About the Measurement of the Hands-Free Transmission Quality

– An Experience Report -

Holger Pastillé, Martin Herrenkind
IAV GmbH, 38518 Gifhorn, Germany; Email: holger.pastille@iav.de

Introduction

Engineers prefer objective and reproducible methods for tests. For the evaluation of speech pattern recognition in cars and the derivation of a consistent dimension, for example, the number of the detected speech samples is countable. Using this method in the laboratory, a comparison of different machines is possible.
Testing of hands-free transmission quality will look differentl, see [1]. The following properties (samples) are objectively and reproducibly measurable (after VDA):

- delay
- loudness
- frequency response
- echo (and damping)
- idle noise et cetera.

However those items do not show reliable results of the transmission quality. A system can pass the VDA test, but there is no conclusion regarding the understanding of the speech quality. Furthermore these measurements can only be realized in stationary automobiles.
One other possibility for receiving an objective value is using tools and modeling a physical hearing analysis. Differences and similarities are describable with a comparison of the test and reference data. Cognition models are herein partly included (e.g. PESQ / TOSQA). Note that these tools are only applicable while the car is not in motion (elimination of noise).
The cognitive human factor is essential. For a final evaluation the statement of a jury is necessary. The results of the auditory tests are transformed into a MOS-value (Mean Opinion Score).

The Evaluation

This part shows testing different hands-free transmission devices using given criteria. It is necessary and more comfortable to judge the speech quality outside the car. The data are anonymised. The first measurement was taken from the car to a far end (FE-Line), see Fig. 1.

The second measurement was taken from one car to another car which had background noise, see Fig. 2.

Car Number 1 is a sedan. The following three systems were tested to a far end: two low-price systems (A and B) and on high price system (C). The second measurement involved the transmission from the sedan to the far end with background noise from the second car $n_2(t)$. In the car2car set-up only the results from

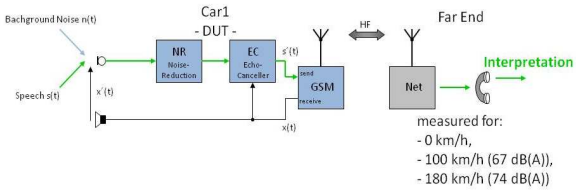

Figure 1: Measurement from car to a far end

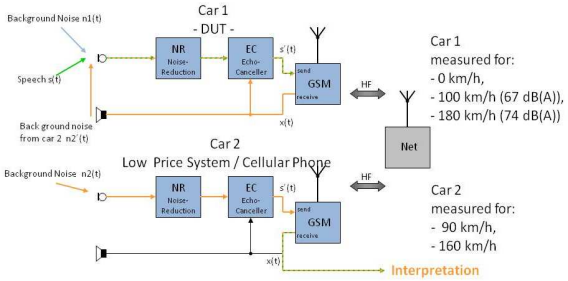

Figure 2: Measurement from car to car

the low-price systems were of interest.
It is always interesting to see the time signals, as shown in Fig. 3. The difference between FE-Line and car2car is

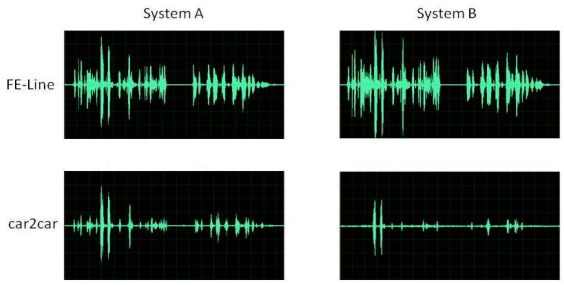

Figure 3: Time signals for FE-Line and car2car at v=0

obvious but the difference between the systems (system "A" compared to system "B") is not. It is impossible to determine if anybody understood the spoken words (system "B" with car2car).
A more common way is a selection of quality parameters as shown below:

- perceived total quality
- speech intelligibility
- background noise

- brilliance (vs. damped, dull)
- hearing effort.

The choice of these parameters and the visualization are based on [2]. According to the ITU Recommendation P.800 [3] the scale of each parameter is defined with up to 7 points, for example the parameter quality has a range from "much worse" via "about the same" to "much better". For practical reasons five parameters where chosen and visualized in a multispeed figure. Only at "180 km/h" is an additional parameter to be found: perceived speech rhythm. The different acoustical conditions inside the car lead to a changed cognitive situation. At lower velocities no change of the speech rhythm is observable.
The result is a pattern that can be easily understood and which makes it possible to compare the behavior of the different systems.

Results

As expected, the communication from car to the far end will give the best results.

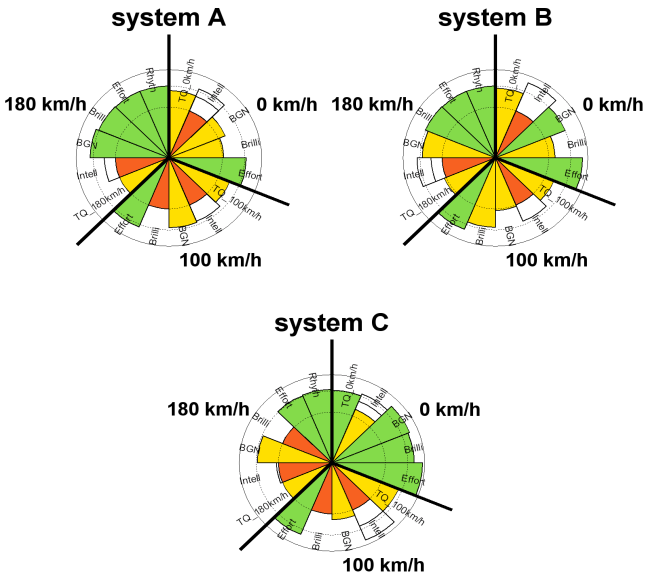

Figure 4: Results for the FE-Line; system A and B above, high-price system C below

The aim of this study was to check the efficiency of the new system "B". Therefore, the additional test from car to car was performed with the systems "A" and "B" only. Figure 5 shows a quick overview of quality and properties of a hands-free system.
This method can be used easily for other benchmark studies.

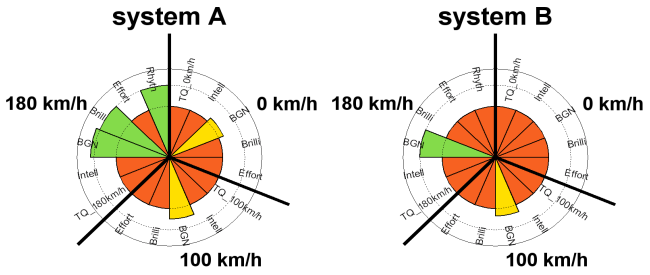

Figure 5: Results for the Car2Car-communication of the systems A and B

Discussion of a single value / single dimension

A difference between system "A" and "B" in Fig. 4 is not easy to identify. Therefore it may be useful to reduce the values to a single dimension or a single value. An example could be the transmission loss of a wall. Like many testing magazines it is our suggestion to weight the parameters as follows:

- perceived total quality with 40%
- speech intelligibility with 25%
- background noise with 10%
- brilliance with 10%
- hearing effort with 15%

After including another parameter the so called "perceived speech rhythm" the above scheme will change. The parameter will be weighted with 5 %, the speech intelligibly will be lowered to 20 % and the rest will remain the same. In this way of weighting the parameters a mirroring of the cognitive effort is given.
The single value will be based on a scale from 1 to 7; best value is the "1". Parameters on a 3 point or a 5 point scale are converted to the 7 point scale.
This single value will be demonstrated by another test of the hands-free transmission quality. A benchmark of five different cars with the top-level hands-free devices is shown in Fig. 6.
This will lead to the following single values:

velocity	Single Value				
	Car A	Car B	Car C	Car D	Car E
0 km/h	2.6	3.9	3.0	2.7	2.4
100 km/h	3.6	3.9	4.3	3.2	3.3
180 km/h	4.1	4.3	4.5	3.4	3.9

Table 1: Single values of top-level hands-free devices

The single value should be used only for one speed and not for all speeds shown above because otherwise an additional weighting would be necessary but should be avoided.

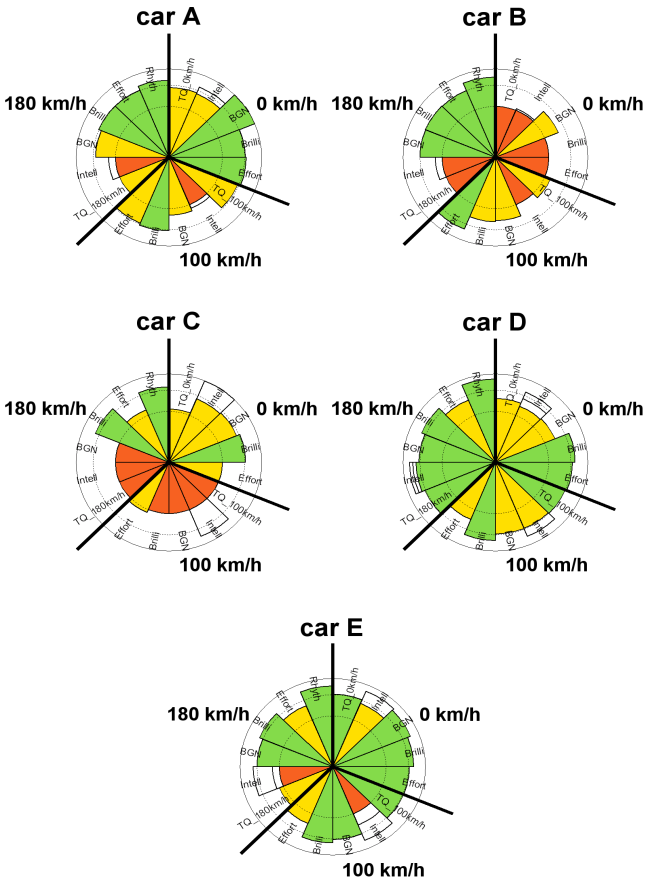

Figure 6: Benchmark of top-level hands-free devices

Summary

Objective tests can not give a satisfactory result of the behavior of a hand-free transmission quality. Releases (resulted through measurements with VDA rules) could gives a different impression of the sound quality.
At this point a decision of the transmission behavior can only be given by a jury assessment. Note that is not necessary to have a trained jury in order to include the so-called cognitive and perceptive factor.
For a swift overview the results are shown in a pie chart. A difference is not noticeable if the properties are very close to each other. The previously presented suggestion of a single value/single dimension for every speed could help to transform the results into a form suitable for higher management.

The authors would like to thank Bobby Haferburg, Anika Baumgart, Ingo Kremer and Mark Behrens.

References

[1] Klasmeyer, G.; Herrenkind, M.; Kreft, K. : *Evaluierung der Übertragungsqualität von Freisprechern im Kfz.* Fortschritte in der Akustik DAGA, Braunschweig 2006

[2] Klasmeyer, G.: *Quality Assement of Speech Enhancement Systems in Vehicles* 2nd ISCA/DEGA Tutorial and Research Workshop on Perceptual Quality of Systems, Berlin 2006

[3] ITU-T P.800: *Methods for subjective Determination of Transmission Quality.* ITU Recommendation P.800, 08/1996