INTERNATIONAL TELECOMMUNICATION UNION

ITU-T

TELECOMMUNICATION=
STANDARDIZATION SECTOR
OF ITU

P.502

(05/2000)

SERIES P: TELEPHONE TRANSMISSION QUALITY, TELEPHONE INSTALLATIONS, LOCAL LINE NETWORKS

Objective measuring apparatus

Objective test methods for speech communication systems using complex test signals

ITU-T Recommendation P.502

(Formerly CCITT Recommendation)

ITU-T P-SERIES RECOMMENDATIONS

TELEPHONE TRANSMISSION QUALITY, TELEPHONE INSTALLATIONS, LOCAL LINE NETWORKS

Vocabulary and effects of transmission parameters on customer opinion of transmission quality	Series	P.10
Subscribers' lines and sets	Series	P.30 P.300
Transmission standards	Series	P.40
Objective measuring apparatus	**Series**	**P.50** **P.500**
Objective electro-acoustical measurements	Series	P.60
Measurements related to speech loudness	Series	P.70
Methods for objective and subjective assessment of quality	Series	P.80 P.800
Audiovisual quality in multimedia services	Series	P.900

For further details, please refer to the list of ITU-T Recommendations.

ITU-T Recommendation P.502

Objective test methods for speech communication systems using complex test signals

Summary

This ITU-T Recommendation describes methods and procedures for the evaluation of complex terminals, network components and transmission systems. The test methods mostly make use of test signals described in ITU-T Recommendations P.50, P.59 and P.501. For various technical implementations and conversational situations, the possible impacts on the speech quality perceived subjectively are given and the relevant measurement procedures are described.

Source

ITU-T Recommendation P.502 was prepared by ITU-T Study Group 12 (1997-2000) and approved under the WTSC Resolution 1 procedure on 18 May 2000.

Keywords

Analysis methods, double talk, single talk, speech quality.

FOREWORD

The International Telecommunication Union (ITU) is the United Nations specialized agency in the field of telecommunications. The ITU Telecommunication Standardization Sector (ITU-T) is a permanent organ of ITU. ITU-T is responsible for studying technical, operating and tariff questions and issuing Recommendations on them with a view to standardizing telecommunications on a worldwide basis.

The World Telecommunication Standardization Conference (WTSC), which meets every four years, establishes the topics for study by the ITU-T study groups which, in turn, produce Recommendations on these topics.

The approval of ITU-T Recommendations is covered by the procedure laid down in WTSC Resolution 1.

In some areas of information technology which fall within ITU-T's purview, the necessary standards are prepared on a collaborative basis with ISO and IEC.

NOTE

In this Recommendation, the expression "Administration" is used for conciseness to indicate both a telecommunication administration and a recognized operating agency.

INTELLECTUAL PROPERTY RIGHTS

ITU draws attention to the possibility that the practice or implementation of this Recommendation may involve the use of a claimed Intellectual Property Right. ITU takes no position concerning the evidence, validity or applicability of claimed Intellectual Property Rights, whether asserted by ITU members or others outside of the Recommendation development process.

As of the date of approval of this Recommendation, ITU had not received notice of intellectual property, protected by patents, which may be required to implement this Recommendation. However, implementors are cautioned that this may not represent the latest information and are therefore strongly urged to consult the TSB patent database.

CONTENTS

Introduction

This ITU-T Recommendation describes methods and procedures for the evaluation of complex terminals, network components and transmission systems. Depending on the various parameters and systems to be measured, test methods are described. The test methods mostly make use of test signals described in ITU-T Recommendations P.50, P.59 and P.501. For various technical implementations and conversational situations, the possible impacts on the speech quality perceived subjectively are given and the relevant measurement procedures are described.

ITU-T Recommendation P.502

Objective test methods for speech communication systems using complex test signals

1 Scope

The aim of this ITU-T Recommendation is the definition of test methods which can be used to evaluate specific artifacts influencing the speech quality transmission of terminals and speech transmission systems. The methods described in this Recommendation are based on test signals as defined in ITU-T Recommendations P.50, P.59 and P.501.

This Recommendation provides a collection of test methods which allow the investigation of various parameters which were found to be important for the assessment of speech communication systems. Each performance parameter is qualified by the speech degradation perceived subjectively and the related objective parameters. For the individual parameters analysis methods are described.

2 References

The following ITU-T Recommendations and other references contain provisions which, through reference in this text, constitute provisions of this Recommendation. At the time of publication, the editions indicated were valid. All Recommendations and other references are subject to revision; all users of this Recommendation are therefore encouraged to investigate the possibility of applying the most recent edition of the Recommendations and other references listed below. A list of the currently valid ITU-T Recommendations is regularly published.

- ITU-T Recommendation G.122 (1993), *Influence of national systems on stability talker echo in international connections*.
- ITU-T Recommendation G.168 (2000), *Digital network echo cancellers*.
- ITU-T Recommendation P.10 (1998), *Vocabulary of terms on telephone transmission quality and telephone sets*.
- ITU-T Recommendation P.50 (1999), *Artificial voices*.
- ITU-T Recommendation P.51 (1996), *Artificial mouth*.
- ITU-T Recommendation P.56 (1993), *Objective measurement of active speech level*.
- ITU-T Recommendation P.57 (1996), *Artificial ears*.
- ITU-T Recommendation P.58 (1996), *Head and torso simulator for telephonometry*.
- ITU-T Recommendation P.59 (1993), *Artificial conversational speech*.
- ITU-T Recommendation P.340 (2000), *Transmission characteristics of hands-free telephones*.
- ITU-T Recommendation P.501 (2000), *Test signals for use in telephonometry*.
- ITU-T Recommendation P.581 (2000), *Use of head and torso simulator (HATS) for hands-free terminal testing*.
- ITU-T Recommendation P.800 (1996), *Methods for subjective determination of transmission quality*.
- ITU-T Recommendation P.810 (1996), *Modulated noise reference unit (MNRU)*.

– ITU-T Recommendation P.830 (1996), *Subjective performance assessment of telephone-band and wideband digital codecs*.

– ITU-T *Handbook on Telephonometry*, 2nd edition; Geneva 1992.

– IEC 60651 (1979), *Sound Level Meters*.

3 Definitions and abbreviations

This ITU-T Recommendation defines the following terms:

AGC characteristics: Characteristics of automatic gain control systems.

attenuation range (a_H): Range in dB of attenuation inserted in sending or receiving direction of a terminal or system.

Send Speech Attenuation During Double Talk (A_{sdt})

Received Speech Attenuation During Double Talk (A_{rdt})

attack time: Time needed to fully activate a transmission path (by a compander).

crest factor: Peak-to-RMS ratio of a signal.

companding: Level dependant attenuation/amplification of a signal.

Composite Source Signal (CSS): Signal composed in time by various signal elements.

Echo Return Loss Enhancement (ERLE): Measure to determine the perceived improvement of disturbance by echo signals.

Echo Return Loss (ERL): Measure to determine the perceived disturbance by echo signals.

Echo Return Loss, double talk (ERL_{dt}): Measure to determine the perceived disturbance by echo signals in double talk conditions.

Echo Return Loss, temporally weighted, single talk (ERL_{tst}): Measure to determine the perceived disturbance by echo signals in single talk conditions taking into account some psychoacoustic effects.

Echo Return Loss, temporally weighted, double talk (ERL_{tdt}): Measure to determine the perceived disturbance by echo signals in double talk conditions taking into account some psychoacoustic effects.

Fast Fourier Transformation (FFT)

Markov Speech Model Process (MSMP)

See ITU-T Recommendation P.501.

Non-Linear Processor (NLP): Processor used typically in echo cancellers to switch off the residual echo.

Pseudo Noise sequence (PN-sequence): Pseudo-random noise with defined frequency-content, derived by inverse Fourier transformation of a predefined frequency spectrum.

RCV: Receiving direction

release time: Time needed to fully deactivate a transmission path (by a compander).

R_{in} (Receive input): (Electrical) receive access point of a device under test.

SND: Sending direction

S_{out} (Send output): (Electrical) send access point of a device under test.

TCL (Terminal Coupling Loss): Echo Loss of a terminal measured from R_{in} to S_{out}, including SLR and RLR.

T_c (Convergence Time)

See 4.3.1.

T_H (hang-over time)

See ITU-T Recommendation P.340.

Tic (Initial Convergence Time)

See convergence time.

T_R (build-up time)

See ITU-T Recommendation P.340.

T_s (switching time)

See ITU-T Recommendation P.340.

V_{TH} (threshold level)

See ITU-T Recommendation P.340.

4 Convergence Performance of Echo Cancellers

This clause describes the convergence performance of echo cancellers. Methods for assessing the subjective effects of various parameters of echo cancellers are described and objective methods for describing these parameters are also suggested.

4.1 Speech Quality Degradation Perceived Subjectively

Depth of convergence, or echo return loss enhancement (ERLE) describes the ability of an echo canceller to cancel signals returned in the opposite transmission direction through an echo path. This can be acoustic echo in the case of a hands-free telephone, or hybrid echo in the case of a two to four-wire conversion. Poor ERLE means that residual echo signals will be more audible.

Convergence time describes how fast the echo canceller reaches a stable state where returned residual echo signal is sufficiently attenuated without inserting loss in either speech transmission path. This is the time required to reach within 3 dB of ERL, *and/or [25] dB loss*. Fast convergence of an echo canceller is needed to prevent echo from reaching a talker at the beginning of a call.

Echo burst may be generated in a condition where an echo canceller may have trouble converging on a particular echo path. Subjective degradation is a function of the echo burst length/level, how close to each other they are, how many there are per minute, and the echo path delay.

4.2 Related Objective Parameters for Single Talk Mode

The quality of the echo control characteristics are determined by the following parameters:

- echo return loss as a function of time, defined as Echo Return Loss (ERL);
- temporally weighted echo return loss (ERL_t);
- time for AEC to converge, defined as Convergence Time (T_c).

In addition the following parameters apply in the double talk situation:

- duplex performance as a function of time;
- response in duplex operation for the above parameters;

- attenuation response in the presence of environmental or network impairments.

The double talk parameters are dealt with in clause 5.

4.3 Analysis Methods

The subclause below suggests some objective measurement techniques that can be used to assess the parameters described above. If network speech echo cancellers are tested, a proper test setup can be found in ITU-T Recommendation G.168. The test conditions for hands-free telephones can be found in ITU-T Recommendation P.340.

4.3.1 Convergence Time (T_c) Test Method

The description of the method to test convergence time of echo cancellers can be found in ITU-T Recommendation G.168. Therefore, the echo signal is measured using a level meter according to IEC 60651. An exponential weighting filter with a time constant of 35 ms (IEC 60651, "Impulse") is applied when integrating the output of the level meter. The measured output signal is displayed as a level versus time diagram. Typically, a limit is given as a function of time which should not be exceeded. A typical example for such a diagram is shown in Figure 1.

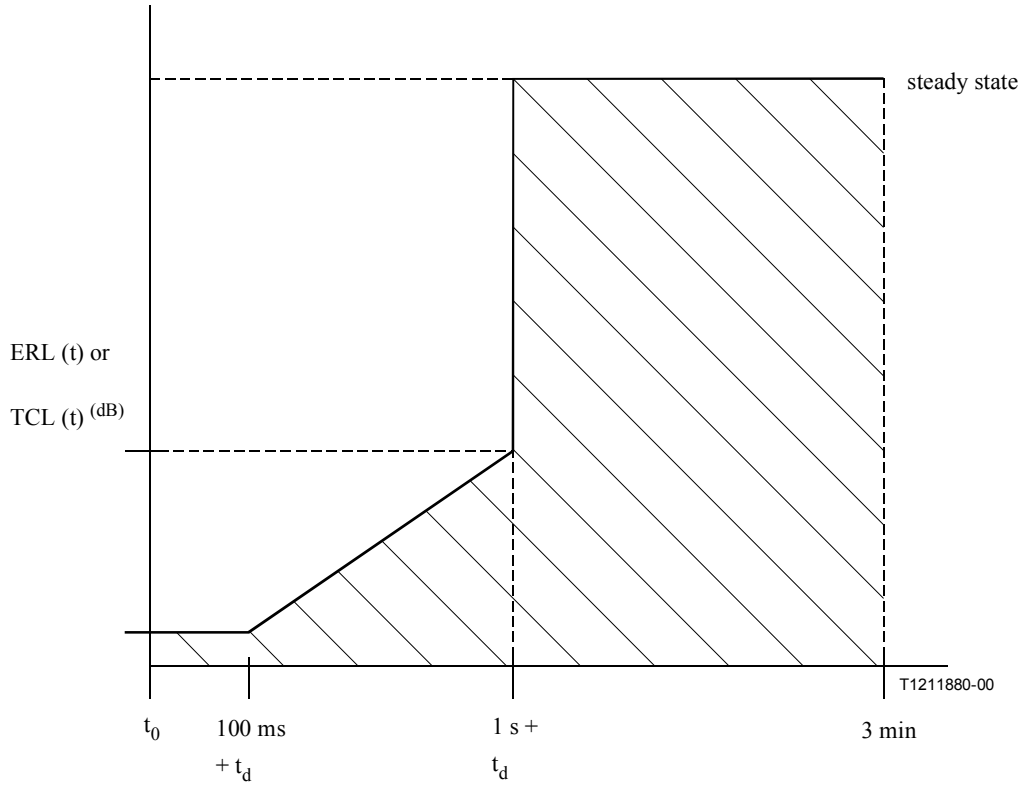


Figure 1/P.502 – Example for a typical limit for ERL or TCL as a function of time

ERL is typically measured when measuring network echo cancellers. TCL defines the coupling loss of a terminal including the acoustical interfaces.

NOTE – An exact definition of T_c is not given. A possible definition for T_c may be the time when an ERL of 3 dB above steady state condition is reached.

4.3.2 Echo return loss, temporally weighted (ERL_{tst}) – single talk

The Echo return loss measurement the methodology as described in ITU-T Recommendations G.122 and G.168 are the traditional methods which are currently the basis for all calculations and planning purposes. The ERL_t methodology is closer to subjective impressions taking into account temporal echo effects is described below.

Temporally weighted echo return loss from the network interface is measured. This method provides a measure for echo bursts, but can also be used instead of any long-term echo return loss measurement.

Pseudo code is provided in Annex A to implement this method. The test signal to be used should be as speech-like as possible. Other test signals can be used, but may produce optimistic results.

Test signal is applied at R_{in} (see e.g. Figure 10) for 30 seconds so that the different functional units (in particular the acoustic echo canceller) reach their steady states. In case of the measurement of acoustic echo cancellers no other signal than the acoustic return from the loudspeaker(s) is applied to the microphone(s).

Record the electrical signals at R_{in} and S_{out} for the next 1 minute. Align the R_{in} and S_{out} recordings in time by adding the system delay between R_{in} and S_{out} to the R_{in} signal. The time dependent value ERL_{tst} is the difference (in dB) between the signal level at R_{in} and S_{out} calculated using the algorithm in Annex A.

NOTE – Echo paths may change during the measurement, they depend on the environment and the use of the equipment.

5 Speech Quality Evaluations during Double Talk

The most critical situation in any conversation is the double talk situation. Equipment involving any kind of non-linear or time variant signal processing may degrade the speech quality, especially parameter like "double talk capability" (perceived subjectively) quite significantly.

5.1 Speech Quality Degradation Perceived Subjectively

The most annoying effects during double talk are:

- sentences, words, syllables interrupted or not transmitted completely during or shortly after/before double talk;
- transmission of speech and/or background noise with time variable level causing annoying "level variations during double talk";
- echo during double talk.

The most critical situations during double talk are the time intervals shortly before and shortly after double talk. During these time periods the self masking of the own voice is no longer effective (see Zwicker [5]).

In case echo cancellation is used, fast convergence of an echo canceller is needed in order to quickly facilitate double talk at the beginning of a call. This means that the non-linear processor can be removed earlier, allowing full double talk to occur. The depth of convergence determines the audibility of residual echo during double talk. If a degree of switched loss is employed to further enhance echo return loss, this can result in audible speech attenuation during doubletalk. Echo bursts are possible during double talk if double talk detection errors cause the echo canceller to diverge. This results in fairly loud bursts of echo.

5.2 Related Objective Parameters

The related objective parameters are:

- build-up times (during double talk);
- hang-over times [switch-off times] (during double talk);
- switching times (during double talk);
- attenuation range (during double talk);
- attenuation distribution (during double talk);
- frequency responses;
- loudness ratings;
- level variation during double talk (companding characteristics).

In addition the following parameters, which are mainly associated to echo canceller implementations are to be considered:

- convergence time during double talk;
- echo return loss (double talk mode): determined from the level of residual echo during double talk;
- sent speech attenuation (see also attenuation range) during double talk: determined from the amount of speech attenuation due the insertion of switched loss;
- temporally weighted echo return loss din double talk conditions: determined from the weighted level of echo bursts.

NOTE – When conducting objective measurements it should be noted in any case whether frequency responses, loudness ratings and levels/level variations are measured shortly before, after or during double talk.

5.3 Analysis Methods

Various methods may be chosen to evaluate the double talk performance of a system. The description below gives an overview about the different technologies. The method described in 5.3.1 is a generalized method which does not specifically assume any technical implementation of the device under test. The methods in 5.3.2 through 5.3.4 assume an echo canceller implementation.

5.3.1 CSS double talk method

5.3.1.1 Signal Construction

A measurement method in a double talk situation can be implemented by using the test signal shown in Figure 2. This test signal consists of a series of uncorrelated composite source signals (ITU-T Recommendation P.501) which are fed in sending and receiving direction simultaneously. The test sequence is constructed that way, that starting with a high level in sending direction, a low level in receiving direction is inserted. The level of each composite source sequence is decreased by 0.5 dB in sending direction and increased by 0.5 dB in receiving direction. The total level difference between a maximum and a minimum composite source "package" in each direction is 20 dB (30 dB for network applications). For hands-free terminals the level ranges may be chosen as follows:

receiving direction: –38 dBm to –18 dBm;

sending direction: –4.7 dBPa to –24.7 dBPa.

For double talk measurements in the network the level variations may be chosen e.g.:

receiving direction: –40 dBm to –10 dBm;

sending direction: –10 dBm to –40 dBm.

Of course different level variations are possible. All level ranges should depend on the desired dynamic range to be evaluated.

The sequence is typically constructed symmetrically, this means when reaching the minimum level in sending direction, the level increases again whereas in receiving direction the signal level decreases again. The symmetric construction of the signal allows also to evaluate the symmetry behaviour of the device under test.

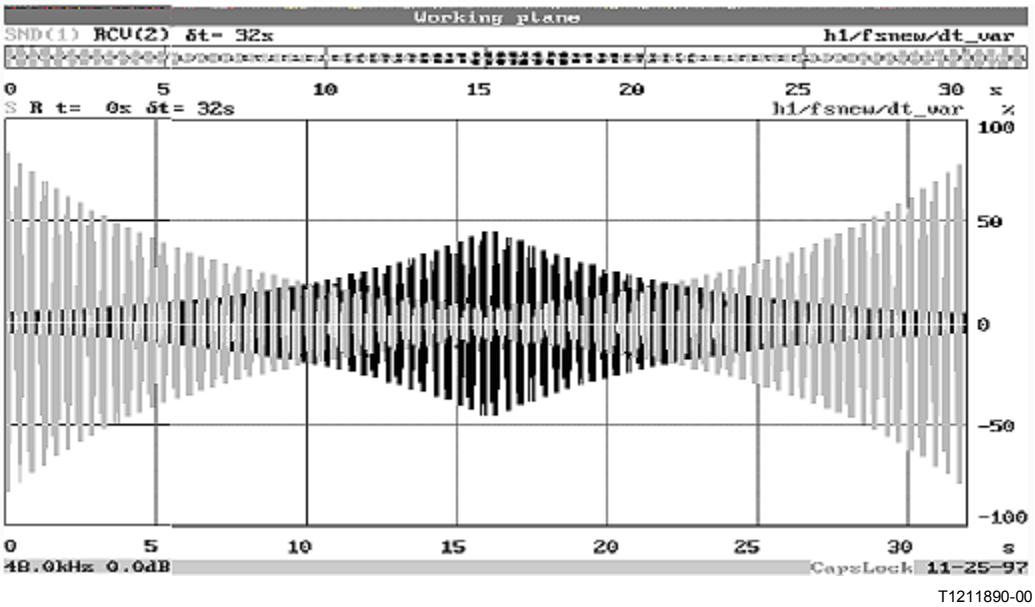
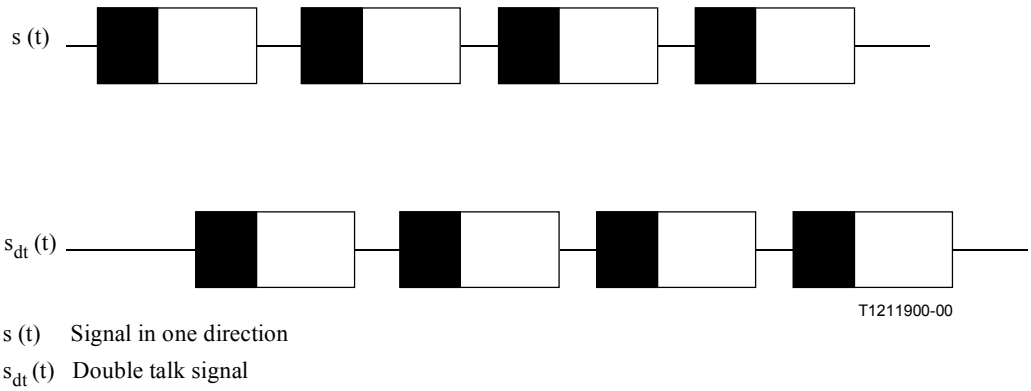


Figure 2/P.502 – Overview of double talk test signal



s (t) Signal in one direction

s_{dt} (t) Double talk signal

Figure 3/P.502 – Cut out of the complete measurement sequence with detailed view on the overlap of sending and receiving direction signal, principle arrangement

Figure 3 shows the construction of the signal in more detail. It can be seen that the overlap of the sequences is only partial. Always the voiced sound (black) overlaps with the end of the pseudo-random noise sequence (white) of the opposite channel. The sequence is constructed in such a way that, during the pauses in receiving direction, the sending direction can be measured; during the pauses in sending direction, the receiving direction can be evaluated. This is useful e.g. in case of analogue devices where a sufficient decoupling between sending and receiving due to limited sidetone capabilities is not possible.

In the same way, a sequence can be constructed which starts with high level excitation in receiving direction and low level excitation in sending direction, in case that different starting points of levels should be evaluated.

In general, it should be noted that the signal construction as shown in Figures 2 and 3 is one example of time relationship between the sending and receiving direction. Of course, other time intervals (e.g. longer pauses, longer pseudo noise (pn)-sequences, different types of CS-signals) can be used, depending on the requirement for the measurement to be fulfilled.

5.3.1.2 Evaluation Procedures

For double talk evaluations the sequence offers a lot of capabilities. Before really evaluating a device, the delay between excitation signal and measured signal needs to be compensated. In the second step the measured signals for both directions are extracted and referred to the excitation signal. The principle of this method when evaluating parameters in the frequency domain (based e.g. on Fourier transformation) is shown in Figure 4.

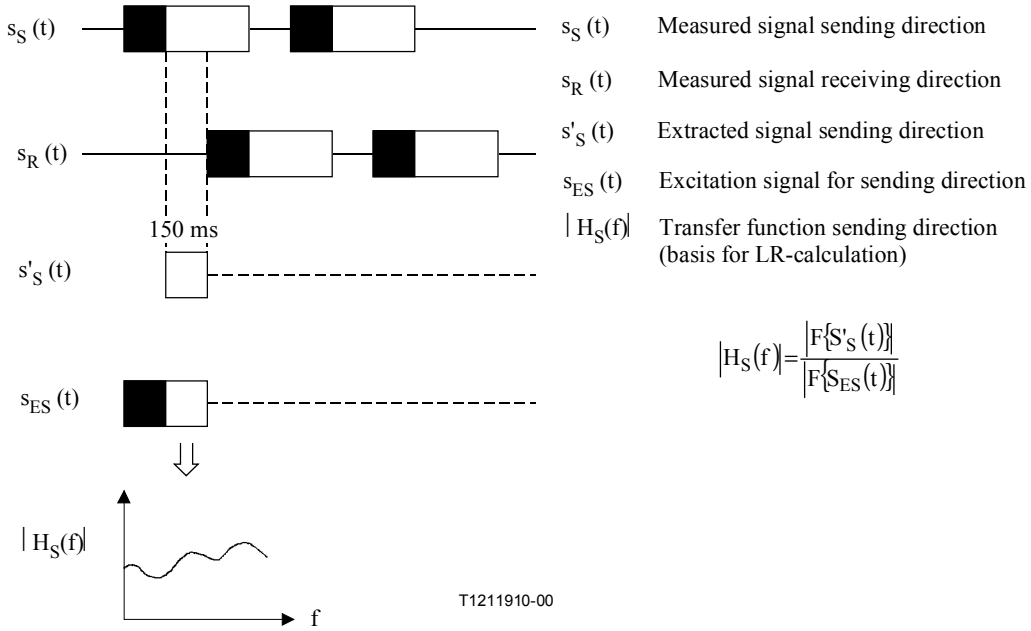


Figure 4/P.502 – Principle of signal extraction and determination of transfer characteristics

NOTE – It always should be ensured, that a valid estimation of frequency responses, loudness ratings etc. is derived from the analysis. Coding algorithms involved may lead to a wrong estimation of transfer functions etc. and may require specific measurement signals and/or analysis procedures.

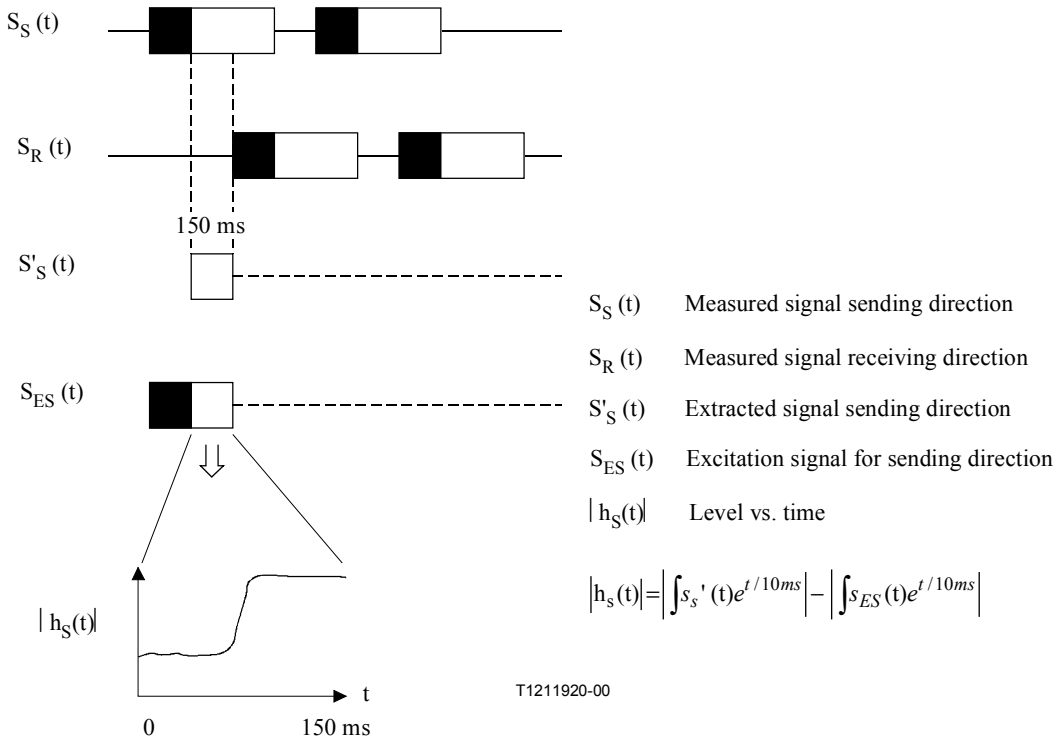
Since the measurement signal is a pseudo random noise of the CSS sequence, it is possible to calculate from this 150 ms measurement signal the following parameters:

- frequency responses,
- loudness rating,
- short-term attenuation (in case of level switching devices),
- long-term attenuation (when measuring at different times of the sequence).

Switching times can be evaluated directly in the time domain due to the exactly defined signal characteristics:

- build-up time (due to the overlapping of the signals only if TR is greater than 50 ms);
- switch-over times.

Switching times are evaluated by determining the level versus time with an adequate short time constant, typically at minimum 10 times shorter than the switching time of the system under test to be evaluated. By monitoring the output signal during the periods, where only one signal is present, switching or level variations can be evaluated in great detail. A general example of the procedure is given in Figure 5. Although the output signal is referenced to the input signal in this example, this referencing is not always required since the expected level during the periods of evaluation is known from the level of the pseudo random part of the CSS.



NOTE – The time constant of 10 ms is shown as one example, a different time constant may be used depending on the analysis requirements.

Figure 5/P.502 – Principle of signal extraction and determination of time constants, the example shows the switching time during double talk

If frequency dependant level evaluation is required, this can be m ade from overlapping Fourier Transformation or Wavelet Transformation of the measured output signal. By this analyses frequency dependant switching or level variation can be evaluated. Care should be taken that windowing and analysis window length are appropriate. The analysis window length should be shorter than the time slot available (due to the signal construction and overlapping) for analysis.

5.3.2 Double talk testing using parallel combined sequences

Double talk testing often imposes conflicting restraints on the type of test signal used. In opposition to the sequential combined sequences as described in 5.3.1, parallel combined sequences allow evaluations during real double talk. As a general principle such signals either should be orthogonal

or should be extractable by means of filters from the signal to be transmitted originally. The methodology for testing is described below.

For echo return loss tests, the double talk signal presented to the canceller at R_{in} (or mouth simulator) should be as similar to the training signal as possible. Cancellers typically freeze adaptation during double talk. For example, if the double talk signal at R_{in} differed from the training signal, residual echo would typically be unrealistically high. This constraint indicates that the double talk signal at R_{in} should be the same as the training signal at R_{in} for echo return loss measurements.

Unfortunately, the use of the speech files alone is not acceptable during double talk. The correlated parts between the two "talkers" would invalidate some test results: parts of the one talker's speech may look like echo of the other talker if the parts are correlated. Another problem is that double talk onset must be very accurately detected for attenuation and clipping tests. This would be very difficult to define over repeated tests using different speech files, but is very easy with tones.

5.3.2.1 Signal Construction Method A

To overcome these issues, both signal types are used; speech as per the training signal, and tones to accurately define the start of double talk. How they are used depends upon the specific test. Speech signals used during training are continued during double talk, as required. A sinusoidal tone is mixed in with the speech or injected on its own to provide an easily measurable reference for attenuation tests or an easily definable start of double talk for clipping tests. By using notch or bandpass filters at S_{out} (or receive output) at the tone frequency, either just the tone or just the speech can be monitored.

When the tone is mixed in with the speech, the power of the tone must be representative of the long-term average power of speech at its frequency, so as not to impact the canceller with any gross deviations in spectral energy from that of the training signal. ITU-T Recommendation P.50 specifies an average spectral relationship (third octave values used). The tones given in Table 1 are recommended. Their power is defined as the number of dB below the average active speech energy in the speech file, when measured as per ITU-T Recommendation P.56.

Table 1/P.502

Tone Frequency	Relative Tone Level (dB) below Nominal Speech Level
500	9
1 000	14
1 750	18
2 500	22

5.3.2.2 Double Talk Attenuation Testing Using Method A

5.3.2.2.1 Send Speech Attenuation During Double Talk (A_{sdt})

The example shown in Figure 6 determines double talk attenuation in the send direction. The concept is easily extended to the receive direction by reversing signals and monitoring at the receive output.

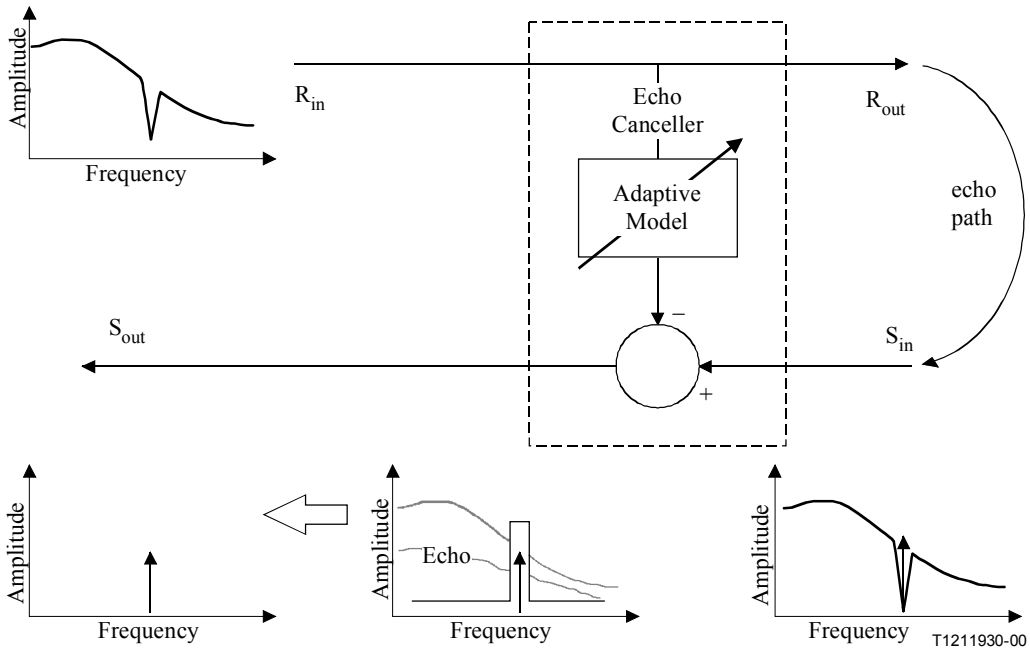


Figure 6/P.502 – Principle of double talk attenuation testing using method A

The methodology is explained below:

In case echo cancellers are involved, the object under test is reset (if possible), and trained as described in Annex C. The "talker active before double talk" is the mouth simulator in case of hands-free telephones or the S_{in}. The "talker initiating double talk" is R_{in}.

Notice that the signals at both the mouth simulator respectively S_{in} and R_{in} are shown notch filtered at the tone frequency. This notch filter is not present during the entire training period, but only just before double talk, and for the remainder of the measurement. The idea is to mix in a tone at the mouth simulator just before double talk (still in single talk), monitor its rms level, have R_{in} enter double talk, and continue monitoring the tone level.

The double talk attenuation is the difference in tone level before double talk and during double talk. The tone is discriminated by applying a bandpass filter at the tone frequency at S_{out}. By continuing measurement during double talk, the switching characteristics including rate of insertion and depth can be determined. The rate of attenuation removal can also be determined by making the activity mask for the "talker initiating double talk" low again after the attenuation depth has stabilized.

Characteristics of the notch filter will now be described. The notch is required on R_{in} to ensure that echo at the tone frequency does not impact the measurement of the tone. The notch filter must show enough attenuation to ensure speech at the tone frequency are adequately repressed so as not to impact the level of the tone at the mouth simulator. The notch filter bandwidth must be tight enough to minimize impact on surrounding frequencies so that the signals are not significantly different than the training signals. Example filter types are described in Annex B.

The bandpass filter has similar constraints. Taken with the notch filter, it must have enough out of band attenuation to ensure that the tone level is not impacted by speech or echo. It must also have a short enough impulse response that the time domain impact is minimized, as the rate of attenuation insertion is also being measured.

The impulse response of the notch filter does not impact the measurement as the tone mixed in at the mouth simulator will be large enough in level to swamp any residual ringing of the notchs.

The rms level of the tone is to be measured using an 8 ms sliding window for smoothing. The window is slid in 4 ms increments for 4 ms of overlap between adjacent points to smooth the results.

The timing of the measurement must be fine tuned to account for any ringing of the bandpass filter. The tone should be injected at 60 seconds, minus the bandpass filter's ringing time (<20 ms assumed), minus 8 ms. The tone reference measurement during single talk is taken starting at 60 seconds minus 8 ms, after the filter ringing has ended. Set delay in the direction of measurement leads to partial measurement during the end of the bandpass filter's ringing. As long as set delay is low (provisionally <5 ms), the amount of ringing effects encountered will be slight and should not impact the measurement.

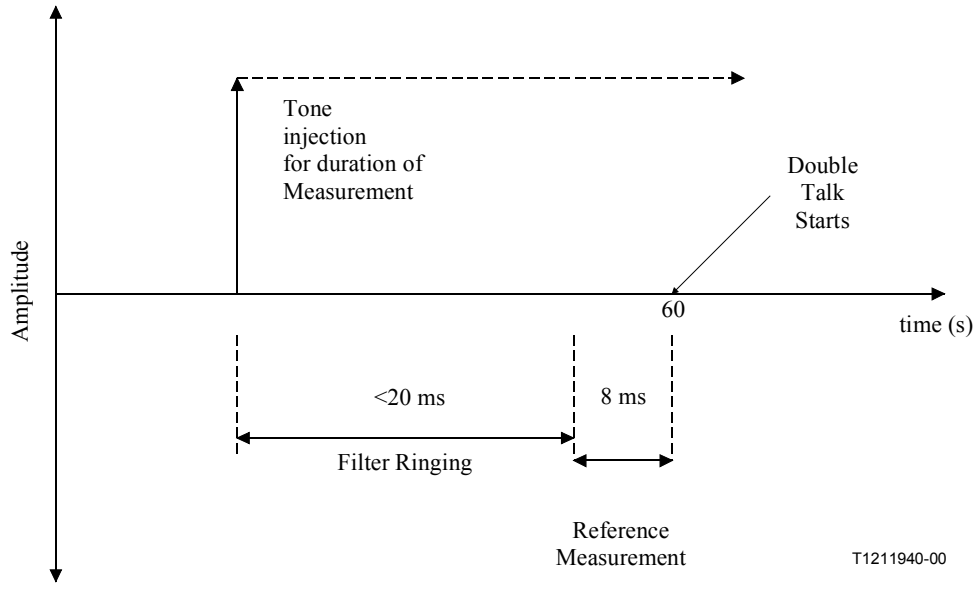


Figure 7/P.502

When sub-banding techniques are used in the AEC, or the technique is unknown, it is advisable to repeat the test for each frequency shown above. In many cases, the attenuation test results at any one frequency may not be indicative of subjective quality. As voice has dominant spectral energy in the lower frequency range, we would expect that the switched loss to be more audible there. If the depth of attenuation is frequency independent, it is advised to use higher test frequencies as the required filters will have less of an impact on the over-all voice levels.

The method given measures the attenuation vs time after entering doubletalk for a specific frequency. The result of this method may depend greatly on the exact nature of the speech signal used, particularly as doubletalk is begun. There may also be a dependence on the frequency of the measurement tone, which is a sine wave embedded in the real speech creating doubletalk.

A detailed description of the steps to be conducted is given below.

The canceller is trained, as described in Annex C, with the activity mask "talker active just before onset of double talk" applied at S_{in} (or at the MRP when terminals are measured). The "talker initiating double talk" mask is applied at R_{in}. Carry out the test described before.

Using the 8 ms sliding averaging window on the sine signal measured at S_{out}, the time dependent value A_{SDT} is the difference (in dB) between the first 8 ms average before double talk and each 8 ms average after double talk.

5.3.2.2.2 Received Speech Attenuation During Double Talk (A_{rdt})

NOTE – In case hands-free telephones are measured, the receive output signal should be measured with the measurement microphone as close to the loudspeaker as possible to provide discrimination in the acoustic domain.

Carry out the test as described before for A_{sdt}, substituting receive for send and vice versa. Therefore, receive and send signals are swapped and results monitored at the receiver output.

5.3.2.3 Echo Return Loss During Double Talk Testing Using Method A

The example shown in Figure 8 determines echo return loss looking towards the terminal from the network. The concept is easily extended to talker echo path loss by reversing signals and monitoring at the receive output.

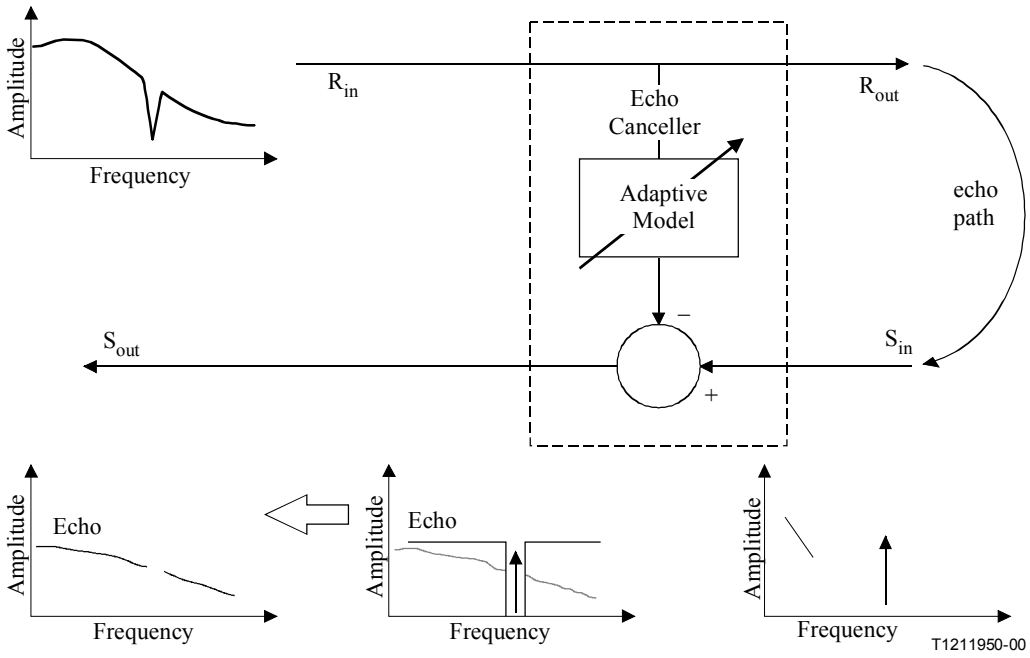


Figure 8/P.502 – Principle of echo return loss testing using method A

The set is reset and trained as described in Annex C. The "talker active before double talk" is R_{in}. The "talker initiating double talk" is the mouth simulator or the S_{in} port.

The notch filter is applied at R_{in} as described in 5.3.2.2. Double talk is entered by applying a tone (only) at the mouth simulator. The tone is injected for 20 seconds. The echo return loss is found by first notching out the tone at then measuring the residual echo using ERL calculations or the ERL_t method as described in Annex A. Using a tone ensures that the precise moment that double talk is entered is known.

Once double talk has ended, the echo return loss measurement may be continued for 10 seconds to measure recovery after double talk. After that time, one second of silence should be played. In this way, the noise in the echo path can be measured. If it can be assumed that the noise and echo are uncorrelated, and that the noise is stationary, the noise measured in the last second may be subtracted from the echo plus noise during double talk to arrive at the echo during double talk.

The timing of the measurement must be fine tuned knowing the echo path delay. This delay properly aligns the source and echo.

In detail the testing procedure is as follows:

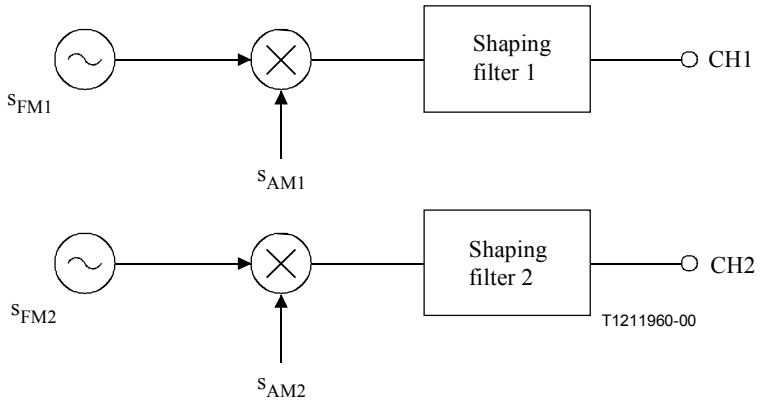
Echo Return Loss – Double Talk (ERL_{dt}), Temporally Weighted – Double Talk (ERL_{tdt})

In case of echo cancellers present, all the EC functional units are initially reset and then enabled. For black box testing, the system is powered up with no acoustic stimuli present at any interface. The canceller is trained and the activity mask "talker active just before onset of double talk" applies to R_{in}. The "talker initiating double talk" mask is applied at S_{in} (e.g. the mouth simulator). The tests are carried out as described in Annex A.

Record the electrical signals at R_{in} and S_{out} during the 20-second tone application. Align the R_{in} and S_{out} recordings in time by adding the system delay between R_{in} and S_{out}. Calculate either ERL_{dt} using the traditional echo loss calculation according to ITU-T Recommendation G.122 or the ERL_{tdt}. The time dependent value ERL_{tdt} is the difference (in dB) between the signal level at R_{in} and the signal at S_{out} calculated as shown in Annex A, for the 20 seconds after initiation of double talk.

5.3.2.4 Signal Construction Method B

Method B uses orthogonal sequences generated by a set of voice like modulated sinewaves, spectrally shaped. The general construction principle which can be found in detail in ITU-T Recommendation P.501 is shown in Figure 9.



$$s_{FM1,2}(t) = \sum A_{FM1,2} * \cos(2\pi t\, n * F_{01,2}); \quad n = 1,2,....$$

$$s_{AM1,2}(t) = \sum A_{AM1,2} * \cos(2\pi t\, F_{AM1,2});$$

Figure 9/P.502 – Two channel test signal generation for double talk evaluations based on AM-FM signals

Typical settings are given in Table 2.

Table 2/P.502

	f_{AM}	f_{FM}	F_0	Shaping filter
Channel 1 (CH 1)	f_{AM1} = 3 Hz	f_{FM1} = 5 Hz	F_{01} = 270 Hz	LP, 5 dB/oct.
Channel 2 (CH 2)	f_{AM2} = 3 Hz	f_{FM2} = 5 Hz	F_{02} = 290 Hz	LP, 5 dB/oct.

For more details see ITU-T Recommendation P.501.

5.3.2.5 Double Talk Attenuation Testing Using Method B

5.3.2.5.1 Send Speech Attenuation During Double Talk (A_{sdt})

The example shown in Figure 10 determines double talk attenuation in the send direction. As for method A, this concept is easily extendable to the receive direction by reversing signals and monitoring at the receive output.

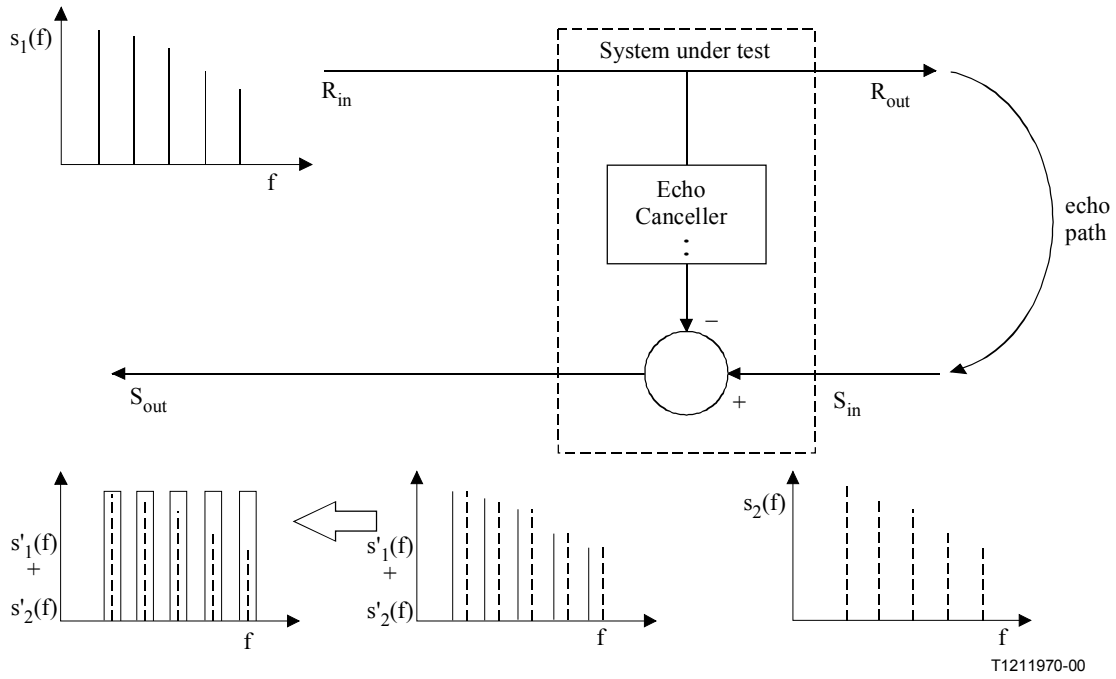


Figure 10/P.502 – Principle of double talk attenuation testing using method B

The methodology is explained below.

In case echo cancellers are involved, the object under test is reset (if possible), and trained as described in Annex A. The "talker active before double talk" is the mouth simulator in case of hands-free telephones or the S_{in}. The "talker initiating double talk" is R_{in}.

The double talk signal is analysed during the double talk period using the following analysis principle.

In order to extract the echo signal from the double talk either a specific filter setting or a specific post processing of the FFT analysis is required, since the spectrum of the signal as well as of the double talk signal is a kind of combfilter spectrum where a specific modulation is applied. The mid frequency f_{mid} of any frequency component, the according frequency modulation f_{mod} as well as the filter shapes or the windowing function of the Fourier transformation need to be taken into account. If the filter approach is used the bandwidth of each filter should be constructed that way that:

$f_u = f_{mid} - f_{mod\ (fm)}$

$f_o = f_{mid} + f_{mod\ (fm)}$

The stopband attenuation should be at least 10 dB higher than the minimum level to be measured within the passband. The same applies for analysis derived from Fourier transformations of the measured echo signal. Here the frequency "smearing" effect of the windowing function needs to be taken into account. In order to have a sufficient separation between the echo signal and the double

talk signal in the low frequency domain, a minimum FFT lengths of 8 k (sampling rate 44.1 or 48 kHz) which amounts to a time window of about 170 ms should be chosen.

A set of typical excitation frequencies for single talk and double talk signal are given in ITU-T Recommendation P.501.

The principle of the analysis is shown in Figure 11.

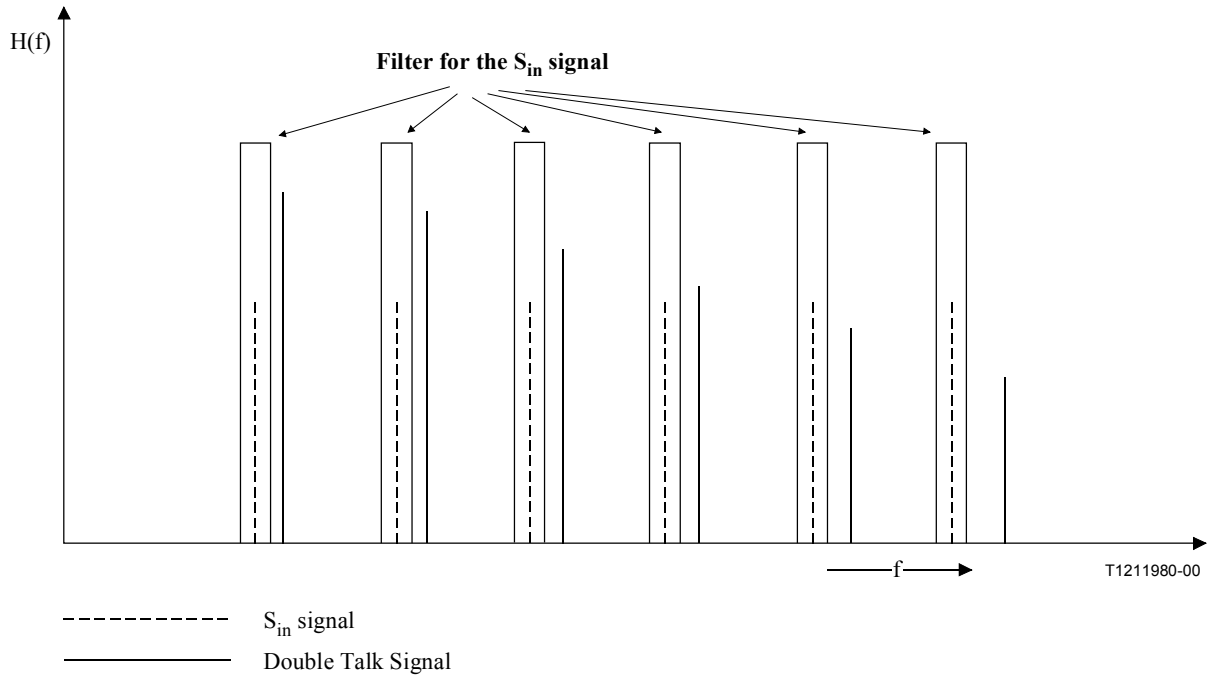


Figure 11/P.502 – Extraction of the S_{in} signal components (schematic)

The double talk attenuation is the difference in tone level before double talk and during double talk. The voiced sounds sequence is discriminated by applying the Filter or FFT procedure as described above. By continuing measurement during double talk, the switching characteristics including rate of insertion and depth can be determined, in time as well as in frequency depending on the analysis method chosen. The rate of attenuation removal can also be determined by making the activity mask for the "talker initiating double talk" low again after the attenuation depth has stabilized.

The timing of the measurement must be fine tuned knowing the echo path delay. This delay properly aligns the source and echo.

5.3.2.5.2 Received Speech Attenuation During Double Talk (A_{rdt})

NOTE – In case hands-free telephones are measured, the receive output signal should be measured with the measurement microphone as close to the loudspeaker as possible to provide discrimination in the acoustic domain.

Carry out the test as described before for A_{sdt}, substituting receive for send and vice versa. Therefore, receive and send signals are swapped and results monitored at the receiver output.

For a detailed description of the different steps (1 to 4) for the evaluation of "sent speech attenuation during double talk (A_{sdt})" and "received speech attenuation during double talk (A_{rdt})" see 5.3.2.2.

5.3.2.6 Echo Return Loss During Double Talk Testing Using Method B

The example shown in Figure 12 determines echo return loss looking towards the terminal from the network. The concept is easily extended to talker echo path loss by reversing signals and monitoring at the receive output.

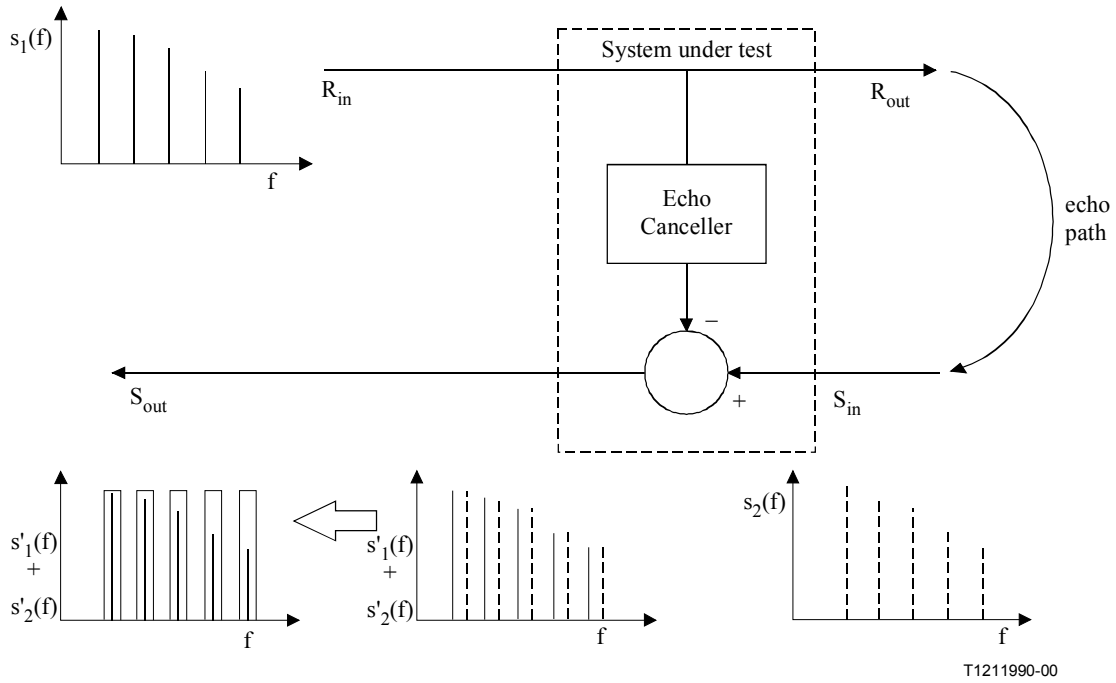


Figure 12/P.502 – Principle of double talk echo return loss testing using method B

The methodology is explained below.

In case echo cancellers are involved, the set is reset, and trained as described in Annex C. The "talker active before double talk" is R_{in}. The "talker initiating double talk" is the mouth simulator or the S_{in} port.

The double talk signal is analyzed during the double talk period using the analysis principle described in the previous subclause. Instead of the double talk signal the echo signal is analyzed by applying the appropriate filter set, see Figure 13.

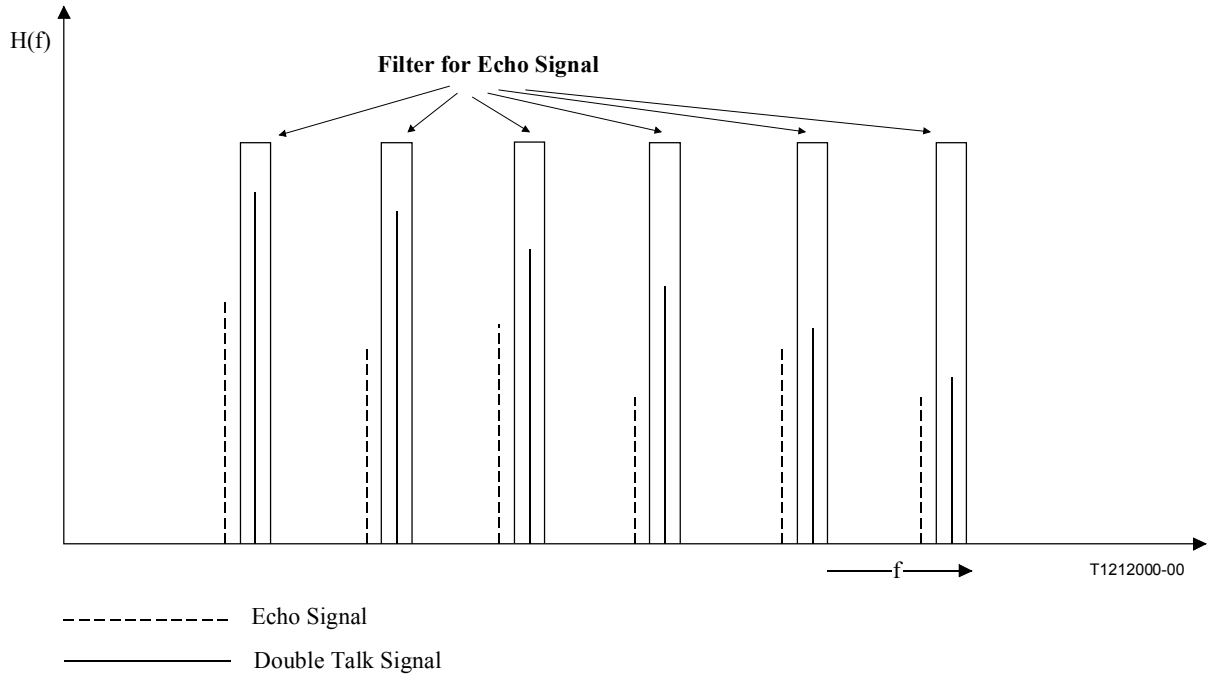


Figure 13/P.502 – Extraction of the echo components of the double talk signal (schematic)

Once double talk has ended, the echo return loss measurement may be continued for 10 seconds to measure recovery after double talk. After that time, one second of silence should be played. In this way, the noise in the echo path can be measured. If it can be assumed that the noise and echo are uncorrelated, and that the noise is stationary, the noise measured in the last second may be subtracted from the echo plus noise during double talk to arrive at the echo during double talk.

The timing of the measurement must be fine tuned knowing the echo path delay. This delay properly aligns the source and echo.

For a detailed description of the different steps (1 to 4) for the evaluation of "sent speech attenuation during double talk (A_{sdt})" and "received speech attenuation during double talk (A_{rdt})" see 5.3.2.3.

6 Companding and AGC Characteristics

Companding or AGC may be used to avoid overload of systems, to compensate for varying speech levels or to equalize speech levels in the network. In any case the aim is to improve either the speech "quality" or to improve input signals for devices operating on speech signals such as echo cancellers speech detectors or others.

6.1 Speech Quality Degradation Perceived Subjectively

Companding or AGC-devices and in general any device introducing time variant amplification to a speech signal lead to speech level fluctuations which may result in degradation of speech quality. In general the influence perceived subjectively depends on the attenuation range, time constants and control characteristics for such a device.

6.2 Related Objective Parameter

The related objective parameters are:

- compander control range;
- attenuation range (range of level adjustments);
- time constants.

6.3 Analysis Methods

In general, different kinds of test signals should be used to determine the performance of companding or AGC devices. Besides artificial test signals speech or nearly speech like signals such as artificial voice (ITU-T Recommendation P.50) or speech like signals (MSMP-signal) as described in ITU-T Recommendation P.501 should be used.

Figures 14 and 15 represent test signals, generated by a periodical repetition of voiced sounds. These signals can be used to measure level adjustments for systems, which have the same reaction on the periodical repetition of a voiced sound and on real speech. Additionally, artificial voice can be used.

The signal given through Figure 14 represents an input signal with a continuously increasing (decreasing) level, whereas the signal levels are adjusted in certain steps for the signal in Figure 15.

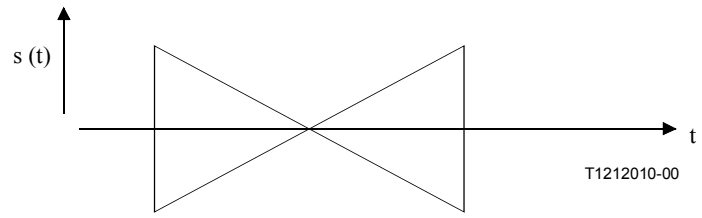


Figure 14/P.502 – Structure of test signal to determine level adjustments (constantly changing input level)

Suggested parameters for the signal in Figure 14 are given in Table 3.

Table 3/P.502

	Signal generation	**Highest level**	**Lowest level**	**Level variation**
SND direction	Voiced sound, periodically repeated	–16 dBm (–3.0 dBPa for terminals)	Below noise floor	Linear
RCV direction	Voiced sound, periodically repeated	–16 dBm	Below noise floor	linear

The complete signal duration can be chosen to 10 s.

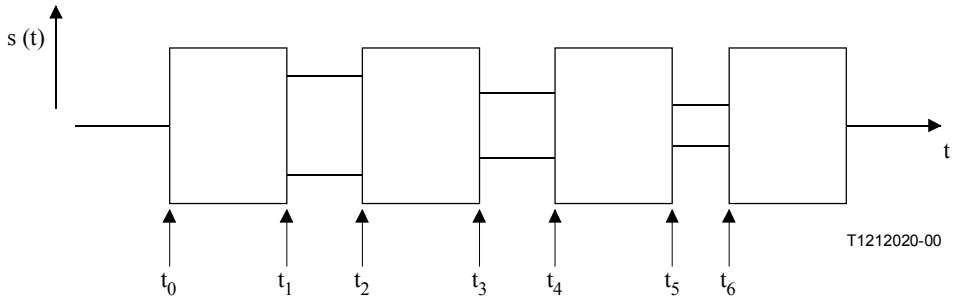


Figure 15/P.502 – Structure of test signal to determine level adjustments

Suggested parameters for the signal in Figure 15 are given in Table 4.

Table 4/P.502

	Signal generation	**Signal level during** $(t_1 - t_0)$ $(t_3 - t_2)$ $(t_5 - t_4)$	**Signal level during** $(t_2 - t_1)$	**Signal level during** $(t_4 - t_3)$	**Signal level during** $(t_6 - t_5)$
SND direction	Voiced sound, periodically repeated	–16 dBm (–3.0 dBPa for Terminals)	–21 dBm (–8.0 dBPa for Terminals)	–26 dBm (–13.0 dBPa for Terminals)	–31 dBm (–18.0 dBPa for Terminals)
RCV direction	Voiced sound, periodically repeated	–16 dBm	–21 dBm	–26 dBm	–31 dBm

The signal duration of the single periods can be chosen to 2.5 s each.

The signal in Figure 14 is suited to determine the:

– **Range of Level Adjustments as a Function of Input Signal Level**

A signal as described above (Figure 14) is applied. The analysis of the output signal is made as a level versus time analysis, referring the measured output signal to the time aligned excitation signal. Time constants should be chosen in a range of 10-125 ms. Care must be taken in order to avoid misleading measurement results due to the non-speech like character of the test signal.

Ideally the output of the analysis is a flat graph versus time. If AGC or companding is involved, the output should not deviate more than ±3 dB from the average measured output. Time constants of an AGC, if involved, should be rather slow (>100 ms). If companding is detected, attack times should be rather short (10-50 ms). Release times, however, should be sufficiently long. If more than ±3 dB companding or AGC is detected, subjective evaluation is required.

The signal in Figure 15 is suited to determine especially the:

– **Time Duration for Level Adjustments**

For this analysis the signal according to Figure 15 should be used. The analysis is made the same way as described above. This requires levels analysis versus time, referring the measured signal to the time aligned excitation signal and display of the result as a graph versus t. The time constant chosen for this analysis should be in a range of 5-10 ms in order to provide a good time resolution.

Ideally no level differences should be noticeable. If AGC or companding is noticed, they should be in a range less than ±3 dB. Therefore, the measured time duration should be as described above.

7 Quality of Background Noise Transmission

When judging the quality of background noise transmission, the background noise is considered as a signal by the listener. Such, in general similar effects, as when applying speech may influence the quality perceived subjectively. This parameter becomes more and more important since modern telecommunication systems are used increasingly in noisy environments.

7.1 Quality Degradation Perceived Subjectively

The most typical influence is found in sending directions for the far end listener when background noise is transmitted. In general, the perceived quality is influenced by:

– level fluctuations in the background noise;

– interruptions in the noise transmission;

– artifacts like modulations produced by signal processing.

The influence may be different in:

– at idle mode;

– with far-end speech;

– with near-end speech,

each situation should be considered separately.

7.2 Related Objective Parameter

For the following analysis descriptions, the background noise is regarded as the test signal. The effects perceived subjectively can be described by the following parameters:

– attenuation range;

– attenuation in SND direction;

– switching characteristics;

– minimum activation level in SND direction;

– frequency response;

– sensitivity of background noise detection (activation level, absolute level, level fluctuations).

In addition, the quality is influenced by the:

– design of NLP or centre clippers in conjunction with echo cancellers;

– design of noise reduction systems.

7.3 Analysis Methods

In general, the simulation of background noise can be a continuous noise signal (with shaped spectrum), or a more sophisticated signal to represent realistic conditions (e.g. office voice babble). In such cases, the background noise should be characterized by its long-term power density spectrum and its average level applied during the measurement.

For the following tests, the background noise signal is regarded as the measurement signal and not as a disturbing component. Consequently, analyses are applied to the noise signal. The transmission quality of background noise (from the near end in SND direction) can be evaluated at idle mode, with far-end speech and with near speech.

In all these cases important parameters are:

• the sensitivity of background noise detection in terms of activation level;

• the absolute level of the transmitted noise signal;

• level fluctuations of the transmitted noise signal.

Since the auditory evaluations of requirements for background noise transmission properties are still in progress, a detailed description of the analysis methods for the time-being is not complete.

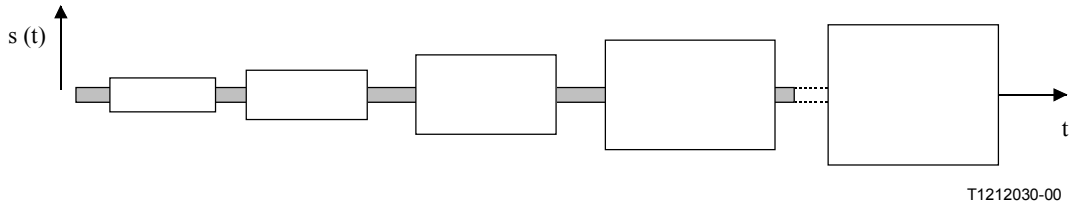
In **idle mode** the background noise transmission may be measured, for example, by applying a diffuse noise field with H_{oth} characteristics or using an appropriate background noise signal such as voice babble, car noise or others representing typical conditions. The signal level is applied, for

example, for a period of 20 seconds, starting with an excitation level of 50 dB_{SPL} (A) or the corresponding level measured at the electrical interface. The level then may be increased by, for example, 3 dB and is again applied for 20 seconds. Such a measurement sequence is constructed which contains blocks of 20 seconds H_{oth} noise increased by 3 dB each. When applying each 20 s portion, no audible background noise variation should be detected. Exact numbers for audibility of time constants are not yet available. In case level fluctuations of more than ±3 dB, as compared to steady state conditions, are measured, subjective evaluation should be conducted.

In general, the background noise signal should be audible all the time.

The lower the transmitted background noise level, the better it is. However, artifacts of noise reduction algorithms need to be avoided. More realistic background noise simulations are still under study.

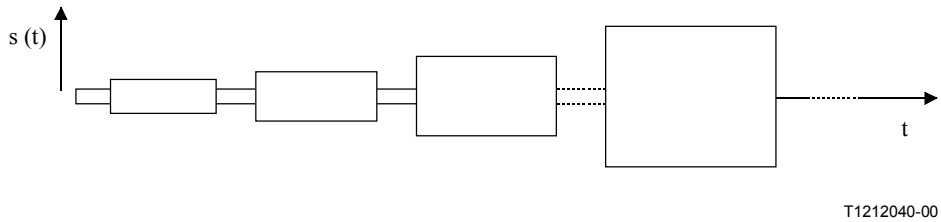
The following signal structure can be used to evaluate the quality of background noise transmission in SND direction coincident **with far-end speech**. Figure 16 represents a continuous noise signal applied at the near end (SND direction, grey color) and a simulation of far end speech in RCV direction (white color, bursts of CSS can be used). The measurement is carried out in SND direction. In Figure 16 the level the CSS bursts vary and the simulation of background noise is applied with a constant level.



NOTE – The dotted line indicates the repetition or elongation of the test signal to achieve the suitable length for the measurement.

Figure 16/P.502 – Example of test signal structure to evaluate the quality of background noise transmission in SND direction (with far-end speech simulation)

A similar signal structure can be used to determine the quality of background noise transmission coincident **with near-end speech** (see Figure 17). In this case the background noise, and the speech signal simulation (again CSS can be used), are applied and measured on the same direction (opposite to Figure 16), e.g. the SND direction.



NOTE – The dotted line indicates the repetition or elongation of the test signal to achieve the suitable length for the measurement.

Figure 17/P.502 – Example of test signal structure to evaluate the quality of background noise transmission in SND direction (with near-end speech simulation)

The background noise signal and the CSS bursts are both given in the same (white) color to indicate, that both components are fed on the same direction.

8 Switching Characteristics

Switching may influence the speech transmission quality in various situations and conditions: during single talk, during double talk while the near-end speaker is active but when the far-end speaker is active as well. In any case, syllables may be truncated or even complete words or sentences may be missing or interrupted.

8.1 Speech Quality Degradation Perceived Subjectively

The quality degradation perceived subjectively can be described as:

– occurrence of speech gaps;

– missing syllables;

– incomplete words or sentences.

Subjects typically name the disturbance introduced as speech gaps. Often interaction between echo and switching is found.

8.2 Related Objective Parameters

The related objective parameters are well known. A description of the basic parameters such as build up time, hangover time, switching time etc. can be found in ITU-T Recommendation P.340. The main important objective parameters are:

– attenuation range;

– switching time (T_S), hangover time (T_H), build-up time (T_R);

– attenuation in SND/RCV direction during double talk;

– minimum activation level to switch over from RCV to SND direction and from SND to RCV direction;

– echo attenuation.

The fundamental voice switching parameters are threshold level (V_{TH}), build-up time (T_R), hangover time (T_H), switching time (T_S) and attenuation range (a_H). A suitable choice of switching parameter values can minimize the degradation of speech quality introduced by voice switching. Improper choice of parameter values, particularly switching times, may lead to serious clipping effects and loss of initial or final consonants in speech.

Threshold levels should be chosen so that switching is not interrupted by random (environmental) noise sources at either end of the call. In addition, ambient room/network noise effects on threshold should not impair performance. Ambient noise levels can be used to improve threshold performance, as talkers tend to speak louder in a noisy environment than in a quiet one.

Build-up time should be short enough so that the initial transient components of speech are not lost, but not so short that insertion loss removal would be noisy.

Hangover time should be long enough to cover average pauses in speech so that intermittent unwanted switching does not occur before the initial talker is finished, but short enough to allow for reasonable break-in from the second talker.

Switching time from one active state to the other active state should be balanced to best simulate full duplex operation. Switching time is also dependent on both build-up time and hangover time.

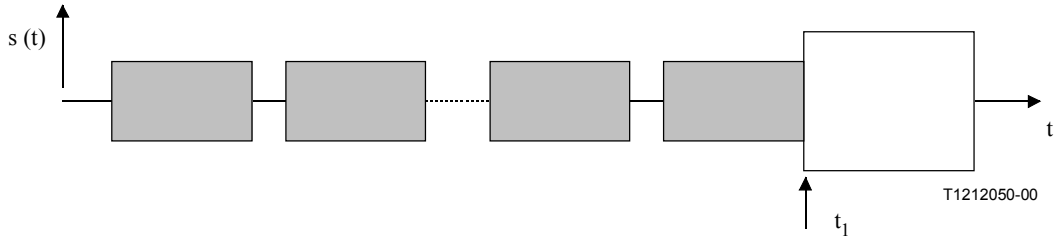
The attenuation range is obtained from the difference between the maximum level at full activation and the minimum level obtained immediately after transmission reversal.

8.3 Analysis Methods

All levels listed in Tables 5 to 7 refer to the MRP in case acoustical levels are given, or to the electrical reference point. In case the access is made electrical instead of the acoustic access, the levels to be used are shown in brackets.

8.3.1 Attenuation Range and Switching Characteristics

One of the most important parameters, especially for implementations with level switching devices is the **attenuation range**. This parameter can be determined with a test signal structure as given in Figure 18.



NOTE – The dotted line indicates the repetition or elongation of the test signal to achieve the suitable length for the measurement.

Figure 18/P.502 – Structure of test signal for attenuation range measurement

A periodical repetition of CSS bursts as a simulation of speech is used to activate one transmission path (grey colour). At the end of one CSS burst, indicated by t_1 on the time-scale, the measurement signal is applied in the opposite path (white colour). This signal consists of a periodical repetition of a voiced sound.

Typical settings are given in Table 5.

Table 5/P.502

	Measurement signal	Measurement signal level	Activation signal (in opposite direction)	Level of the activation signal (in opposite direction)
Switching **RCV -> SND**	Voiced sound in SND direction, period. repetition	–3 dBPa –16.7 dBm	CSS in RCV	–18.3 dBm (incl. pauses)
Switching **SND -> RCV**	Voiced sound in RCV direction, period. repetition	–16.7 dBm	CSS in SND	–4.7 dBPa (–18.3 dBm) (incl. pauses)

The following parameters can be measured:

- **Attenuation range**

  The attenuation range is measured by activating the opposite direction first before measuring the attenuation range of the direction under test. The attenuation range is described as the difference between minimum level and maximum level of the transmitted test signal referred to as the excitation signal. The measurement is conducted simply by evaluating the level

versus time. The time constants to be chosen for this measurement are typically in the range of 5 ms.

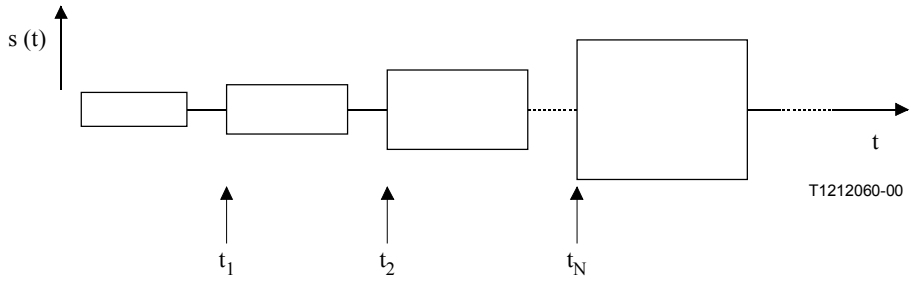
The limits for the attenuation range for the different types of hands-free terminals can be found in ITU-T Recommendation P.340. These limits could be applicable for other devices introducing switching.

- **Switching characteristics (for speech like signals), e.g. switching times**

  The general definitions (and limits) for the switching characteristics can be found in ITU-T Recommendation P.340. The measurements are conducted basically the same way as described above. This means first the opposite direction is activated and afterwards the direction under test is measured (see test signal Figure 18). The level of the transmitted test signal is measured as a function of time. The time constants to be applied for the measurements again are in a range of 1 ms.

8.3.2 Threshold Level and Build-Up Time (for Minimum Activation Level)

The signal structure as given through Figure 19 represents signal parts with increasing levels. The **minimum activation level** to switch on the RCV or SND direction from idle mode can be determined using these sequences. Periods of the CSS (as a simulation of speech) with increasing levels are suited for this signal.



NOTE – The dotted line indicates the repetition or elongation of the test signal to achieve the suitable length for the measurement.

Figure 19/P.502 – Structure of test signal to determine the minimum activation level

Typical settings can be chosen as given in Table 6.

Table 6/P.502

	Active duration/ pause duration	**Level of the first period**	**Level difference between two periods**
CSS for switching in **SND direction**	248.62 ms/ 451.38 ms	–23 dBPa (Note) (–36.7 dBm)	1 dB
CSS for switching in **RCV direction**	248.62 ms/ 451.38 ms	–36.6 dBm (Note)	1 dB
NOTE – These levels should be sufficiently low, to ensure that a wide level range is measured.			

If the transmitted signals are measured and referred to the original measurement signal, the minimum activation level can be determined. The activation can be analysed at the beginning of each signal burst (t_1, t_2, ..., t_N).

The parameters which can be determined using this signal are:

- **Threshold level (for speech like signals)**

  The measurement sequence is shown in Figure 19. The analysis required to find the minimum activation level is a simple level analysis versus time. The time constant for the measurement is chosen between 1 and 5 ms and the level of the measured signal versus time. The excitation signal is displayed. As such, the minimum excitation level, needed to activate the device under test, can be found just by evaluating the level difference during the active parts of the composite source signal. Since the excitation level is known, the minimum threshold level can be determined.

- **Build-up times (for speech like signals, level dependent)**

  The analysis is basically the same as the one described in 8.3.1 except for the time constant which is changed to 1 ms. The switching time is then determined by evaluating the level versus time graph.

8.3.3 Hangover time

The transition from activation to idle can be represented by feeding in an activation signal (e.g. voiced sound of CSS) in one direction, followed by a second signal in the same direction, but of lower level, which does not activate the hands-free telephone (noise signal) (see Figure 20). The second part of the signal measured thus indicates the attenuation, from which the Hangover Time (switch-off time) can be determined.

The duration of the voiced sound is 0.5 s in order to reach a final stable system condition. The level corresponds to standard levels. If level dependant evaluation is needed, the levels as defined in 8.3.2 can be chosen. The second part of the signal (noise signal) has a duration of 1 s. The level must be selected low enough so as not to activate the equipment. The suggested levels to be applied are: −34.7 dBPa for terminals in sending, −50 dBm for electrical access in sending, and −50 dBm for receiving.

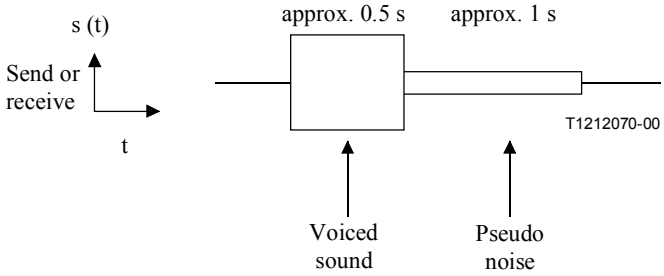
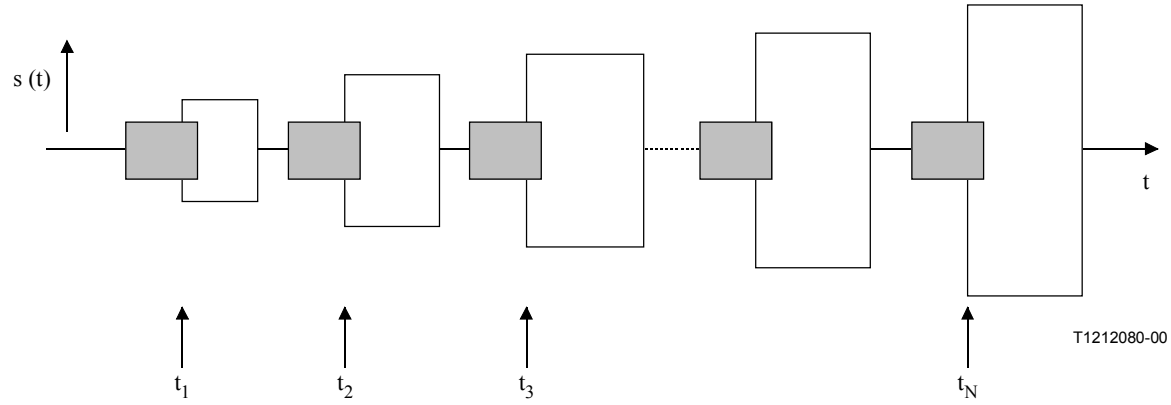


Figure 20/P.502 – Switch-off response measurement

8.3.4 Threshold Level and Switching Time to Switch Over from RCV to SND (SND to RCV)

If the **threshold levels to switch over** from RCV to SND direction (or vice versa, i.e. from SND to RCV direction) shall be measured, the given test signals can be used with slight modifications. As shown in Figure 21, an additional signal is needed in the opposite transmission direction (grey colour). The level of the measurement signal (white colour) increases again periodically. Periods of the CSS are suited for both signals in Figure 21, if the switching characteristics shall be determined applying speech like signals. Again, the signals should be chosen to be uncorrelated.



NOTE – The dotted line indicates the repetition or elongation of the test signal to achieve the suitable length for the measurement.

Figure 21/P.502 – Structure of test signal to determine the minimum activation level to switch over

Suitable settings are given in Table 7.

Table 7/P.502

	Active duration/ pause duration	**Level of the first period**	**Level difference between two periods**	**Level (active part) in opposite transmission direction**
CSS to switch over to **SND direction**	248.62 ms/ 451.38 ms	–13 dBPa (–26.7 dBm)	1 dB	–16.7 dBm (RCV)
CSS in to switch over to **RCV direction**	248.62 ms/ 451.38 ms	–26.7 dBm	1 dB	–3 dBPa (SND) (–16.7 dBm)

Again the activation can be analyzed at the beginning of the signal bursts (t_1, t_2, ..., t_N).

In addition, the same tests be can performed with a simulation of background noise applied at the opposite transmission path.

Assessable parameters are:

- **The minimum activation level (for speech like signals) to switch over**

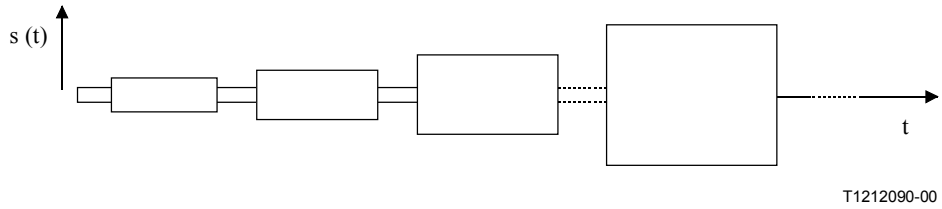
  The minimum activation level to switch over is determined the same way as described for the minimum activation level, the only difference is that the evaluation is made during the pauses of the double talk sequence (see times t_1, t_2 .... in Figure 19).

- **The switching times (switch over)**

  The analysis is conducted as described for the switching times needed for minimum activation (8.3.2).

8.3.5 Switching Characteristics in the Presence of Background Noise

The signal structure given in Figure 22 can be used to determine the **switching characteristics in the presence of background noise**. In this case, a speech like signal (CSS) and a background noise simulation are applied simultaneously on the same channel. The parameters for the CSS can be taken from the tables above.



NOTE – The dotted line indicates the repetition or elongation of the test signal to achieve the suitable length for the measurement.

Figure 22/P.502 – Structure of test signal to determine the minimum activation level in the presence of background noise

The parameters to be determined are:

- **The minimum threshold level in the presence of background noise**

  For this measurement, a background noise is applied (in addition to the excitation signal, see Figure 22). H_{oth} noise or typical background noise (preferably with no high level fluctuations) is chosen with a level according to the typical use of the device (e.g. hands-free phone) under test. For office type telephone, the typical level is in the range –54 to –44 dBPa(A). For other applications, other types of background noise, e.g. car noise with different levels, are suitable.

  The minimum activation level is determined in the same way as described for the minimum activation level without background noise. The sequences of the measurement signals are chosen the same way.

- **The build up time in the presence of background noise**

  Here again the same procedure is used as described before when evaluating the switching times for the minimum threshold level (8.3.2). The only difference is the presence of background noise which is applied in the same manner as described above.

ANNEX A

Detailed Test Methodology for Temporally Weighted ERL_t

A.1 Echo Return Loss Algorithm

The temporally weighted echo return loss ERL_t measurement method is described. This method requires that the echo and the source signal be recorded over the duration of the measurement, and post processing to be used. Real-time measurement techniques are possible, but are not described in this ITU-T Recommendation.

Freezing the canceller is not recommended for ERL tests. Some results with non-stationary signals have shown that convergence times and subsequent converged ERL when "thawed" depend upon the point in time at which the canceller was frozen.

A.1.1 Echo Return Loss, Temporally Weighted (ERL_t)

Temporally weighted ERL, ERL_t, is intended to:

- Provide a measure of time dependent on echo return loss with peaky behaviour, psycho-acoustically weighted; the ERL_t.
- Provide an estimate of the number of potentially objectionable echo bursts, and the psycho-acoustically weighted echo return loss during the bursts.

The echo signal is first filtered to model the frequency selectivity of human hearing at loudness levels of 30 Phons, as described in A.1.2. This weights the echo power in a way that the human hearing response would.

Noise reduction may then be applied and the echo and stimulus files synchronized. Noise reduction is where the noise is measured and subtracted from the echo plus noise to arrive at a better estimate of the echo alone. Such a measurement should occur for at least two seconds after all stimulus activity has stopped. Echo and source are converted into 4 ms power averaged frames allowing adequate resolution and immunity to synchronization errors.

If the stimulus is inactive, the algorithm simply skips that frame, and moves on to the next echo and stimulus frames. If the stimulus is declared active, the echo frame is compared with a threshold to determine if an echo event occurs. The period of echo activity between inactive echo states is termed an echo "event". These events are then weighted using psycho-acoustic modelling.

By using a threshold of −65 dB (5 dB above A-law or μ-law noise floor), ERL_t can be determined. Similarly, for A law, the threshold must also be 5 dB above the noise floor. The actual test algorithm in pseudo code and it is detailed in A.1.4.

A.1.2 Modelling Echo Audibility

In modelling echo audibility, the algorithm accounts for 3 fundamental aspects of human hearing behaviour:

1) The frequency selectivity of human hearing at a loudness level of 30 Phons ("Fletcher-Munson" response equivalent to 30 dB at 1 kHz) [1].

   Thirty Phons was chosen as it represents echo levels that result from terminals that just fail handset terminals coupling loss specifications (determined using loss planning analysis). Variance from 20 to 50 Phons provide essentially the same weighting within the telephony band. An A weighted filter is used.

   Note that the use of this exact weighting characteristic assumes headphone/handset type listening, or "mean audible pressure" (MAP) response. Free-field listening such as over a hands-free would require the Robinson and Dadson "mean audible field" (MAF) weighting, but the difference is slight. MAP weighting will be used to better reflect the more common use of handset.

   The average loss of the filter with white noise is 1.3 dB when measured using ERL_s or ERL_t. With non-stationary signals, the loss will be time dependent.

2) The ear's tendency to combine the loudness of sequential signals even though they may be discrete in time ("temporal combination"). This typically occurs when the two signals are separated by a silent period, which is less than 20 ms [2], [3], [4]. If two bursts of echo are separated by a period of inactivity less than 20 ms, they are considered as one longer echo event as far as loudness is concerned. This continues until the gap between events is at least 20 ms, at which time the echo event is declared over. This can be thought of as a 20 ms hangover for the current echo event. During this hangover period, echo and stimulus powers are not included as part of the event. An example of temporal combination is given in Figure A.1.

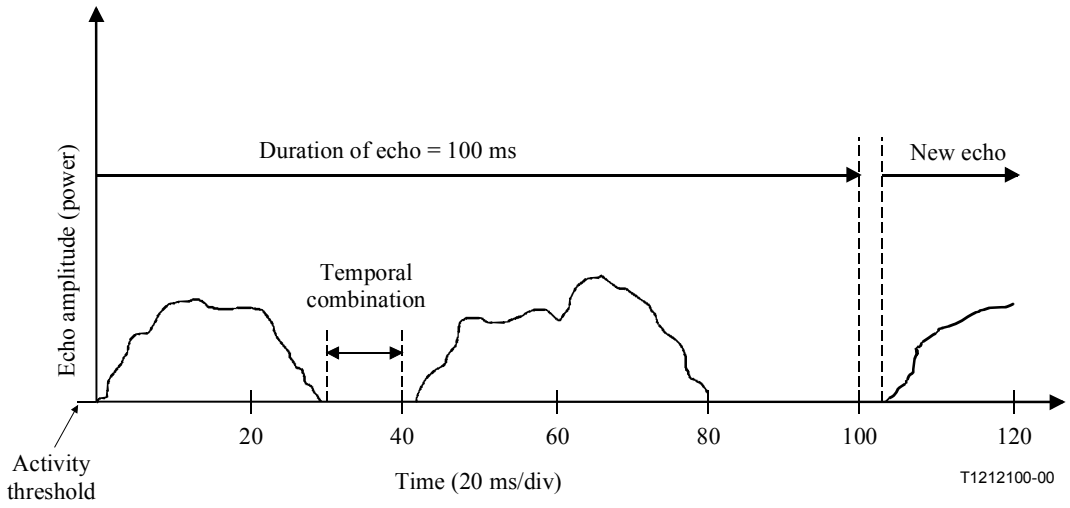


Figure A.1/P.502

3) The duration of the total echo event after temporal combination is measured based on the ear's natural temporal integration behaviour. The total duration includes any gap(s) between events that are captured by temporal combination, but not the final 20 ms hangover.

If the duration is less than 750 ms, the level of the event is reduced to account for the temporal integration behaviour of human hearing. An equation describing the relationship was derived based upon audition studies with noise:

Temporal integration weighting = $-23 + 8 \log(t)$ in dB

where t = total duration of echo event (ms), $t < 750$ ms

Note that tones result in a slightly different relationship, but it was felt that noise was a much closer approximation to the true nature of the echo than a sine.

A graphical representation of temporal weighting is shown in Figure A.2.

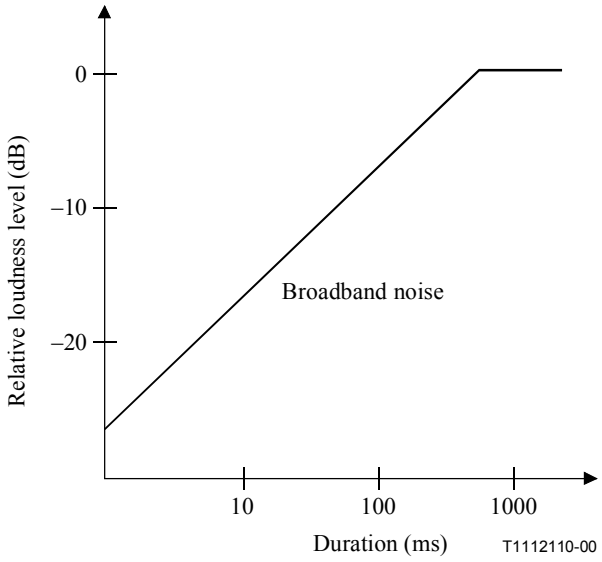


Figure A.2/P.502

If the duration is longer than 750 ms, the level of the total event is left unweighted. Note that test results have shown echo bursts less than 750 ms to be common occurrences from cancellers.

A.1.3 Expressing ERL_t Results

Traditional ERL methods refer the echo power during the duration of measurement to the source power during the duration of measurement to arrive at the echo return loss. In this method, the final weighted power of echo during each event is referred to the power of the source signal during the same event, to arrive at the "Active ERL_t", $AERL_t$, of each event. The echo is referred to the source signal during the event only, as this is the way in which our ear would compare the echo.

A long-term average of the weighted active echo return loss is found by summing the power of all weighted echo during active events, and comparing to the power of the source as seen during all events only. The result is the "Active Long Term ERL_t".

For comparison with traditional ERL methods, the power of all weighted echo during events is summed, then referred to the total source power as measured for the entire duration of the measurement. The result is the "Long-Term ERL_t".

Note that the terminology for ERL_t results was chosen to be consistent with ITU-T Recommendation P.56.

Other statistics compiled include minimum and maximum $AERL_t$, standard deviation ("sigma") of $AERL_t$, the mean of $AERL_t$ and the total number of echo events (combined events due to the "Haas" effect are considered one total event). Also included are: the number of echo events per minute, the percentage of echo event free speech, the number of events < 750 ms, the average length of an event and the duration of source inactivity.

A.1.4 ERL_t Test Algorithm

ERL_t is method for evaluating the echo return loss of a terminal using psychoacoustic modelling and for predicting the occurrences of potentially objectionable echoes. It incorporates 3 fundamental aspects of human audition:

- frequency selectivity of human hearing ("Fletcher Munson" response);
- temporal addition of level for events within 20 ms of each other ("Haas" effect);
- temporal integration for stimuli below 750 ms.

The implementation details of the algorithm follow.

A source signal as described in clauses 4 and 5 is used. Speech based stimulus signals are recommended as their results are most representative of real world usage. The system output is always some echo or noise making its way through the system uncancelled.

The stimulus and echo should be recorded and made available in digital format. User inputs regarding set type (analogue or digital), EPD_n and double talk or single talk tests should be available. Calibration parameters should be used to scale echo and stimulus frames to absolute values, and hybrid processing should have removed hybrid echo for 2 wire analogue sets.

The stimulus and the echo files will be processed as power values averaged over 4 ms frames. The successive stimulus file frames will be termed x_i, the echo frames will be denoted y_i, where i = 1, 2, 3.... is the actual frame index. Intermediate frames conforming to an "echo event" will be noted as x_k, and y_k, where k = 1, 2, 3... is the echo event index, and is reset when the event ends a new one commences.

Statistics compiled during the ERL_t measurement include the Active Long-Term ERL_t ($ALERL_t$), Long-Term ERL_t ($LERL_t$), minimum and maximum Active ERL_t (MINERL, MAXERL), its sigma and mean, the total number of echo events (combined events due to the "Haas" effect are considered one total event) (NEVENTS). Also included are: the number of echo events per minute (NEVMIN), the percentage of echo event free speech (PER), number of events < 750 ms (N750), the average length of an event (AVGEVENT), and the duration stimulus was inactive (DUR). The terminology for ERL_t results was chosen to be consistent with ITU-T Recommendation P.56. The duration of stimulus inactivity is not included in the time based results.

ERL_t Algorithm

- *Step 1* (Optional but recommended)

  Calculate the correlation of stimulus and echo file to fine tune EPD_n. Use the criteria that the present correlation peak occurs at EPD_n unless a following correlation peak has a magnitude at least 10 dB greater. This approximate guideline is based upon subjective studies on delay detection with multiple impulses.

- *Step 2*

  Align the echo and stimulus files in time by removing delay equal to EPD_n from the echo file.

- *Step 3*

  The individual echo samples are processed through a filter approximating the mean audible pressure equal loudness contour for 30 Phons. This can be accurately approximated (within ±1 dB from 200 Hz to 2500 Hz) by a first order high pass filter with a −3 dB point of 800 Hz.

- *Step 4*

  If it can be assumed that the noise in the echo path is stationary and uncorrelated with the echo, the noise is measured for 2 seconds after the stop of source and echo activity. The noise is then subtracted from the echo plus noise to arrive at a better estimate of the echo alone.

- *Step 5*

  Samples are converted to absolute numbers using the calibration data. The stimulus samples are combined into 4 ms power averaged frames denoted as x_i. The weighted, noise filtered echo samples are combined into 4 ms power averaged frames denoted as y_i.

- *Step 6 Begin Echo Return Loss Calculations*

  Initialize variables:

  i = 0 (frame counter);

  j = 0 (frame counter for inactive signal duration);

  $n_{k=0}$ = 0 (number of frames in current echo event);

  NSAMPS = 0 (accumulated number of frames for all events);

  HAAS = 0 (counter up to 20 ms);

  $e_{i=0}$ = 0 (running summation of all echo power for all events after weighting, as seen at frame counter i);

  $p_{i=0}$ = 0 (running summation of all stimulus power during the measurement, as seen at frame counter i);

  $e_{k=0}$ = 0 (running summation of echo power during the particular echo event after weighting, as seen at event frame counter k);

$s_{k=0}$ = 0 (running summation of stimulus power during the particular echo event after weighting, as seen at event frame counter k);

WEIGHT = 0 (temporal based weight of most recent event);

LEVENT = 0 (echo return loss level of most recent event, after weighting);

NEVENT = 0 (total number of echo events);

N750 = 0 (total number of echo events < 750 ms);

MINERL = 75 (minimum echo return loss level of all events);

MAXERL = 0 (maximum echo return loss level of all events);

EVENT[NEVENT] = 0 (initialize array for all event loss levels (in dB) to zero; used to calculate sigma);

TEMPSK = 0 (running sum of stimulus power during all events);

SUM = 0 (used in calculating sigma);

SQ = 0 (used in calculating sigma);

- *Step 7*

Increment frame counter and read in 4 ms averaged echo power y_i, and 4 ms averaged stimulus power, x_i; if there are no more valid inputs and either measurement file is complete, go to step 8.

1 i = i +1 (unless last i, then go to step 8).

Sum stimulus powers:

$p_i = p_i + x_i$

Is stimulus loud enough for a valid echo loss calculation? If not, disregard present frame and move to next frame.

4 If x_i < (long-term stimulus rms level – 25 dB)

j = j + 1

i = i + 1

Go to 4.

Else:

Test echo against threshold:

If y_i < –65 dB {5 dB above A-law or μ-law noise floor}

Increment frame event counter:

k = k +1

Increment frame event length including any gaps < 20 ms:

$n_k = n_k + 1$ + HAAS

Reset "Haas kicker":

HAAS = 0

Accumulate echo power of event:

$e_k = e_k + y_i$

Accumulate stimulus power during event:

$s_k = s_k + x_i$

Go to 1.

Else:

Has there been no event within last 20 ms?

If k = 0

HAAS = 0

Go to 1.

Else:

There has been an event within the last 20 ms:

HAAS = HAAS + 1

Has 20 ms without an event elapsed after a recent event?

If HAAS*4 < 20

Go to 1.

Else:

An event is over, add an event to the event counter:

NEVENT = NEVENT + 1

Increment the total events duration counter by adding the duration in frames of the most recent event:

$NSAMPS = NSAMPS + n_k$

Was the most recent event duration < 750 ms?

If $n_k*4 < 750$

Calculate temporal integration weighting for most recent echo event:

$WEIGHT = 8*\log10(n_k*4) - 23$

Increment the counter for the number of events that were temporally weighted:

N750 = N750 +1

Else:

Calculate weighted echo return loss of the most recent event in dB:

$LEVENT = 10*\log10(s_k/e_k) - WEIGHT$

Store the minimum and maximum echo return losses in dB:

IF LEVENT < MINERL; MINERL = LEVENT

IF LEVENT > MAXERL; MAXERL = LEVENT

Store the echo return loss of the most recent event in dB for future sigma calculation:

EVENT(NEVENT) = LEVENT

Reconvert the echo return loss of the most recent event into linear; recalculate weighted linear echo power:

$e_k = s_k/(10**(LEVENT/10))$

Accumulate all the echo event powers for future use in calculating $ALERL_t$ and $LERL_t$:

$e_i = e_i + e_k$

Accumulate all the stimulus powers during events for future use in calculating $ALERL_t$:

$TEMPSK = TEMPSK + s_k$

Reset echo event variables:

k = 0

$n_k = 0$

WEIGHT = 0

HAAS = 0

$e_k = 0$

$s_k = 0$

Go to 1.

- *Step 8*

Calculate Active Long-Term ERL_t ($ALERL_t$), Long-Term ERL_t ($LERL_t$), the number of echo events per minute (NEVMIN), the percentage of echo event free speech (PER), the average length of an event (AVGEVENT) and duration during which speech was inactive (DUR).

NOTE – Zero check e_i before computing; if $e_i = 0$, set $ALERL_t$ and $LERL_t$ to 100 dB.

$ALERL_t = 10*\log10(TEMPSK/e_i)$

$LERL_t = 10*\log10(p_i/e_i)$

NEVMIN = 60*NEVENT/((i-j)*0.004) {number of events per minute}

PER = 100*((i-j) – NSAMPS)/(i-j) {percentage of echo free speech)

AVGEVENT = NSAMPS*4/NEVENT {average length of an event in milliseconds}

DUR = j**0.004

Calculate sigma by analysing the EVENT array which contains the echo return loss of each event; each event, regardless of duration, is given equal weighting in the sigma calculation; the suggestion is that it is the transition between discreet events and not their duration that is most objectionable.

Loop j from 1 to NEVENT:

SUM=SUM+EVENT(j)

SQ=SQ+EVENT(j)**2

ENDLOOP

SIGMA = SQRT(SQ/NEVENT – [SUM/NEVENT}**2)

Calculate mean of the events:

MEAN = SUM/NEVENT

- *Step 9*

Output statistics:

Print $ALERL_t$, $LERL_t$, MINERL, MAXERL, NEVENT, NEVMIN, PER, N750, AVGEVENT, DUR,

SIGMA, MEAN

ANNEX B

Double talk measurement filters for Method A

Double talk testing requires the use of notch and bandpass filters at various frequencies. A recommended implementation is tabulated below.

The terms described are:

- fpl: lower frequency at which the bandpass or bandstop is at –3 dB;
- fpu: upper frequency at which the bandpass or bandstop is at –3 dB;
- fsl: lower frequency at which the bandpass or bandstop is at –atten dB;
- fsu: upper frequency at which the bandpass or bandstop is at –atten dB;
- atten: the specified full attenuation of the filter;
- atten (actual): the actual full attenuation of the filter;
- ripple: ripple of the filter in dB (±);
- gain: gain of the bandpass filter (linear) in the pass band;
- order: filter order in taps (8 kHz sample rate) for the bandpass. The bandpass ringing time is order times 125 μs. For the bandstop, order refers to the order of the biquad (elliptical).

Filter type	500 Hz FIR bandpass	1 kHz FIR bandpass	1.75 kHz FIR bandpass	2.5 kHz FIR bandpass	500 Hz IIR bandstop	1 kHz IIR bandstop	1.75 kHz IIR bandstop	2.5 kHz IIR bandstop
fpl	495	990	1 733	2 475	400	800	1 450	2 100
fpu	505	1 010	1 767	2 525	610	1 250	2 100	2 950
fsl	435	900	1 611	2 302	435	900	1 610	2 300
fsu	570	1 100	1 900	2 715	570	1 100	1 900	2 715
atten	30	30	30	30	30	30	30	30
atten (actual)	31	29.5	34	34	30	40		
ripple	1	1	3	1	1.5	1.5	1.5	1.5
gain	0.92	0.9	0.78	0.99				
order	160	100	80	60	6	6	6	6

The bandpass filter's ringing time will impact the measurement if not accounted for. Measurements must commence only after the filter had stopped ringing due to initial application. This is necessary so that a clean reference measurement can be made for attenuation tests and clipping tests. The longest ringing time is 20 ms for the 500 Hz bandpass filter. Since the averaging window for measurement in attenuation testing is 8 ms, the bandpass filter must be inserted at 20 + 8 = 28 ms before the onset of double talk, or 60 – 0. (See Figure 7.)

ANNEX C

Training Sequence Description

C.1 Canceller Training prior to Double Talk

Basic information about the timing in a conversation can be found in ITU-T Recommendation P.59: talk spurts, pauses, double talk, mutual silence. The training masks given below are derived from this Recommendation.

C.1.1 Double Talk Training Activity Masks

The exact amplitude masks will now be specified along with signal amplitude characteristics during double talk. Each type of double talk test has special requirements for signal duration and amplitude during double talk. For echo return loss, the duration must be long enough to capture any divergence, but not so long as to be a burden on test system memory resources or so long as to result in an unacceptable computation time. Tests have shown a 20 second double talk duration to be acceptable for double talk echo return loss testing. Once double talk has ended (the talker initiating double talk becomes inactive), the echo return loss measurement may continue for 10 seconds (the talker active just before double talk remains active) to measure recovery after double talk. After that time, two seconds of silence should be played. In this way, the noise in the echo path can be measured. If it can be assumed that the noise and echo are uncorrelated, and that the noise is stationary, the noise measured in the two seconds may be subtracted from the echo plus noise during double talk to arrive at a more precise measure of the echo during double talk.

The duration of double talk during double talk attenuation and clipping tests may be much shorter. As all time constants under study should be less than 200 ms, the double talk duration is set at 200 ms. Analysis is continued (the talker active just before double talk remains active) for one second after the end of double talk for the attenuation tests in order to measure any loss removal as single talk is re-entered. There is no need to estimate and correct for noise in the double talk attenuation and clipping tests.

The recommended masks are given on Figures C.1 and C.2.

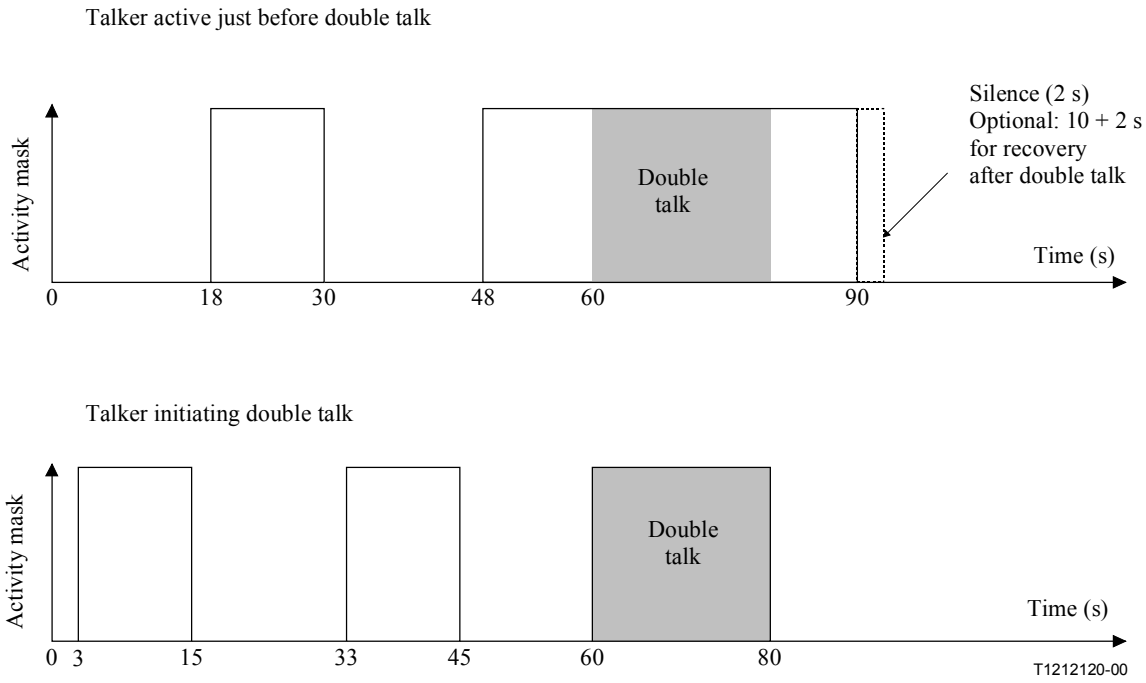


Figure C.1/P.502 – Echo Return Loss Test Activity Mask

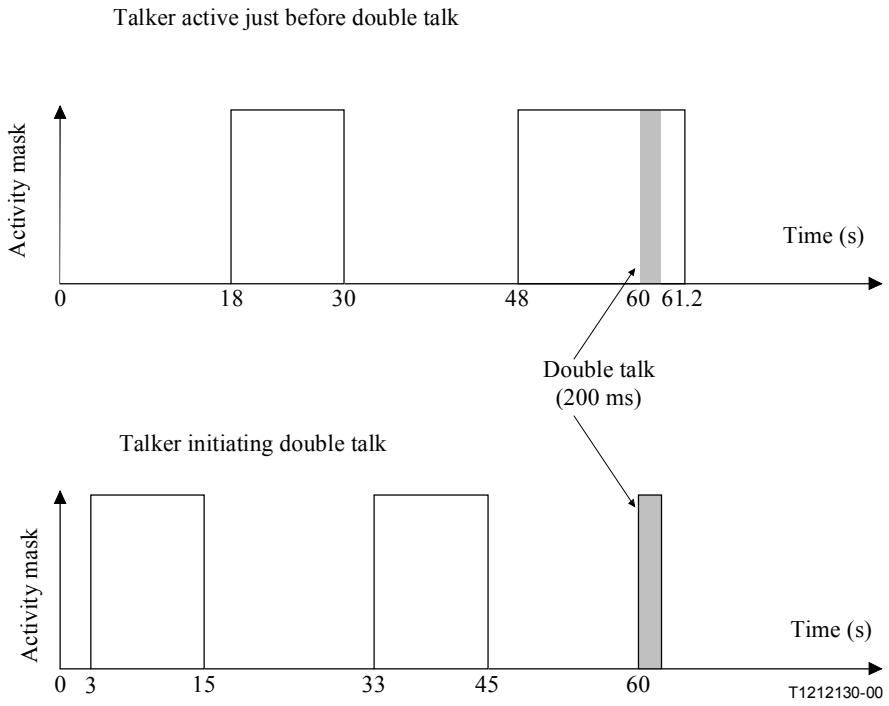


Figure C.2/P.502 – Attenuation Test and Clipping Test Activity Mask

C.1.2 Synchronizing the Double Talk Training Activity Masks

The timing of the masks must be synchronized to avoid pre-mature double talk. This involves accounting for the 1.5 ms air path delay between the artificial mouth and the HFT. The mouth simulator signal should be initiated 1.5 ms before the R_{in} signal by delaying the stimulus file used on receive by 1.5 ms. The start of double talk is defined as occurring when the microphone location sees valid send activity and R_{in} sees valid receive activity.

In the activity mask diagrams, the t = 0 starting point refers to the beginning of the file applied at R_{in}. The starting point for the mouth simulator signal can be thought to be t = –1.5 ms, but the terminal will see them synchronized. At the 60 second mark, double talk is entered and double talk testing begins.

C.1.3 Compensating for Measurement Filters

The double talk test methods require the use of filters injected in the audio path. These filters will have an impact on the time domain resolution and the precise moment at which double talk testing can begin. By using filters of known ringing time, the measurement can be put in a wait state while the filter ringing settles.

APPENDIX I

Bibliographic references

[1] HEARING, GULICK, GESCHEIDER, FRISNA: *Oxford University Press*, 1989.

[2] DAVIS (D.), DAVIS (C.): The LEDE Concept, *JAES*, 1985.

[3] OLIVE (S.): The Detection of Reflections, *JAES*, 1987.

[4] OLIVE (S.): Modification of Timbre by Resonance, *JAES*, 1988.

[5] ZWICKER (E.), FASTL (H.): Psychoacoustics, *Springer Verlag*, 1990.

[6] Enhancements of hands-free telecommunications, *Esprit Consortium, Annals of telecommunications*, 49 Nos. 7-8, 1994.

[7] Methodology of Evaluation and Standards, Deliverable 1.2, *Freetel*, July 1993.

[8] GIERLICH (H.W.): The auditory perceived quality of hands-free telephones: auditory judgements, instrumental measurements and their relationship. *Speech Communication 20*, pp. 241-254, October 1996.

[9] Subjective valuation procedures for hands-free telephones – Double talk performance. *ITU-T Contribution COM 12-5*, Geneva, April 1997.

[10] Subjective evaluation of hands-free telephones using conversational tests, specific double talk tests and listening only tests. *ITU-T Contribution COM 12-6*, Geneva, April 1997.

[11] Double talk measurements for hands-free telephones: Measurement proposals and measurement results. *ITU-T Contribution COM 12-32*, Geneva, February 1998.

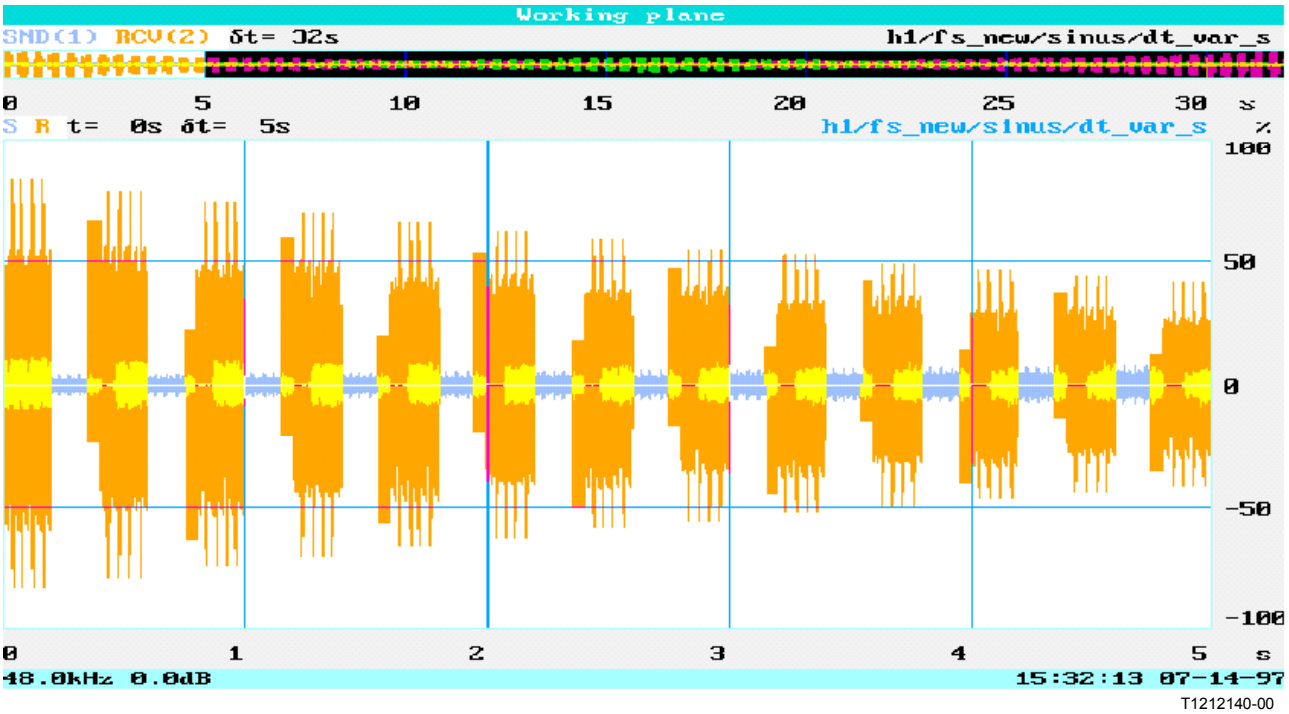
APPENDIX II

Example Evaluations

II.1 Some Example Evaluations according to clause 5

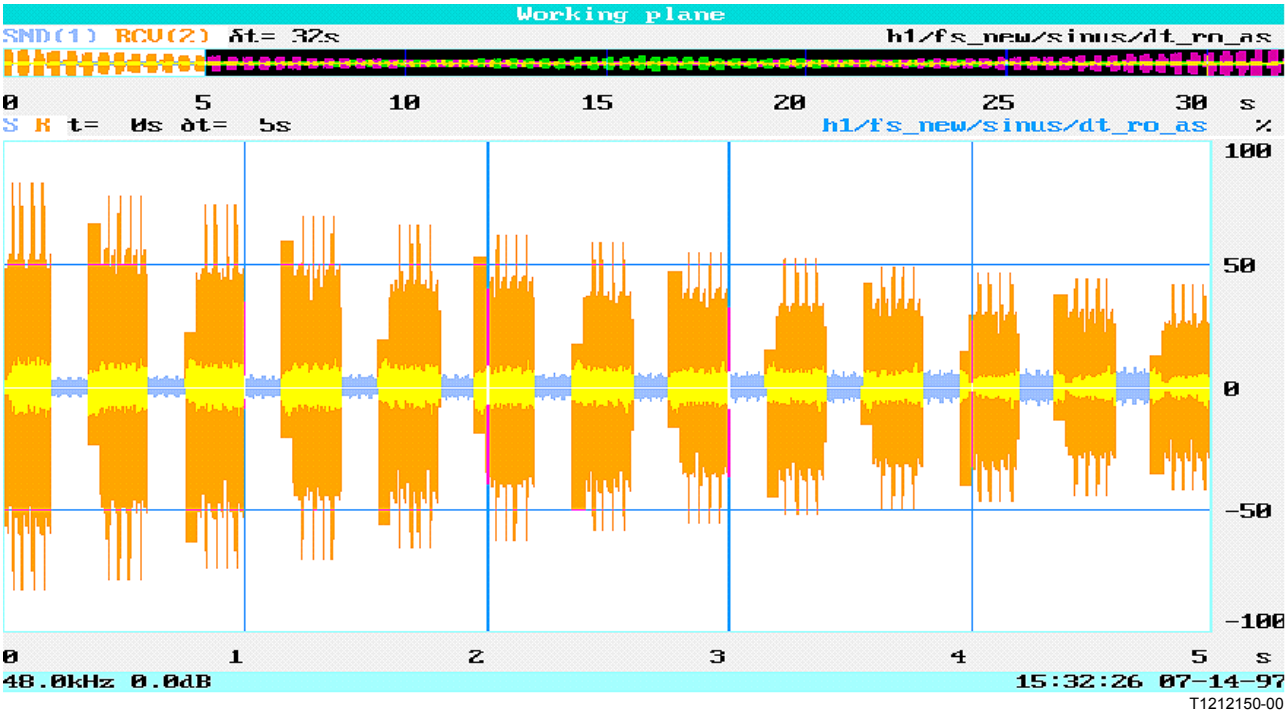
II.1.1 Frequency Responses During Double Talk

Based on the evaluations of hands-free telephones some examples for the application of the procedure are shown in Figure II.1.



Dark: excitation signal in sending direction.
Light: measured signal in sending direction.
Pauses filled by double talk signal transferred from HFT loudspeaker to HFT microphone.

Figure II.1/P.502 – Sending direction of a level switching HFT

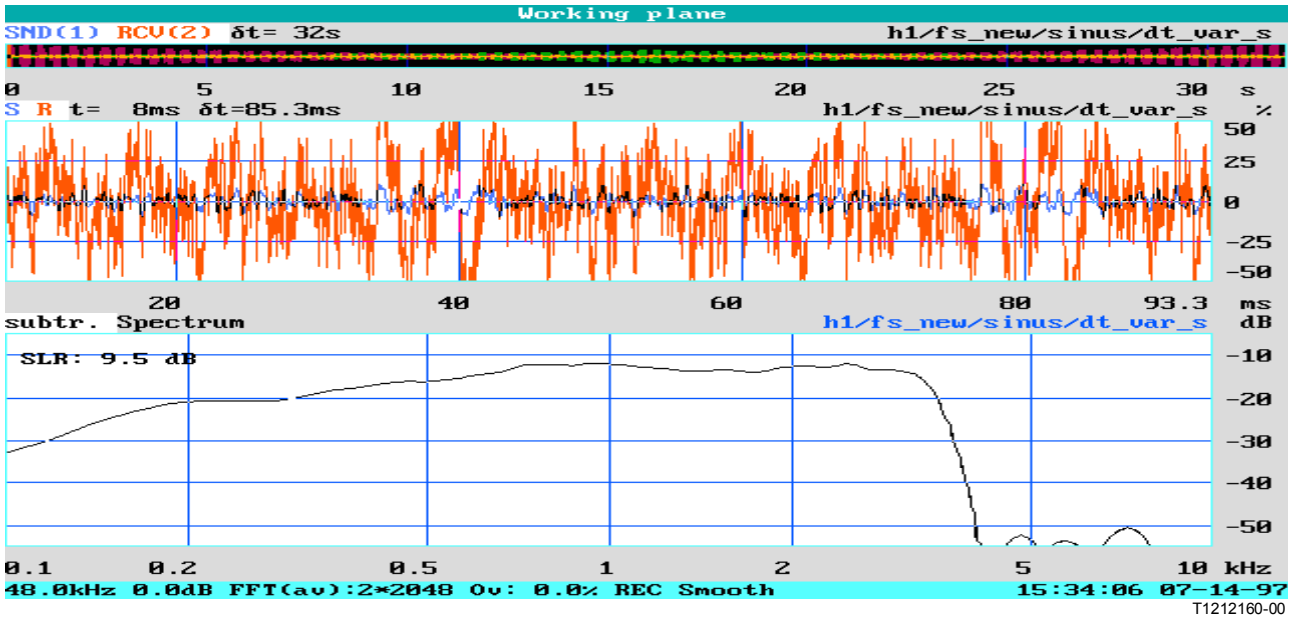


Dark: excitation signal in sending direction.
Light: measured signal in sending direction.
Pauses filled by double talk signal transferred from HFT loudspeaker to HFT microphone.

Figure II.2/P.502 – Sending direction of an echo cancelling and level switching HFT

Figures II.1 and II.2 show the measured result when using this type of test signals for the evaluation of the sending direction of two individual hands-free telephones. Relevant for the measurement itself are only the periods where just the sending direction signal is present. This is according to Figure 3, a period of 150 ms which starts 50 ms after the voiced sound activation signal. During that time interval only the sending signal is present. Figures II.1 and II.2 show the measurement results just as a time sequence for the first 5 s.

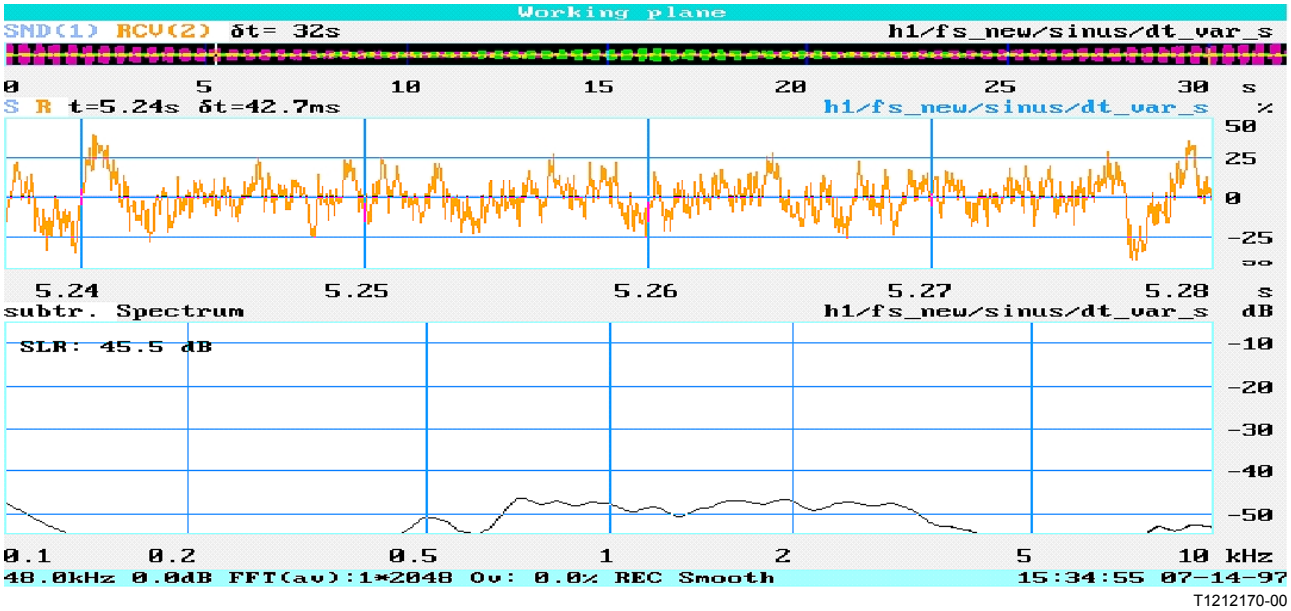
Figures II.3, II.4 and II.5 now show the result of frequency response and SLR measurements for the level switching hands-free telephone in sending direction.



SLR = 9.5 dB

Measured in the beginning with –4.7 dBPa signal level.
Upper part of the picture: time history signal.
Lower part: frequency response and loudness rating.

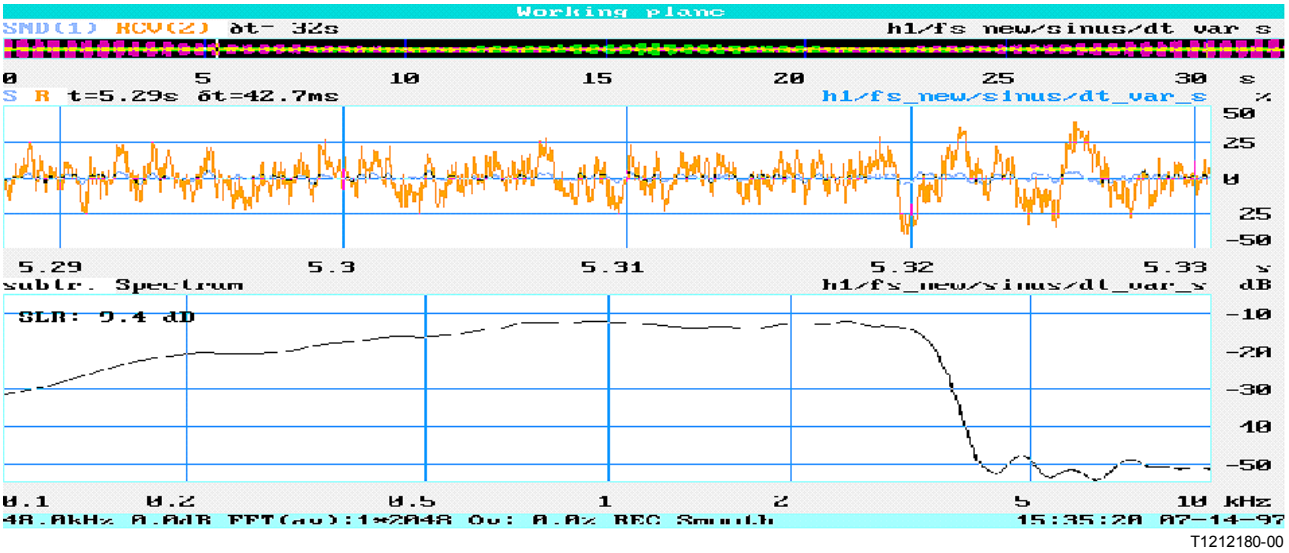
Figure II.3/P.502 – Level switching HFT: transfer characteristics and loudness ratings



SLR = 45.5 dB

Measured after ~5 s, excitation signal level –9.5 dBPa, measurement before activation of the telephone.
Upper part of the picture: time history signal.
Lower part: frequency response and loudness rating.

Figure II.4/P.502 – Level switching HFT: transfer characteristics and loudness ratings



SLR = 9.4 dB

Measured after ~5 s, excitation signal level –9.5 dBPa, measurement before activation of the telephone.
Upper part of the picture: time history signal.
Lower part: frequency response and loudness rating.

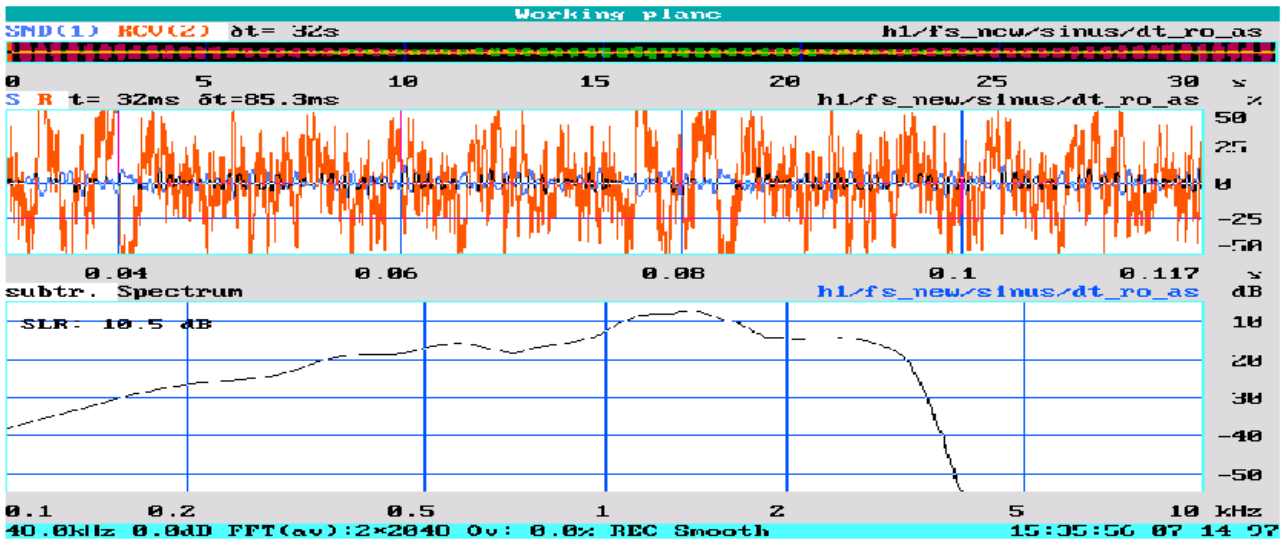
Figure II.5/P.502 – Level switching HFT: transfer characteristics and loudness ratings

From Figures II.3 to II.5 it can be seen how level switching device in sending direction reacts on this type of signal. From the analysis it is clear that the sending loudness rating in the beginning of the sequence, where the sending direction is fully activated, is 9.5 dB. During periods where the activation of the system is not complete, the sending loudness rating is 45.5 dB (see also frequency response measured at that time). From this it can be said, that the telephone attenuation is about 30 dB, when comparing the frequency responses, no frequency dependent characteristics can be seen.

The activation time for this device, as a function of sending and receiving level, can be measured as well. For example, after ~5 s, it takes about 90 ms. The level in sending at this point in time is –9.5 dBPa, the level in receiving is –31.5 dBm.

Transfer characteristics, Loudness Ratings and Switching times for other level combinations can be evaluated accordingly.

Figures II.6 to II.8 show the same type of evaluation, but now for an echo cancelling/level switching device. Since this telephone does not show any switching during double talk sequences of the beginning, after 5 s and after 16 s have been evaluated.

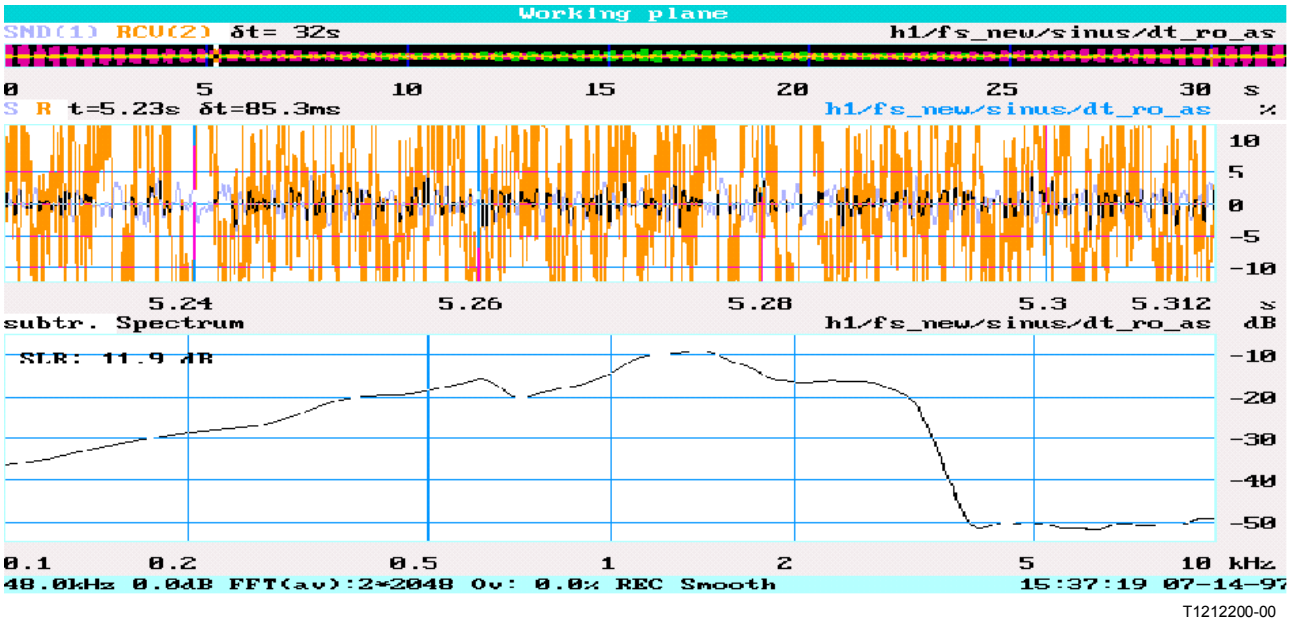


T1212190-00

SLR = 10.5 dB

Measured in the beginning with –4.7 dBPa signal level.
Upper part of the picture: time history signal.
Lower part: frequency response and loudness rating.

Figure II.6/P.502 – Echo cancelling and level switching HFT: transfer characteristics and loudness ratings

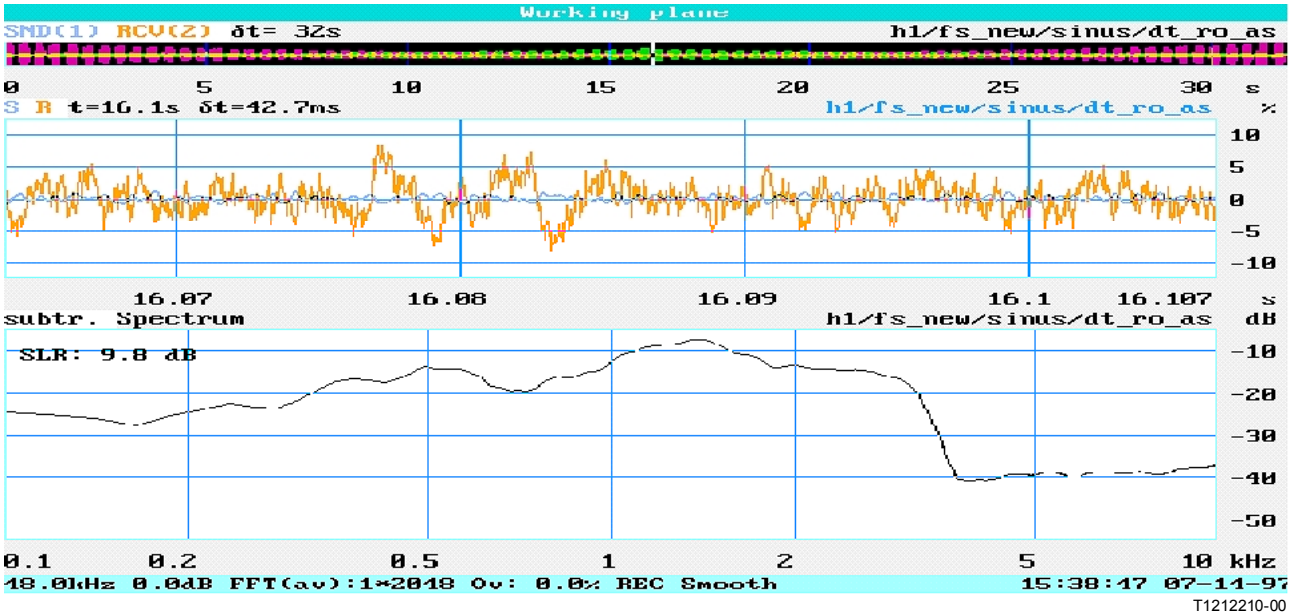


T1212200-00

SLR = 11.9 dB

Measured after 5 s, excitation signal level –9.5 dBPa, measurement before activation of the telephone.
Upper part of the picture: time history signal.
Lower part: frequency response and loudness rating.

Figure II.7/P.502 – Echo cancelling and level switching HFT: transfer characteristics and loudness ratings



SLR = 9.8 dB

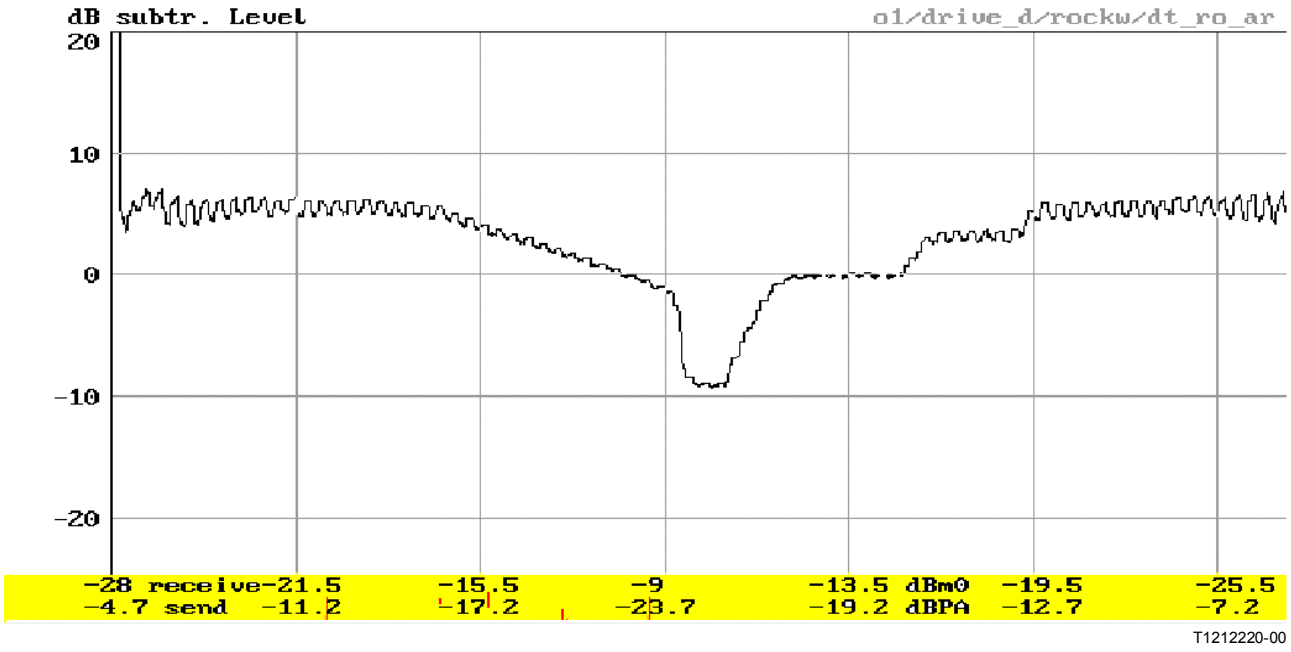
Measured after 5 s, excitation signal level –24.7 dBPa, measurement before activation of the telephone.
Upper part of the picture: time history signal.
Lower part: frequency response and loudness rating.

Figure II.8/P.502 – Echo cancelling and level switching HFT: transfer characteristics and loudness ratings

The evaluation of this set shows that transfer characteristics and loudness ratings under all conditions in the whole level range of –4.7 dBPa down to –24.7 dBPa in the presence of the double talk signal are stabile, no switch off is realizable. However, from the loudness ratings and frequency responses it is clear, that some kind of companding effects can be noted. The loudness rating is not always stable, there is a variation (signal dependent) of about 2 dB in loudness rating.

II.1.2 Level Variations During Double Talk

A general view on the behaviour of a HFT in double talk conditions can be seen in Figure II.9. Here the receiving path amplification depending on the receiving as well as on the sending signal level can be seen. The test signal used for this evaluation is shown in Figure 2.

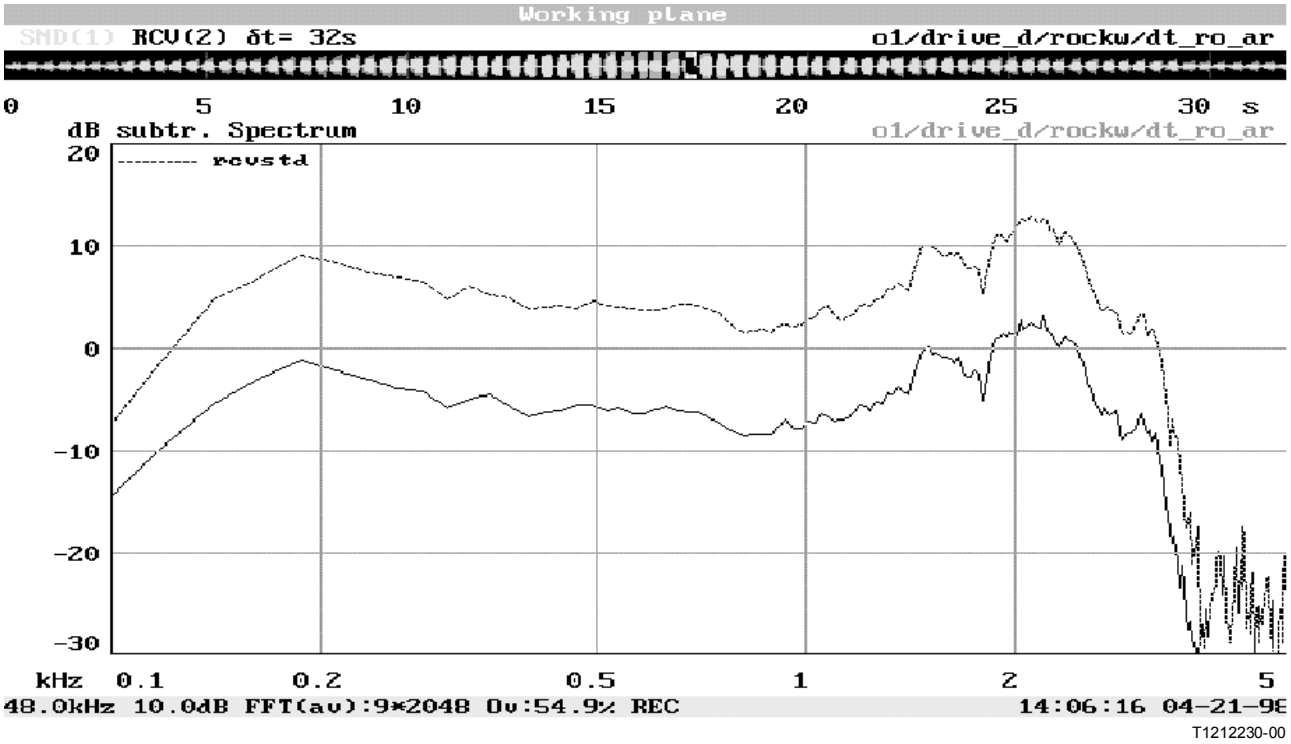


Level variation:
Receiving: –28 dBm0 ... –8 dBm0 ... –28 dBm0
Sending: –4.7 dBPa ... –24.7 dBPa ... –4.7 dBPa

Figure II.9/P.502 – Variation of amplification in receiving direction during double talk

It is obvious, that an asymmetric AGC is in operation which for higher receiving signal levels will have a noticeable effect on the loudness fluctuation perceived subjectively. For increasing receiving levels first a smooth decrease in amplification can be found followed by a sudden attenuation of about 10 dB. While decreasing the receiving level again, 3 steps in the attenuation can be found: 10 dB, 12.5 dB and 15 dB were specifically the 10 dB step will have a noticeable impact and the perceived quality. The level variation is quite high (15 dB) and incorporates a rapid change of amplification.

A more detailed evaluation using this type of test signal is shown in Figure II.10. Figure II.10 shows frequency responses and loudness ratings (at –28 dBm0 and –8 dBm0 receiving signal level) which are measured in the receiving direction during double talk using the pause in the double talk signal. So, the measured frequency response and loudness rating was measured using 88 ms of signal for analysis. In this example the level variation as a function of frequency of the HFT under test can be seen.



RLR = 6.3 dB with:
high level signal (–4.7 dBPa) excitation in sending and low level signal excitation in receiving (–28 dBm0)

RLR = 19.5 dB with:
low-level excitation in sending (–24.7 dBPa) an high level excitation (–8 dBm0) in receiving.

Figure II.10/P.502 – Frequency responses and loudness ratings in receiving direction during double talk

II.1.3 Switching During Double Talk

One example derived from the sending direction is shown in Figure II.11. In sending directions for sending levels less than −10.7 dBPa level switching during double talk can be found. In order to determine the subjective relevance of this switching, the time constants and switching gain need to be determined as well as the level dependent change of these parameter. Switching time and gain can be seen in Figure II.11. 35 ms after double talk the amplification is raised by about 10 dB. The switching time is about 10 ms.

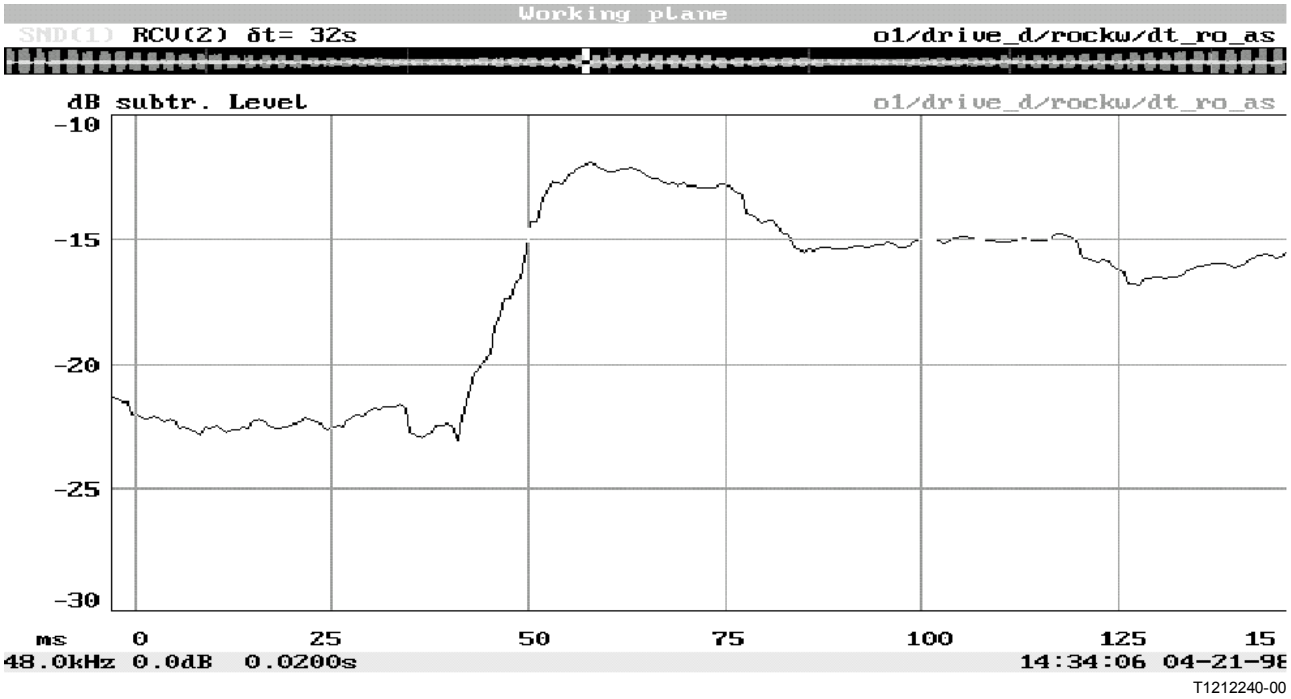


Figure II.11/P.502 – Switching time and level variation during double talk

SERIES OF ITU-T RECOMMENDATIONS

Series A	Organization of the work of ITU-T
Series B	Means of expression: definitions, symbols, classification
Series C	General telecommunication statistics
Series D	General tariff principles
Series E	Overall network operation, telephone service, service operation and human factors
Series F	Non-telephone telecommunication services
Series G	Transmission systems and media, digital systems and networks
Series H	Audiovisual and multimedia systems
Series I	Integrated services digital network
Series J	Transmission of television, sound programme and other multimedia signals
Series K	Protection against interference
Series L	Construction, installation and protection of cables and other elements of outside plant
Series M	TMN and network maintenance: international transmission systems, telephone circuits, telegraphy, facsimile and leased circuits
Series N	Maintenance: international sound programme and television transmission circuits
Series O	Specifications of measuring equipment
Series P	**Telephone transmission quality, telephone installations, local line networks**
Series Q	Switching and signalling
Series R	Telegraph transmission
Series S	Telegraph services terminal equipment
Series T	Terminals for telematic services
Series U	Telegraph switching
Series V	Data communication over the telephone network
Series X	Data networks and open system communications
Series Y	Global information infrastructure and Internet protocol aspects
Series Z	Languages and general software aspects for telecommunication systems

Geneva, 2001