Influence of Driving Noise Characteristics on Speech Quality in the Presence of Background Noise in Vehicles

Udo Müsch, Silvia Poschen, Frank Kettler
HEAD acoustics GmbH, 52134 Herzogenrath
Udo.Muesch@head-acoustics.de; Silvia.Poschen@head-acoustics.de; Frank.Kettler@head-acoustics.de

Introduction

Since several years objective tests regarding the speech quality in the presence of background noise are part of the ITU-T Recommendation P.1100 and P.1110 standards [1] [2]. A considerable amount of hands-free implementations that are currently in the market have been tested recently. This contribution presents results that existing hands-free terminals can achieve and compares these results to the limits given in ITU-T Recommendations P.1100 and P.1110. Furthermore, the influence of the driving noise characteristics on speech quality is investigated.

Instrumental assessment of speech quality in the presence of background noise

An objective method for measuring speech quality in the presence of background noise has been introduced in ETSI EG 202 396-3 [3]. The fundamental system is presented in **figure 1**.

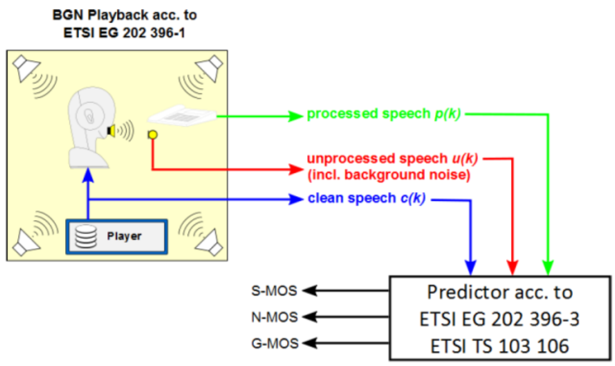

Figure 1: Fundamental setup for objective testing of speech quality in the presence of background noise.

The practical realization is shown in **figure 2**. A Head And Torso Simulator (HATS) according to ITU-T Recommendation P.58 [4] is placed on the driver's seat inside the vehicle. The vehicle is equipped with a background noise simulation system according to ETSI EG 202 396-1 [5], which plays back the driving noise. Speech is played back by the artificial mouth of the HATS. The speech signal is transmitted by the hands-free terminal (HFT) to the measurement system via a simulated mobile network.

In order to objectively measure the speech quality three different signals are required: First is the so called "processed speech" signal which contains the signal transmitted by the device under test (DUT) including all HFT signal processing. Also included in this signal is the

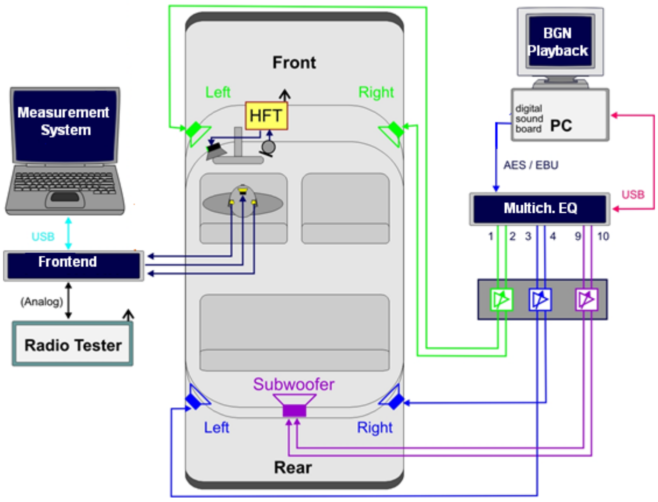

Figure 2: Principle Test setup for objective testing of speech quality in the presence of background noise.

coder of the mobile network and the estimated frequency characteristic of the far end terminal. The second input signal used for calculation is the so called "unprocessed speech" signal which captures the signal inside the vehicle with a reference microphone close to the HFT microphone. This signal contains speech and background noise, in this case the driving noise of the car. As a third signal the "clean speech" signal is used. This is the raw signal which is played back by the artificial mouth of the HATS.

The output values of the calculation are three Mean Opinion Scores called S-MOS, N-MOS and G-MOS. The S-MOS is a measure of the distortion of the speech introduced by the signal processing. The N-MOS describes the intrusiveness of the transmitted noise. Finally, the G-MOS is a measure for the overall quality. All these MOS values refer to a five point listening quality/ACR scale according to ITU-T Recommendation P.800 [6].

It has to be noted that the ETSI EG 202 396-3 method is validated for hands-free communication in cars. This is in contrast to the similar ETSI TS 103 106 [7]. The given results have been calculated with the current version 1.4.1 of the standard.

Current noise scenarios and limits

Table 1 shows the limits given in ITU-T Recommendation P.1100/P.1110 (01/15). It should be noted that the recommendations propose the same limits for both bandwidths (narrowband and wideband). Limits are given for two different speed ranges.

Table 1: ITU-T P.1100/P.1110 (01/15) 11.13.3.1
Requirements for N-MOS, S-MOS and G-MOS

Background noises for driving speed ≤ 80 km/h	Background noises for driving speed > 80 km/h and ≤ 130 km/h
N-MOS ≥ 3.0	N-MOS ≥ 2.5
S-MOS ≥ 3.0	S-MOS ≥ 2.5
G-MOS ≥ 3.0	G-MOS ≥ 2.5

In addition, Annex D of the recommendations defines four different noise scenarios which are differentiated not only by the driving speed but also by the setting of the heating, ventilation and air conditioning fan (**table 2**). The user scenarios 2 and 4 require that the fan is adjusted in a way that the fan noise is 6 dB above the driving noise. This means in practical application that the fan is almost always set to its maximum value.

Table 2: ITU-T P.1100/P.1110 (01/2015) Annex D
Standard set of user scenarios

User scenario	Description	Vehicle speed	HVAC settings
1	Stationary vehicle with low HVAC noise	0 km/h (at idle)	FAN=Lowest setting
2	City driving with high HVAC noise	60 km/h (37 mph)	FAN= Setting closest to 6 dB(A) above driving noise with FAN = lowest setting; AIRFLOW=Directed to windows
3	Highway driving with low HVAC noise	120 km/h (75 mph)	FAN=Lowest setting
4	Highway driving with high HVAC noise	120 km/h (75 mph)	FAN=Setting closest to 6 dB(A) above driving noise as in condition 3; AIRFLOW=Directed to windows

Results for current implementations

44 hands-free systems that are currently in the market have been tested. **Figures 3** and **4** show an overview of the results as box plots. The lower border of the boxes represents the 75% percentile, the upper border the 25% percentile. The median is indicated as a red or green line. **Figure 3** shows the results for speeds ≤ 80km/h. The scores for speeds > 80 km/h are given in **figure 4**. The number of HFTs used in each condition is given in the headline of the diagrams.

The plots for speeds ≤ 80 km/h indicate a large deviation of the achieved results for narrowband (**figure 3**, left plot) and to a lesser extent also for wideband (**figure 3**, right). In narrowband, the median (red line) of the results is below the proposed limits for S-MOS and G-MOS. For speeds > 80 km/h (**figure 4**) the deviation of the results is much narrower. However, it can be also observed that at least in narrowband most of the tested devices do not fulfill the requirements for S-MOS and G-MOS.

To further investigate the root cause of these deviations, the results are presented for all four user scenarios in **figure 5** for narrowband and **figure 6** for wideband. In addition to the S-MOS, N-MOS and G-MOS values also the level of the driving noise is given in the box plots.

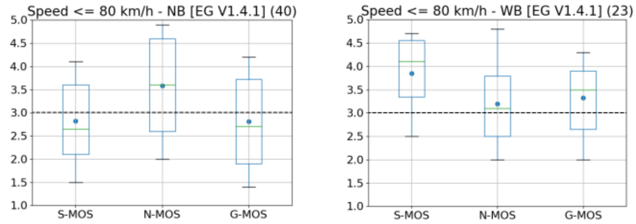

Figure 3: Test results for speeds ≤ 80 km/h, narrowband (left) and wideband (right)

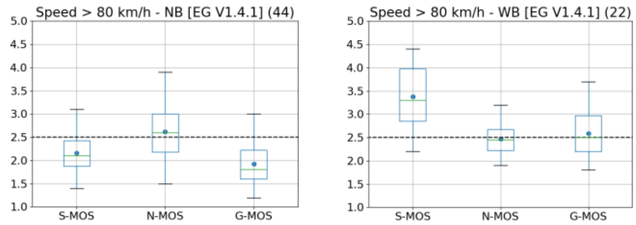

Figure 4: Test results for speeds > 80 km/h, narrowband (left) and wideband (right)

Comparing the upper two plots in **figure 5** shows a completely different performance for scenario 1 (engine idle at 0 km/h, left plot) and scenario 2 (60 km/h, fan set to maximum, right plot); these two scenarios describe totally different noise environments. The difference of the average noise levels for scenarios 1 and 2 is approximately 20 dB. Accordingly the limits for scenario 1 are met for almost all devices under test, whereas for scenario 2 all DUT (apart from one outlier) do not fulfill the given requirements.

Looking at the lower two plots in **figure 5** shows that for

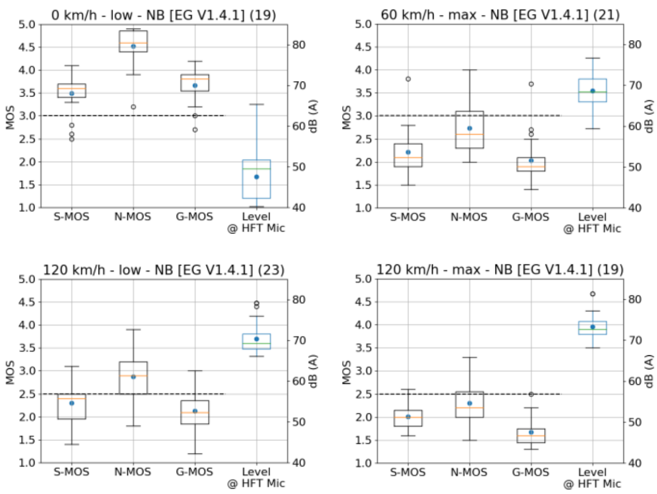

Figure 5: Test results for hands-free terminals in the market, narrowband, User Scenario 1-4

scenario 3 narrowband only the N-MOS is fulfilled for nearly all DUTs. For scenario 4 most devices do not meet any of the required limits.

The results for wideband (**figure 6**) show similar tendencies. However, the achieved MOS values are generally higher due to the better signal-to-noise ratio for frequencies higher than 4 kHz compared to narrowband.

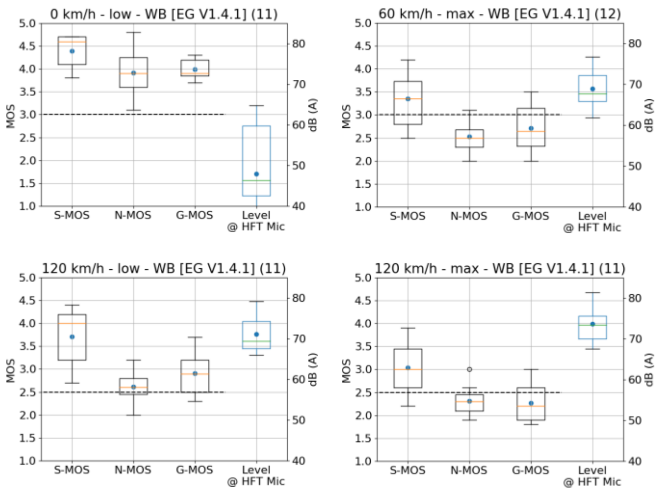

Figure 6: Test results for hands-free terminals in the market, wideband, User Scenario 1-4

To investigate the influence of the background noise level on the overall quality, the G-MOS values have been plotted versus background noise level in **figure 7** for narrowband and wideband.

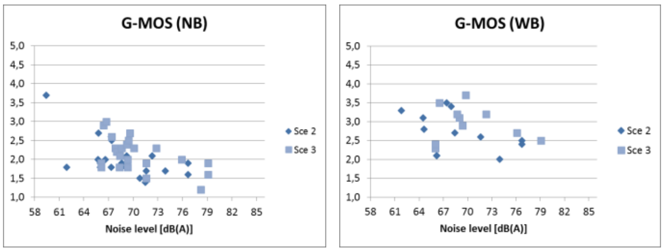

Figure 7: G-MOS versus background noise level for narrowband (left) and wideband (right), scenario 2 (60 km/h, fan high) and scenario 3 (120 km/h, fan low).

For narrowband (**figure 7**, left plot) a weak correlation can be observed between the noise level and the G-MOS, i.e. the lower the noise level, the higher G-MOS values can be expected. In wideband (**figure 7**, right plot) good G-MOS scores can be achieved also for higher noise levels. More remarkable is the big overlap between the noise levels of scenario 2 (60 km/h, fan high) and scenario 3 (120 km/h, fan low). This shows that adjusting the fan to the required level leads to an almost identical level for scenario 2 compared to scenario 3. This is important because the proposed limits for N-MOS, G-MOS and S-MOS according to ITU-T Recommendation P.1100 and P.1110 are significantly different for these two scenarios.

Detailed view on noise levels and spectral characteristics

Figures 8 and **9** show comparisons of noise spectra measured with a reference microphone close to the DUT microphone. The A-weighted levels and the corresponding narrowband N-MOS, S-MOS and G-MOS are given in **table 3**.

Table 3: Comparison of driving noise level and G-MOS, N-MOS and S-MOS (narrowband) for the four different test cases shown in figure 8 and 9

	Car 1, scenario 3	Car 2, scenario 3	Car 3, scenario 2	Car 3, scenario 3
Level	67 dB(A)	79 dB(A)	69 dB(A)	70 dB(A)
N-MOS	3.3	2.5	2.5	2.7
S-MOS	2.3	1.6	2.4	2.6
G-MOS	2.3	1.7	2.1	2.2

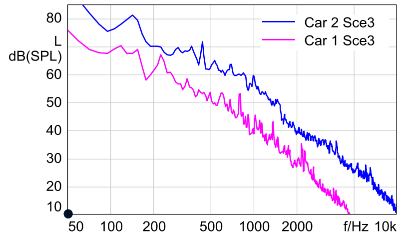

Figure 8: Comparison of driving noise spectra of two different cars for scenario 3 (120 km/h, fan low)

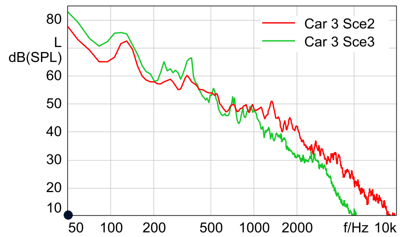

Figure 9: Comparison of driving noise spectra of scenario 2 (60 km/h, fan high) and scenario 3 (120 km/h, fan low) in the same vehicle

Figure 8 shows exemplarily the average noise spectrum of two different cars at 120 km/h with low fan setting (scenario 3). Both DUTs do not fulfil the requirement of ≥ 3.0 for narrowband. However, the G-MOS, N-MOS and S-MOS values for car 2 (blue curve) are significantly lower due to the 12 dB higher noise level.

As a different example, **figure 9** shows the noise spectra for two different noise scenarios in the same car. In this case noise scenario 2 (60 km/h, fan high) and scenario 3 (120 km/h, fan low) are compared. It can be observed that even though the overall A-weighted level is almost identical (69 dB(A) vs. 70 dB(A), see **table 3**), the noise of scenario 2

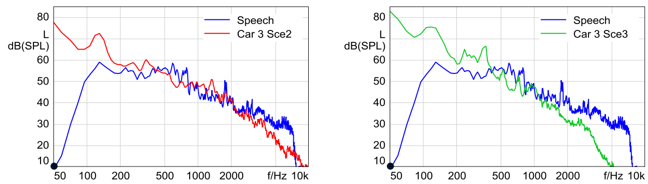

Figure 10: Comparison of driving noise and speech spectra for scenario 2 (left) and scenario 3 (right)

shows a significantly higher level for frequencies above 1 kHz. Comparing these noise spectra with the speech spectrum present at the hands-free microphone shows that the signal-to-noise ratio in the higher frequencies is significantly worse for the noise scenario 2 (**figure 10**).

In the narrowband frequency range from 300 Hz to 3.4 kHz the signal-to-noise ratio for scenario 2 is approximately 0 dB, whereas for scenario 3 a positive SNR can be achieved, which helps to achieve a good performance of the noise reduction of the DUT.

Conclusion and current activities of ITU-T

As shown before many of the existing implementations do not fulfil the limits in ITU-T Recommendations P.1100 and P.1110 (01/15). This is due to the fact that the limits were not based on data of existing implementations. The motivation was to define performance limits for the transmitted speech and noise independent of the vehicle characteristics. However, the existing measurements show a wide range of noise levels for the given noise scenarios. These different noise levels are related to the vehicle type and not the HFT implementation itself, but have an effect on the MOS results.

A further influence is the requirement for the fan noise which can be regarded as unrealistically high. The spectral overlap with a speech spectrum leads to a degraded performance of the algorithms regarding speech quality. As a result of these findings the ITU-T Recommendations have been updated in order to allow the fulfilment of requirements. Several steps have been undertaken to achieve more realistic conditions and limits.

- New limits were generated for driving speed > 80 km/h and ≤ 130 km/h. These limits are based on the median of MOS of the available test data and replace the previous limits which were unrealistically high for narrowband. The limits reflect the different performance in narrowband and wideband.
- Separate requirements have been introduced for user scenario 1 (0 km/h, engine idle). These limits are more challenging than before.
- In Annex D, the requirements of the fan noise level for user scenarios 2 and 4 have been relaxed. It is now not any longer required to have a fan noise 6 dB higher than the driving noise. Instead the fan should be adjusted to its medium setting.

Tables 4, 5 and **6** sum up the changes in the requirements proposed. These changes are the result of the ITU-T meeting from January 2017 and are in force since March, 2017 [8][9].

Table 4: ITU-T P.1100 (01/2017) narrowband 11.13.3.1 Requirements for N-MOS, S-MOS and G-MOS

Background noises in quiet, low fan (see Annex D)	Background noises for driving speed ≤ 80 km/h	Background noises for driving speed > 80 km/h and ≤ 130 km/h
N-MOS ≥ 4.6	N-MOS ≥ 3.0	N-MOS ≥ 2.9
S-MOS ≥ 3.6	S-MOS ≥ 3.0	S-MOS ≥ 2.4
G-MOS ≥ 3.8	G-MOS ≥ 3.0	G-MOS ≥ 2.1

Table 5: ITU-T P.1110 (03/2017) wideband 11.13.3.1 Requirements for N-MOS, S-MOS and G-MOS

Background noises in quiet, low fan (see Annex D)	Background noises for driving speed ≤ 80 km/h	Background noises for driving speed > 80 km/h and ≤ 130 km/h
N-MOS ≥ 3.9	N-MOS ≥ 3.0	N-MOS ≥ 2.6
S-MOS ≥ 4.6	S-MOS ≥ 3.0	S-MOS ≥ 3.0
G-MOS ≥ 3.9	G-MOS ≥ 3.0	G-MOS ≥ 2.9

Table 6: ITU-T P.1100/P.1110 (03/2017) Annex D Standard set of user scenarios

User scenario	Description	Vehicle speed	HVAC settings
1	Stationary vehicle with low HVAC noise	0 km/h (at idle)	FAN=Lowest setting
2	City driving with high HVAC noise	60 km/h (37 mph)	FAN= Medium setting; AIRFLOW=Directed to windows
3	Highway driving with low HVAC noise	120 km/h (75 mph)	FAN=Lowest setting
4	Highway driving with high HVAC noise	120 km/h (75 mph)	FAN= Medium setting; AIRFLOW=Directed to windows

References

[1] ITU-T Recommendation P.1100. (01/2015). Narrow-band hands-free communication in motor vehicles.

[2] ITU-T Recommendation P.1110. (01/2015). Wideband hands-free communication in motor vehicles.

[3] ETSI EG 202 396-3 v1.4.1. (06/2014). Speech Quality performance in the presence of background noise Part 3: Background noise transmission - Objective test methods.

[4] ITU-T Recommendation P.58. (05/2013). Head and torso simulator for telephonometry, May 2013. International Telecommunication Union.

[5] ETSI EG 202 396-1. (2008). Speech Processing, Transmission and Quality Aspects (STQ); Speech quality performance in the presence of background noise; Part 1: Background noise simulation technique and background noise database. V1.2.2. ETSI.

[6] ITU-T Recommendation P.800. (08/1996). Methods for subjective Determination of Transmission Quality. International Telecommunication Union.

[7] ETSI TS 103 106 v1.4.1. (11/2016). Speech Quality performance in the presence of background noise: Background noise transmission for mobile terminals - Objective test methods.

[8] ITU-T Recommendation P.1100. (03/2017). Narrow-band hands-free communication in motor vehicles.

[9] ITU-T Recommendation P.1110. (03/2017). Wideband hands-free communication in motor vehicles.