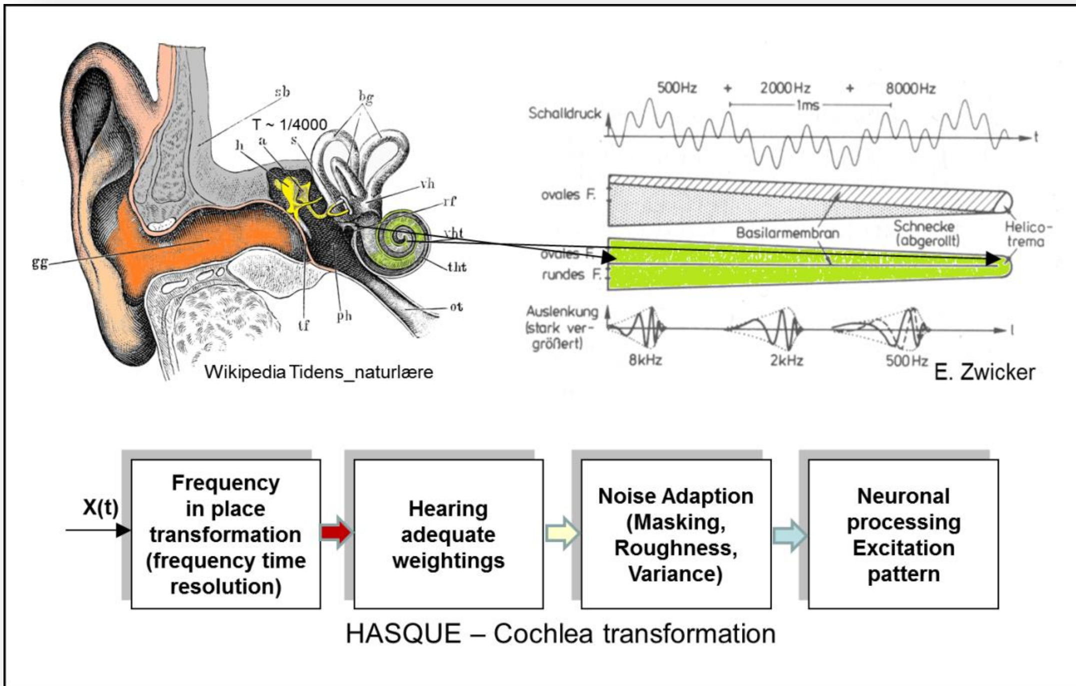
Sound acoustics

Research, Development, Implementation

HASQUE



HASQUE – Cochlea transformation

Objective quality evaluation of audio and telecommunication systems by listening test simulation based on HASQUE

Inhalt

Introduction

HASQUE (*Hearing Adequate Signal Quality Evaluation*) is a listening test simulation [1] principle for quality evaluation of audio and telecommunication systems. HASQUE replaces time consuming listening tests with strong variant results by objective measurement technique with reproducible and plausible results.

Essential features of the HASQUE principle are the adaptability to different listening tests by the programmable listening test conditions (playback loudness, bandwidth, threshold of acceptance) and the hearing adequate weighting of signals and background noise even so during speech pauses.

HASQUE measurement systems are used among others by the „Bundesanstalt für den Digitalfunk der Behörden und Organisationen mit Sicherheitsaufgaben" (BDBOS) for certification of digital radio devices and control centres.

Development

Main rules for the simulation of subjective listening test properties were indicated with the aid of the research results by Eberhard Zwicker [2, 3].
The correct simulation of the frequency into place transformation in the inner ear (Cochlea) was among others a challenge for the development of new algorithms because the frequency dependent time resolution of the human ear couldn't be correctly simulated by the Fast Fourier Transformation (FFT) or by known band filter approaches. Therefore a sliding spectral analysis [4] was proposed even in the 1990th for an enhanced time resolution. In the following years the BARK - transformation [5] and the CFT [**Fehler! Verweisquelle konnte nicht gefunden werden.**] were developed for improved simulation of the frequency into place transformation with hearing adequate time resolution.

The development of the HASQUE principle started in 2003 based on the CFT, after it didn't draw any useful solution within the International Telecommunication Union (ITU) for a hearing adequate quality evaluation of signals with background noise and new principles for noise reduction, bandwidth extension or echo cancellation.

HASQUE was filed as patent in 2005[8] and continuously enhanced by the experience with different applications in the technical fields as multimedia (PC), defense (acoustic positioning), telecommunication (evaluation of noise reduction, bandwidth extension, speech codecs) and professional audio technique (evaluation of TV spots). The continuous research and the collaboration with companies of these technical fields led to an improvement of the correlation with subjective quality measures clear above 90%.

After completion of the HASQUE principle the development of HASQUE based measurement system was started in order to reach an easy to serve user interface with extended test functions for time saving error analysis and error recognition.

Listening test simulation with HASQUE

A correct listening test simulation requires the correct simulation of both the human hearing system and the frame conditions of the subjective listening test.

Simulation of the listening test frame condition

Listening tests are carried out according to ITU recommendation [8, 10] at a defined **listening loudness** - e.g. as a rule at a sound pressure level of 79 dB (A) - and an application dependent **bandwidth**. The **threshold of acceptance** of the test persons shows strong variations - individually and dependent on the days form and current mood - but will point in any case to the application.
Hence telecommunication systems are weighted with higher threshold of acceptance than professional studio systems.
Test persons are weighting audio samples at a **quality scale** which might differ dependent on the applied recommendation.

An essential feature of the HASQUE principle is, that the listening test parameters **listening loudness, bandwidth** and **threshold of acceptance** as well as the **quality scale** can be set according to different listening tests and quality scales.

A distortion will be received differently dependent on the **listening loudness**. Especially weak distortions are more clear perceived at high than medium or low volume although the undistorted signal contents are changed dependent on the volume too [3 , 4].

The indication of the **bandwidth** is necessary in order to make quality measures of different listening tests comparable. A reduction of the bandwidth causes in any case an increase of the objective quality measure due to the fact that distortions outside the observed bandwidth are suppressed. Measurement systems without bandwidth limitation operate with the maximum possible bandwidth, which is only limited by the sampling rate of the application. A bandwidth limitation is introduced according to ITU-T P.862 [11] for the simulation of the telephone quality.

The **threshold of acceptance** reflects the maximum accepted loudness of distortions and is indicated in Sone.

The **quality scale** can be defined individually. A scaling according to ITU-T on a five point scale (e.g. STI-CISScale) is also possible as a scaling in percentage.

Simulation of the human hearing system

The simulation of the human hearing system is carried out within the HASQUE principle with a so called Cochlea transformation within the whole available frequency range without band limitation. The name Cochlea was chosen as most listening properties are determined in the Cochlea of the inner ear.
Figure 1 indicates the basic functions of the Cochlea transformation. The frequency in place transformation is operating with optimized time frequency resolution and computes the place coefficients which are weighted with hearing adequate weighting coefficients. The following noise adaptation function is used to compute masking effects with the aid of adaptive algorithms in time and frequency domain in order to receive a hearing adequate masked threshold. This means that the masked

threshold will be adapted to the background noise of the system under test dependent on the noise properties. Hence constant noise sources as e.g. a motor vehicle with constant speed are differently weighted than variant noise sources as e.g. the rattling of a passing train. The analysis of the noise adapted place coefficients according to the signal properties into different excitation patterns makes a fine approximation to subjective perceived excitations possible and can be seen as a neuronal process [18].

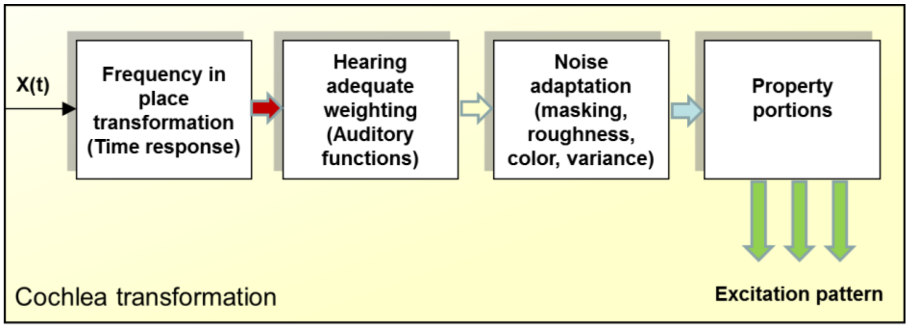


Figure 1 : Simulation of subjective perceived excitations

HASQUE is simulating the listening properties of the human hearing system with the aid of the Cochlea transformation, compares the simulated excitation pattern of the input and output signal of the system under test and computes with the resulting audible error of the system a quality measure.

Comparison with subjective listening tests

If we compare the results of different listening tests, we can observe that the quality measures (MOS) decrease with increasing distortions (Figure 2). Distortions might be different noises, signal interrupts, linear and nonlinear distortions. Hence the abscissa in Figure 2 is indicated without unit. Speech samples with following properties were applied for these tests:

- Bandwidth 4-8kHz
- Distortions SNR 6-50dB
- Noise sources
  - Street
  - Car
  - Kitchen
  - Restaurant
  - Office
- Record length 5...60 Seconds

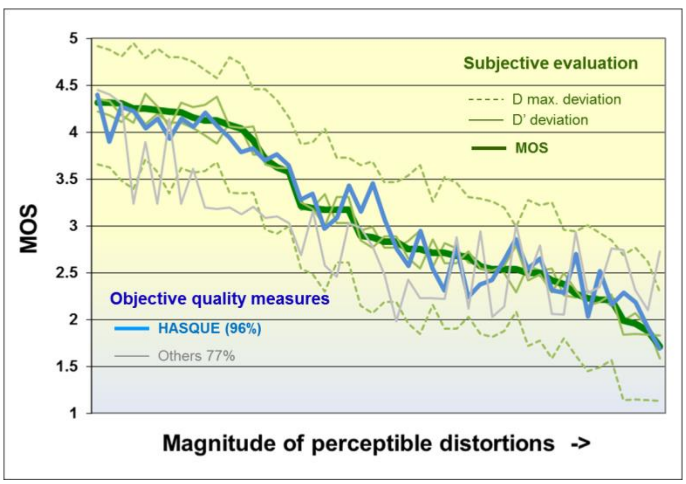


Figure 2: Subjective and objective quality measures

The maximum deviation of the subjective scores (D=dotted green lines)from the mean opinion scores (green) result in the individual threshold of acceptance of the test persons which might also vary dependent on the days form of a test person. An interesting fact is, that the quality differentiation - what sounds better, what worse - between different test cases show high correlation with the MOS curve. Hence the mean deviation of the threshold of acceptance which leads to a deviation from the MOS can be interpreted as offset. In order to achieve an offset free curve (D’) making the quality differentiation between MOS and test person dependent deviations visible, the offset can be subtracted from the test person dependent scores according to G 1 with N = number of tests, D = maximum deviation and n current test sample.

$$D'(n) = D(n) - \frac{\sum_{n=0}^{N-1} MOS(n) - \sum_{n=0}^{N-1} D(n)}{N}$$

G 1: Offset compensated scores D’

The offset compensated scores according to G 1 are indicated in Figure 2 with the green thin lines and show the high correlation of pure quality differentiation between test persons with different threshold of acceptance.

The blue curve indicate the objective scores of the HASQUE principle,
Essential errors in the objective quality evaluation might occur by fault simulation of the human hearing properties. Especially a band limitation, which might be introduces by e.g. IRS - filters as applied in other principles might lead to strong deviations from hearing adequate mean opinion scores.

Band limitations and nonlinear distortions are comparable with a damage of the hearing system or with ear muffs during listening tests. Hence for plausible evaluation clean reference signals are mandatory [11].

Measurement systems

HASQUE measurement systems are made available for real-time and offline evaluation of audio- and telecommunication systems.

Real-time measurement systems

HASQUE real time measurement systems (Figure 3) operate with bi-directional analogue interfaces which are sampled with high quality 24 Bit AD and DA converters at 48 kHz sampling rate.

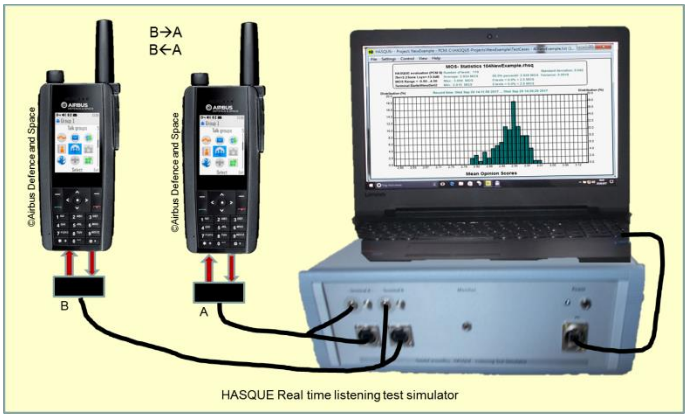


Figure 3: Real-time measurement system

A **hardware test program** is made available for the user providing both, a **function test** and an **automatic level adaptation** between system under test and measurement system.

Both functions are easy to use, need not more than some minutes to be carried out and ensure a correct quality evaluation of the system under test.

The hardware test program applies among others a special test signal (Bartlett burst) which is suitable to measure the signal transfer properties (level, distortions, SNR) of most systems under test. The Bartlett burst was developed in order to make measurements at systems with automatic sine wave suppression or signal degrading codecs possible.

What kind of test samples [13] and how many tests are to be carried out, or if test cases should be repeated are application dependent decisions [14], which can be determined by the user with the aid of a task editor belonging to this measurement system. The task editor provides easy tasks for the **measurement control** and a compiler for error check and task script creation.

The measurement control operates after start according to the user defined tasks and creates recordings, file lists and measurement results with general results over all, histograms and single results.

Function test

The function test indicates immediately after start of the hardware test program the graphically representation of the Bartlett burst at the signal interfaces and in the legend measurement results about current level, distortion and SNR (Figure 4).

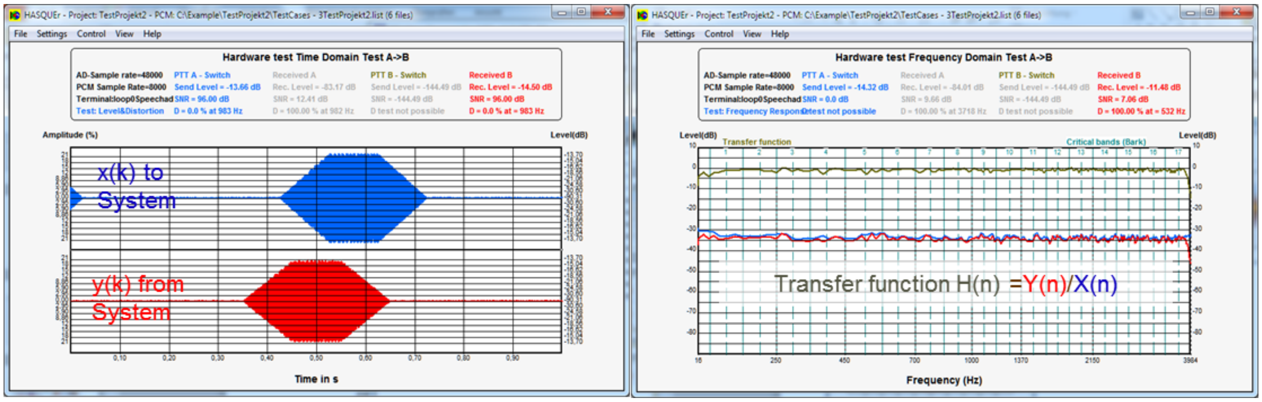


Figure 4: Function test of the connected system under test by simple mouse click

The reproduction of the received Bartlett burst (y (k)) makes assumptions about the system properties possible. Deformed edges of the indicated sextant point to an AGC influence and a concave dent in the middle of the sextant points to an error or mismatch in the system.

The test of the transfer function of the connected systems is carried out very quickly by mouse click in the legend and shows the current spectra at the interfaces and the corresponding transfer function of the system under test.

All necessary measurement functions were developed especially for these quick tests, in order to ensure a reliable error free operation between measurement system and system under test for a reliable quality evaluation.

Automatic level adaptation

The interface levels between measurement system and system under test can be well adapted with the aid of the automatic level adaptation through which a low measurement tolerance (typ. < 1%) and thus reproducible measurement results are achievable.

Automatic level adaptation takes place with the aid of a wizard as shown in Figure 5 through which nearly any manual possible mismatch can be excluded.
The applied measurement programs behind the wizard apply above mentioned Bartlett burst for the finding of the maximum interface levels and adjust the necessary send and receive gain of the measurement system to reach a distortion free measurement operation.

The wizard opens a dialog box with the found settings and provides edit fields for the description of the connected hardware after automatic level adaptation.

The whole content of the indicated dialog box can be saved after all in a terminal parameter file with the user defined name for future measurements to be repeated with the same system under test.

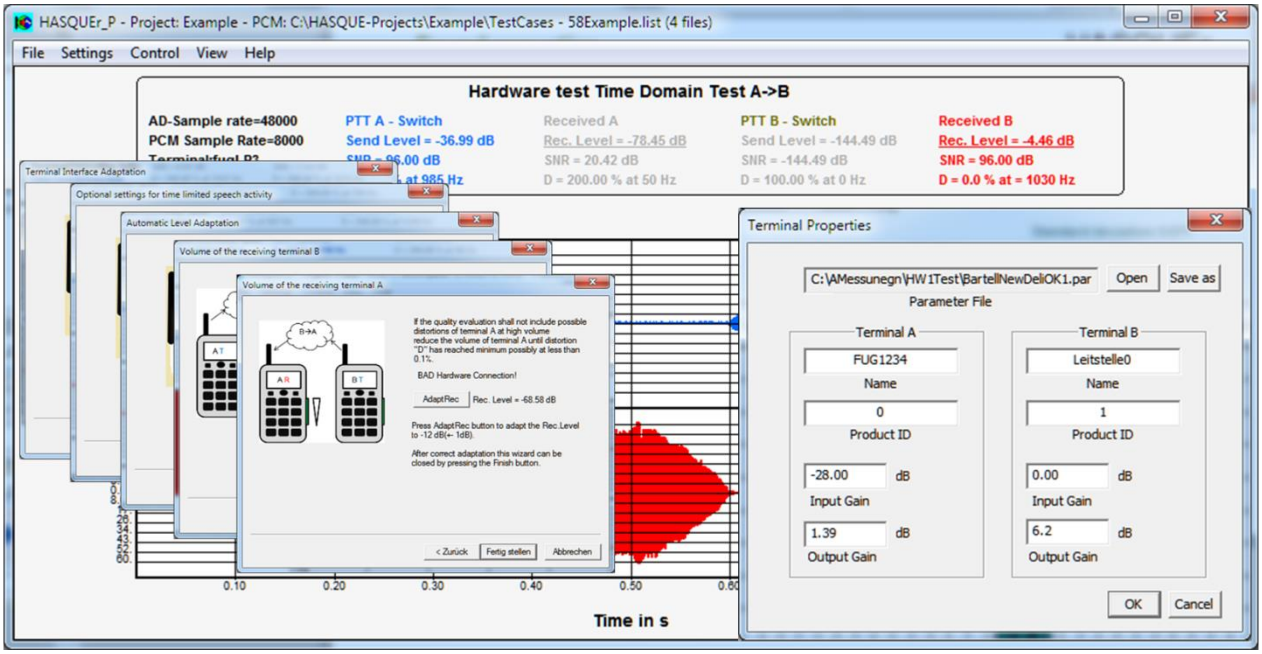


Figure 5: Automatic level adaptation between measurement system and system under test

Task editor for measurement control

Real time measurements are started with the aid of a programmable task interpreter. The task interpreter is used to control the measurement system, the measurement direction, the connected system under test, the individual selection of the reference samples and the recordings from the system under test.

A task editor for the coding of the task interpreter is made available, by which the measurement control is checked during compilation before start of the real time measurement. Following programming versions can be made available:

TIP- Individual Programming
Open a Popup Window = Please confirm connection
Start a Loop = 10
 Reference file =C:\ITUTestCases\Ref1.wav
 Press PTTA
 Wait =1500
 Evaluate
 Release PTTA
 Wait =1500
 Press PTTB
 Wait =1500
 Evaluate
 Release PTTB
 Wait =1500
End of Loop

TOP- Optimized Programming
Open a Popup Window = Please confirm connection
Start a Loop = 10
 Reference file =C:\ITUTestCases\Ref1.wav
 EvaluateAB
 EvaluateBA
End of Loop

TUP- Unidirectional Programming
Open a Popup Window = Please confirm connection
Start a Loop = 10
 Send Ref1
 Receive Ref1
End of Loop

Figure 6: Programming versions - Task Editor

Figure 6 shows programming example of different task editors, which carry out the same measurement control.
The TIP-programming makes individual definitions of the hardware response time which is needed e.g. for the control of Push To Talk switch (PTT) possible. The TOP programming combines the PTT control with a fixed response time, which can be determined in the settings.
The TUP is used for distributed measurement systems, offers the same programming of the hardware response time as TOP and operates with a defined data base from where the test cases can be selected by synonyms (Ref1…Ref2).

Offline measurement system

The offline measurement system offers in addition to the listening test simulation of post processed recordings numerous measurement functions for error tracing and error analysis of audio and telecommunication systems and thus is even so suitable for quality enhancements of the systems under test.

Offline measurements are carried out with windows conform PCM signals (*.wav files). It is possible to carry out single and series measurements. The sample rate of the measurement system is adapted to the sample rate indicated in the header of the audio files. Statistics about quality measures, latency and speech interrupts are computed in conjunction with series measurements.

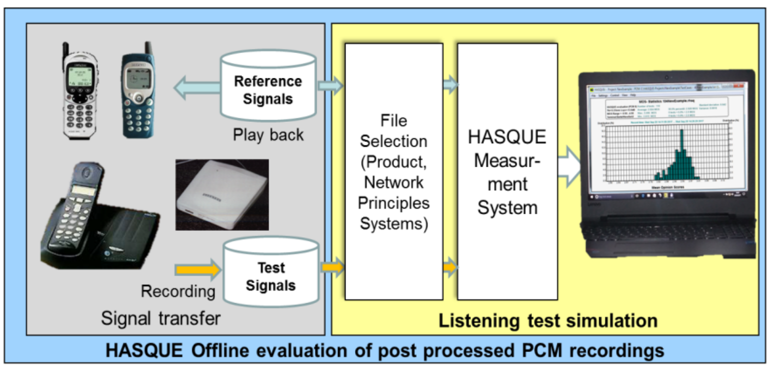


Figure 7: Offline measurement system

Series measurements are carried out with file lists, which can be created with the aid of a special editor for easy composition of different test cases. An import function for compressed RST-files extracts the belonging reference and test cases, creates the corresponding file list and releases the selections for evaluation.

The entire functions of the offline measurement system as well the following indicated measurement results and representations are part of the real time measurement system.

Measurement results

Statistics

The statistic evaluation of measurement results (Figure 8) of the system under test provides valuable hints about the system properties with one view.
Statistics with belonging histograms are computed and indicated for quality measures, latency and speech interrupts making clear conclusions about the system behavior of the system under test possible.

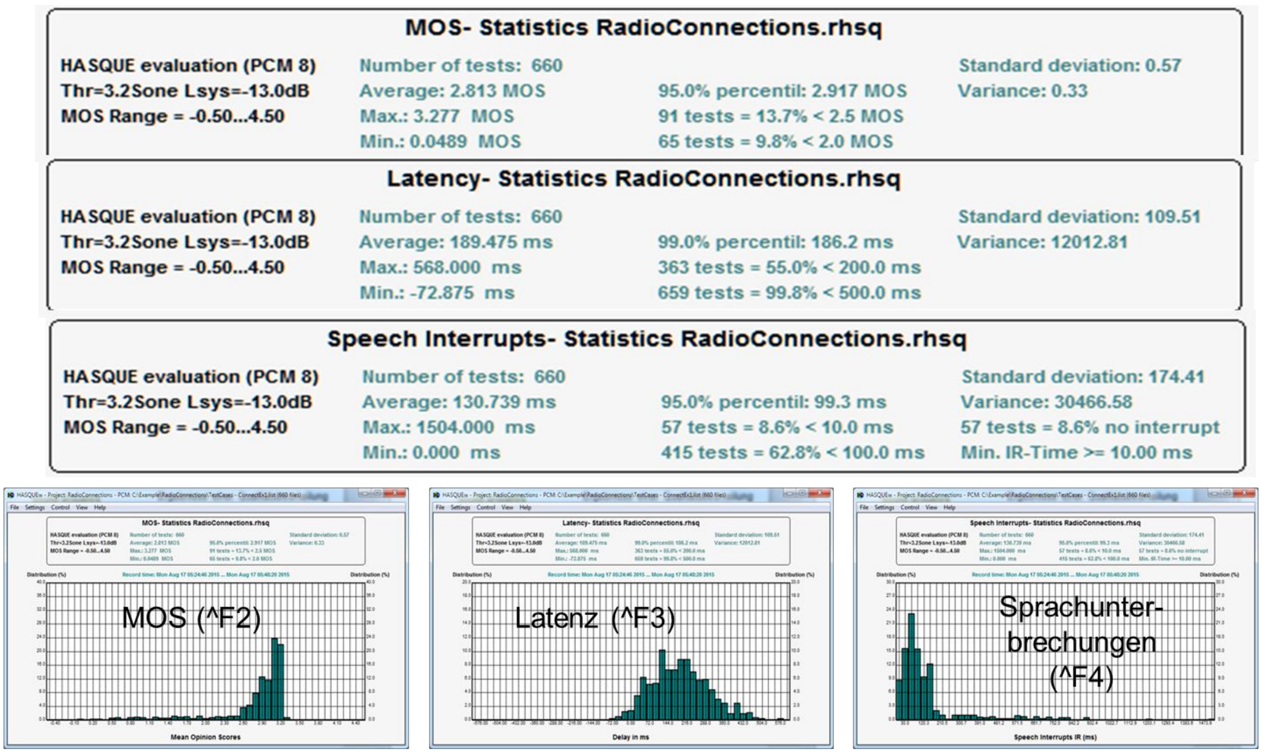


Figure 8: Statistics about the system under test

The number of tests, the arithmetic average magnitude, the maximum and the minimum Value, the standard deviation and variance are indicated in the legend of the histograms. The percentile and the number of tests which exceed an upper and lower programmable threshold are indicated in the legend in addition.

The percentile indicates the average magnitude of most frequently appeared measurement results within the indicated percentage of measurements. The indicated percentage of the percentile is programmable making the exclusion of runaways for the average magnitude derivation possible.

Also upper and lower thresholds are programmable making statements about the number of test cases which are in the range of still and not accepted measurement values.

General view of measurement results

HASQUE measurement systems provide general views of results by which the user can recognize when and where critical test cases occurred with one view (Figure 9).

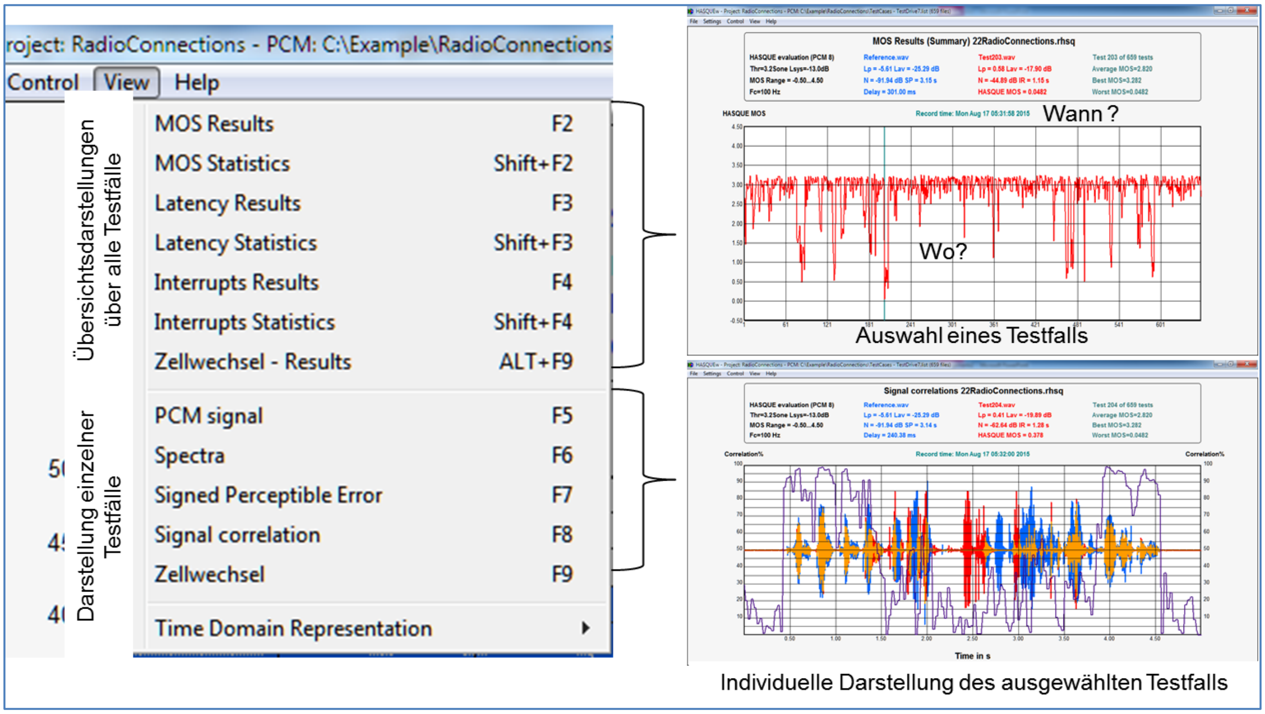


Figure 9: Time saving selection of single results - easy error recognition

General views of results are made available for different properties as speech interrupts, MOS, latency and an individual programmable error type (e.g. “Zellwechsel”) for individual examinations.
The choice of single test cases takes place by mouse click in the represented general view making the access to conspicuous test cases possible for detailed examinations in time and frequency domain.

Single test results

Easy access to single test results of interest is made possible by mouse click in general views and with the indicated minimum and maximum statistics in the legends by the graphical user interface.

Measurement results of selected test cases are indicated in the legend of each individual representation. Single results in time domain might be zoomed and time shifted in order to make conspicuous passages visible for enhanced examinations. With the aid of the playback function it is possible to evaluate and compare selected zoomed passages by listening.

Following representations and measurement functions are made available for examinations.

Perceptible errors

Figure 10 shows a selected test case with belonging perceptible error (brown curve). Selected passages can be listened and compared with the belonging reference signal in order to hear what we see.

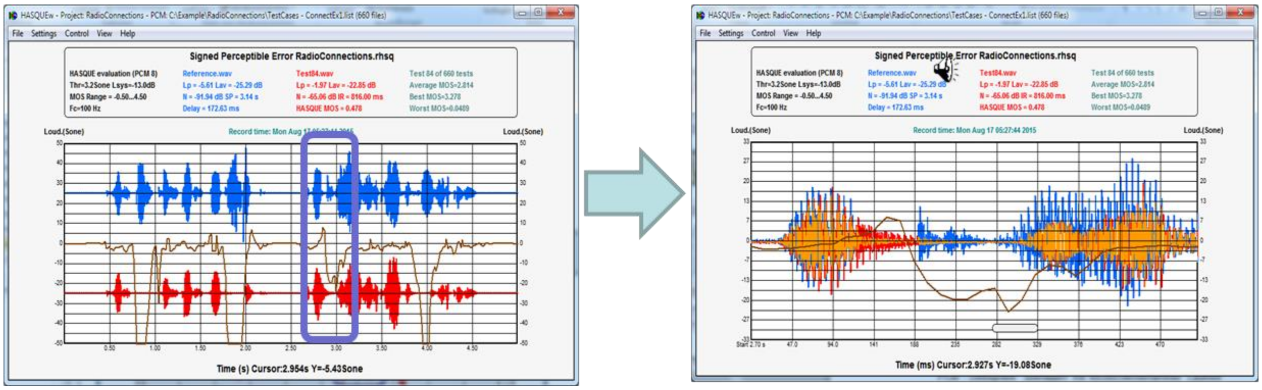


Figure 10: Perceptible error in time domain in order to see what we hear

Signal correlation

Differences between input and output signals of the system under test which are not recognized precisely by the perceptible error due to natural masking effects can be examined with the aid of the correlation analysis (Figure 11)

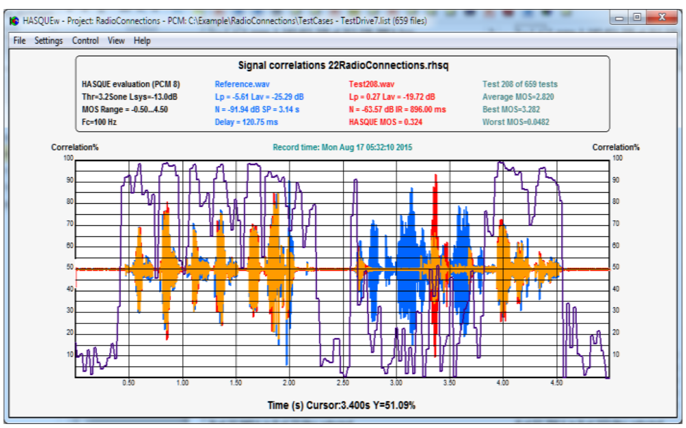


Figure 11: Signal correlation

Undistorted signals - i.e. the output signal equals the input signal - are indicated with 100% correlation. The indicated correlation over the time axis in percentage makes statements about the similarity and thus about the probability for the existence of artefacts or other signal distortions possible.

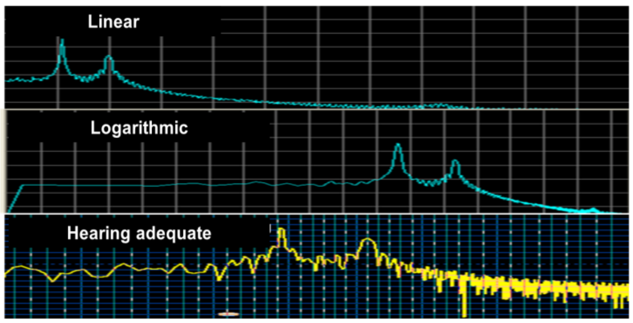
Variations of correlations over the time axis show the time and frequency dependent precision of the signal transfer.

Signal correlations are applied as error property for the evaluation of classified errors leading to a high error recognition rate.

Latency shifts and jitter which might be introduced by the system under test are eliminated for the derivation of signal correlation and so in the graphically representation as shown in Figure 11.

Spectral representation

HASQUE measurement systems indicate spectra with hearing adequate frequency distribution and the corresponding critical bands on the Bark scale. This representation relieves examinations by enhanced interpretation -what we see is that we can hear. Figure 12 makes clear that a double tone with some background noise which is placed in the middle of the audible frequency range cannot be represented correctly with linear or logarithmic distribution.



Double tone in the middle of the Cochlea with background noise

Figure 12: Hearing adequate vs. lin. and log. representations

Most interesting frequencies in the lower frequency range are such strongly compressed, that a reliable examination is not possible with linear representation. In contrast to this, the logarithmic representation takes more than 50 % of the X axis for the representation of the less interesting frequencies in the range between 0 and 50 Hz into account and the remaining more important frequencies are such compressed, that the examination becomes difficult.

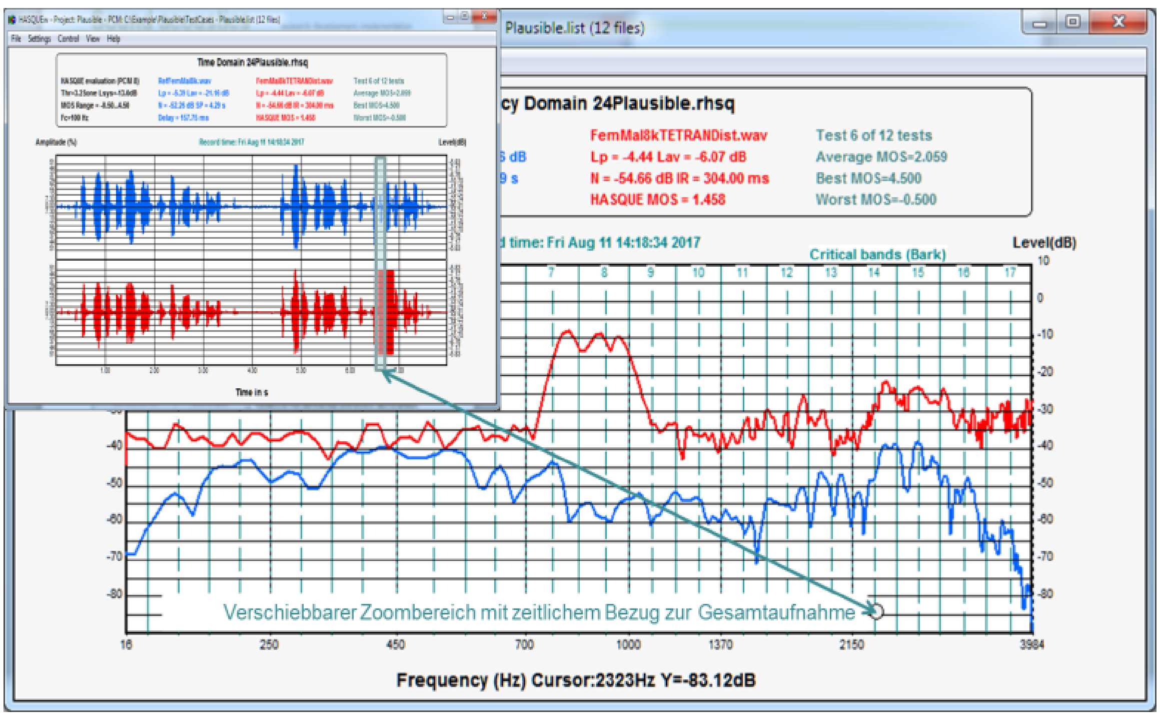


Figure 13: Frequency representation of a zoomed time section

The spectral analysis Figure 13 is always carried out in the selected time window and will be continuously renewed during shifts on the X axis of a zoomed time signal. Hence signal errors can be traced and examined in the frequency domain most precisely.

Error classification and error recognition

Signal errors might be caused by different reasons. The error classification allows individual determination of an error type as e.g. signal interrupts by weak radio connection, artefacts by cell reselection or acoustic distortion by alarm signals.

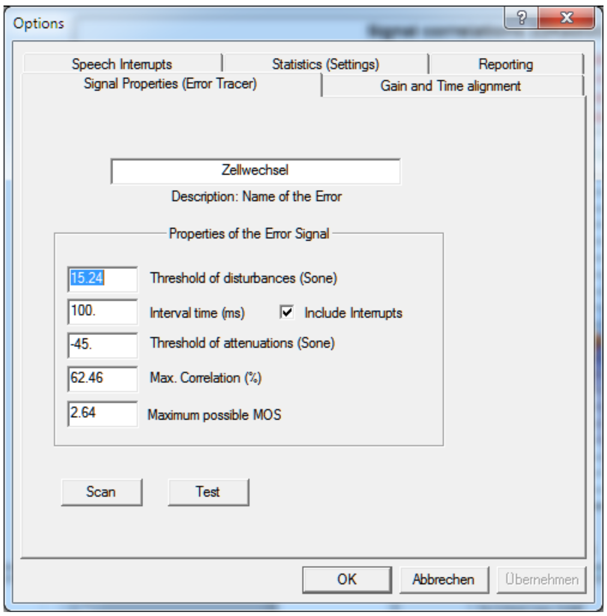


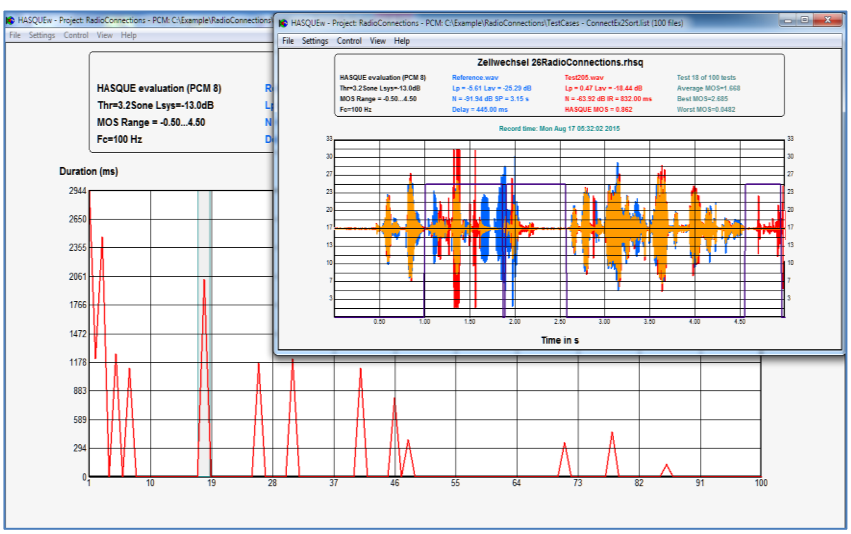
Figure 14: Error description and classification

The error classification is carried out with the aid of the dialog box “Signal Properties” Figure 14. The name of the error can be defined individually and is indicated in the graphics and corresponding menu items as shown in Figure 9.

The description of the error type properties can be found out automatically with the aid of a scanner.

The user can mark the error of interest by zooming in the desired graphically representation. Error properties are scanned from the representation and are taken over with finishing the error scanner wizard

It can be checked if the error recognition operates correctly by a test button which opens the graphic for the indication of the new defined error.



The error tracer is finding out the defined error and the duration of the error for each test case.

Individually defined errors can be examined in the general view with duration and time indication and so for each test case as shown in Figure 1 .

Figure 15: General view and single test case of an individually classified error

The recognition rate for individual defined errors is dependent on the error type and can be improved by iterative fine tuning of the properties with wrong interpreted test cases.
Tests of today for the recognition of cell reselection showed a recognition rate of 99% with and 100 % without latency jitter at 5708 test cases.

Conclusion and outlook

The complex tasks as quality evaluation, error recognition and quality enhancement of audio and telecommunication systems can be carried out reliable and efficient with the aid of the HASQUE principle, some new test functions and the well-engineered graphical user interface of the HASQUE measurement systems. The general view with direct access to conspicuous results makes remarkable time savings in contrast to classical search methods possible.

The error classification of the HASQUE measurement system offers new examination capabilities at a system under test, through which necessary means for the enhancement of the corresponding system can be recognized and putted into action.

Offline systems are suitable for retroactive examinations on post processed recordings or former evaluations in order to proof up to now not recognized system properties, or to check if the system under test reaches the necessary audio quality for the certification at the BDBOS.

Real time measurement systems are quickly and reliable adaptable to various systems under test for reproducible measurement results with low tolerance deviation.

The implementation of the HASQUE principle in distributed measurement systems, in products and in networks makes a reliable and continuous quality observation of audio and telecommunication systems with high correlation (>90%) between subjective and objective quality measures possible.

Our continuous research and development of today is concentrated on the automatic recognition and assignment of new classified signal errors for diagnostic systems.

Literature

1. Wikipedia: Hörtestsimulation
2. E. Zwicker: *Psychoakustik*. Springer, Berlin 1982, ISBN 3-540-11401-7
3. E. Zwicker, H. Fastl: *Psychoacoustics*. Springer, Berlin 1999, ISBN 3-540-65063-6
4. E. Terhardt, “*Fourier Transformation of Time Signals*,” Acustica, Vol. 57, 1985.
5. Rolf Kapust: „*Qualitätsbeurteilung codierter Audiosignale mittels einer BARK – Transformation*“. Dissertation, Technische Fakultät der Universität Erlangen, 1993.
6. Michael Walker: *Verfahren und Anordnung zur Durchführung einer an die Übertragungsfunktion menschlicher Sinnesorgane angepassten Fourier Transformation sowie darauf basierende Vorrichtungen zur Geräuschreduktion und Spracherkennung*, EP20020360079, 07.03.2002
7. Michael Walker: *Gehöradäquate digitale Sprachsignalverarbeitung*. Funkschau 04/2003, ISSN 0016-2841. Seite 57-58
8. Michael Walker: Vorrichtung und Verfahren für eine gehöradäquate objektive Qualitätsschätzung von Audiosignalen, P. 10 2005 019 903.8, 29.4.2005.
9. ITU-T P.800 Recommendation, *Methods for subjective determination of transmission quality*
10. ITU-T P.835 Recommendation, *Subjective test methodology for evaluating speech communication systems that include noise suppression algorithm*
11. ITU-T P.862 Recommendation, *Perceptual evaluation of speech quality (PESQ), an objective method for end-to-end speech quality assessment of narrow band telephone networks and speech codecs*
12. B. Ettinger/E. Herter, *Sprachverarbeitung*, ISBN 3-446-16076-0, Seite 178
13. Sound acoustics Literatur: *Anforderungen an Referenzsignale für eine korrekte Hörtestsimulation*
14. Sound acoustics Literatur: Empfehlungen für eine zuverlässige Qualitätsbeurteilung von TETRA-Endgeräten mit dem HASQUE Messsystem