Psychoacoustic Audio Quality Measurements Using R&S®UPV Audio Analyzer

Application Note

Products:

- R&S®UPV
- R&S®UPV66
- R&S®UPV-K61
- R&S®UPV-K62

This application note gives background information about the psychoacoustic measurements PESQ® (Perceptual Evaluation of Speech Quality) and PEAQ® (Perceptual Evaluation of Audio Quality). It also describes how to perform these measurements using the Rohde & Schwarz Audio Analyzer.

Application Note

Klaus Schiffner

04.2009-1GA49_0E

ROHDE&SCHWARZ

Table of Contents

1 Overview

For a long time, subjective listening tests have been the only method to evaluate the audio quality of perceptual coding processes. Since such tests are often expensive or impractical, objective measuring methods have been developed which are based on the psychoacoustic behavior of the human ear.

International standardization of perceptual audio measurement techniques was mainly driven by two expert groups within the International Telecommunication Union (ITU).

Within the ITU's Telecommunication Standardization Sector (ITU-T), **PESQ®** [1]) **(perceptual evaluation of speech quality)** was finalized in 2001 as **Recommendation P.862**. PESQ has meanwhile proven to be the most reliable algorithm for objective speech quality assessment, and has reached a strong market penetration.

The ITU's Radiocommunication Sector (ITU-R) established a task group to recommend an objective, perception-based model to evaluate the quality of wideband audio codecs. The model was recommended as a "method of objective measurements of perceived audio quality" under **ITU-R Rec. BS.1387** in late 1998. It is nowadays known as **PEAQ®** [2]) **(perceptual evaluation of audio quality)**.

Both standards, ITU-T P.862, and ITU-R BS.1387, today represent the state-of-the-art technique for the objective evaluation of the perceived audio quality of audio codecs.

This application note gives background information about the theory of PESQ and PEAQ and describes how to perform these measurements using the Rohde & Schwarz Audio Analyzer.

The test setup requires an R&S®UPV Audio Analyzer with PESQ measurement option R&S®UPV-K61 and/or PEAQ measurement option R&S®UPV-K62

[1]) *PESQ® is a registered trademark of OPTICOM Dipl.-Ing. M. Keyhl GmbH, Germany and of Psytechnics Ltd., UK*

[2]) *PEAQ® is a registered trademark of OPTICOM Dipl.-Ing. M. Keyhl GmbH, Germany*

2 Psychoacoustic Audio Quality Measurements

Compression has become the state-of-the-art technology in modern communications and audio applications such as mobile phones, VoIP, MP3 players, DVD, and many more. However, the economic benefit of lowering data rates further to a minimum is contradictory to clear sound. Sound quality and speech intelligibility have become issues again, especially with all the new digital sound technologies.

Data compression algorithms take advantage of the properties of the human auditory system by controlling the spectrotemporal distribution of resulting coding distortions so that they are below the threshold of hearing.
Traditional audio analysis tools such as THD+N or S/N measure the physical parameters of the audio signals and the overall level of distortion, including inaudible components. Therefore very often the results do not correspond to the perceived audio quality.

For a long time, subjective listening tests have been the only method to evaluate the audio quality of perceptual coding processes. Since such tests are often expensive or impractical, objective measuring methods have been developed which are based on the psychoacoustic behavior of the human ear.

2.1 Theory of Audio Quality Testing

2.1.1 ITU-T P.800 Listening Tests

Useful methods of testing telephone band speech signals were first standardized within the ITU-T. Recommendation P.800 defines the absolute category rating test method using a five-grade quality scale as shown in Table 1.
Testing is done without a comparison to an undistorted reference. This copes with the typical situation of a phone call, where the listener has no access to a comparison with the original voice of the other party. However, it should be noted that the listening tests in accordance with P.800 could be regarded as a comparison between a test signal and a reference “in the mind” of the listener. This is because of the fact that a listener is very familiar with the natural sound of a human voice.

Speech quality	Grade
Excellent	5
Good	4
Fair	3
Poor	2
Bad	1

Table 1: The ITU-T five-grade quality scale

For comparison reasons, and in order to be able to merge the results of different individuals, it is necessary to adjust the listeners' opinions to an absolute scale. For this purpose, predefined examples with well-defined noise insertions of fixed modulated noise reference units are presented at the beginning of a test. Each sample represents an example distortion corresponding to the ITU-T version of the five-grade quality scale. Based on these test conditions, a group of typically 20 to 50 test subjects will be presented with an identical series of speech fragments. Every test subject will be asked to rate each sample by applying the quality scale. After statistical processing of the individual results, a mean opinion score (MOS) can be calculated. Of course, the effort needed in terms of subjects and time is tremendous and therefore such test methods cannot be applied in the daily life.

2.1.2 ITU-R BS.1116 Listening Tests

The ITU has also recommended a test procedure to assess wideband audio codecs on the basis of subjective tests. It should be noted that subjective assessments of low-bit-rate audio codecs in the past always targeted at almost transparent quality. For this reason, the test method focuses on the comparison of the coded/decoded signal to the unprocessed original reference. The relevant recommendation is known as BS.1116. The used test method is referred to as "double-blind triple-stimulus with hidden reference". It is extremely sensitive and allows for the accurate detection of small impairments. The grading scale used should be treated as continuous with "anchors" derived from the ITU-R five-grade impairment scale.

The analysis of the results from a subjective listening test is in general based on the subjective difference grade (SDG), which is defined as follows:

$$SDG = Grade_{signal\ under\ test} - Grade_{reference\ signal}$$

The SDG values range from 0 to −4, where 0 corresponds to an imperceptible impairment and −4 to an impairment judged as very annoying as shown in the last column in Table 2.

Impairment	Grade	SDG
Imperceptible	5.0	0.0
Perceptible, but not annoying	4.0	−1.0
Slightly annoying	3.0	−2.0
Annoying	2.0	−3.0
Very annoying	1.0	−4.0

Table 2: The ITU-R five-grade impairment scale

In contrast to the listening test in accordance with ITU-T P.800, an explicit comparison between the test signal and a reference signal is needed in the case of BS.1116, since the listener never knows how the original signal sounds.

This method was applied in a variety of international verification tests in the past. However, it should be kept in mind that because of the scope of the recommendation it can be applied to small impairments only, which means a practical limitation to almost "transparent" studio quality. Another issue is the fact that it recommends using the scale at a resolution of one decimal place, resulting in 41 (!) discrete steps. There are indications that for some subjects this is too much of a choice, and in addition the meaning of the impairment anchors is interpreted differently.

2.1.3 Audio Measurements Employing "Perceptual Modeling"

For many years in the development of compression schemes, assessing quality was a pending issue. Consequently, the idea of replacing the subjective tests with objective, computer-based methods has been an ongoing focus of research and development, and meanwhile several methods have been introduced.
The underlying concepts of the proposed algorithms for perceptual techniques are all quite similar. The common structure of these algorithms is depicted in Fig. 1. The process of human perception is modeled by employing a measurement technique which compares a reference signal (i.e. the "input" signal to a codec) with a test signal (i.e. the "output" signal of the codec). First, the algorithms process an ear model for the reference and the test signal in order to calculate an estimate for the audible signal components. The result can be considered the internal representation inside the human auditory system. The comparison of the internal representations of the reference and the test signal leads to an estimate of the audible difference. In order to derive an overall quality figure, this information, which is a function of time, must be processed accordingly, as the brain of a human subject would do in a listening test. The respective part of processing within an algorithm is referred to as cognitive modeling. In the end, a total quality figure will be derived, which can be compared to a mean opinion score (MOS) resulting from a listening test.

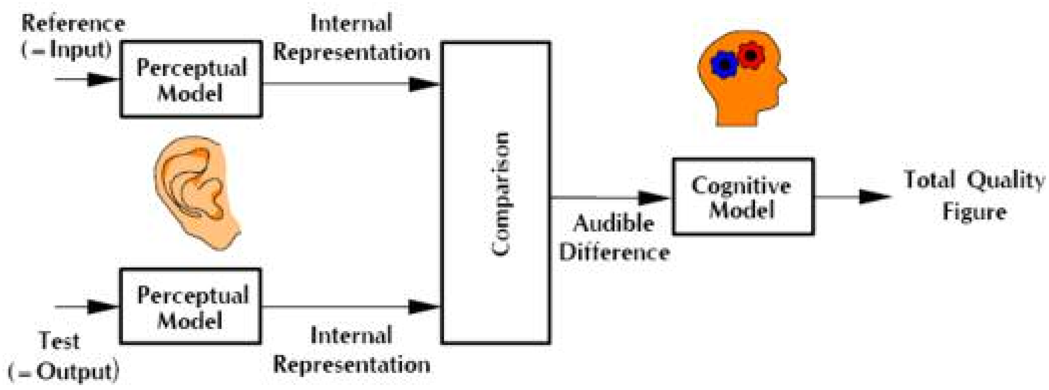

Figure 1: The underlying concept for perceptual measurement

The evaluation of the internal representation is often related to an estimate of the masked threshold. This estimate is based on data found in a number of psychoacoustic experiments. Most of these experiments model certain isolated effects of the human auditory system. One way to design a perceptual measurement algorithm is to generalize this model data and apply it to complex audio signals.

2.1.4 International Standardization

International standardization of perceptual audio measurement techniques was mainly driven by two expert groups within the International Telecommunication Union (ITU).

Within the ITU's Telecommunication Standardization Sector (ITU-T), study group 12 finalized in 1996 Recommendation P.861 for the objective analysis of speech codecs based on the PSQM (perceptual speech quality measure) algorithm. Since then, considerable progress has been made on an update to ITU-T P.861. Driven by the demand for a verified test procedure for VoIP, an expert group within ITU-T SG12 has been working on an improved speech quality model. This has resulted in a new model called **PESQ (perceptual evaluation of speech quality)**. In February 2001, PESQ was accepted as **ITU-T Rec. P.862**. PESQ has meanwhile proven to be the most reliable algorithm for objective speech quality assessment, and has reached a strong market penetration.

Within the study period from 1994 to 1998, the ITU's Radiocommunication Sector (ITU-R) established task group 10/4 with the authority to recommend an objective, perception-based model to evaluate the quality of wideband audio codecs. After collecting a set of proposals, including the most popular ideas such as NMR, PAQM, PERCEVAL, POM and others, the group of model proponents opted for a joint collaboration to derive an improved model. In 1998, two versions of this new model were presented: a "basic" version, featuring a low complexity approach, and an "advanced" version for higher accuracy but at the cost of higher complexity. After thorough verification, the model was recommended as a "method of objective measurements of perceived audio quality" under **ITU-R Rec. BS.1387** in late 1998. It is nowadays known as **PEAQ (perceptual evaluation of audio quality)**.

Both standards, ITU-T P.862, and ITU-R BS.1387, today represent the state-of-the-art technique for the objective evaluation of the perceived audio quality of audio codecs. It should be noted, however, that both techniques were derived from modeling the corresponding subjective experiment by an algorithm-based approach. It is therefore essential to understand the scope of the modeled subjective experiment when trying to interpret the calculated results. Fig. 2 summarizes the subjective test procedures and their corresponding objective counterpart in the context of typical data rate limits. The threshold between both worlds − broadcasting and telecommunications − is floating due to the ongoing attempt to further reduce the bit rates by means of more efficient coding schemes. Consequently, the overall data rate scale depicted in the figure should be taken as a coarse indicator only.

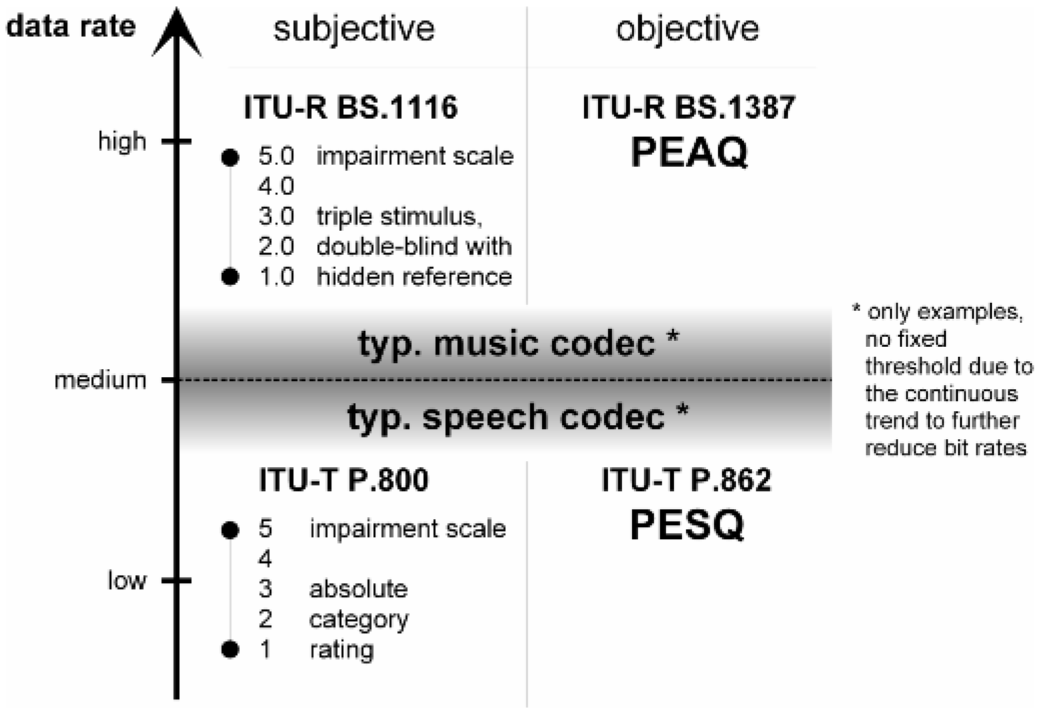

Figure 2: Overview on subjective and objective recommendations

In summary, one could say the following:

- ITU-T P.800 uses speech test signals only, and the untrained listeners are asked, "How good do these samples sound?"
- ITU-R BS.1116 uses music test signals plus reference signals, and trained listeners are asked, "How different do these samples sound?"
- Thus, PESQ and PEAQ are designed for different applications using different algorithms.

3 Perceptual Evaluation of Speech Quality (PESQ)

3.1 Standards

The basic standards for PESQ are the ITU-T Recommendations P.862, P.862.1 and P.862.2.

ITU-T P862.3 is an application guide which provides important information for obtaining stable, reliable and meaningful PESQ measurement results in practice. Studying this paper is strongly recommended (*http://www.itu.int/rec/T-REC-P/e*).

3.2 Source Material

It is important that test signals for use with PESQ are representative of the real signals carried by communications networks. Networks may treat speech and silence differently, and coding algorithms are often highly optimized for speech. Consequently, these algorithms may give meaningless results if they are tested with signals that do not contain the key temporal and spectral properties of speech.

- Reference signals must be in single-channel format (mono).
- The sample rate is defined to be 8 kHz or 16 kHz.
- There must be at least 0.5 seconds of silence at the beginning and at the end of the signal.
  Silence of up to 2 seconds is recommended for devices under test with large delay.
- In realtime mode of the R&S®UPV, the length of the test signal is identical to the length of the reference signal. The delay of the device under test must not be longer than the silence at the end of the reference signal, because otherwise the end of the speech signal is not measured.
- For offline measurements, the length of the recorded signal must not be shorter than that of the reference signal.
- Test signals should be representative of both male and female speakers.

Speech signals can be obtained from different databases and authorities. ITU-T Rec. P.501 amendment 1, for example, includes a CD-ROM with a series of recorded sequences in different languages.

3.2.1 Production of Source Material

If natural speech recordings are used, the guidelines given in ITU-T Rec. P.830 clause 7 should be followed; also P.501 amendment 1 gives some hints on how to make recordings.

Test signals should include speech bursts separated by silent periods, to be representative of natural pauses in speech. As a guide, 1 to 3 seconds is a typical duration of a speech burst, although this varies considerably between languages. Certain types of voice activity detectors are sensitive only to silent periods that are longer than 200 ms. Speech should be active for between 40 % and 80 % of the time, though again this is somewhat language-dependent. Recordings made for use with PESQ should have a duration of about 8 seconds; it is recommended that at least two male and two female speakers are used for each testing condition.

If a condition is to be tested over a long period, it is most appropriate to make a number of separate recordings of around 8 to 20 seconds of speech and to process each file separately with PESQ. This has additional benefits: If the same original recording is used in each case, time variations in the quality of the condition will be very apparent; alternatively, several different speakers and/or source recordings can be used, allowing more accurate measurement of speaker or material dependence on the condition.
Note that the nonlinear averaging process in PESQ means that the average score over a set of files will usually not equal the score of a single concatenated version of the same set of files.

3.2.2 Filtering and Level Calibration

Signals should be passed through a filter with appropriate frequency characteristics to simulate sending frequency characteristics of a telephone handset, and level-equalized in the same manner as real voices. ITU-T recommends the use of the modified intermediate reference system (IRS) sending frequency characteristic as defined in Annex D/P.830. Level alignment to an amplitude that is representative of real traffic should be performed in accordance with 7.2.2/P.830.

In some cases, the measurement system used (for example, a two-wire analog interface) may introduce significant level changes. These should be taken into account to ensure that the signal passed into the network is at a representative level.

The prepared source material after handset (send) filtering and level alignment is normally used as the original signal for PESQ.

3.2.3 How to Deal with Comfort Noise Insertion

Many modern telephony systems use noise substitution or comfort noise insertion. Especially the second case may lead to measurement results that are worse than expected when reference signals are used that contain digital zero during the silent intervals. PESQ will compare this digital silence to the comfort noise and detect the comfort noise as a distortion. Users will notice this if the PESQ score for speech is significantly better than the PESQ score for silence. P.862 does not explicitly mention this problem, but it recommends that the reference material be prepared in accordance with P.830. P.830 in turn recommends that Gaussian noise equivalent to −68 dBmp be inserted before material is presented to the test subjects. If such a noise is inserted in the silent intervals of the reference signal, this will effectively solve the problem described here.

3.3 Fundamentals of the PESQ Measurement Algorithm

3.3.1 PESQ Measurement

When PSQM was standardized as ITU-T P.861, the scope of the standard at that time were state-of-the-art codecs as they were mainly used for mobile transmission, such as GSM. VoIP was not yet a topic. The requirements for measurement equipment have changed dramatically since then. As a consequence, the ITU set up a working group to revise the P.861 standard in order to cope with the new demands arising from modern networks such as VoIP. With these networks, the measurement algorithm has to deal with much higher distortions than with GSM codecs, but maybe the most eminent factor is that the delay between the reference and the test signal is not constant anymore.

With the new ITU standard P.862 (PESQ), this problem is eliminated. PESQ combines the psychoacoustic and cognitive model of PSQM with a time alignment algorithm that perfectly handles varying delays. Fig. 3 gives an overview of the structure of the PESQ algorithm.

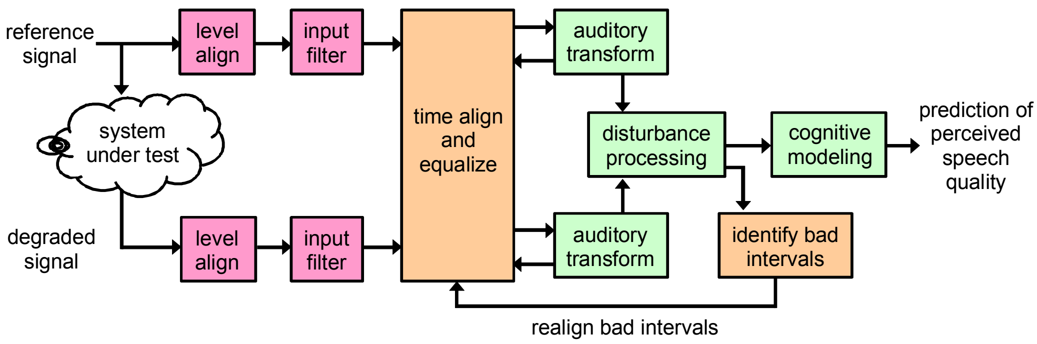

Figure 3: Structure of PESQ algorithm

3.3.2 Wideband PESQ

PESQ also offers a special operating mode which allows the assessment of wideband speech signals. This wideband extension of PESQ uses a flat input filter and a different mapping function to the MOS scale than used with narrowband speech. It has been standardized by the ITU as ITU-T Recommendation P.862.2. This recommendation defines that for wideband speech, the PESQ score must not be used any longer. Instead, one has to refer to the PESQ score's mapped counterpart.

3.4 PESQ Measurements Using the R&S®UPV Audio Analyzer

This measurement function is only available if the R&S®UPV-K61 option is installed. It is only provided in the analog analyzer in single-channel mode.

The "perceptual evaluation of speech quality" (PESQ) measurement method, which was published by the International Telecommunication Union in 2001 as Recommendation ITU-T P.862, enables measurements to be made on speech signals that are transmitted at low bit rates using highly compressive psychoacoustic coding methods. PESQ employs an algorithm that enables these signals to be evaluated by comparing them with reference signals. The R&S®UPV provides this measuring method, which is licensed by OPTICOM Dipl.-Ing. M. Keyhl GmbH of Erlangen, Germany. PESQ® is a registered trademark of OPTICOM and of Psytechnics Ltd., UK.

A common feature of all psychoacoustic coding methods is that they utilize the properties of human hearing to modify the transmitted signal so that the portions of the signal that would in any case not be perceived are removed from the signal. Compression can be made even simpler with speech signals, since they occupy considerably less bandwidth than other types of signals such as music. When speech compression is used, it must be possible to determine objectively, with the aid of psychoacoustic measuring methods, whether such speech transmission paths produce unacceptable changes to the signal.

PESQ was developed using a large number of recordings containing sentences spoken by a variety of speakers in a variety of languages. The recordings were made using several different speech encoders having different levels of quality and with typical network transmission disturbances. In a series of listening tests, an adequate number of test listeners classified these examples on the well-known speech quality scale ranging from 1 (bad) to 5 (excellent).

The goal in the development of PESQ was to develop a method for determining an objective measurement value (mean opinion score, MOS) that correlates very well with the listening test results, based on comparing the original, undegraded speech signal (the reference signal) with the degraded signal (the measured signal). This means that in order to perform a PESQ measurement, the reference signal must be connected to the input of the system under test and the measurement signal must be taken from the output of the system under test.

3.4.1 Electrical PESQ Measurements

Producers of new psychoacoustic coding methods wish to optimize their algorithms for speech transmission, and network operators need suitable test tools that supplement the already commonly used signal and distortion measurements. This type of measurement is typically performed electrically, directly at the input and output ports of the R&S®UPV.

Fig. 4 shows how the reference signal is routed directly from the generator output to the device under test and how the output of the device under test is fed back to the analyzer as the measurement signal. At the same time, the original reference signal is routed internally to the analyzer, allowing it to access both signals for the evaluation.

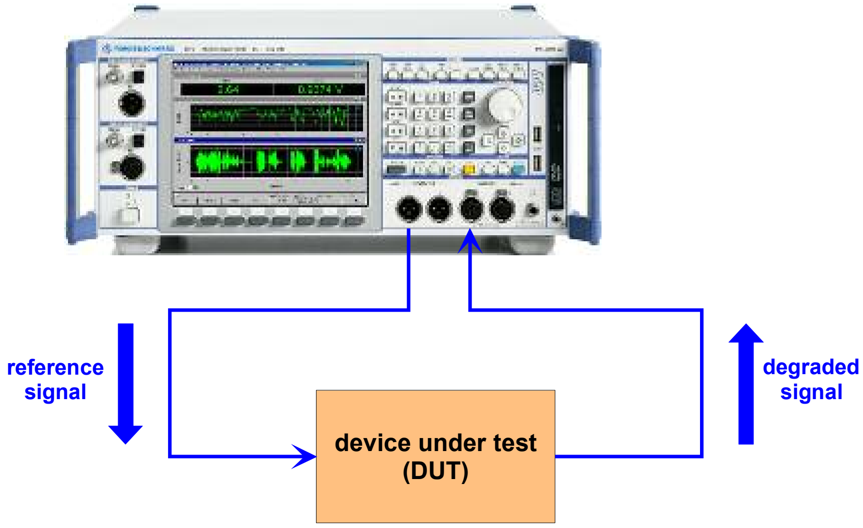

Figure 4: Principal of electrical PESQ measurement

3.4.2 Acoustic PESQ Measurements

Manufacturers of mobile phones increasingly wish to use speech quality to make their products stand out from the competition. Speech quality is essentially influenced by the coding method used in the mobile phone. In this case, PESQ measurements are made under realistic conditions by using an acoustic link to the mobile phone. The measurement setup consists of a combination of the R&S®CMU radio communication tester and the R&S®UPV audio analyzer.

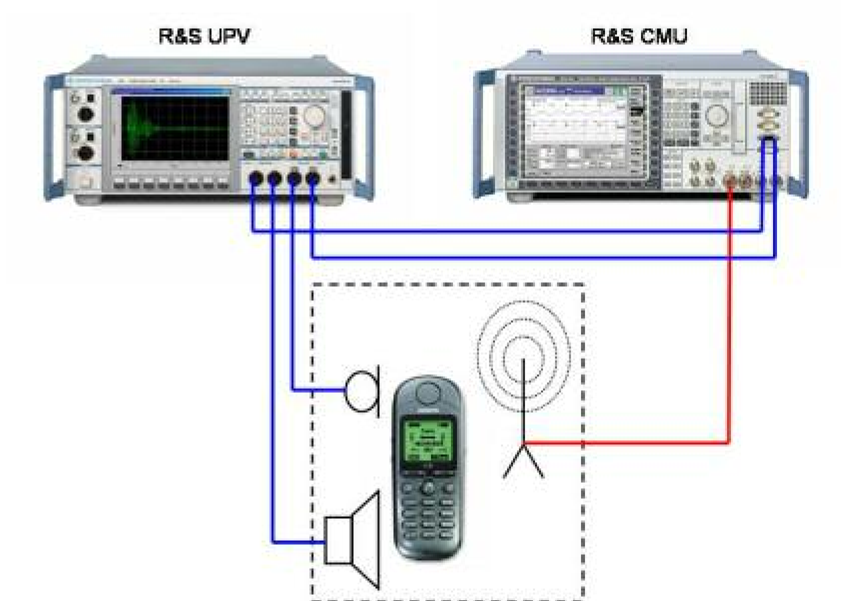

Figure 5: Principal of acoustic PESQ measurement

The reference signal is sent to the mobile phone via an acoustic transducer (artificial mouth). It is coded in the mobile phone and then sent to the R&S®CMU via a radio link. It is decoded in the R&S®CMU and sent to the input of the analyzer as the measurement signal. As with electrical measurement, the original reference signal is also routed internally to the analyzer so that it can access both signals for the evaluation. In the opposite direction, the reference signal is sent to the R&S®CMU, where it is coded and transmitted to the mobile phone via the radio link. Another acoustic transducer (artificial ear) converts the speech signal emerging from the speaker of the mobile phone into the measurement signal.

Good acoustic coupling between mobile phone and artificial ear is needed for measurements in the receiving direction of the phone. Any background noise should be as low as possible. Acoustic measurements should therefore be conducted in a sound-isolated chamber. In order to provide the correct spectrum of the reference signal to the microphone of the device under test, a filter equalizing the frequency response of the artificial mouth should be set in the ***Generator Function*** panel of the R&S®UPV.

PESQ measurements should not be used for handsfree measurements or to assess the influence of background noise.

The advantages of the R&S®UPV can be fully exploited with acoustic measurements. Due to the outstanding characteristics of the analog hardware, no PESQ-relevant quality impairments arise in the measured signal even when there are large differences in signal level between the output and input signals. In addition, a simple application running in the R&S®UPV can utilize the dual-channel capability of the R&S®UPV to make measurements in both directions (transmit and receive) without any changes to the test setup. The results can then be displayed in a single graph.

3.4.3 Offline Measurements

If inputs and outputs of the device under test are not accessible at the same time, offline measurements are the solution. In this case, the output signal of the DUT is recorded in WAV file format. This can be done either by the R&S®UPV audio analyzer (function ***Record***) or by any other recording device. However, it is recommended to use recording devices with a sample frequency accuracy of at least 50 ppm because the PESQ algorithm is not optimized to compensate slowly diverging pitches. Possible sample rates of the stored degraded WAV file are 8 kHz, 16 kHz, 48 kHz or 96 kHz. The length of the recorded signal must be at least the length of the generator reference signal. However, it should not differ by more than 20 %, otherwise the level align process of the PESQ algorithm might become sub-optimal.
To run the PESQ algorithm, the WAV file containing the test signal is played in the analyzer section internally while the reference file is played by the generator section, allowing both signals to be accessed by the analysis.

Typical applications for offline operation are tests of different types of coding algorithms where input and output signals are available in files only.

3.5 Results Obtained from PESQ Measurement Function

3.5.1 PESQ Value and MOS Value

The most important result of PESQ is the PESQ score. It directly expresses the voice quality on a MOS-like scale. The PESQ score as defined by ITU Recommendation P.862 ranges from −0.5 (worst) up to 4.5 (best).

MOS values are nowadays used with different extensions. MOS-LQS (listening quality subjective) represents results from listening tests. PESQ measurement results are shown in MOS-LQO (listening quality objective). The additional suffixes N and W stand for narrowband and wideband.

The PESQ score can be mapped to the ITU P.800 scale by applying a simple mapping function as standardized by the ITU in P.862.1. A graphical representation of mapping from PESQ score to MOS-LQO is shown in Fig. 6.

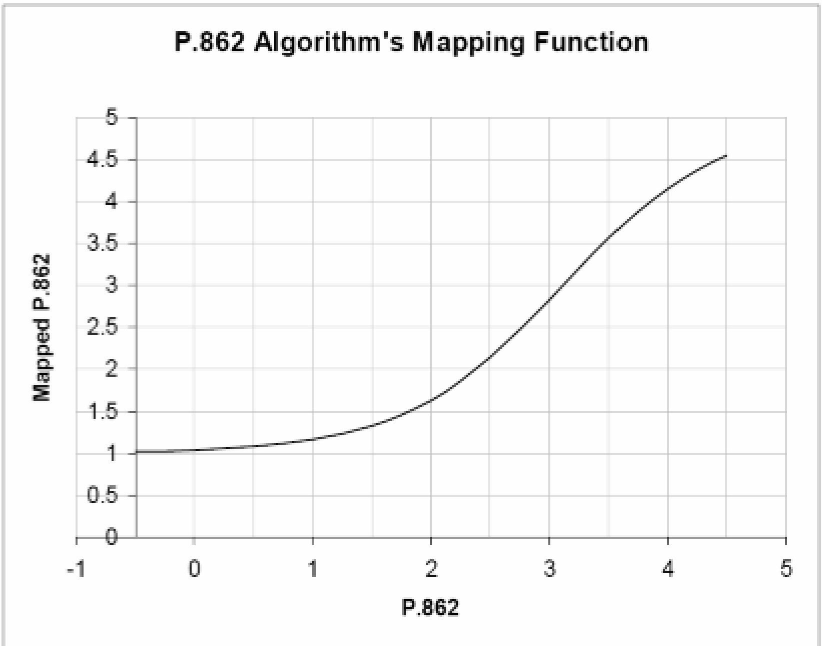

Figure 6: Mapping from PESQ score to MOS-LQO

The average PESQ score values from the listening tests are plotted on the Y-axis, and the associated MOS-LQO values in accordance with ITU P.862.1 are shown on the X-axis.

Over the course of time, the ITU has developed several different methods for calculating objective measurements from the average values of the listening test results.

The R&S®UPV implements the three most important measuring methods, which differ only slightly from each other and have been approved by the ITU:

- ITU P.862: The measured value is the "PESQ score". The range of values extends from −0.5 (worst) to +4.5 (best).
- ITU P.862.1: The measured value is the "MOS-LQON" (listening quality objective narrowband). The range of values extends from −0.5 (worst) to +4.5 (best).
- ITU P.862.2: This is the wideband extension of P.862. The measured value is the "MOS-LQOW" (listening quality objective wideband). The range of values extends from −0.5 (worst) to +4.5 (best). Note that measurements obtained using this option cannot be compared with results obtained in accordance with P.862 or P.862.1.

In addition to the overall results, the measurements can be made by calculating the PESQ values for the active speech intervals or for the silence intervals of the signal. The latter is of particular interest because it shows how well the codec may substitute background noise.

Note that for wideband PESQ the PESQ score shall not be used, but only the mapped MOS-LQO. In order to use common terms in both narrow- and wideband mode, it is strongly recommended to always use the MOS-LQO, which is mapped either by Recommendation P.862.1 for narrowband or by P.862.2 for wideband speech.

The R&S®UPV audio analyzer uses the short forms "PESQ" for PESQ score and "MOS" for MOS-LQO.

3.5.2 Amplitude of Test Signal

Speech signals show a large variation in amplitude over time. The R&S®UPV analyzes the test signals in small portions of time. After the analysis routine is finished, one of the two following amplitude values can be displayed by the ***Level Monitor***:

- RMS value integrated over the whole test signal, or
- maximum peak value within the time interval of the test signal

Figure 7: Numeric display of PESQ results

3.5.3 Attenuation

Especially all analog equipment introduces an attenuation into the speech signal. A high attenuation generally leads to a worse perception of speech. PESQ, however, does not weight this as a degradation, since it has no absolute reference level available. Also, in real world systems a low speech level on the electrical side does not mean that the signal sounds very quiet, since the transducers used have a significant impact on the final loudness. For PESQ it is therefore generally impossible to weight the attenuation in terms of quality of speech. Knowing the value of the attenuation, however, is important for optimizing the overall system design.

3.5.4 Delay/Latency

As soon as a signal is processed by any piece of equipment, it will be slightly delayed. The resulting delay is also frequently called latency. During the transmission of a speech signal these delays may add up and become intolerable. The longer the delay is, the more discipline is required from both parties involved in a conversation. Delays larger than approx. 300 ms are generally unacceptable. The delay for VoIP systems is typically around 150 ms, sometimes even much longer.

Since the PESQ library has to compensate for the delay between the signals before it can compare them, it provides the average delay as a result value, as well as the delay of each individual frame.

The PESQ tool automatically compensates for delays between the reference signal and the measured signal in the range of up to 5 seconds. If measurements have to be done with delay times longer than 0.5 s, the used reference file played by the R&S®UPV audio analyzer needs to have a longer silence part at the end. While executing a measurement, R&S®UPV records the test signal as long as the reference file runs. If there is a delay in the device under test, the silence part at the end of the reference signal must be longer than this delay to make sure that the speech portion is completely recorded.

The reference and the measurement signals are each internally divided into small portions of time. The time offset between the two signal paths is measured for each of these increments; it can be displayed in the ***PESQ Graph*** displays as a graphical plot of time delay versus time (see "PESQ Graphical Displays" section for more details). The line ***Avg Delay*** in the ***Analyzer Function*** panel displays a value which is the average of all of the measured offset values for the entire signal.

Figure 8: Average delay

3.5.5 Waveform Display

With PESQ measurement, the ***Waveform*** display is used to check the signal in the time domain. In particular, the specified silence periods at the start and end of the test signal can be checked, as well as correct signal amplitude.
The ***Waveform*** is activated in the ***Analyzer Function*** panel and configured with regard to the trigger condition and storage depth. The ***Waveform*** graphics dialog can also be opened from here. The scaling – including the limit check and reference basis – is selected in the ***Waveform Config*** panel. Fig. 9 shows an example.

When a DUT is measured online, the test signal from the DUT has to be recorded before it is fed to the internal PESQ algorithm. The ***Waveform*** analyzer gets the data in parallel and displays the result as soon as the recording is finished. In offline mode, the reference signal and the test signal can be processed by the PESQ algorithm directly from the files without delay. If the ***Waveform*** function is activated, this function runs first and therefore delays the start of the internal PESQ algorithm.

3.5.6 PESQ Graphical Displays

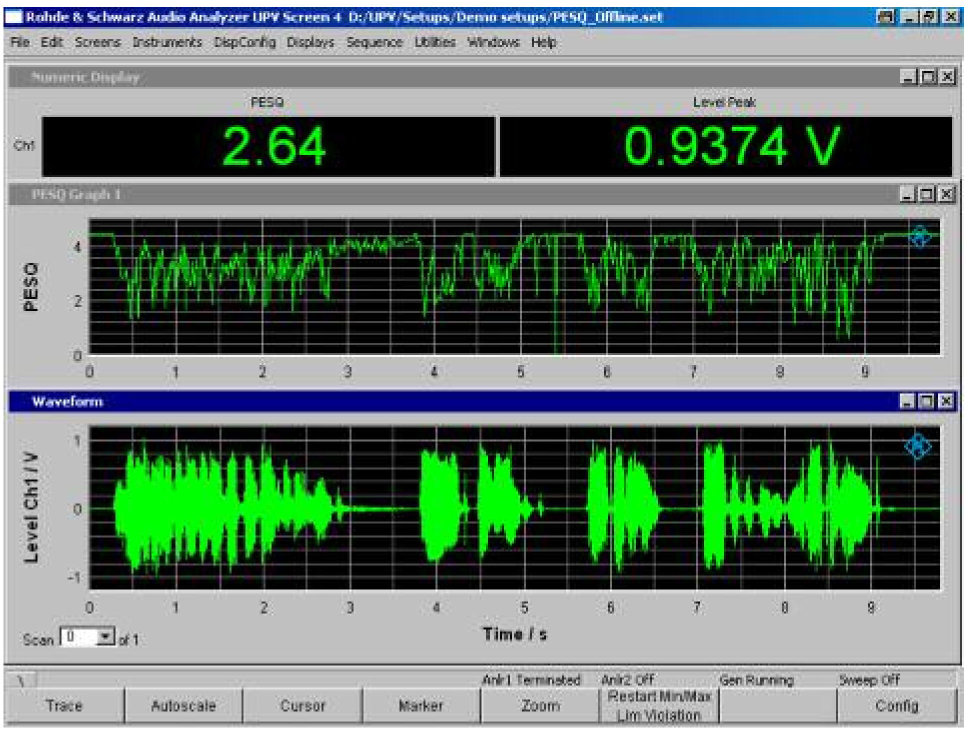

Figure 9: Graphical displays show PESQ values and/or the waveform versus time

With the ***PESQ Graph*** windows, the R&S®UPV provides two additional types of graphs associated with PESQ measurements. The measurement process analyzes the signals in small increments and determines the measured values (PESQ score or MOS-LQO) and the time delay between the reference signal and the measured signal for each increment. While the numeric displays show the selected measurement values referenced to the entire signal, the graphical displays show the results versus the time increments. Using these graphs, one can see the regions of the signal where the measured codec deviates from the expected behavior. Fig. 9 shows a combination of numeric values referenced to the entire signal and results versus time.

Note: In wideband mode (P862.2), the MOS value versus time cannot be displayed with the current version of the PESQ software.

In addition, the traces of reference and/or measurement signal, as well as display of the dropouts, can be shown in the ***PESQ Graph*** windows. Depending on the signal selected on the other axis, the resolution on the X-axis is selected automatically so that both signals can always be displayed in the same graphic. For example, the resolution of "MOS versus time" is much lower than "reference signal versus time", so the graph shows the reference signal in reduced resolution – but without swallowing samples. In contrast, in the display of "delay versus time" vis-à-vis "reference signal versus time" there is a much higher resolution on the X-axis; it corresponds to the resolution of the WAV file.

The ***PESQ Graph*** is configured in the ***PESQ Graph Config*** panel. In line ***Y-Source*** the type of results to be displayed graphically can be selected:

- ***PESQ & MOS***: The PESQ or the MOS value is displayed versus time
- ***Delay***: The delay versus time shows the time offset between the signal paths of reference signal and measured signal for each of the measured time increments
- ***Ref Signal***: The reference signal versus time corresponds to the waveform of the reference signal
- ***Deg Signal***: The measurement signal versus time corresponds to the waveform of the measurement signal. This waveform can also be shown in full resolution in a separate graphic via the "Waveform Monitor" function.
- ***Dropouts***: Dropouts versus time corresponds to a hypothetical representation in which the values of an amplitude- and time-corrected measurement signal and the reference signal are subtracted from each other in proper phase relation and thus only the stray signals deviate from the zero line.

Figure 10: Different results of the PESQ measurement can be displayed versus time

3.6 Introduction to Instrument Operation

3.6.1 Setting the Generator Instrument

Whenever PESQ analysis is performed, the generator part of the R&S®UPV is needed to replay the reference signals. In accordance with ITU-T P.862, reference files with sample frequencies of 8 kHz or 16 kHz have to be used. For wideband applications (ITU-T P.862.2) sample frequency always has to be 16 kHz.
WAV files can be replayed by setting the line ***Function*** to ***Play*** in the ***Generator Function*** panel. Using the analog interfaces, ***Bandwidth*** has to be set to ***Play Auto*** in the ***Generator Config*** panel.

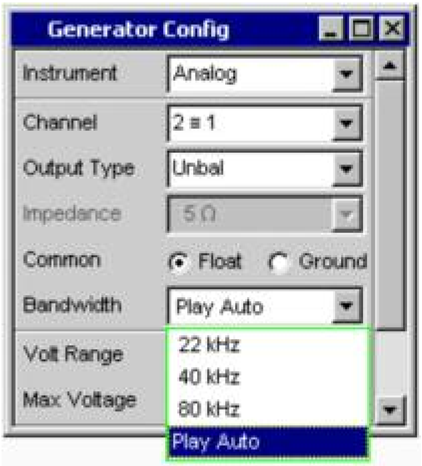

Figure 11: Setting the Generator Config panel

3.6.2 Setting the Analyzer Instruments

To run PESQ, the analog interface has to be selected by setting the ***Instrument*** in the ***Analyzer Config*** panel. The analyzer needs to be set in single-channel mode; ***Range*** has to be set to ***Fix***.

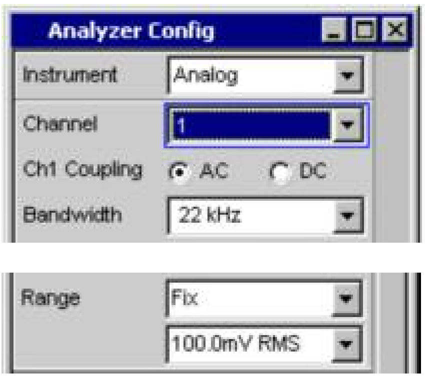

Figure 12: Setting the Analyzer Config panel

To activate the PESQ measurement function, in the ***Analyzer Function*** panel the line ***Function*** has to be set to ***PESQ***.

The setting in the line ***According to*** specifies whether the PESQ value in accordance with ITU-T P.862 or the MOS value in accordance with ITU-T P.862.1 (narrowband) or ITU-T P.862.2 (wideband) will be displayed. It also specifies whether the measured values should be referenced to the entire signal or only to the speech portion or only to the silence portion.

Changing the setting of ***According to*** does not trigger a new measurement but calculates the desired PESQ or MOS value from the last logged record in accordance with the selected process. The respective measurement result is thus immediately available.

The line ***Meas Mode*** defines whether the measurement is to be performed in realtime (***DUT***) or with an offline analysis using previously stored WAV files (***Offline***).

In ***DUT*** mode, the reference WAV file will be played back, and the measurement signal will be recorded by the analyzer at the same time. The measurement signal has the same length as the reference signal. If there is a delay in the device under test, the silence at the end of the reference signal must be longer than this delay to make sure that the speech portion is completely recorded. The evaluation is performed immediately after playback of the reference signal is finished. The recorded WAV file is always stored in the same sample rate in which the reference signal is present.

In ***Offline*** mode, the PESQ measurement is to be performed on previously stored WAV files. After the files containing the reference signal and the measurement signal are specified, the evaluation can be started by pressing the START button.
Possible sample rates of the stored WAV files are 8 kHz, 16 kHz, 48 kHz or 96 kHz.
The length of the recorded signal must be at least the length of the generator reference signal.

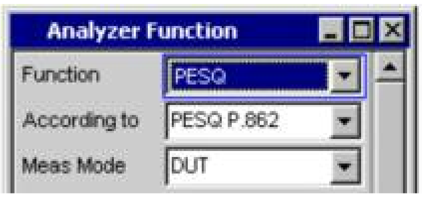

Figure 13: Setting the measurement mode

3.6.3 Selecting Levels and Measurement Ranges

To measure PESQ, correct levels and measurement ranges must be set.

The reference file, which is stored in a WAV file in digital format, is output by the R&S®UPV generator. The output level can be set in the line ***Volt Peak***; for example, to set it to ***0.1 V*** would mean that samples stored in the file with full-scale amplitude will be output with 0.1 V peak level at the analog outputs.
Quite often, stored test signals do not reach full-scale amplitude, resulting in lower output levels at the generator, and therefore may have amplitudes that are too low to drive the DUT properly.

To overcome this situation, R&S®UPV offers the possibility to rescale the samples in order to reach digital full-scale levels when replaying the files (line ***Scale Pk to FS*** in ***Generator Function*** panel).
For PESQ measurements, it is advisable to use this function to make sure that the set output level ***Volt Peak*** really is reached.

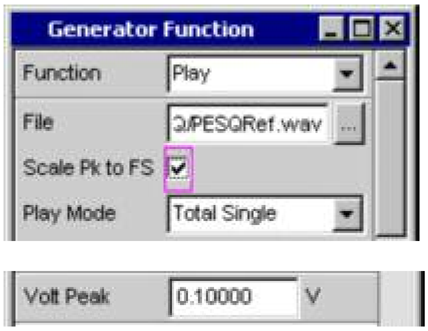

Figure 14: Setting the output level

Because speech signals show a large variety in amplitude over time, an autorange function in the analyzer would switch between ranges quite often. To prevent interruption of the PESQ measurement, the analyzer must therefore be set to a fixed measurement range. Setting the range according to the set output level in the ***Generator Function*** panel would be the first approach. But if the device under test introduces some gain or attenuation, this must be compensated. To find the best measurement range setting, the ***Level Monitor*** function can help. If the reference signal has been played to the DUT and fed into the analyzer from the device's output, the ***Level Peak*** indication will display the maximum peak level which has been output by the DUT. The measurement range should be set as close as possible to a value of measured peak value divided by √2, but must be bigger than this value.

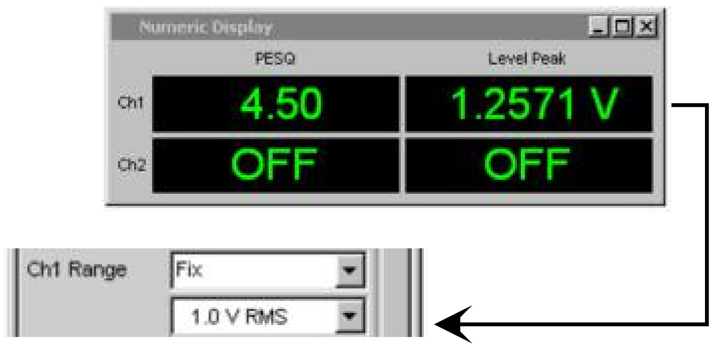

Figure 15: Correct setting of measurement range

The internal algorithm of PESQ is done in digital domain. The analysis requires that the reference signal and test signal have the same levels. The used PESQ implementation automatically corrects static level differences of up to 30 dB; however, to get reliable results, level differences should be kept below 10 dB.

In the ***Analyzer Function*** panel, the lines ***Ref Level*** and ***Deg Level*** display, respectively, the RMS values of the reference signal and the test signal at the input of the internal PESQ routine (i.e. after the measurement input stages but before the PESQ internal level correction). The values are shown in dB referenced to the theoretical full-scale values. The measurement setup is correct as long as the difference between ***Ref Level*** and ***Deg Level*** does not exceed 10 dB (5 dB is recommended). The screenshot below shows an example of a correct setting.

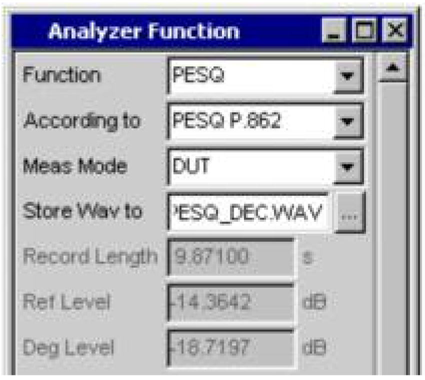

Figure 16: Example of Ref Level and Deg Level results

3.6.4 Saving of measured signals

If the PESQ measurement is performed in realtime (***Meas Mode*** set to ***DUT***), the measured signal can be stored as a WAV file. This done by activating the ●●● button in the line ***Store Wav to*** (see Fig. 16) after the measurement is terminated. The ***File Selector*** opens to define file name and directory.
The WAV file is always stored in the same sample rate in which the reference signal is played.

4 Perceptual Evaluation of Audio Quality (PEAQ)

4.1 Source Material

The reference file should be a signal that comes as close as possible to the kind of signal that will be applied to the device under test in real life. For example, real music should be used for the assessing audio codecs that are used for the transmission of high quality music. Especially with wideband music codecs, a variety of at least six to ten different test samples should be selected, since the performance of audio codecs differs widely depending on the test material.

The duration of the test sequence should be within the range of approximately four to eight seconds. Longer tests will lead to averaging effects (short distortions may be averaged down by a long but almost perfect transmission), and shorter sequences may not be long enough to contain representative parts of the signal. The sample rate of the reference and test files is defined by ITU R BS.1387 to be 48 kHz sample rate. The sample format is defined to be PCM 8 or 16 bit linear with mono or stereo content. Note that no coded audio format is supported, as decoding is not part of the measurement process. A set of typical wideband audio examples is mentioned in ITU-R Rec. BS.1387.

The signal should begin and end with at least 0.5 seconds of silence.
In realtime mode of the R&S®UPV, the length of the test signal is identical to the length of the reference signal. The delay of the device under test must not be longer than the silence at the end of the reference signal, because otherwise the end of the music signal is not measured.
For offline measurements, the sample rates of both the reference file and the stored degraded file must be 48 kHz. The length of the recorded signal must not be shorter than that of the reference signal selected in the generator.

Scientific proposals as well as international standards such as ITU-R BS.1387 (PEAQ) usually describe a measurement algorithm and do not take into account all the constraints of a realistic measurement situation. For example, in ITU-R BS.1387 it is assumed that the input signals are time aligned. Signals that have been acquired in real systems might not be time aligned. The PEAQ tool used in the R&S®UPV audio analyzer offers a sophisticated delay compensation. As a result, time alignment of the signals will be provided, together with values that are suitable to characterize the delay which was introduced by the device under test. The automatic delay compensation module makes it possible to automatically compensate for delays up to ±1000 ms. However, the delay compensation cannot handle varying delays, as they might occur due to missing synchronization of the input signals, or in IP-based systems.

4.2 Fundamentals of the PEAQ Measurement Algorithm

As mentioned earlier, there are two versions of PEAQ: a "basic" version, featuring a low complexity approach resulting in fast measurements, and an "advanced" version for highest possible accuracy but at the cost of higher complexity. The following paragraphs provide a brief overview of the fundamental principles involved.

The structure of both versions is very similar. The major difference between the "basic" and the "advanced" version is hidden in the respective ear models and the set of model output variables (MOV) used. Both versions comprise artificial neural networks for the cognitive modeling.

The "basic" version implements an FFT-based ear model featuring fundamental psychoacoustic principles. Following the signal flow from the input signal to the final calculation of the excitation pattern, the processing starts by a transformation of the input signal to the frequency domain. A 2048 point FFT is applied along with subsequent scaling of the spectra. This results in a frequency resolution of approximately 23.4 Hz and a corresponding temporal resolution of 23.4 ms (at 48 kHz sample rate). In the consecutive block, the effects of the outer and middle ear are modeled by weighting the spectrum with the appropriate filter functions. Afterwards the spectra are grouped into critical bands, achieving a resolution of 1/4 bark per band. The subsequent adding of "internal noise" is intended to model effects such as the permanent masking of sounds in our auditory system caused by the streaming of blood and other physiological phenomena. This step is followed by the calculation of masking effects. Simultaneous masking is modeled by a frequency- and level-dependent spreading function. Temporal masking is modeled only partly, since the temporal resolution is in the same range as the timing of any backward masking effects, which therefore cannot be modeled.
Using the feature extractor, eleven MOVs are extracted from the ear model output and processed by the cognitive model to form the PEAQ values.

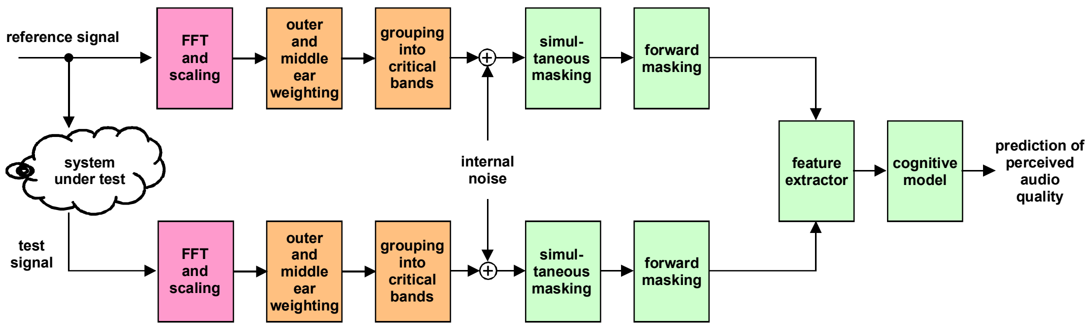

Figure 17: Perceptual model, PEAQ "basic" version

The "advanced" version uses some MOVs derived by implementing the ear model of the "basic" version and also introduces a second ear model with improved temporal resolution.
Compared to the "basic" version, the "advanced" version model performs the time to frequency warping using a filter bank, consequently grouping the signal into 40 auditory bands with a temporal resolution of approximately 0.66 ms. This allows for a very accurate modeling of backward masking effects. After the calculation of backward and simultaneous masking, the signal is subsampled by a factor of 1:6 in order to improve the computational efficiency. Following the addition of the internal noise to the subsampled signal, finally the forward masking effects are calculated. A feature extractor is used to extract the five model output variables of this second ear model. Combined with the MOVs from the "basic" model, the cognitive model forms the PEAQ values.

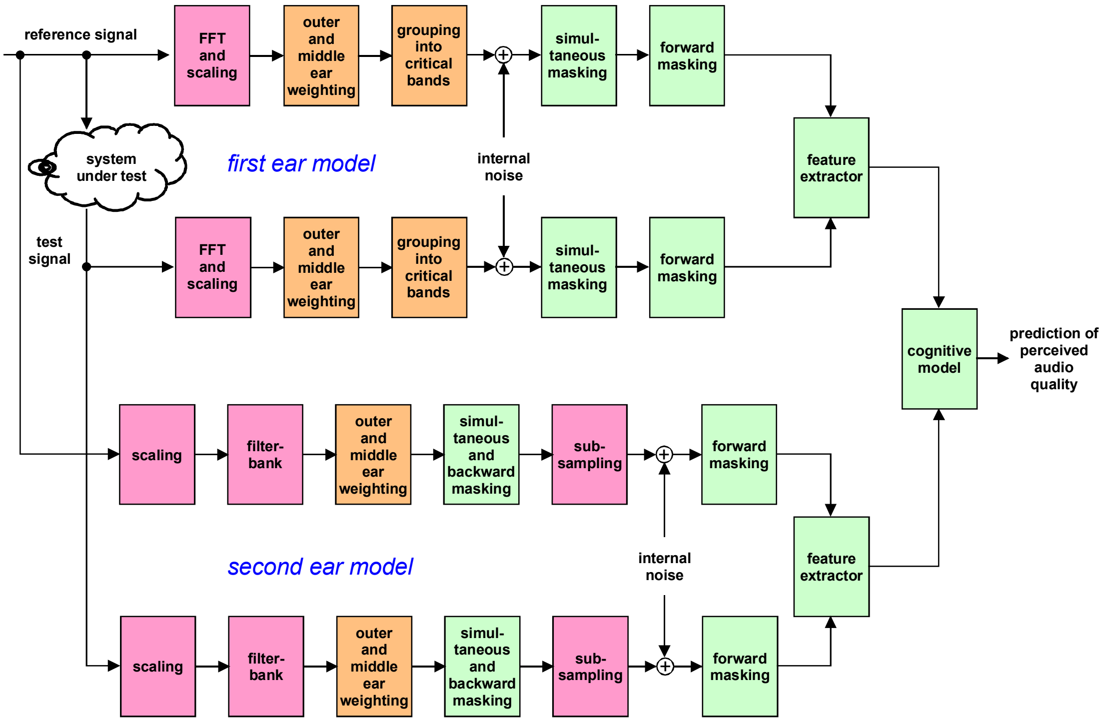

Figure 18: Perceptual model, PEAQ "advanced" version

In comparison to the FFT-based "basic" approach, the temporal resolution is improved, thus allowing for a better simulation of temporal effects, at the cost of frequency resolution and computational complexity.

It should be remembered that PEAQ is an approach to measure stereo music signals, meaning all signal paths in the above figure are dual-channel.

4.3 PEAQ Measurements Using the R&S®UPV Audio Analyzer

The measurement function PEAQ is only available if the R&S®UPV-K62 option is installed.
Reference files can be played from any of the available generator instruments. The measurement function PEAQ is performed in any of the available dual-channel analyzer instruments, but it is not provided in the multichannel analyzers.
That means PEAQ can be measured using analog interfaces as well as digital formats such as AES/EBU or I²S or even the programmable formats as offered by the R&S®UPV-B42 universal serial interface.

The PEAQ measurement process, which was published by the International Telecommunication Union in 1998 as Recommendation ITU-R BS.1387, permits the measurement of the quality-of-sound signals transmitted at a low bit rate by compressing psychoacoustic coding processes. The PEAQ algorithm allows these signals to be analyzed by comparing them with the reference signal. The R&S®UPV provides this measurement, which is licensed by OPTICOM Dipl.-Ing. M. Keyhl GmbH of Erlangen, Germany. PEAQ® is a registered trademark of OPTICOM.

All psychoacoustic coding processes have in common that they use the characteristics of human hearing to save those parts of the signal which are not discernable during transmission. Whether such a transmission link for general sound signals such as music causes impermissible changes to the signal must be objectively verifiable with the assistance of suitable psychoacoustic measurement procedures.

A large number of recordings containing music signals of different kinds were used for developing PEAQ. The recordings were overlaid with interference typically caused by filters and also by using various coders and thus qualities. A sufficiently large number of test listeners classified these examples in a series of listening tests on the familiar scale for audio quality from 1 (bad) to 5 (excellent). In contrast to the measurement of speech quality also available in the R&S®UPV (PESQ), with PEAQ the listeners have always had the comparison with the original signal. A "negative scale", referred to as the subjective difference grade (SDG) value, is obtained by determining the difference between the quality of the test signal and the quality of the reference signal. The worse the test signal becomes compared with the reference signal, the more negative the SDG value.

$$SDG = Grade_{Signal\ Under\ Test} - Grade_{Reference\ Signal}$$

The reason for developing PEAQ was to develop a process that compares the original, undistorted music signal (reference signal) with the distorted signal (measurement signal) to establish an objective measurement value that correlates very well with the mean value of the listening results. The PEAQ measurement can therefore only be performed in such a way that the system to be tested is connected at its input to the reference signal and the measurement signal is measured at its output.

Though the results of the subjective and the objective assessment correlate very well, it is not possible to directly compare the numeric results of SDG and ODG.

Originally the PEAQ measurement process was designed for signals with small impairments (almost transparent studio quality) and a constant delay. But it is also suitable for the assessment of signals with larger impairments.

It cannot be used, for example, for analyzing loudspeakers, cables, speech codecs, and IP-based interference (variable delay and packet losses).

It was not yet validated for codecs with extremely low bit rates, e.g. those ones using the method of spectral band replication.

4.3.1 Online Measurements (*Meas Mode DUT*)

PEAQ analysis can be done by measuring the performance of the device under test directly. The illustration shows how the reference signal is guided to the DUT from the generator output and fed into the analyzer again from the device's output as a measurement signal. The original reference signal is simultaneously fed internally to the analyzer so that the analysis can access both signals.

Measurement on encoder circuitries are typical applications to be analyzed online. Depending on the DUT, a generator and an analyzer might be used in different domains. For example, the evaluation of D/A converter equipment requires the digital generator combined with the analog analyzer of the R&S®UPV.

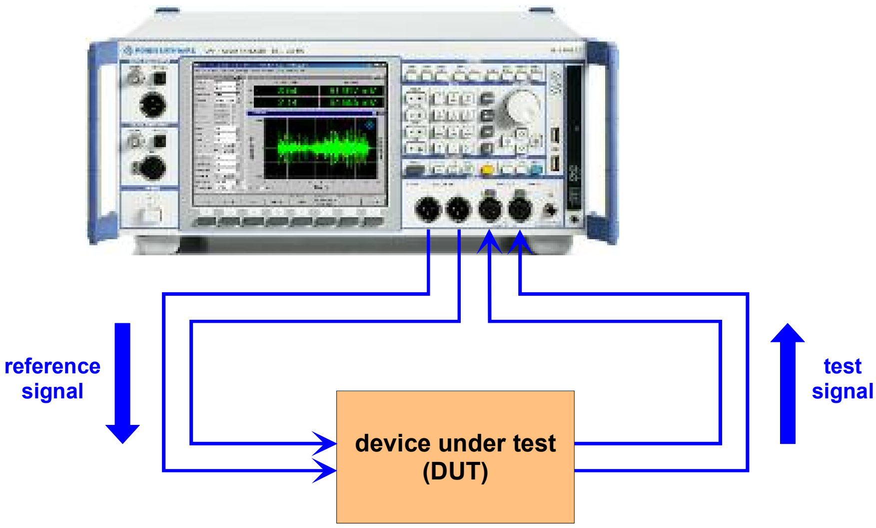

Figure 19: Principal of PEAQ measurement

4.3.2 Offline Measurements (*Meas Mode Offline*)

If inputs and outputs of the device under test are not accessible at the same time, offline measurements are the solution. In this case, the output signal of the DUT is recorded in WAV file format. This can be done either by using the R&S®UPV audio analyzer (***Record*** function) or by any other recording device using 48 kHz sampling frequency. However, it is advisable to use recording devices with a sample frequency accuracy of at least 50 ppm because the PEAQ algorithm is not optimized to compensate slowly diverging pitches.
To run the PEAQ algorithm, the WAV file containing the test signal is played in the analyzer section internally while the reference file is played by the generator section, allowing the analysis to access both signals in this case as well.

Applications for offline operation are measurements on typical play-only devices such as MP3 players. The DUT plays the MP3 file, and the resulting test signal is recorded. To analyze the overall quality of such a music reproduction, the original, uncoded reference file is played by the R&S®UPV generator while the recorded test file is measured by the analyzer section. Other applications are tests of different types of coding algorithm where input and output signals are available in files only.

4.4 Results Obtained from PEAQ Measurement Function

4.4.1 Objective Difference Grade (ODG) and Distortion Index (DI)

The most immanent output of PEAQ is the indication of the perceived quality. This is offered in two ways: by the distortion index (DI) and by the final objective difference grade (ODG).
The ODG is the output value from the objective measurement method that corresponds to the SDG in the subjective domain. The resolution of the ODG is limited to one decimal. However, be cautious and do not generally expect that a difference between any pair of ODGs of a tenth of a grade is significant. This also applies when looking at results from a subjective listening test. As seen in the right-hand diagram shown in Fig. 20, the ODG can also show positive values. Such values can occur, since PEAQ uses the cognitive model to map the MOVs to the results of subjective listening tests. In the case of subjective listening tests, the SDG can assume a positive value when a test person has incorrectly assigned the reference and test signal.

The DI has the same meaning as the ODG. However, DI and ODG can only be compared quantitatively, but not qualitatively. Fig. 20 shows two curves that represent the relation between the quality and the DI value (left diagram) and the relation between the quality and the ODG value (right-hand diagram). As the diagram on the left demonstrates, the DI is characterized by a saturation that is less than the saturation of the ODG curve. Furthermore, the range of values is different.

As a general rule, use the ODG as the quality measure for ODG values greater than approx. −3.6. The ODG correlates very well with subjective assessments in this range. When the ODG value is less than −3.6, use the DI.

ODG is mainly suitable for measurements on high-quality equipment. For devices with higher compression rates showing annoying impairments, DI is the better choice.

Note: Never compare the ODG value of one measurement with the DI value of another.

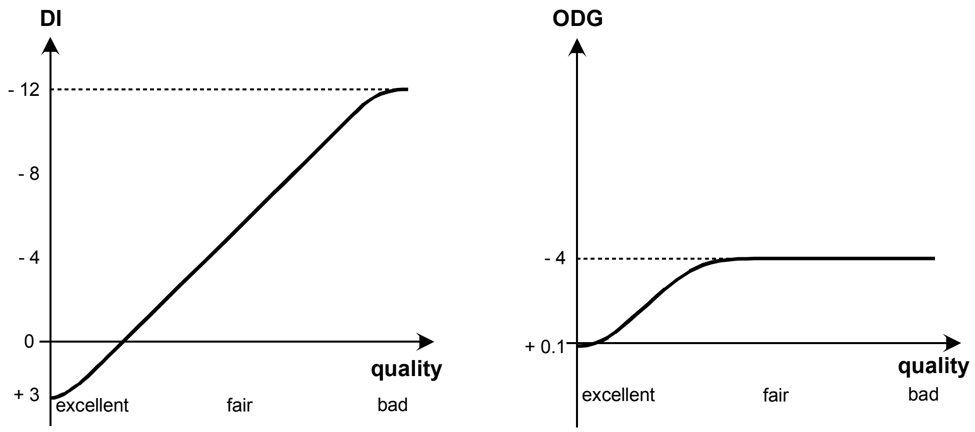

Figure 20: Distortion index (DI) and objective difference grade (ODG)

4.4.2 Amplitude of Test Signal

Music and speech signals show a large variety in amplitude over time. The R&S®UPV analyzes the test signals in small portions of time. After the analysis routine is finished, one of these two amplitude values can be displayed by the ***Level Monitor***:

- RMS value integrated over the whole test signal, or
- the maximum peak value within the time interval of the test signal

Figure 21: Numeric display of PEAQ results

4.4.3 Delay/Latency

As soon as a signal is processed by any piece of equipment, it will be slightly delayed. The resulting delay is also frequently called latency.
The PEAQ tool automatically compensates for delays in the signal in the range of up to ±1000 ms. To give the user a feedback about the delay present in the signals, the delay is returned in ms. Note that negative values for the delay are possible, indicating that the signal started earlier in the test signal than the reference signal. While this is not possible in real systems under test, it might occur in file-based operations (offline

measurements) if the files have been edited and therefore the latency has been modified.

Only signals with a constant time offset can be processed correctly by PEAQ. The value ***Avg Delay*** displays the average temporal offset between the reference and the measurement signal. The following figure shows an example of results as they are indicated by the R&S®UPV audio analyzer in the ***Analyzer Function*** panel.

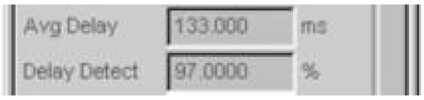

Figure 22: Average delay

As an indication of the performance of the delay compensation module, a value called the ***Delay Detect*** is returned. This value has a range from 0 % to 100 %, with values close to 100 % indicating that delay compensation was successfully carried out. For signals which, for example, contain severe distortions, this value can become significantly lower. If the value of ***Delay Detect*** drops below 70 %, subjective verification of the objective results is strongly recommended, as this indicates relevant impairments.

4.4.4 Automatic Detection and Correction of Swapped Channels

Some transmission systems swap the left and right channel, but this may also happen if a recording device is not properly connected. In order to conduct proper measurements, this must be corrected before running PEAQ. The PEAQ tool used in the R&S®UPV does this automatically.

4.4.5 *Waveform* Display

With PEAQ measurement, the ***Waveform*** display is used to check the signal in the time domain. Above all, the prescribed periods of silence at the beginning and the end of the signal and also the correct level modulation can be checked here. The ***Waveform*** display always shows the signal that is connected to the analyzer's inputs, and both channels can be displayed in the same graph.

The ***Waveform*** function is activated in the ***Analyzer Function*** panel and configured with regard to the trigger condition and storage depth. The ***Waveform*** graphics dialog can also be opened from here. The scaling – including the limit check and reference basis – is selected in the ***Waveform Config*** panel.

When measuring a device under test (online measurement), the test signal from the DUT has to be recorded before it is fed to the internal PEAQ algorithm. The ***Waveform*** analyzer gets the data in parallel and displays the result as soon as the recording is finished. In offline mode the reference signal and the test signal can be processed by the PEAQ algorithm directly from the files without delay. If the ***Waveform*** function is activated, this function runs first and therefore delays the start of the internal PEAQ algorithm.

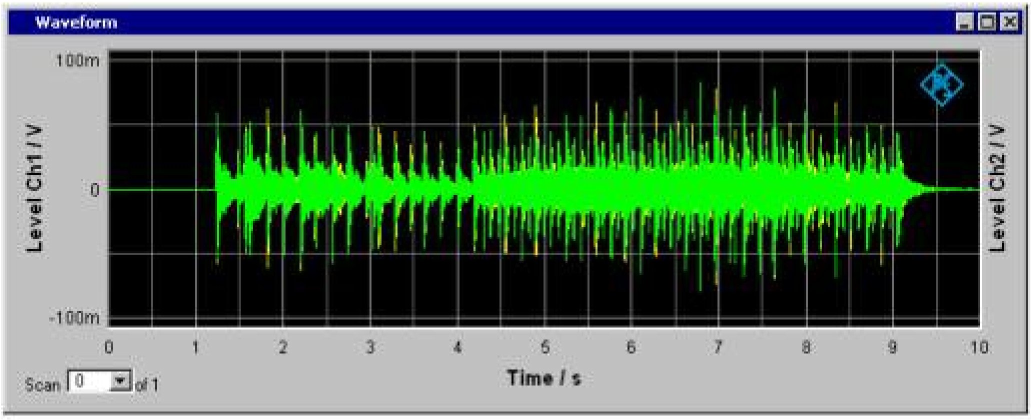

Figure 23: The Waveform display shows both channels in time domain (channel 1 in green, channel 2 in yellow color)

4.5 Introduction to Instrument Operation

4.5.1 Setting the Generator Instrument

Whenever PEAQ analysis is performed, the generator part of the R&S®UPV is needed to replay the reference signals. In accordance with ITU-R BS.1387, reference files with sample frequencies of 48 kHz have to be used.
WAV files can be replayed by setting the line ***Function*** to ***Play*** in the ***Generator Function*** panel. Using the analog interfaces, ***Bandwidth*** has to be set to ***22 kHz*** or to ***Play Auto*** in the ***Generator Config*** panel. For all digital generator instruments, the line ***Sample Freq*** in the ***Generator Config*** panel has to be set to ***48 kHz***.

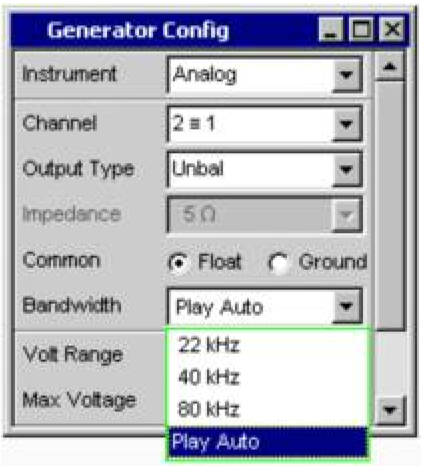

Figure 24: Setting the Generator Config panel

4.5.2 Setting the Analyzer Instruments

To select analog or digital interfaces is done by setting the ***Instrument*** in the ***Analyzer Config*** panel. Using the analog analyzer, dual-channel mode and ***Bandwidth 22 kHz*** has to be selected and ***Range*** has to be set to ***Fix***.
With any of the digital analyzer instruments ***Sample Freq*** needs to be ***48 kHz***.

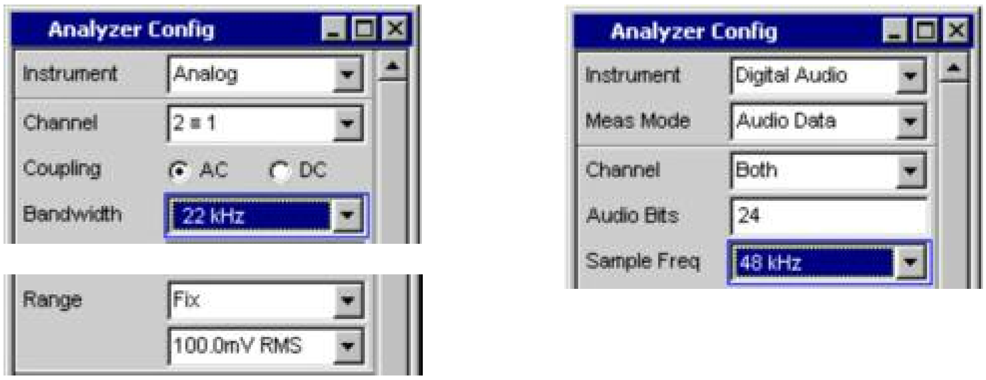

Figure 25: Setting the Analyzer Config panel for analog (left) and digital applications (right)

To activate the PEAQ measurement function, in the Analyzer Function panel the line ***Function*** has to be set to ***PEAQ***.
The setting in the line ***Version*** specifies whether the basic or the advanced ear model is to be used.

The line ***Meas Mode*** defines whether the measurement is to be performed in realtime (***DUT***) or with an offline analysis using previously stored WAV files (***Offline***).
In ***DUT*** mode, while the reference WAV file is run, the measurement signal is simultaneously recorded at the analyzer. The analysis takes place as soon as the reference signal has been run completely. The WAV file is always stored with a sampling rate of 48 kHz.
In ***Offline*** mode, the PEAQ measurement is to be performed on previously stored WAV files. After the files containing the reference signal and the measurement signal are specified, the evaluation can be started by pressing the START button.
The sample rate of the stored WAV files must be 48 kHz. The length of the recorded signals must not be shorter than that of the reference signal selected in the generator.

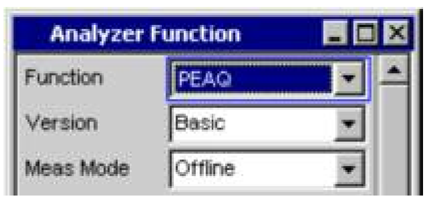

Figure 26: Setting the Analyzer Function panel

4.5.3 Selecting Levels and Measurement Ranges

Especially for measuring PEAQ in analog domain, correct levels and measurement ranges must be set.

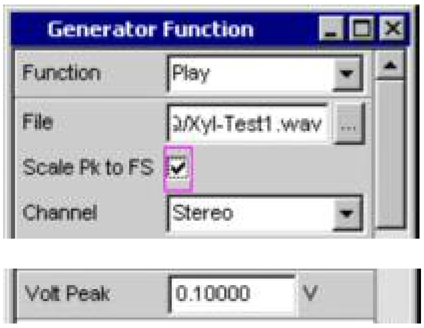

Figure 27: Setting the output level

The reference file, which is stored in a WAV file in digital format, is output by the R&S®UPV generator. The output level can be set in the line ***Volt Peak***; for example, setting the output level to ***0.1 V*** means that samples stored in the file with full-scale amplitude will be output with 0.1 V peak level at the analog outputs.
Quite often, stored music or test signals do not reach full-scale amplitude, resulting in lower output levels at the generator, and therefore may have amplitudes that are too low to drive the DUT properly.
To overcome this situation, R&S®UPV offers the possibility to rescale the samples in order to reach digital full-scale levels when replaying the files (line Scale ***Pk to FS*** in ***Generator Function*** panel).
For PEAQ measurements, it is advisable to use this function to make sure that the set output level ***Volt Peak*** really is reached (see Fig. 27).

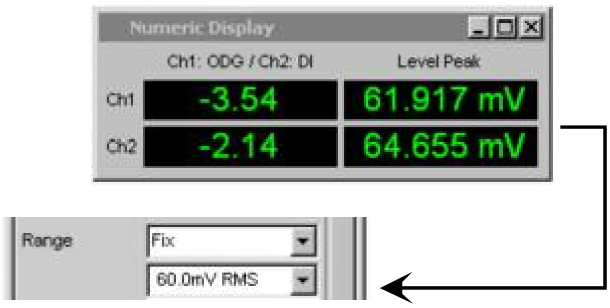

Figure 28: Correct setting of measurement range

Because music and speech signals show a large variety in amplitude over time, an autorange function in the analyzer would switch between ranges quite often. To prevent interruption of the PEAQ measurement, the analyzer must therefore be set to a fixed measurement range. Setting the range according to the set output level in the ***Generator Function*** panel would be the first approach. But if the device under test introduces some gain or attenuation, this must be compensated. To find the best measurement range setting, the ***Level Monitor*** function can help. If the reference signal has been played to the DUT and fed into the analyzer from the device's output, the ***Level Peak*** indication will display the maximum peak level which has been output by the DUT. The measurement range should be set as close as possible to a value of measured peak value divided by √2, but must be bigger than this value.

The internal algorithm of PEAQ is done in digital domain. In accordance with the standard, the analysis requires that the reference signal and test signal have the same level. The used PEAQ implementation automatically corrects static level differences of up to 30 dB; however, to get reliable results, level differences should be kept below 10 dB.

In the ***Analyzer Function*** panel, the lines ***Ref Level*** and ***Deg Level*** display, respectively, the RMS values of the reference signal and the test signal at the input of the internal PEAQ routine (i.e. after the measurement input stages but before the PEAQ internal level correction). Both channels are averaged.
To set the measurement properly, the difference between ***Ref Level*** and ***Deg Level*** should be as small as possible (less than 10 dB is recommended).
The screenshot below shows an example of a correct setting.

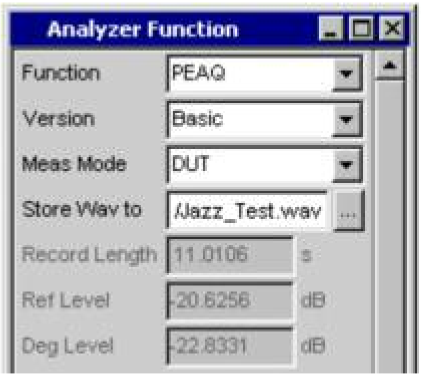

Figure 29: Example of Ref Level and Deg Level results

4.5.4 Saving of measured signals

If the PESQ measurement is performed in realtime (***Meas Mode*** set to ***DUT***), the measured signal can be stored as a WAV file. This done by activating the ●●● button in the line ***Store Wav to*** (see Fig. 29) after the measurement is terminated. The ***File Selector*** opens to define file name and directory.
The WAV file is always stored with a sampling rate of 48 kHz.

5 Summary

Psychoacoustic measurements are used more and more to get information about the perceived quality of compressed speech and music signals.

PESQ (perceptual evaluation of speech quality), in accordance with ITU-T P.862, and PEAQ (perceptual evaluation of audio quality), in accordance with ITU-R BS.1387, today represent the state-of-the-art technique for the objective evaluation of the perceived audio quality of audio codecs.

The R&S®UPV Audio Analyzer offers both analysis tools by using the PESQ measurement option R&S®UPV-K61 and the PEAQ measurement option R&S®UPV-K62.

Due to its PESQ and PEAQ measurement capability, the R&S®UPV audio analyzer makes it possible to handle new applications in all areas of compressed audio transmission including consumer audio and the mobile radio sectors. The R&S®UPV is the first audio analyzer which provides these functions in a standard test set, which means that users do not need to invest in any additional, expensive test equipment.

6 Abbreviations

DI	distortion index
LQO	listening quality objective
LQON	listening quality objective narrowband
LQOW	listening quality objective wideband
LQS	listening quality subjective
MOS	mean opinion score
MOV	model output variables
PEAQ	perceptual evaluation of audio quality
PESQ	perceptual evaluation of speech quality
PSQM	perceptual speech quality measure
SDG	subjective difference grade

7 References

ITU-R Recommendation BS.1116, Methods for the subjective assessment of small impairments in audio systems including multichannel sound systems (1997)

ITU-R Recommendation BS.1387, Method for objective measurements of perceived audio (1998-2001)

ITU-T Recommendation P.501 Amendment 1, Test signals for use in telephonometry (2004)

ITU-T Recommendation P.800, Methods for subjective determination of transmission quality (1996)

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

ITU-T Recommendation P.862, PESQ an objective method for end-to-end speech quality assessment of narrowband telephone networks and speech codecs (2001)

ITU-T Recommendation P.862.1, Mapping function for transforming P.862 raw result scores to MOS-LQO (2003)

ITU-T Recommendation P.862.2, Wideband extension to P.862 for the assessment of wideband telephone networks and speech codecs (2005)

ITU-T Recommendation P.862.3, Application guide for objective quality measurement based on Recommendations P.862, P.862.1 and P.862.2 (2005)

Operating Manual R&S®UPV Audio Analyzer

Some information in this application note is based on publications by OPTICOM

8 Ordering Information

Designation	Type	Order No.
Base unit		
Audio Analyzer	R&S®UPV	1146.2003.02
Audio Analyzer without Display	R&S®UPV66	1146.2003.66
Software options		
Software for PESQ Measurement	R&S®UPV-K61	1401.7309.02
Software for PEAQ Measurement	R&S®UPV-K62	1401.7750.02

For additional information about audio analyzers and other measurement equipment see the Rohde & Schwarz website www.rohde-schwarz.com.

About Rohde & Schwarz

Rohde & Schwarz is an independent group of companies specializing in electronics. It is a leading supplier of solutions in the fields of test and measurement, broadcasting, radiomonitoring and radiolocation, as well as secure communications. Established 75 years ago, Rohde & Schwarz has a global presence and a dedicated service network in over 70 countries. Company headquarters are in Munich, Germany.

Regional contact

Europe, Africa, Middle East
+49 1805 12 42 42* or +49 89 4129 137 74
customersupport@rohde-schwarz.com

North America
1-888-TEST-RSA (1-888-837-8772)
customer.support@rsa.rohde-schwarz.com

Latin America
+1-410-910-7988
customersupport.la@rohde-schwarz.com

Asia/Pacific
+65 65 13 04 88
customersupport.asia@rohde-schwarz.com

Certified Quality System
ISO 9001
DQS REG. NO 1954 QM

Certified Environmental System
ISO 14001
DQS REG. NO 1954 UM

Rohde & Schwarz GmbH & Co. KG
Mühldorfstraße 15 | D - 81671 München
Phone + 49 89 4129 - 0 | Fax + 49 89 4129 – 13777

www.rohde-schwarz.com