One-view Visualization of Speech Quality Parameters for Mobile Hands-free Devices

F. Kettler, H.W. Gierlich, S. Poschen

HEAD acoustics GmbH, Ebertstr. 30 a, D-52134 Herzogenrath, Frank.Kettler@head-acoustics.de

1. Abstract

The ITU-T Recommendation P.505 [1] provides a new quality representation methodology. The focus is to summarize test results in a graphical representation - easy to use and easy to understand for experts and non experts. It can also serve as a basis for commercial decisions. Nevertheless it still comprehends enough detailed technical information to discuss possible next optimization steps. Various mobile hands-free implementations were measured and the results are represented this way. Although interaction aspects are explicitly not considered, the representation is suitable for decision making about the next steps presuming detailed background knowledge about these interactions.

2. Introduction in “Quality Pie” Visualization

In order to provide a quick overview about the results of a mobile hands-free implementation for all conversational speech quality aspects a graphical result representation in the manner of a “quality pie” was derived. The focus of this representation is to provide

- a “quick and easy to read” overview about the implementation including strength and weakness,
- a comparison to recommended values (or average results from benchmarking tests) or other implementations
- detailed information for development to improve the performance.

The circle segments and displayed parameters can be selected and adapted to the application, i.e. the device under test. An example of a hands-free implementation with parameter selection acc. to the VDA Specification [2] is shown in figure 1.

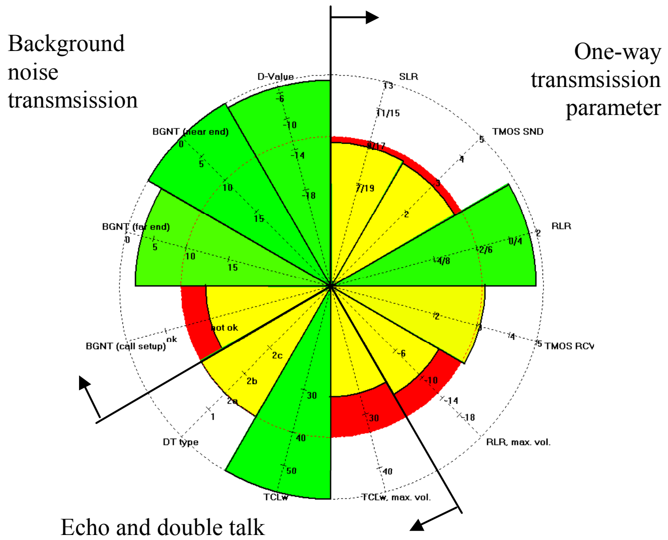

Fig. 1: Hands-free “Quality Pie”, parameter acc. to VDA [2]

The following general assumptions are made for the quality pie representation: Each parameter is represented by a pie slice. The size of each slice directly correlates to quality. In addition the size is color-coded. The green color indicates a higher quality for this specific parameter. Interaction aspects between single parameters are not considered. An inner-red circle indicates the requirement for each parameter. For those parameters that should be within a range, like the sending loudness rating (SLR) of 13 ± 4 dB the axis is double scaled. It raises from the origin of the diagram radial to the outside up to the recommended value (13 dB for the SLR in this example) and in addition radial to the inside. Other axes like the background noise transmission quality after call setup (BGNT call setup) are scaled only between two values (ok, not ok).

3. Interpretation of HFT “Quality Pies”

The hands-free “quality pie” shown in figure 1 does not represent an existing implementation. It is only used here for explanation purposes. In general the 12 segments - which can be regarded as a maximum suitable number being visualized in one diagram - can be subdivided into three groups covering different conversational aspects. The first 5 segments -clockwise arranged- represent one-way transmission parameters. The sending direction is covered by the sending loudness rating and the TMOS result using the TOSQA2001 analysis method [3]. The following two slices represent the receiving loudness rating (RLR) and the TMOS result in receiving direction. The fifth segment represents the RLR value at maximum volume.

The following 3 segments indicate the echo attenuation expressed through the parameter weighted terminal coupling loss according to ITU-T Recommendation G.122 [4] (TCL_W) measured at maximum volume, at nominal volume and the double talk performance. The last 4 segments represent parameters concerning the quality of background noise transmission.

4. Examples

One advantage of this representation can be seen in the possibility to trace different development phases. Figure 2 and 3 show two results measured during a system development.

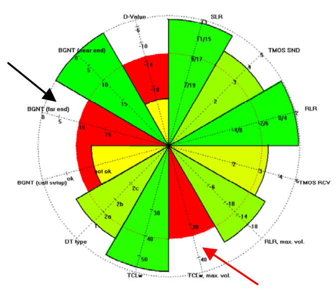

Fig. 2: Before …

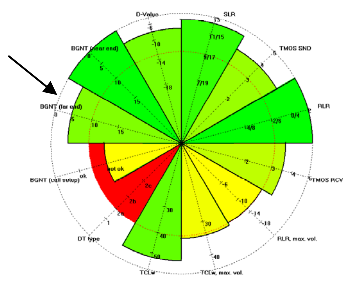

Fig. 3: ... and after first optimization step

Significant impairments could be observed in background noise transmission during the application of far end signals (see black arrow in Fig. 2). The background noise is completely attenuated. Other issues are the insufficiently low D value and significant echo disturbances at maximum playback volume (red arrow in figure 2). The reason for these impairments could be found in an insufficient performance of the implemented noise reduction and non-optimized choice of control parameters for the echo suppression unit. This parameter needs to be addressed, because a mobile hands-free phone will likely be used at relatively high playback volumes. Non-linearities in the echo path introduced by speaker distortions and temporal saturation in the microphone path can not be covered by the echo canceller. A "fall-back solution" relying on more simple attenuation techniques may be considered as an appropriate way to solve the residual echoes disturbances in this operation mode.

On the other hand Fig 2 also demonstrates a type 1 double talk performance characterization. The noise reduction was then optimized. The background noise modulation during the application of a far end signal could be minimized to a 5 dB attenuation (see black arrow in figure 3). The D value was also significantly improved. The corresponding slices in the quality pie clearly indicated these improvements. A more aggressive echo suppression unit as a "backup" of the algorithmic echo canceller for non-linear echoes leads to a sufficient attenuation at high playback levels (TCL_W max. volume). On the other hand the more aggressive echo suppression unit now leads to a type 3 double talk performance as a trade-off of these development steps.

Potential for improvement for such an implementation can be found in a more sophisticated speech level controlled attenuation. An important parameter for hands-free implementations is the monitoring capability of the play-back level provided by the loudspeaker amplifiers. This control parameter is extremely important, because the amplifier is typically embedded in the echo path for the implemented echo canceller. The double talk capability tests according to the VDA specification are carried out under "nominal" conditions, i.e. without background noise play-back and at nominal play-back volume. Any fall-back solution as described above by the mentioned attenuation at high play-back volumes should therefore be level-controlled in order to maintain a full duplex capability at lower or nominal levels.

Nevertheless the overall performance can already be regarded as improved. In a next step the optimization of double talk performance needs to be addressed for this implementation.

Figure 4 shows an analysis example of the background noise modulation together with an application of a near end test signal. This test leads to good results for the implementation represented in Fig. 2 and 3 shown above. This near end signal is played-back via the artificial mouth of the HATS (Head and Torso Simulator [5], positioned on the drivers' seat). The grey curve represents the level vs. time of the periodical repetition of composite source signal bursts used as test signal.

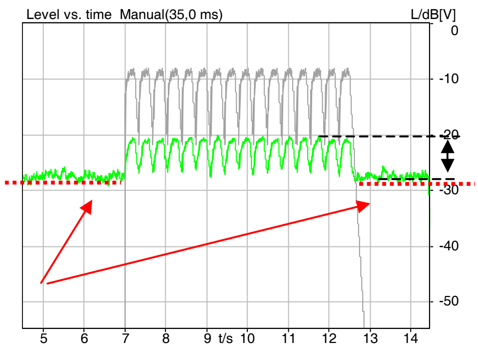

Fig. 4: Test result for the background noise transmission during the application of a near end signal (*BGNT (near end)*)

The green curve represents the measured signal in uplink direction consisting of the transmitted background noise (applied in the car cabin via a sound-simulation system) and the near end signal from the drivers' position. The red arrows indicate that the near end signal does not lead to a modulation or a "divergence" of the noise reduction system. The modulation of approximately 0 dB is represented by the parameters slice named "*BGNT (near end)*" in Fig. 2 and 3.

A relatively low signal-to-noise ratio is indicated by the black arrow. Note that this signal-to-noise estimation from Fig. 4 confirms the low D value for this implementation (see figure 2), although it needs to be stated that the D value measurement is based on two independent tests of background noise and speech-like signals in isolation.

Another example for the application of this P.505 result representation is shown Figure 6. It represents a commercially available after-market hands-free kit. The background noise transmission quality can be regarded as high for this implementation. Double talk performance and echo suppression expressed by the parameter TCL_W under nominal test conditions indicate also a high quality.

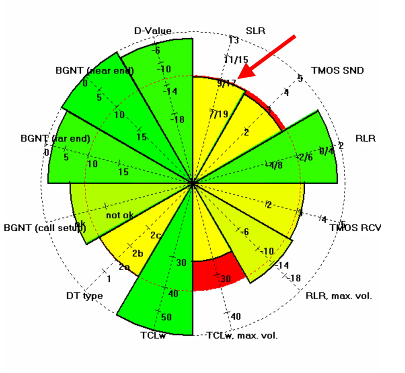

Fig. 6: Commercially available hands-free kit

The relatively high SLR of 17 dB (see red arrow) leads to a relatively low but acceptable signal to noise ratio in the GSM coded signal even without noise in the vehicle. The TMOS of 2.8 in sending direction indicates this. This is a good example to point out a possible interaction between different parameters like the SLR and TMOS.

The TOSQA2001 algorithm simulates a subjective listening test according to ITU-T Recommendation P.800 [6]. The tests described in here are based on the assumption of equalized playback levels for all listening examples. A user would probably also tend to increase the playback volume of his phone. In order to reproduce this, models like TOSQA2001 also equalize the level of the transmitted signal. A low level speech sequence (resulting after a transmission over a system with a high SLR) is therefore amplified. This leads to a higher idle noise level. Consequently the TMOS rating degrades, even the speech itself is undistorted, mainly because noise gets audible and disturbing after this amplification.

On the other hand a high SLR is often implemented as a trade-off between a still acceptable uplink speech level and a high acoustical attenuation of the echo signal. The example shown in figure 6 represents an implementation which is optimized under these circumstances. The only point of critic for this implementation is the occurrence of temporal echo at maximum volume. The playback volume is high, thus providing additional room for optimization.

5. Conclusions

The comprehensive visualization of complex speech quality parameters in one figure provides an efficient method to describe the overall quality of a complex telecommunication device. The parameters selection according to the VDA specification for hands-free telephones covers all conversational aspects thus providing this overview. Moreover, assuming a detailed background knowledge and experience on the interaction of parameters it also provides sufficient information for the decision of next optimization steps.

6. References

[1] ITU-T P.505, One-view visualization of speech quality measurement results

[2] VDA Specification for Car Hands-free Terminals, Version 1.5, Dec. 2004

[3] EG 201 377-1: Speech Processing, Transmission and Quality Aspects (STQ); specification and measurement of speech transmission quality; part 1: Introduction to objective comparison measurement methods for one-way speech quality across networks

[4] ITU-T G.122, Influence of National Systems on Stability and Talker Echo in International Connections

[5] ITU-T P.58, Head and Torso Simulators for Telephonometry

[6] ITU-T P.800, Methods for Subjective Determination of Transmission Quality