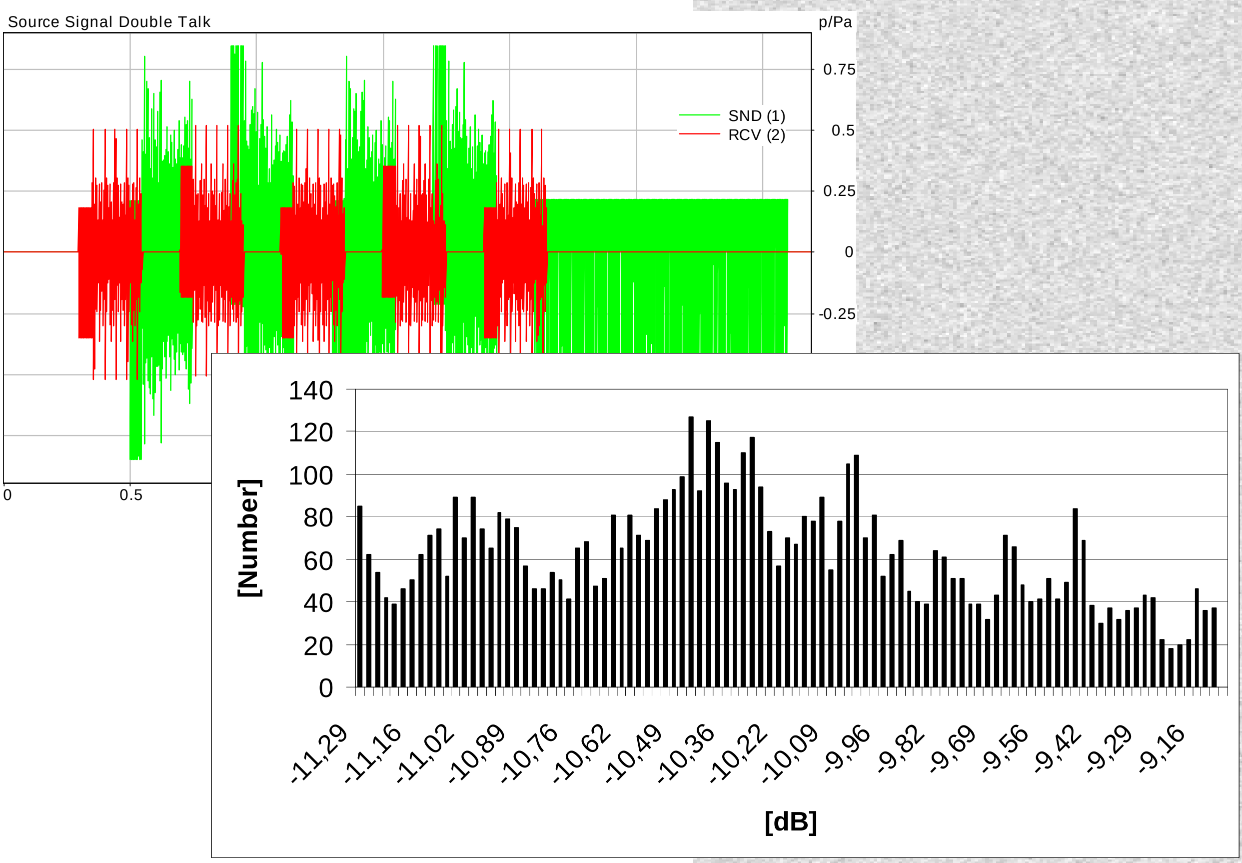
Automated Double Talk Analysis



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

Automated Double Talk Analysis

Contents

1. Introduction

The analysis methods for double talk are mainly described in ITU-T P.340 [1] and P.501 [2]. Based on interlaced double talk composite source signals (CSS) the analysis is realized by subtracting level vs. time analyses of transmitted signal vs. source signal during double talk. Ideally, this results in a flat level vs. time graph, meaning to no difference in attenuation range between single talk and double talk.

However, the established method is often regarded as the "expert's method", since it depends on the interpretation of a trained operator. It is further influenced by time misalignments, non-linear and time variant signal processing and other factors. Thus, an objective and unambiguous analysis that leads to identical results between different labs is not always possible.

The above characterized problem is described in the appendix of ITU-T P.502 [3] which outlines a new approach that allows the automated objective analysis in a clearly defined way. The procedure is based on the use of histograms for double talk analysis.

This application note portrays the method of using histograms for double talk analysis as per Appendix III of ITU-T P.502 [4] using typical test cases.

2. Background and Method

2.1. Current Status of Double Talk Analysis

Figure 1 shows a common test signal (left) as well as a typical result in time domain (right) of a hands-free terminal in sending direction. The corresponding analysis result of attenuation range is given in Figure 2.

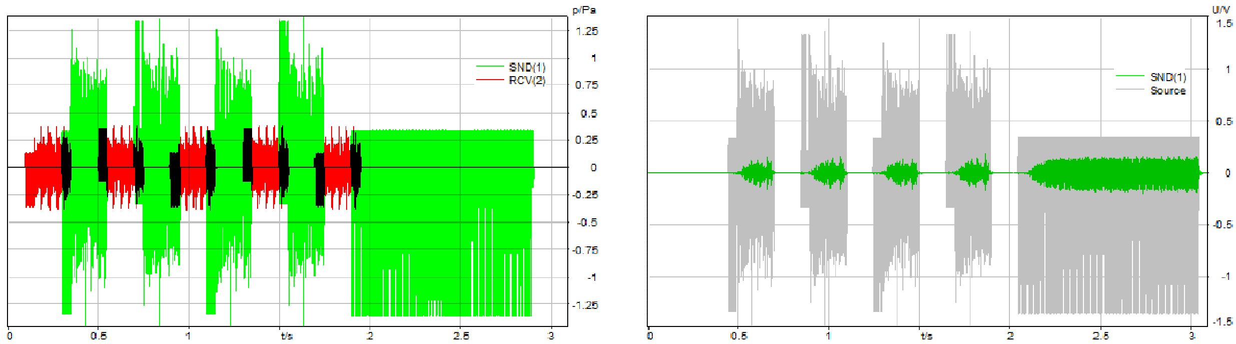


Figure 1: Double talk test signal (left) and example test result time data (right)

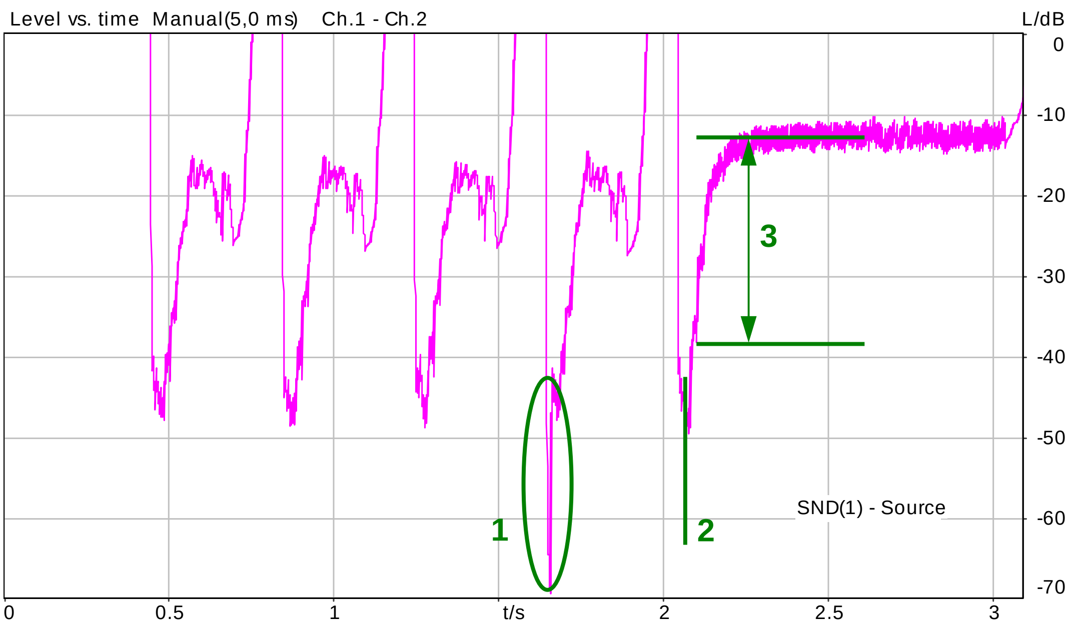


Figure 2: Typical result of a double talk test based on the attenuation range measurement

Figure 2 shows that an unambiguous analysis of the level vs. time curve is almost impossible:

- Different attenuations are inserted at different times. The attenuation range progress is highly time-variant.
- Inclusion of single peaks of the resulting attenuation range curve (see mark 1 in Figure 2) is suspect for two reasons:
    - The common value for the time weighting of the level vs. time analysis is 5 ms, so any time interpretation less than 5 ms can be seen as not meaningful;
    - It is questionable that such extremely short attenuation events would cause any perception (and resulting annoyance) for subscribers – a judgment should also take into account that the subscriber is talking at this moment, so he is masking any attenuation range by his own voice.
- The validity of including these single peaks will need to be regarded when defining the resulting dB score of the attenuation range, see marks 2 and 3 of Figure 2.
- The effect of automatic gain control (AGC) of the device may be superimposed on the effect of attenuation: AGC may add attenuation to the value of attenuation range. However, the pure attenuation range test does not separate between AGC (that would also appear in single talk) and attenuation range due to double talk situation, thus the result might be misinterpreted as impaired double talk performance if not explicitly compensated from AGC behavior.

In addition to the previously described problems in determining the resulting attenuation range, it should be noted that the analysis must be done manually by the operator – so, there is no automated analysis procedure that delivers an objective result score. Thus, the analysis is partially dependent upon the operator's knowledge and experience. In practice, the time signal as well as additional subjective verification (e.g. by listening to double talk speech samples) is often also taken into account.

Consequently, it is very difficult to obtain repeatable results between different operators and different labs.

2.2. New Objective Test Method Based on Histograms

The principle of the automated double talk analysis is shown in Figure 3. The double talk signal played back by the HATS artificial mouth is time aligned by inserting the delay of the DUT in receiving direction and the test system delay ("Δt"). The signal is then fed into the DUT as described by ITU-T recommendation P.340 [1].

The reference signal is not the unfiltered double talk test signal but the transmitted single talk signal for the sending direction: this method will include both the frequency as well as possible AGC characteristics of the DUT. This reference (single talk) signal is determined in a test run prior to the double talk test run and then subtracted from the measured double talk signal. The resulting signal no longer contains AGC effects observed in single talk and is compensated for the spectral characteristics of the DUT.

Double talk analysis is then performed by interpreting each individual CSS burst: the signal is subsequently subdivided into its single time sequences (see Figure 3, in this case 5 sequences).

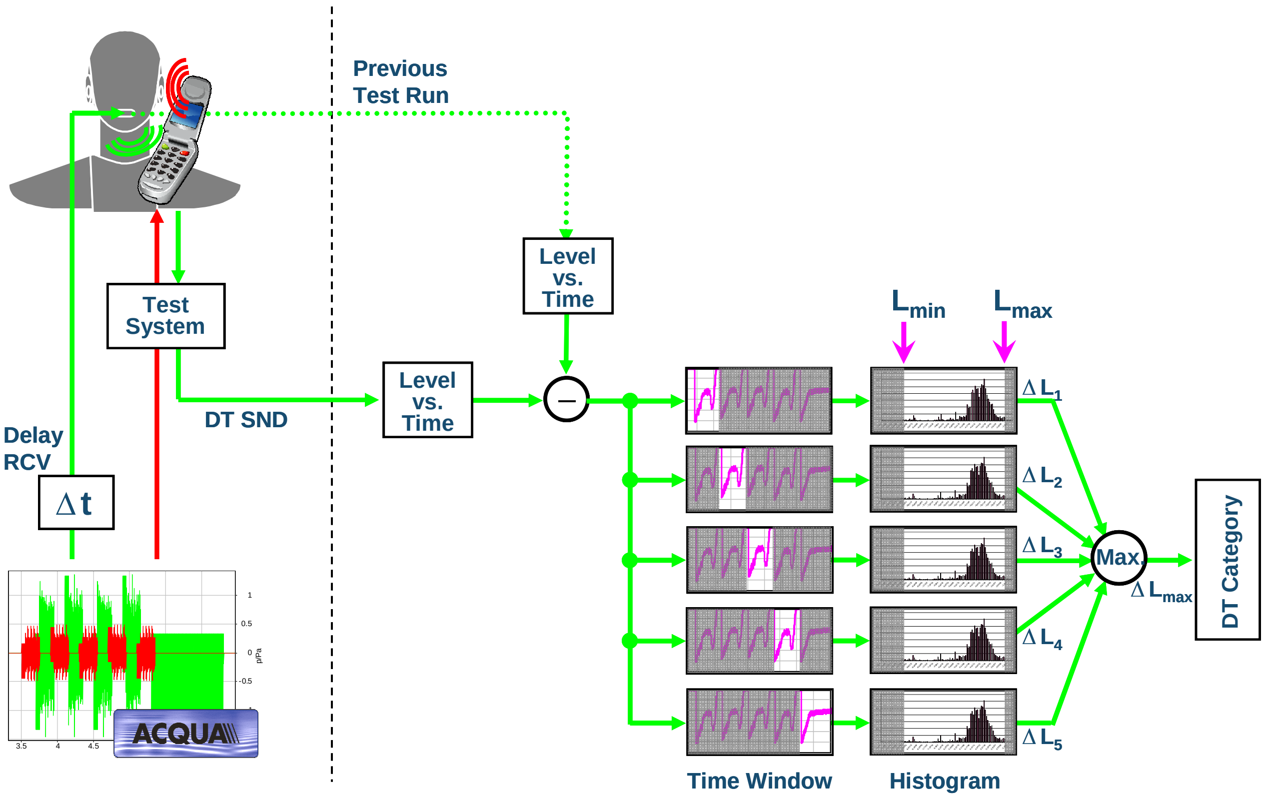


Figure 3: Procedure of automated double talk analysis

A level histogram is created for each CSS burst, and from this level histogram the degree of attenuation is determined ($a_{h,DT,SND}$ and $a_{h,DT,RCV}$ respectively, according to [1]). The resulting maximum $a_{h,DT}$ determines the double talk category (in this example for the sending direction).

Figure 4 presents further details of the histogram creation and the attenuation range result calculation (sending direction, level of transmitted signal referred to level of reference signal):

- The level versus time L(k) is calculated according to IEC 61672 [5] with a time constant of 5 ms for both signals ($L_{DT,SND}(k)$ and $L_{Ref}(k)$).
- The difference between both signals ΔL(k) is calculated as $\Delta L(k) = L_{DT,SND}(k) – L_{Ref}(k)$.
- Minimum and maximum limits for the histogram are derived from minimum and maximum level difference ($\Delta L_{min} = \min\{\Delta L(k)\}$ and $\Delta L_{max} = \max\{\Delta L(k)\}$).
- Division of histogram in 100 equally spaced bins between the minimum and the maximum histogram limits, ΔL_{min} and ΔL_{max}.
- Deletion of the lower 20% and the upper 15% histogram values. New, "effective" histogram limits are given by $\Delta L_{min20\%}$ and $\Delta L_{max15\%}$. This can be interpreted as a smoothing of the curve, which allows the suppression of slight level variations not important for the subjective perception, as well as the suppression of some strong peaks which last only for a short period of time and also are not important for the subjective double talk quality perception.
- Calculation of attenuation range $a_{h,DT,SND}$ as the difference between $\Delta L_{min20\%}$ and $\Delta L_{max15\%}$, i.e., $a_{h,DT,SND} = \Delta L_{max15\%} – \Delta L_{min20\%}$.

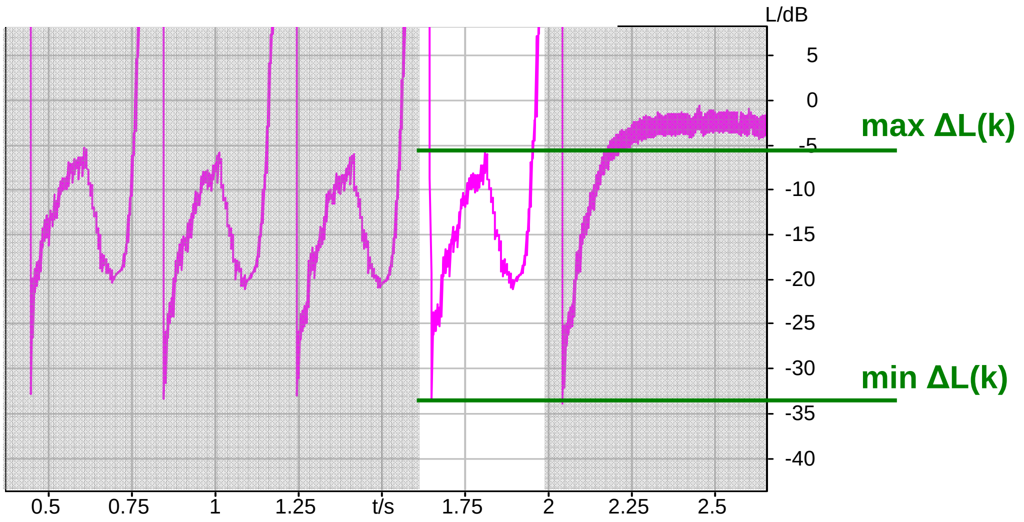


Figure 4: Result of a double talk analysis displayed as level difference versus time

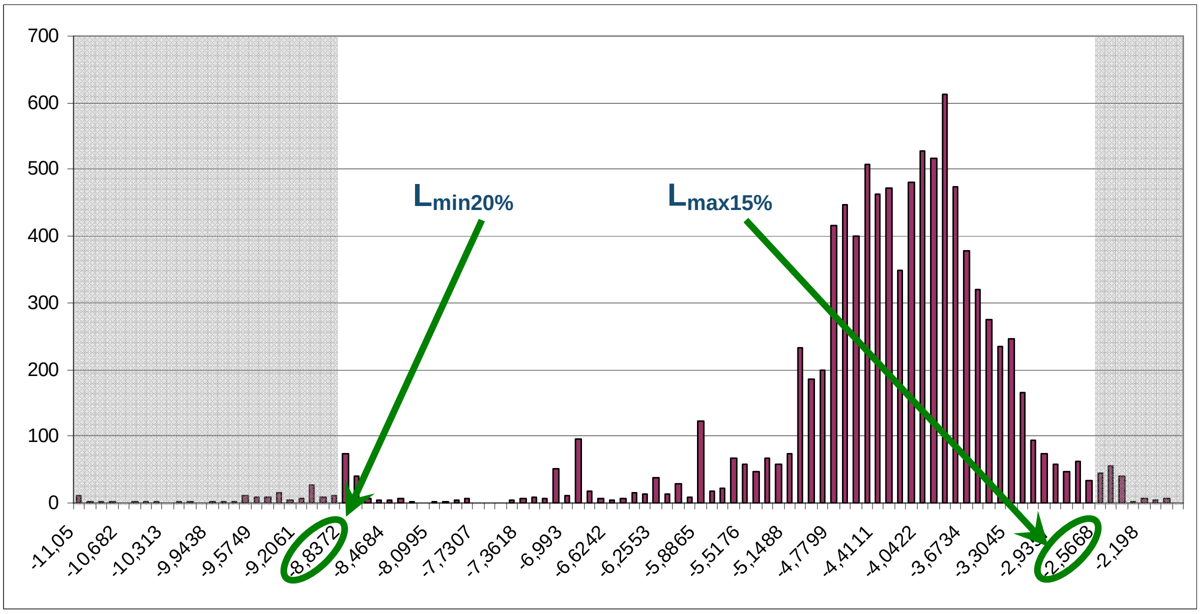


Figure 5: Histogram representation (principle)

The result of this processing is a histogram representation as shown by Figure 5. This histogram is different for each CSS burst.

The deletion of the lower 20% and upper 15% histogram values was developed empirically based on subjective experts' evaluation of 65 different types of mobile phones. By this evaluation, an average attenuation error between the subjective experts evaluation and the described objective procedures is 1,68 dB, which is equivalent to an average classification error of 0,292. Type classification errors have never been greater than 2 classes, 97% of all type classifications match by a difference of no more than 1 class (cf. Figure 6).

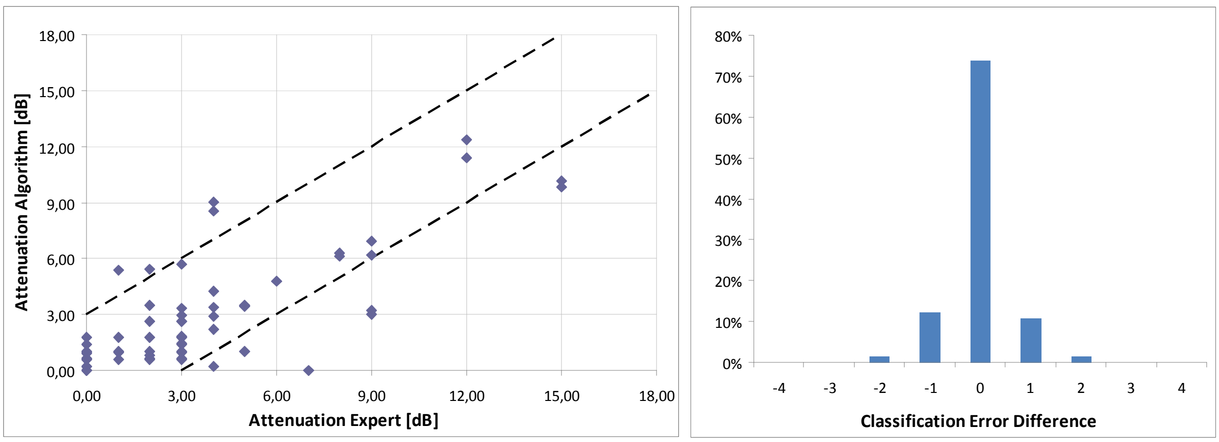


Figure 6: Scatter plot (left) and error difference (right) of automated double talk evaluation

The deletion of a percentage of histogram values may lead to similar Attenuation Range values in spite of widely varying level distributions. An example of this is the level vs. time analysis of a mobile phone shown in Figure 7. The extremely short front end clipping of the first CSS burst (mark 1 of Figure 7) results in “wide” level bins with most samples tightly bunched in a relatively small number of bins. Elimination of the lower 20% of values and upper 15% of values leads to an Attenuation range value that is much closer to that of the third CSS burst (mark 2 of Figure 7), which starts out with a much smaller level distribution.

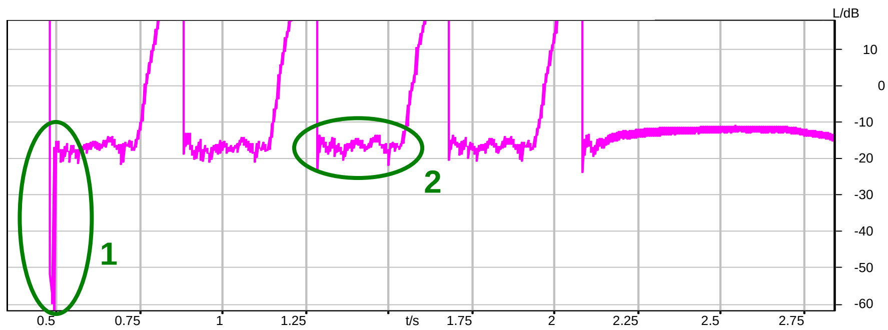


Figure 7: Result of a double talk analysis of a mobile phone

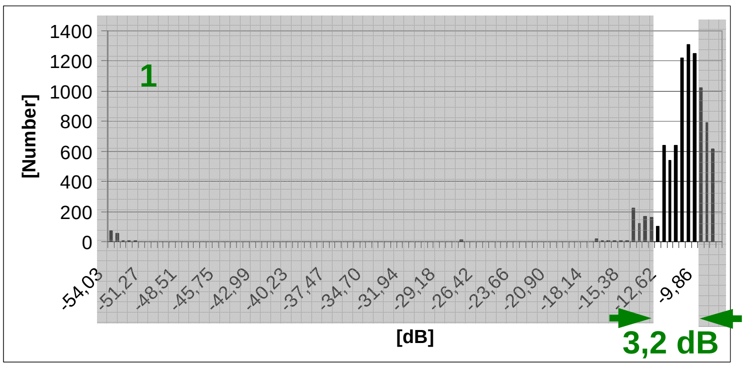


Figure 8: Histogram of first CSS burst (cf. mark 1 of Figure 7)

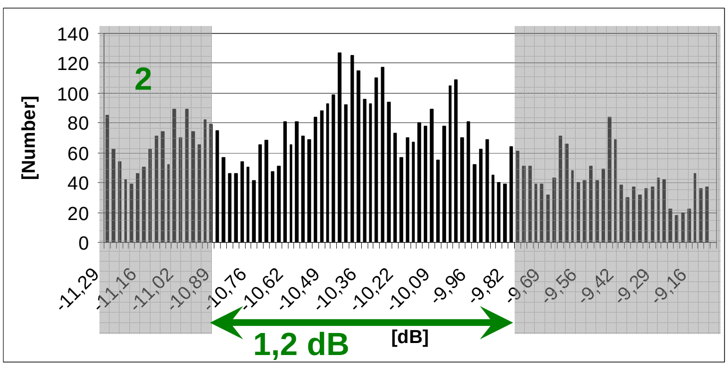


Figure 9: Histogram of third CSS burst (cf. mark 2 of Figure 7)

3. Application, Test Setup & Results

3.1. Realization in ACQUA

When using the new Automated Double Talk Analysis in ACQUA, the Advanced Communication QUality Analysis system created by HEAD acoustics, the following prerequisites must be fulfilled:

- The ACQUA Software must be version 3.1 or higher.
- The used standard must include the new SMD type *Automated Double Talk*. Existing standards must be supplemented accordingly.

A typical setup for automated double talk tests is shown in Figure 10. The system represents a standard setup for handset double talk tests.

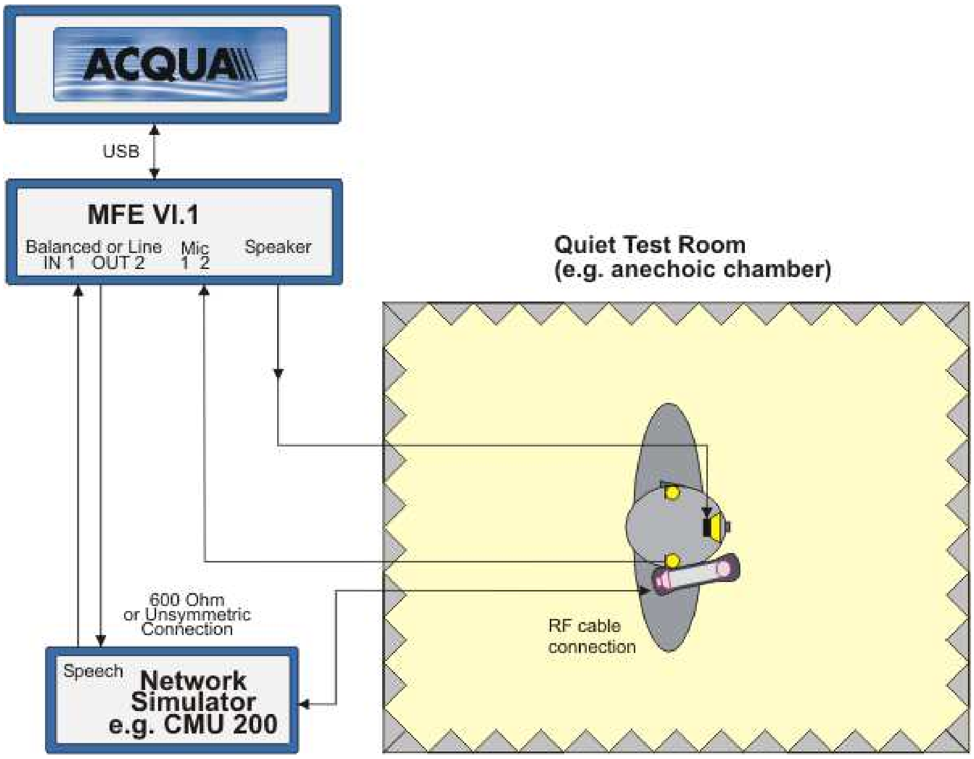


Figure 10: ACQUA setup for double talk tests on mobile handsets

Figure 11 demonstrates a sample SMD (Single Measurement Descriptor) of automated double talk in sending direction. The content of the main SMD entries is as follows:

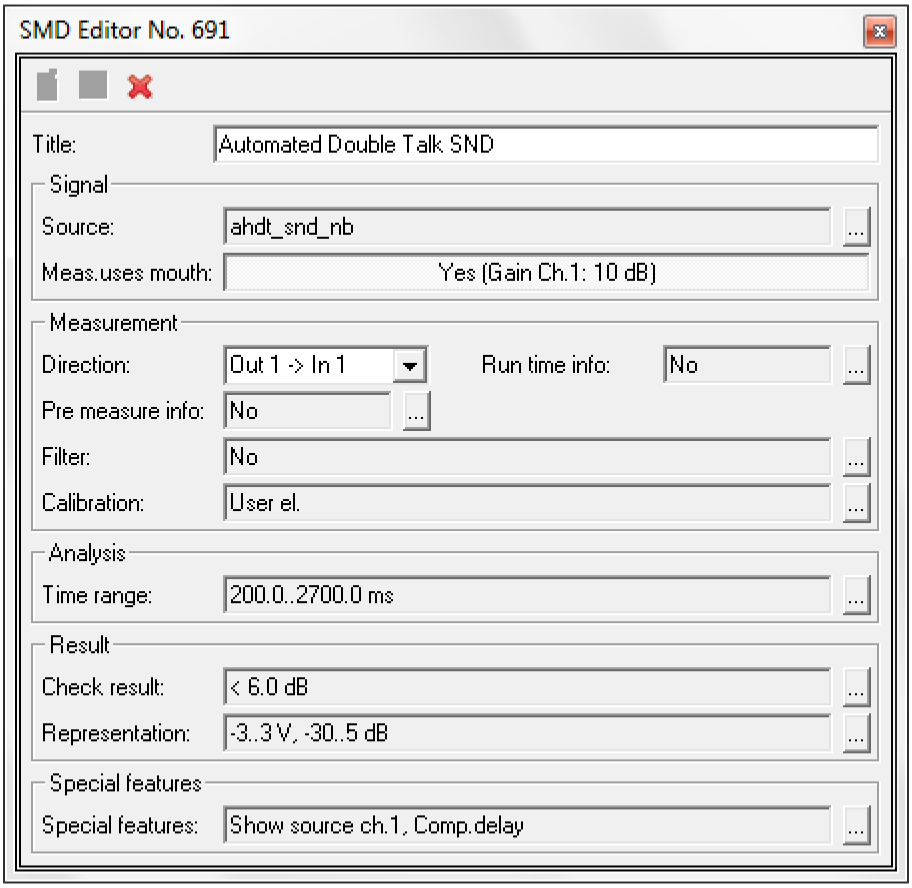


Figure 11: Screenshot of automated double talk SMD

- ***Source***: The stimulus listed by Figure 11 is a standard signal for double talk. The bandwidth of the source signal needs to match to the bandwidth of the DUT (narrow band or wide band) for the RCV channel.
  In case where an adaptation sequence is planned before the test signal, the adaptation sequence needs to be appended to the beginning of the source signal, adjusting the *Time range* accordingly.
- ***Direction*** is applied as standard direction of ACQUA tests, in this case *Out 1 -> In 1* for sending direction. For receiving direction, *Out 2 -> In 2* would be used.
  Direction also determines the channel of the single talk run that is transmitted as previous test run (run 1) to identify single talk level vs. time (cf. Figure 3) before playing back the double talk sequence as run 2.

- ***Time range*** needs to be set to the entire analysis area, i.e. the complete time window and not a time window of a certain CSS burst sequence. Thus, the SMD analyses the whole source signal and also delivers the rating of the resulting double talk type – it is not necessary to set up several SMDs for each single burst.
  As noted above, if an adaptation sequence is included in the source signal before the double talk sequence, make sure to exclude the adaptation sequence time from the time range since so that it is not included in the automated double talk analysis.
- ***Check result***: The requirement can be entered as dB value. Figure 11 illustrates a requirement of maximum 6 dB attenuation range SND (which represents a double talk type 2a).
- ***Special features***: Here, *Compensate delay* needs to be regarded.

For the entire test, two delay compensations have to be considered:

- The delay in receiving direction $\Delta t_{OUT\ 2}$ that occurs as the signal passes through the test system, including the coder and decoder of a radio communication tester (cf. Figure 12, MFE VI.1 Out 2). This delay is added to the sending direction so that the resulting double talk output signal is synchronized at the MRP of the HATS. This delay compensation is executed by the entries *Send channel delayed* of the SMD submenu *Source*.
- The delay in sending direction $\Delta t_{IN\ 1}$that occurs as the signal passes through the test system (cf. Figure 12, MFE VI.1 In 1), is compensated by the entry *Compensate delay* of the SMD submenu *Special features*. By this compensation, the measured signal is delay compensated to its source signal.

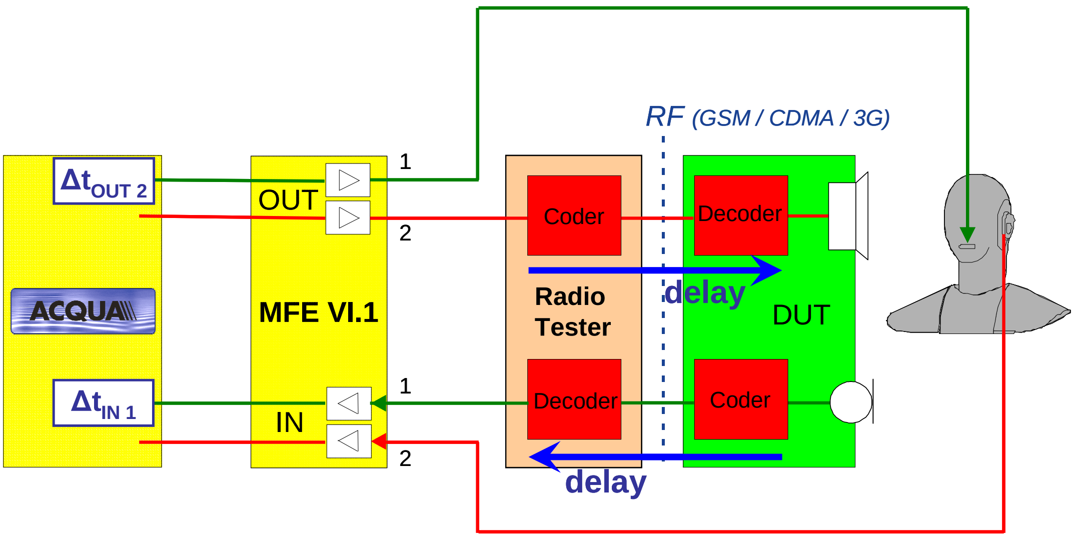


Figure 12: Delay compensation in measurement setup

The automated double talk analysis method is not fundamentally restricted to the use of CSS bursts as source signals: As any source signal could be used by the SMD, the analysis method could also be applied for other double talk signals, e.g. real speech. However it must be stated that the histogram method has not (yet) been verified based on signals other than CSS, and it can be assumed that the histogram treatment (deletion of value percentages etc.) needs to be adapted. This work is under study.

Although the automated double talk analysis is principally applicable for both signal directions (sending and receiving), the method has been validated for sending direction only: For receiving direction, all interpretation of double talk needs to consider the presence of sidetone, thus the overlay of sidetone causes difficulties in any interpretation – subjective as well as automated. However, the SMD structure itself allows the use of both signal directions.

ACQUA does not provide automated double talk analysis as a post-analysis option in ACQUAlyzer: The method of automated double talk analysis is only applicable in cases where a DUT has been measured under single talk and double talk conditions by using the same test conditions and the identical test signal. Thus, a result signal cannot be analyzed and interpreted in (later) post processing.

3.2. Sample Test Results

The following test results represent typical findings of double talk behavior of modern mobile phones. The tests have been carried out using a Radio Communication Tester in 3G (WCDMA) narrowband mode (AMR NB, 12.2 kbps) and wideband mode (AMR WB, 12.65 kbps). The test setup follows the example shown in Figure 10.

Test 1

Figure 13 illustrates the automated double talk analysis result window of a mobile phone in sending direction (SMD design as per Figure 11). A full set of corresponding result values is shown in Figure 14.

The automated double talk analysis window is the result of the double talk analysis curve offset by the single talk analysis curve. Thus, the analysis window is generated based on two time signals and is therefore not only correlated to the time signal window of the double talk run. The plot is calculated as level vs. time and shows the attenuation range of the DUT. Periods of no signal in sending direction or of negative attenuation (which would be equal to amplification) are set to 0 dB (cf. Figure 13).

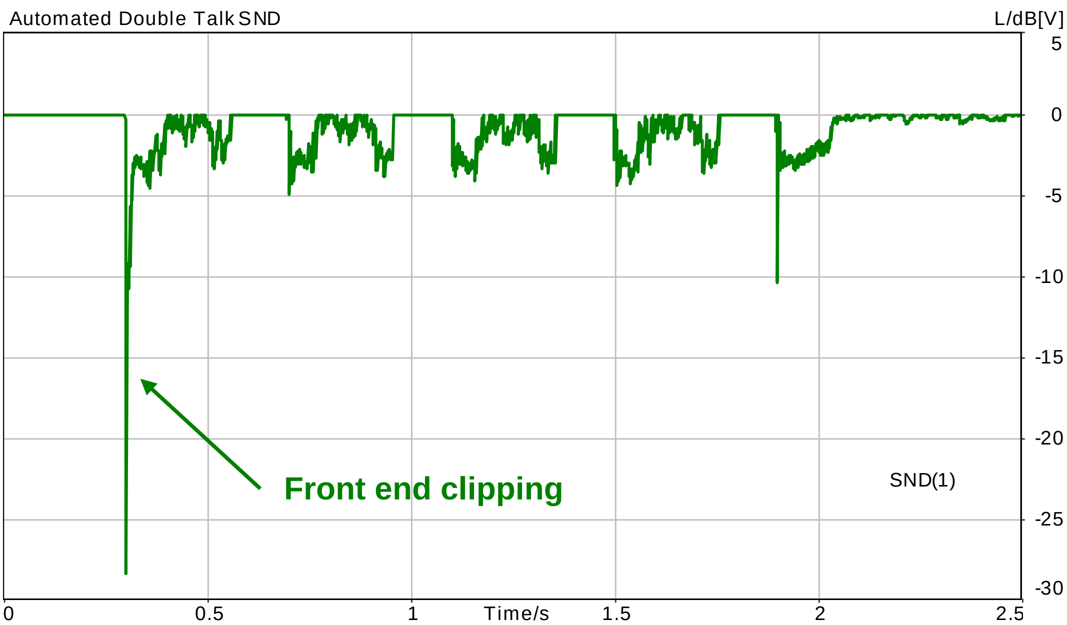


Figure 13: Analysis result signal of automated double talk test

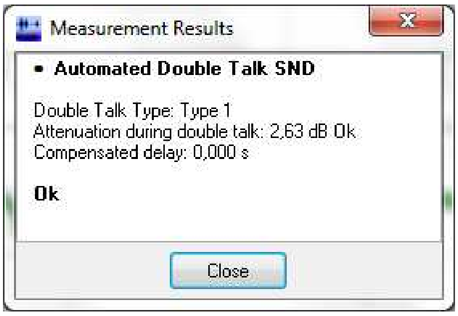


Figure 14: Automated double talk result values

The three results values of an automated double talk test in detail:

- ***Double Talk Type*** is based on ITU-T P.340 categories [1]. Table 1 lists the double talk types for sending direction.
- ***Attenuation during double talk*** expresses the resulting maximum attenuation of the calculation and determines the *Double Talk Type*.
- ***Compensated delay*** indicates delay jitter that may appear during playback of the test sequence: This event may especially happen when testing e.g. VoIP applications. Any delay jitter is compensated by the algorithm.

Double Talk Type	**1**	**2a**	**2b**	**2c**	**3**
Attenuation SND	≤ 3 dB	≤ 6 dB	≤ 9 dB	≤ 12 dB	> 12 dB

Table 1: Double Talk Types for sending direction

The analysis window (Figure 13) as well as the resulting double talk type (type 1, cf. Figure 14) indicates that the measured mobile phone provides an excellent performance: Overall attenuation range is low, initial peaks are too short to cause annoyance. Due to the histogram calculation that deletes certain percentages of values the initial peaks (front end clipping) are not taken into account for the attenuation range calculation.

Test 2

The sending direction of a second mobile is tested in handset mode by using WCDMA wide band codec (Figure 15). Detailed analysis is available in ACQUA by displaying the level vs. time curves of the single talk test run, double talk test run and resulting curve in the ACQUA diagram (Figure 16).

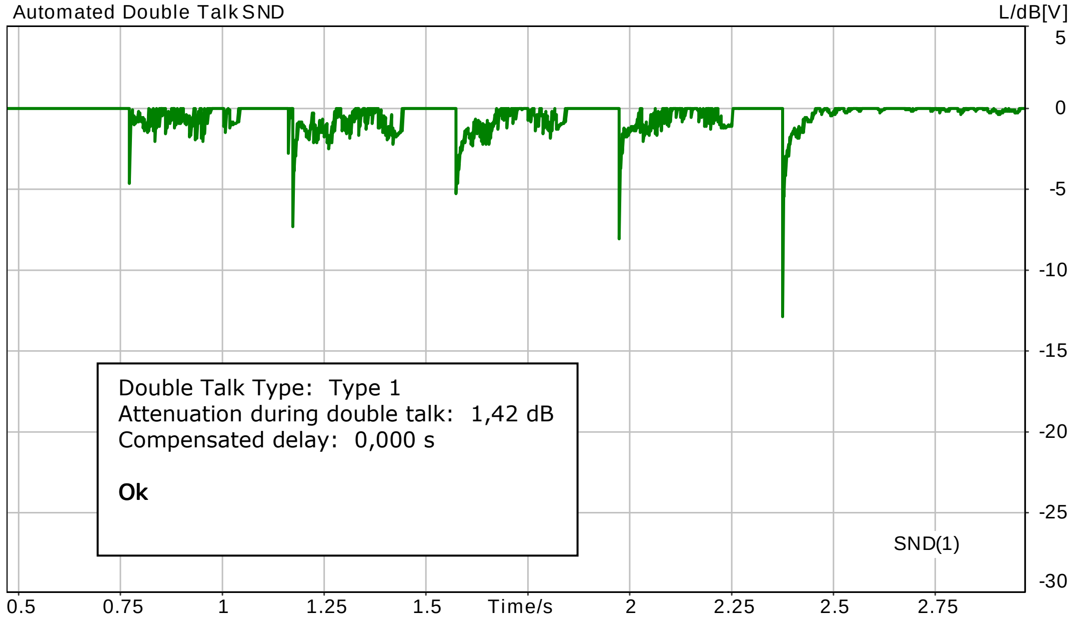


Figure 15: Analysis & result values of automated double talk test

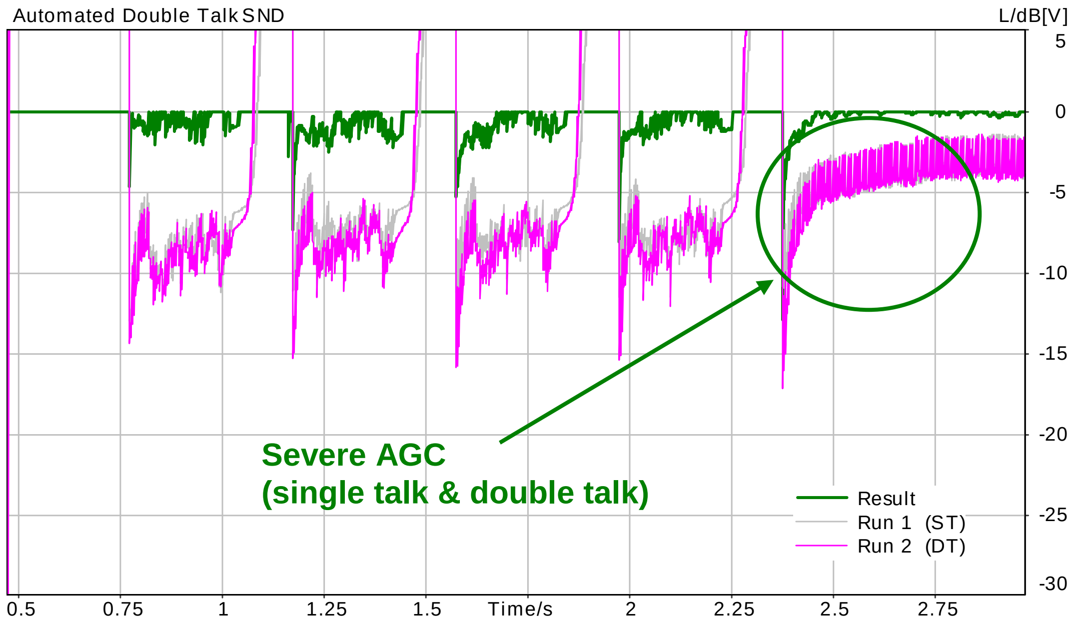


Figure 16: Resulting (green), single talk (grey) & double talk (magenta) analysis

Figure 16 illustrates strong AGC behavior of the DUT. However, the AGC performance appears in single talk as well as in double talk, so it must not be misinterpreted as double talk attenuation range. Consequently, the result is (correctly) adjusted by this effect and leads to an attenuation range of 1,42 dB only – which represents excellent double talk performance.

Test 3

The mobile phone of Test 2 will now be examined in handheld hands-free mode (again AMR WB 12.65 kbps). Time signals of the two single runs (single talk / double talk) are shown in Figure 17. The corresponding analysis level vs. time of single talk and double talk runs as well as main results are presented by Figure 18.

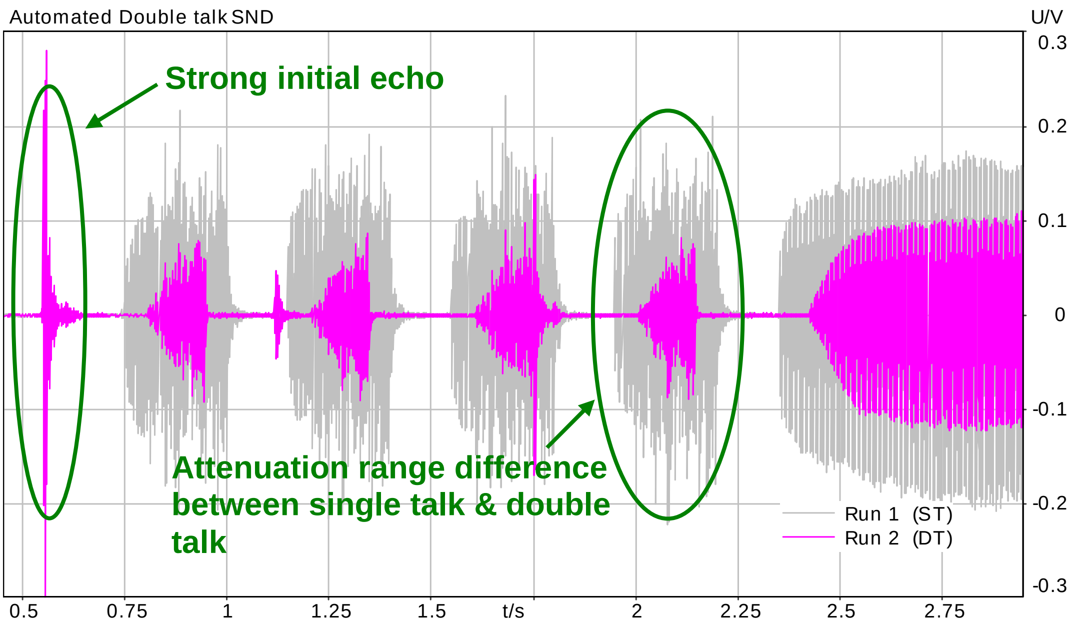


Figure 17: Time signals of single talk (grey) and double talk (magenta) runs

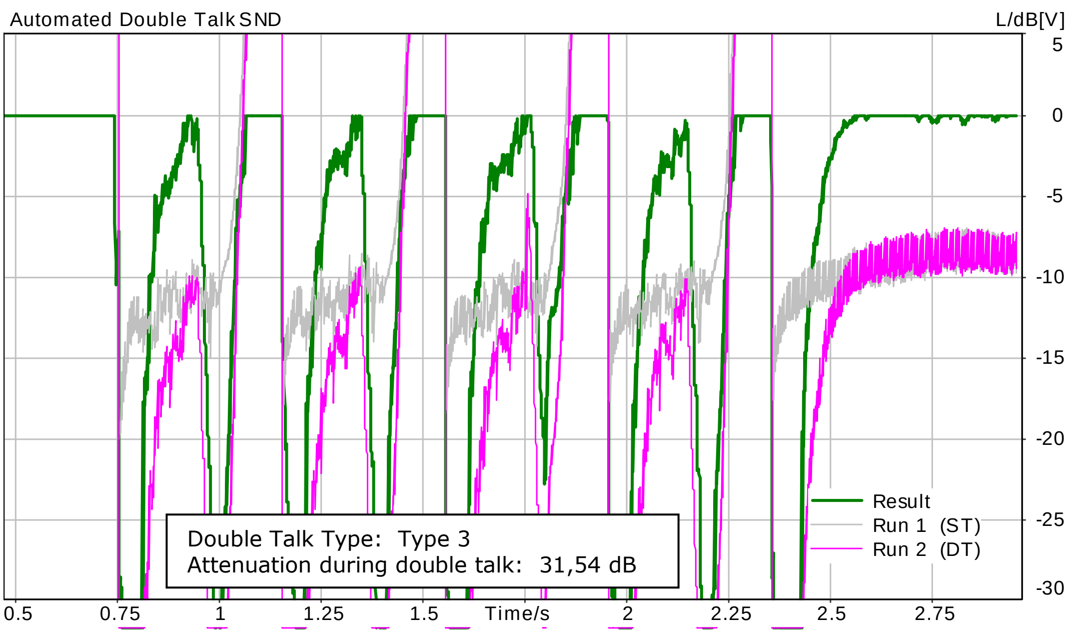


Figure 18: Resulting (green), single talk (grey) & double talk (magenta) analysis

In contrast to the handset performance, the handheld hands-free performance of the DUT is bad: The resulting attenuation range is 31,54 dB, consequently the double talk type is type 3. The resulting curve (cf. Figure 18) shows a strong increase and also decrease of attenuation gain which is typical for very limited double talk behavior.

Moreover, the underlying time signal of run 2 (cf. Figure 17, double talk run) indicates a strong but short initial peak at the beginning of the test signal playback. As the used source signal only provides signal content in receiving direction at this moment, the peak can be identified as an echo component – which can be seen as sort of opposite performance to attenuation. Thus, it does not represent attenuation and is therefore not of interest for judging the double talk attenuation performance.

Test 4

A new setup based on a different mobile phone is now examined: Figure 19 presents the level vs. time curves of the single talk test run, double talk test run and resulting curve as well as the main result values.

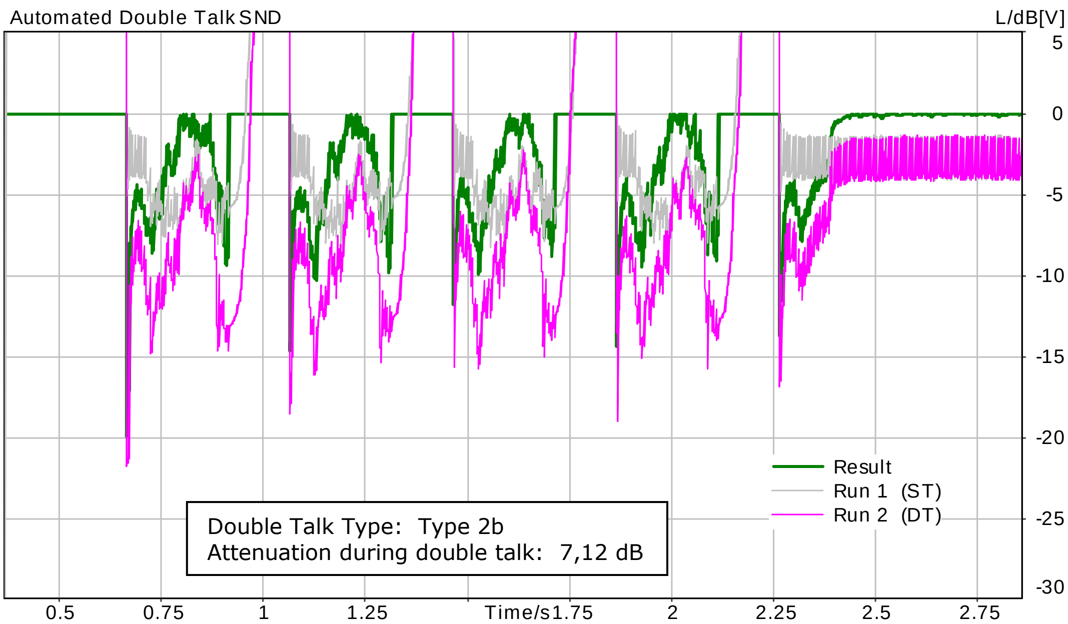


Figure 19: Resulting (green), single talk (grey) & double talk (magenta) analysis

The test result shows a type 2b double talk device which represents typical "average" double talk quality: A certain attenuation range is given, however attenuation appears comparatively smooth and short enough to be only slightly annoying.

4. Summary

Compared to the established method of manual level vs. time analysis, the automated double talk analysis based on histogram interpretation offers a number of benefits:

- The implementation of two runs (single talk and double talk) provides adequate consideration of AGC performance of the DUT;
- The result value is stable, reproducible and therefore fully objective, as it is based on an automated and purely mathematical algorithm and no longer takes the operator's rating into account;
- Due to this fact, complete conformity between different test labs is ensured;
- The method provides a constantly high correlation to the subscriber's viewpoint;
- The necessity of an "expert's viewpoint" for judging double talk results is replaced or at least supplemented, thus the result quality is not entirely dependent on the operator's experience.

The automated double talk analysis can therefore be seen as sophisticated, reliable and easy-to-use approach to determine double talk attenuation range of telecommunication terminals.

5. References

[1] ITU-T Recommendation P.340, Transmission characteristics and speech quality parameters of hands-free terminals

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

[3] ITU-T Recommendation P.502, Objective test methods for speech communication systems using complex test signals

[4] ITU-T Recommendation P.502 Amendment 1, Objective test methods for speech communication systems using complex test signals, Amendment 1: New Appendix III – Automated double talk analysis procedure

[5] IEC 61672, Electroacoustics – Sound level meters – Part 1: Specifications