range: 1,0 .. 5,0 s	
MOS	1,2
Delay	438,0 ms
Echo Level	-39,50 dB
Avg. Delta Rel.App.	1,32 cPa
Max. Correlation	8,42
Time range: 5,0 .. 9,0 s	
MOS	1,4
Delay	438,0 ms
Echo Level	-24,55 dB
Avg. Delta Rel.App.	2,23 cPa
Max. Correlation	4,70
Time range: 9,0 .. 13,0 s	
MOS	1,0
Delay	438,0 ms
Echo Level	-29,29 dB
Avg. Delta Rel.App.	2,97 cPa
Max. Correlation	3,13
Time range: 13,0 .. 17,0 s	
MOS	1,0
Delay	438,0 ms
Echo Level	-27,25 dB
Avg. Delta Rel.App.	3,60 cPa
Max. Correlation	5,81

Close

Measurement Results

- **EQUEST - WB Mixed (Sequenced Time Range)**

MOS (Average): 4,7

Time range: 1,0 .. 5,0 s	
MOS	4,3
Delay	416,0 ms
Echo Level	-75,72 dB
Avg. Delta Rel.App.	0,06 cPa
Max. Correlation	15,84
Time range: 5,0 .. 9,0 s	
MOS	4,8
Delay	416,0 ms
Echo Level	-79,02 dB
Avg. Delta Rel.App.	0,00 cPa
Max. Correlation	5,07
Time range: 9,0 .. 13,0 s	
MOS	4,8
Delay	416,0 ms
Echo Level	-79,03 dB
Avg. Delta Rel.App.	0,01 cPa
Max. Correlation	3,62
Time range: 13,0 .. 17,0 s	
MOS	4,9
Delay	416,0 ms
Echo Level	-79,19 dB
Avg. Delta Rel.App.	0,00 cPa
Max. Correlation	4,68

Close

Figure 26: Result values of EQUEST test (left: handset mode / right: hands-free mode)

The comparison shows that the DUT echo canceller obviously works sufficiently well in handset mode. The average EQUEST MOS of 4.7 represents almost maximum achievable quality.

However, the time signal and EQUEST scores of the hands-free mode reveal severe quality impairments: MOS scores for the single speech sentences are down to MOS DCR 1.0 which in fact represents the minimum achievable score.

An additional analysis of the FFT vs. time is shown in figure 27. The timing of all echo disturbances correlate with the source signal (round trip delay shift of mobile connection needs to be considered), but all echo peaks show a bandwidth from 100 Hz to 4 kHz. Upon subjectively listening to the measured response the echo does not sound like human speech, but rather noise which could be perceived as annoying during a conversation.

In summary it can be concluded that although the spectral correlation between echo and source is low for this case, the time correlation is high. Together with the high echo level the measured signal provokes a disturbance that is very annoying (-> MOS DCR 1).

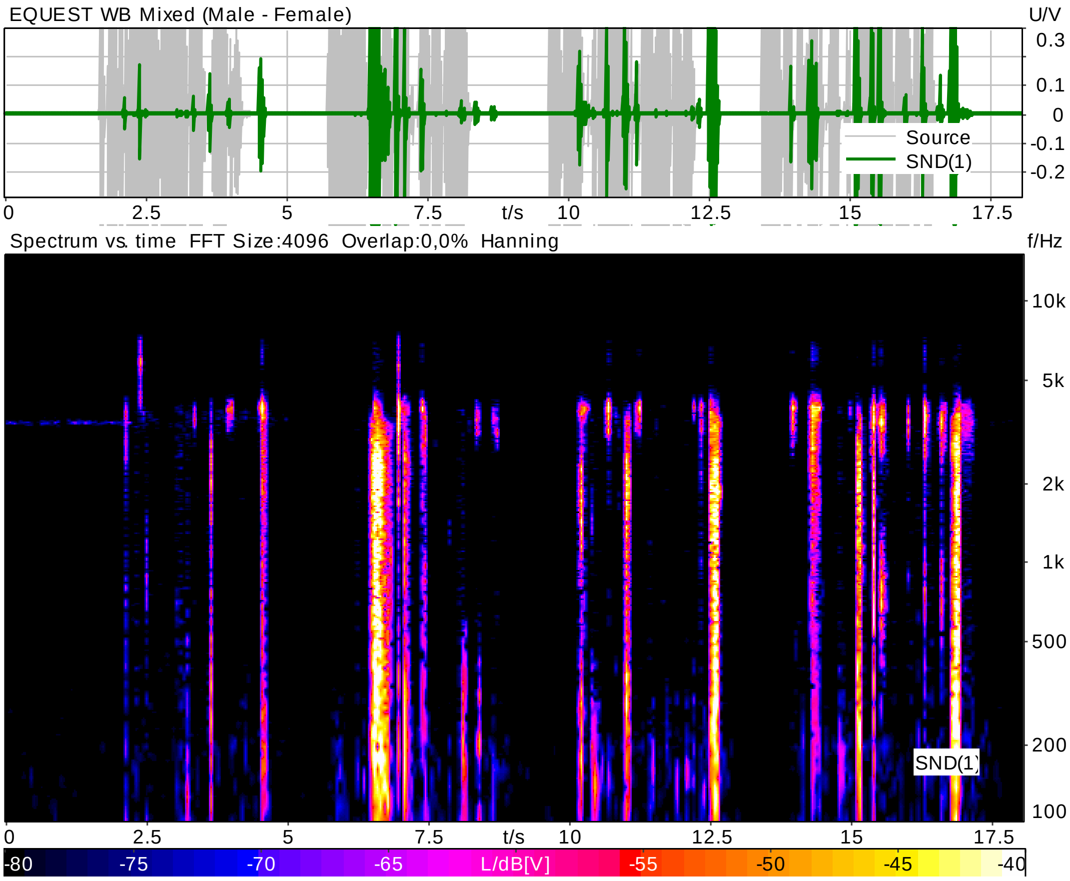

Figure 27: FFT vs. time analysis of EQUEST test result (WB, hands-free)

5. Summary

The extension of the standard parameters echo attenuation and echo delay by the hearing model based EQUEST bears a high potential for echo analyses. The incorporation of the sidetone into the model and its value assessment as well as the approach of using a psychoacoustic parameter leads to a high correlation between EQUEST rating and subscriber's viewpoint. The initial evaluation presents the usefulness of EQUEST for in-depth evaluation of echo disturbances of modern telephone terminal equipment.

It can be noted that the analysis of EQUEST replaces or at least supplements an expert's analysis as shown in the various sample test results. Consequently, EQUEST can be seen as valuable tool for automated echo testing.

The significance of EQUEST particularly manifests in conjunction with the challenges of wide band telephonometry and its specific claims regarding echo performance.

6. 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] ITU-T Recommendation 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] F. Kettler, H.W. Gierlich, E. Diedrich, J. Berger, Echobeurteilung beim Abhören von Kunstkopfaufnahmen im Vergleich zum aktiven Sprechen, DAGA 2001, Hamburg, Germany

[10] Horst Kuchling, Taschenbuch der Physik, 13. Auflage, 1991

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

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

[13] ITU-T Recommendation P.501, Test signals for use in telephonometry, Jan. 2012

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

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