Mobile Phone Performance Evaluation in Background Noise

H.W. Gierlich, Frank Kettler, Silvia Poschen, Stephane Kengne
HEAD acoustics GmbH ,Herzogenrath, Germany

Introduction

The performance of mobile terminals (cell phones, mobile hands-free telephones) in the presence of background noise is crucial because this represents the typical use case today. This also motivates the development and conduction of appropriate speech quality tests.

Test methods for performance evaluation in background noise

The D-value calculated from DELSM according to ITU-T Recommendation G.111 [1] like the ANR represents a one-dimensional score indicating the sensitivity differences of a terminal for speech and background noise. For both, the transmission of the signal is separately determined form the transmission of the noise, any enhancement technique which is applied to improve the speech quality while the noise is present is not taken into account appropriately.

The D-value is defined for narrowband scenarios. In order to provide similar results for wideband devices as well the unweighted signal to noise ratio SNR' calculated from the wideband speech level S and the wideband noise level N is used for comparison instead. SNR' is calculated as follows:

$$SNR' = \frac{10 * \log 10\left(\sum_{f_k 1}^{f_k 2} \frac{S(f_k)^2}{(1Pa)^2}\right)}{10 * \log 10\left(\sum_{f_k 1}^{f_k 2} \frac{N(f_k)^2}{(1Pa)^2}\right)} dB$$

$S(f_k)$ – Spectrum of transmitted speech signal at the terminal output (rms);

$N(f_k)$ – Spectrum of transmitted noise at the terminal output (rms);

$f_k 1 = 100$ Hz; $f_k 2 = 8$ kHz

The SNRI calculation according to G.160 [2] is purely focusing on SNR improvement during periods where speech is present simultaneously with the background noise. It is suitable to determine the SNR improvement achieved by a noise cancellation algorithm. However, it is just signal based and does not take into account other parameters which might influence the speech or noise transmission quality. Any auditory impact on the quality as perceived by the user is not considered. The SNRI is determined as follows:

$$SNRI = SNRout - SNRin$$

$$SNRout = 10 \cdot \text{Log}\left\{ \max\left[\xi, \frac{10^{\frac{1}{K_{sph}} \sum_{l=1}^{K_{sph}} \text{Log}\left(\xi + \sum_n y^2(l,n)\right)}}{10^{\frac{1}{K_{nse}} \sum_{m=1}^{K_{nse}} \text{Log}\left(\xi + \sum_p y^2(m,p)\right)}} - 1 \right] \right\}$$

$$SNRin = 10 \cdot \text{Log}\left\{ \max\left[\xi, \frac{10^{\frac{1}{K_{sph}} \sum_{l=1}^{K_{sph}} \text{Log}\left(\xi + \sum_n d^2(l,n)\right)}}{10^{\frac{1}{K_{nse}} \sum_{m=1}^{K_{nse}} \text{Log}\left(\xi + \sum_p d^2(m,p)\right)}} - 1 \right] \right\}$$

S	Clean speech
D	Input / unprocessed signal
Y	Output / processed signal
K_{sph}, K_{spm}, K_{spl}	Number of speech-frames with high (h), mid (m) and low (l) speech power
K_{nse}	Number of all frames without speech activity / noise
ξ	Limiting constant for SNR; G.160 old: 10e-5; new: 0.0631

ETSI EG 202 396-3 (3QUEST) [3], [4] describes an algorithm which can be used to determine quality parameters related to this situation. This algorithm calculates three quality scores, the N-MOS representing the quality of transmitted noise, the S-MOS as a quality measure for the transmitted speech and the G-MOS, the general impression combining quality of noise and speech. The analysis is based on a large number of subjective tests and especially covers latest noise reduction algorithms and speech coders.

Background noises

Tests were carried out with four different realistic background noise scenarios. The café noise is played back with a level of approximately 64 dB(A) but shows a quite instationary characteristics. The car noise is applied with a level of 70 dB(A), it is mostly stationary. The level of train station noise is approximately 70 dB(A) but slowly fluctuates over time. The road noise condition represents the most critical background noise scenario among the four test situations. It is applied with a high level of approximately 75 dB(A) and shows high fluctuations vs. time. All background noises were taken from the ETSI EG 202 396-1 [5], the noise playback setup was according to [5].

Analysis of the results

In all figures the results of D-value analysis (narrowband) the SNR' (wideband) are given by the yellow dots. The relevant scale is on the right hand y-axis. The four background noise scenarios can be found on the x-axis. The MOS scores calculated from ETSI EG 202 396-3 (3QUEST) are color coded and given on the left hand y-axis (blue: S-MOS, red: N-MOS, green: G-MOS).

The measurement results for device no. 1 (2-microphone solution) indicate a very strong noise reduction. The D-value is extremely high for the three mostly stationary background noises (car, the train station, the road noise). For the non-stationary café noise a D-value of approximately -3 dB indicates a less aggressive noise cancellation in this scenario. The SNRI also shows very high values (strong noise suppression), it value also indicates a less good performance for the instationary café noise. The N-MOS from 3QUEST indicates a very high quality for all transmitted background noises neither seen in the D-value nor in the SNRI. The lower D-value and the somewhat lower SNRI cannot be confirmed. The quality of the transmitted noise is high as indicated by the high N-MOS of approximately 4.5 also for the café noise scenario. The speech signal is significantly impaired by strong artifacts for all four background noise scenarios as indicated by the S-MOS (blue) and also seen in the general quality impression expressed by the G-MOS (green). Especially for stationary noises the cancellation is very aggressive impairing the speech quality.

The same analysis is carried out for a different cell phone also providing a two microphone solution (**figure 2**). The D-value is nearly constant for all background noises. The SNRI does not show as much improvement as for device 1 but is high for all background noises. It shows the capability of high noise reduction for 2 microphone solutions. The quality perceived subjectively however can be measured by 3QUEST. Noise quality (N-MOS, red) and speech quality (S-MOS, blue) are relatively high and decrease slightly with higher background noises. Consequently the global quality impression is slightly decreasing as shown by the G-MOS.

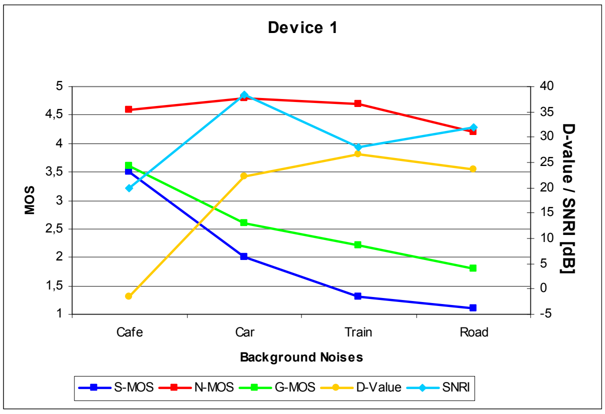

Fig. 1: Narrowband device 1, D-value/SNRI vs. S-, N- and G-MOS (left)

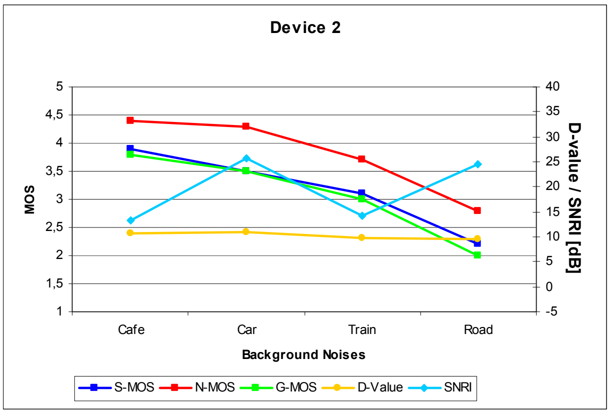

Fig. 2: Narrowband device 2, D-value/SNRI vs. S-, N- and G-MOS (left)

The analysis results of 2 wideband phones are given in **figure 3** and 4. Both provide in general a good quality. The performance measured with 3QUEST show that both phones are mostly comparable. Comparing the SNR' results to these quality scores from the ETSI model it can be stated that the differences are very high for the two devices although the quality of the two devices is quite comparable. This again demonstrates that the calculated SNR' does not reliably correlate to quality estimation. The SNRI results of the two phones are more comparable between each other and indicate, that the ability of noise reduction of the two phones is comparable. As in Narrowband, a higher SNRI can be achieved for more stationary noises than for dynamic background noises. But as the SNR' the SNRI does not correlate to the N-MOS, S-MOS and G-MOS.

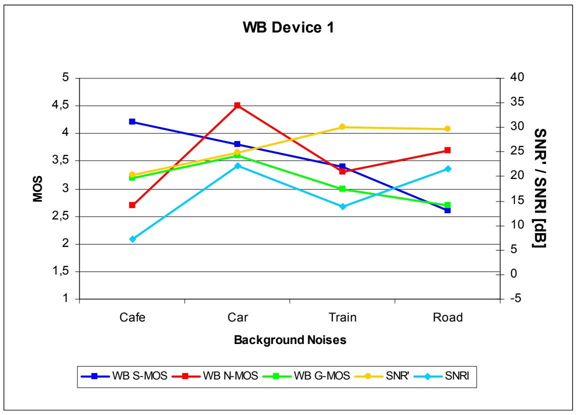

Fig. 5: Wideband device 1, SNR'/SNRI vs. S-, N- and G-MOS (left)

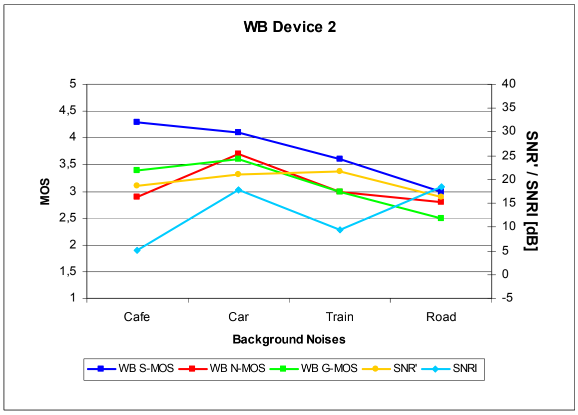

Fig. 6: Wideband device 2, SNR'/SNRI vs. S-, N- and G-MOS (left)

Conclusions

The most appropriate method to analyze the speech quality in the presence of background noise in uplink for modern mobile phones is the model 3QUEST as described ETSI EG 202 396-3. Neither D-Value nor SNR' or SNRI cover all impacts on the quality as finally perceived by the user.

References

[1] ITU-T Recommendation G.111: Loudness ratings (LRs) in an international connection (03/93)

[2] ITU-T Recommendation G.160: Voice Enhancement Devices (06/08)

[3] ETSI EG 202 396-3: Speech Quality performance in the presence of background noise; Part 3: Background noise transmission - Objective test methods (11/08)

[4] Reimes, J.; Gierlich, H.W.; Ketler, F.; Poschen, S.; Lepage, M.: The Relative Approach Algorithm and Its Applications in New Perceptual Models for Noisy Speech and Echo Performance, ACTA ACUSTICA 2010, accepted

[5] ETSI EG 202 396-1 Speech Quality performance in the presence of background noise Part 1: Background noise simulation technique and background noise database (06/08)