Instrumental Evaluation of In-Car Communication Systems
Anne Theiß, Gerhard Schmidt, Jochen Withopf, Christian Lüke
Digital Signal Processing and System Theory, CAU Kiel, Germany
Email: {ath,gus,jow,cl}@tf.uni-kiel.de
Web: www.dss.tf.uni-kiel.de

Abstract
The aim of an in-car communication system is to improve the communication inside a vehicle. This communication is impaired, due to the high background noise occurring at higher velocities and can be improved by recording, processing, and playing back the speech signal of the talking passenger using loudspeakers. Due to this closed electro-acoustic loop further issues arise, requiring more sophisticated algorithms, e.g., feedback control. After the development of an ICC system, the quality of the system is of great interest. Not only to analyze and to improve a single ICC system an evaluation is beneficial, but also to compare different ones. However, the term communication quality is not easy to fit into one instrumental measure, since quality depends on a variety of criteria. This contribution gives some first instrumental measures classifying the quality of an ICC system and a first approach to combine these to obtain an overall evaluation strategy.

1 Introduction
The communication within a vehicle might be impaired due to high background noise and the unusual communication situation. Unusual, because the passengers of the front seats are sitting with their backs towards the passengers on the rear seats. To improve this situation, the passengers start to raise their voices or to turn towards their communication partners. A solution is the implementation of an In-Car Communication system (ICC system) inside the vehicle [1-3]. Such a system records the speech signal of the talking passenger by means of microphones, enhances the speech signal, and plays it back via the loudspeakers integrated in the vehicle. The enhancement of the microphone signal includes various algorithmic components such as noise suppression, noise-dependent gain control, microphone selection, etc. [2]. Besides these basic components, the issue of the closed loop has to be solved and all algorithmic components need to be realized with low-delay processing structures. All in all, the ICC system requires a set of sophisticated algorithmic components in order to provide a good quality.

After the development of such a system, the question about the quality arises. In [3,4] some first evaluation results have been published. The communication quality is composed of various different factors (e.g. quality of transmission channel, expectations of communication partners, conversational effort, etc.). On top of that, all factors are of different importance to different persons as communication quality has an individual significance. Hence, communication quality is a complex concept, which requires a well-defined evaluation strategy. The proposed evaluation strategy is split into three evaluation groups consisting of several measures.

- The first group determines the characteristics of the vehicle, as some support the processing of the ICC system and others not. For example, the distances between the loudspeakers and the microphones are of interest as they classify the feedback paths. Another example is the noise level, occurring inside the vehicle, which the ICC system has to overcome.
- The second group evaluates the characteristics of the ICC system. An evaluation measure belonging to this group is, e.g., the delay introduced by the ICC system. Three measures out of this group will be introduced within this contribution.
- The third evaluation group contains measures which are deriving the communication quality, for example the signal-to-noise ratio at the ears of the listening passenger or the logarithmic spectral distance.

The first part of this contribution deals with the evaluation environment. Section 3 and 4 describe three measures and their quality classification. Section 5 gives an initial approach of the overall quality and the conclusions are drawn in the last one.

2 Evaluation Environment
The evaluation of an ICC system requires a reproducible test environment. Fig. 1 shows a basic structure of such an environment and Tab. 1 the used indices.

![Figure 1: Basic structure of the ICC system evaluation.](image)

<table>
<thead>
<tr>
<th>Abbr.</th>
<th>Meaning</th>
</tr>
</thead>
<tbody>
<tr>
<td>N</td>
<td>Noise simulation paths</td>
</tr>
<tr>
<td>L</td>
<td>Loudspeaker</td>
</tr>
<tr>
<td>M</td>
<td>Microphone</td>
</tr>
<tr>
<td>E</td>
<td>Ear microphone (artificial listener)</td>
</tr>
<tr>
<td>T</td>
<td>Torso loudspeaker (artificial talker)</td>
</tr>
</tbody>
</table>

Table 1: List of used indices.

The evaluation environment consists of three parts, the ICC system, which is considered as a black box during the evaluation, the noise simulation and the ICC system evaluation.

The noise scenarios that are necessary to excite the ICC system accordingly, are implemented by means of an acoustic ambiance simulation. This simulation is capable...
of creating the same power spectral densities at the ears of the passengers as in a reference scenario (recordings during a real drive) [5]. An artificial talking and a listening passenger are placed inside the passenger compartment to simulate the communication participants, see Fig. 1. In order to obtain a concise system to develop evaluation methods, some simplifications are made. The considered ICC system only supports the direction front to rear, as this direction is the important one [6]. In addition, only one ear signal of the listening passenger is considered, since all concepts can be directly transferred to the other channel. The test environment can be simplified to the signal flow graph of Fig. 2.

![Signal flow graph of the evaluation structure.](image)

Figure 2: Signal flow graph of the evaluation structure.

All measures explained here are so-called reference-based instrumental test methods. The reference signal is reproduced by an artificial mouth positioned at, for example, the driver position. The utilized reference signals may vary in their structure, due to the specified evaluation scenario to achieve signals including the Lombard effect and speech rate. The utilized reference signals are sampled with 44.1 kHz. Details on the creation of the reference noise passages must be relatively short, e.g., 200 ms, in order to achieve an almost time-invariant ICC system. Since the reference signal should consist of more than one noise passage, this structure is repeated several times. The PSD at the output is then estimated by averaging over the noise-only sections, hence,

$$S_{xx}^{(off)}(\mu) = \sum_{k=1}^{K-1} |X(\mu, v)|^2 \sum_{v=0}^{K-1} a_t(v)$$

with

$$a_t(k) = \begin{cases} 1, & \text{if } X(\mu, k) \text{ consists of reference noise,} \\ 0, & \text{otherwise.} \end{cases}$$

Considering the signal flow graph of Fig. 2, all contributions to $S_{xx}^{(off)}(\mu)$ can be identified. The signals $b_{\text{TE}}(n)$ and $b_{\text{ICC}}(n)$ are corresponding to the noise signals recorded by the ear microphone of the artificial listener and the ICC microphone, respectively.

Assuming that the output signal of the noise simulation and the reference signal are orthogonal and the auto PSDs of the noise signals at the microphones are equal to $S_{bb}(\mu)$, it can be concluded for the PSD of the listener ear signal $x(n)$ with an active ICC system

$$S_{xx}^{(on)}(\mu) = \left| H_{\text{TE}}(\mu) + H_{\text{TM}}(\mu)H_{\text{off}}(\mu)H_{\text{LE}}(\mu) \right|^2 S_{yy}(\mu) + \ldots$$

Reference signal contribution

$$\left| H_{\text{off}}(\mu)H_{\text{LE}}(\mu) \right|^2 S_{bb}(\mu) + S_{bb}(\mu)$$

Noise ICC mic., Noise ear mic.

Three components contribute to the estimated PSD: the reference signal transmitted through the room and the ICC system, the noise part recorded by the ICC microphone, and the noise captured by the ear microphone of the artificial listener (see Fig. 1). By means of the estimated PSD of the ear microphone signal when the ICC system is deactivated and the reference PSD $S_{yy}(\mu)$, derived analogously to Eq. 3, the following ratio is derived:

$$R^{(off)}(\mu) = S^{(off)}(\mu) = \left| H_{\text{TE}}(\mu) \right|^2 + S_{bb}(\mu)$$

(6)

The ratio is also computed with an active ICC system

$$R^{(on)}(\mu) = \left| H_{\text{TE}}(\mu) + H_{\text{TM}}(\mu)H_{\text{off}}(\mu)H_{\text{LE}}(\mu) \right|^2 + \ldots$$

$$\left| H_{\text{off}}(\mu)H_{\text{LE}}(\mu) \right|^2 S_{bb}(\mu)$$

Assuming that $S_{yy}(\mu) \gg S_{bb}(\mu)$, the system support

$$G_S = \frac{10}{M} \sum_{\mu=0}^{M-1} \log_{10} |1 + \frac{H_{\text{TM}}(\mu)H_{\text{off}}(\mu)H_{\text{LE}}(\mu)}{H_{\text{TE}}(\mu)}|^2$$

(8)

is obtained by averaging the logarithm of the ratio defined in Eq. 3 over the $M$ frequency bins of interest. Not the whole frequency range is considered, since an ICC system only supports a certain frequency range. Evaluating Eq. 3 leads to a single value representing the system support. In order to link this value to a quality statement, quality levels need to be defined.

### 3.1 Quality Assessment

The quality classes are based on the ITU performance classification [7], where class 1 corresponds to an exceptionally good implementation with special effort, class 2 characterizes an implementation which most likely fulfills the...
requirements, class 3 corresponds to an system which might have some weak components which can be compensated by other modules, and class 4 characterizes a weak system under test.

In order to derive a quality classification for the system support, two limiting cases are considered. The first one is the case, where no support of the ICC system is present. The second case corresponds to the case, where the maximum gain of the ICC system is reached before a significant quality loss becomes audible. Unlike the evaluation procedure, a cooperative ICC system is used for the development of the quality classes. This cooperative ICC system offers the possibility to parametrize different systems in order to create the limiting cases.

Figure 3: Result of the system support evaluation for the limiting cases. System support $R_S(\mu)$ in blue, $R_{\text{off}}(\mu)$ in red, and $R_{\text{off}}(\mu)$ in green.

To find the lowest quality class (class 4), the system support evaluation is performed with an ICC system. The result in Fig. 3 shows that some small variations of $R_S(\mu)$ are derived, hence, the boundary of class 4 should not be set to 0 dB. For the maximum gain case a system support of approximately 10 dB is derived, which gives the boundary of the best quality class. The remaining classes are distributed evenly over this range and are given in Tab. 2. This specific quality classification is an initial approach to evaluate this particular system and to attract a first impression of the quality classes. This cooperative ICC system offers the possibility to parametrize different systems in order to create the limiting cases.

Table 2: Quality classes of the system support evaluation.

<table>
<thead>
<tr>
<th>Class</th>
<th>$G_S$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>$&gt; 10$ dB</td>
</tr>
<tr>
<td>2</td>
<td>6 - 10 dB</td>
</tr>
<tr>
<td>3</td>
<td>2 - 6 dB</td>
</tr>
<tr>
<td>4</td>
<td>&lt; 2 dB</td>
</tr>
</tbody>
</table>

4 Delay

The delay is derived by means of a cross-correlation analysis between the reference signal $y(n)$ and the ear microphone signal $x(n)$ of the artificial listener. The reference signal reproduced by the artificial talking passenger consists of speech only. In addition this signal is amplified to reduce the complexity, the computation is done in the frequency domain by means of an analysis filter bank as mentioned before in section 2.

After the estimation of the cross-correlation, the point in time at which the function reaches its maximum is derived. This is done with and without an active ICC system. The difference between these two time points gives the delay between direct sound and the signal reproduced by the ICC system:

$$D = \arg \max_\kappa \{ |s_{xy}(\kappa)| \} - \arg \max_\kappa \{ |s_{xy}(\kappa)| \}.$$  (10)

4.1 Quality Assessment

Basically, it is reasonable to assume that the less delay is introduced by the ICC system the better the system is. Additionally the quality in terms of the delay is also depending on the so-called precedence effect, which describes the localization between two sound sources. In [5] a psychoacoustic experiment combining gain and delay was presented and the range between 10 and 20 ms was identified as the best range, which leads to the chosen range of the quality classes. If the delay exceeds 30 ms the second sound source is perceived as an echo [8]. These considerations are leading to the quality classification as shown in Tab. 3. Of course, this classification is just an initial approach which need to be verified by means of additional subjective tests.

Table 3: Quality classes of the delay evaluation.

<table>
<thead>
<tr>
<th>Class</th>
<th>$D$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>&lt; 12 ms</td>
</tr>
<tr>
<td>2</td>
<td>12 - 16 ms</td>
</tr>
<tr>
<td>3</td>
<td>16 - 20 ms</td>
</tr>
<tr>
<td>4</td>
<td>&gt; 20 ms</td>
</tr>
</tbody>
</table>

5 Residual Noise

The residual noise measure determines the noise introduced by the ICC system due to the playback of the processed microphone signal. The noise incrementation is measured by estimating the PSD at the ear of the artificial listener with and without an ICC system. Therefore, a test signal similar to one of the system support evaluation is utilized. In contrast to the system support evaluation, the residual noise evaluation utilizes a reference signal with speech passages and pauses in between. The pauses are used to estimated the noise power by

$$N_{xx}(\mu) = \frac{\sum_{v=0}^{K-1} |X(\mu,v)|^2 a_v(v)}{\sum_{v=0}^{K-1} a_v(v)},$$  (11)

with

$$a_v(k) = \begin{cases} 1, & \text{if } X(\mu,k) \text{ corresponds to a pause}, \\ 0, & \text{otherwise}. \end{cases}$$  (12)

The ratio

$$R_N(\mu) = \frac{N_{xx}(\mu)}{N_{xx}(\mu)}/N_{xx}(\mu)$$  (13)

$$= |H_{\text{tot}}(\mu)H_{LO}(\mu)|^2 + 1$$  (14)

between both PSDs gives the amount of the noise level added by the ICC system. Finally, the average logarithmic noise increment is derived by

$$G_N = \frac{10}{M} \sum_{\mu=0}^{M-1} \log_{10} \left( |H_{\text{tot}}(\mu)H_{LO}(\mu)|^2 + 1 \right).$$  (15)
5.1 Quality Assessment

Again, two limiting cases are evaluated and used to derive the limits of class 1 and 4. In order to find class 1, the ICC system was deactivated and for class 4 the system was parametrized in such a way that a significant increase in noise was observed. The result for both cases are depicted in Fig. 4.

The remaining quality classes are distributed evenly over the specified range, see Tab. 4.

<table>
<thead>
<tr>
<th>Class</th>
<th>$G_N$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>$&lt; 2$ dB</td>
</tr>
<tr>
<td>2</td>
<td>$2 - 7$ dB</td>
</tr>
<tr>
<td>3</td>
<td>$7 - 12$ dB</td>
</tr>
<tr>
<td>4</td>
<td>$&gt; 12$ dB</td>
</tr>
</tbody>
</table>

Table 4: Quality classes of the residual noise evaluation.

6 Overall Quality

Once the quality level $q(l, i)$ for each measure out of the four-stage quality classes was assigned, a combination of these values is performed in order to create an overall quality of the ICC system.

The quality of each evaluation group $q(l)$ is derived and, subsequently, the overall quality by

$$
\bar{Q} = \frac{1}{L} \sum_{l=0}^{L-1} q(l) = \frac{1}{L} \sum_{l=0}^{L-1} \sum_{i=0}^{l-1} q(l, i) \cdot a_q(l, i),
$$

where $i$ is the index of the single measure within the group $l$ and $a_q(l, i)$ is the corresponding weight. In a first intuitive approach, all values of $a_q(l, i)$ are equal to one.

In Fig. 5 a quality representation by means of the quality pie was presented. In style of this, the graphical representation of the overall quality, as shown in Fig. 5, was developed. This representation should summarize the evaluation of the ICC system and is intended to provide the results made by only one-view. Fig. 5 illustrates the achieved quality class of the single measures by the color (green equals class 1, yellow class 2, orange class 3, and red class 4) of the corresponding bar and the derived value by the filling of this bar. In addition the filling color of the groups equals the achieved group quality class $q(l)$ and the background color corresponds to the overall quality $\bar{Q}$.

7 Conclusions

An evaluation strategy containing three main groups (characteristics of the vehicle, characteristics of the ICC system, and communication quality) where each group consist of several single measures was presented. In detail three measures considering the behavior of the ICC system have been described (system support, delay, and residual noise) and a first quality classification has been presented. In order to derive more reliable boundaries of the quality levels further research, such as subjective tests, is required.

Further measures should be included into the evaluation groups to cover most of the factors influencing the communication quality. For example, the noise level inside the vehicle and the change of the signal-to-noise ratio due to the ICC system would be of great interest. Furthermore, the weightings of the single quality values have to be defined, as different measures could have different impact on the overall quality. Therefore, the impact factor of each measure has to be determined by further subjective tests and investigations.

References