Evaluation of Systems that Improve the Communication in Passenger Compartments

Tim Haulick, Gerhard Schmidt

Harman/Becker Automotive Systems, 89077 Ulm, Deutschland, Email: thaulick/geschmidt@harmanbecker.com

Introduction

Due to the background noise and sound absorbing materials the communication between front and rear passengers in sedans and vans is often difficult. Especially as driver and front passenger are talking towards the windshield, they are hardly intelligible for those sitting behind them. To improve the speech intelligibility in-car communication systems – shortly called *intercom systems* – can be used. These systems record the speech of the speaking passengers by means of microphones and improve the communication by playing the recorded signals via those loudspeakers located close to the listening passengers.

Intercom systems operate in a closed electro-acoustic loop. Each microphone picks up at least a portion of a loudspeakers signal. If this portion is not sufficiently small sustained oscillations appear – which can be heard as howling or whistling. The howling margin depends on the output gain of the intercom system as well as on the gains of the analog amplifiers. For this reason all gains within the system need to be adjusted carefully. To improve the stability margin signal processing techniques, such as beamforming, feedback and echo cancellation, adaptive notch filtering, adaptive gain adjustment, and nonlinear processing can be applied. Details about signal processing units can be found, e.g., in [1]. In this paper we will focus on how to evaluate the quality of these algorithms and how to compare two competing approaches.

Evaluation of Intercom Systems

The fairest way of answering this question are, of course, subjective tests. For this reason we will present two subjective tests in the next sections: a rhyme test and a comparison mean opinion score. These tests are, however, quite expensive and time consuming since a lot of test subjects need to be involved. Objective tests, on the other hand, are much simpler and – if well designed – can also give a good indication about the quality of intercom systems. All tests presented in the following are based on a real intercom system.

Subjective Methods

For evaluating the improvements concerning speech quality and speech intelligibility two subjective tests can be utilized. Improvements or degradations of the speech intelligibility can be measured with a so-called *diagnostic* or *modified rhyme test*. In these tests pairs or even larger groups of rhyming words are used to focus on the intelligibility of each syllable or even on each phoneme of a word. A good speech intelligibility is one of the basic requirements of any communication system. If this quantity has reached a certain level, communication is possible and people focus also on other aspects such as the naturalness of speech or the amount of reverberation. These aspects can be analyzed with a so-called *comparison mean opinion score* (CMOS).

Comparison Mean Opinion Scores

For the CMOS evaluation of the intercom system a list with 50 phrases consisting of popular song texts, proverbs, and advertisements was used. Each sentence was played back using an artificial mouth loudspeaker (loudspeaker with approximately the same radiation pattern as a human mouth) at the passenger's seat. Binaural recordings were made on one of the rear seats. These recordings were used as audio examples for all participants of the CMOS test. Per scenario 12 subjects were asked about the speech quality of the presented signals. Per subject 25 pairs of audio examples were presented. Each pair consists of the same stimulus sentence – once recorded with an activated intercom system and once without. The order of presentation was chosen randomly. After listening to each pair of signals the subjects were asked to rate the differences between both signals.

Table 1: Results of the CMOS test.

Voting	0 km/h	130 km/h
On is much worse than off	1.7 %	0.0 %
On is worse than off	5.7 %	1.8 %
On is slightly worse than off	12.3 %	2.5 %
On and off are about the same	29.7 %	7.1 %
On is slightly better than off	31.3 %	25.8 %
On is better than off	18.0 %	46.8 %
On is much better than off	1.3 %	16.0 %

In Tab. 1 the results of the CMOS test are depicted. No significant difference in the speech quality can be observed for the low noise condition (0 km/h). However, in noisy driving conditions (130 km/h) only 4.3 % of the subjects preferred the system to be switched off, 7.1 % of the subjects had no preference, and 88.6 % preferred an activated intercom system. This shows a clear (and significant) preference for the intercom system.

Rhyme Tests

Furthermore, we performed four rhyme tests at the following conditions: intercom system on and off at 0 and 130 km/h. Again 12 listeners participated in each test. For each listener 40 pairs of rhyming words were selected randomly from a prerecorded database. The recording conditions were the same as in the CMOS test. Both words were presented visually first. Secondly, one of the examples was selected (again randomly) and played via a headset.[1] Afterwards the listeners had to decide which

[1] The recording and playback devices were calibrated in such a way that a true binaural impression with calibrated output levels could be achieved.

of the two stimulus words was acoustically presented.

Since the intercom system adjusts its gain automatically according to the background noise it is not surprising that no or nearly no difference (95.0 % for the activated system and 95.2 % for the deactivated system) was measured at 0 km/h. The conditions of those two tests were more or less optimal – meaning that all stimulus words were clearly understandable. Most of the errors were made such that the word that was presented on the left of the computer monitor (that was read first) was also selected by the listeners even if the second was actually presented acoustically. As a result one can conclude that the intercom system does not improve the speech intelligibility when the car is standing. This was not surprising since the intelligibility of the speech was already quite high. Under noisy conditions, however, the amount of correct results could be increased impressively: from 85.4 % without the intercom system to 92.1 % with an activated system. The relative error rate has been reduced by about 50 %.

Objective Methods

Subjective tests have two main drawbacks: on one hand they are quite time consuming – and thus expensive – and on the other hand small differences between different systems or algorithmic versions are quite hard to evaluate with a small group of listeners. For this reason, objective evaluation methods should be applied not for replacing but as a supplement to subjective tests.

Improvements for the Listening Passengers

One way of measuring the improvement of the speech quality due to an intercom system is to measure the impulse or frequency responses from the mouth of the speaking passenger, say the driver, to the ears of the listening passenger, e.g. the left rear passenger. Such measurements should be performed with and without the intercom system. If the background noise is not increased due to the intercom system the ratio between the absolute values of the frequency responses is a good indicator for the signal-to-noise ratio improvement. If we denote the frequency responses with $H(e^{j\Omega})$ we can compute for each measurement pair the ratio

$$R\big(e^{j\Omega}\big) = |H_{\mathrm{on}}(e^{j\Omega})| \,/\, \big(|H_{\mathrm{off}}(e^{j\Omega})| + \epsilon\big)\,. \qquad (1)$$

The subscripts *on* and *off* indicate whether the intercom system should be activated or not. The constant ϵ avoids division by zero. In Fig. 1 two frequency responses are depicted. Both have been measured at a speed of about 70 km/h[2] between the front passenger's mouth reference point and the right ear of the right rear passenger.

Usually four measurements are made for each frequency response (left/right speaking passenger to left/right ear of the listening passenger) and the corresponding ratios are averaged afterwards. For the measurements an artificial mouth loudspeaker and a head-and-torso simulator with ear-microphones should be used.

[2]The speed, respectively the corresponding background noise level, has influence on the output gain of the system and thus on the frequency response.

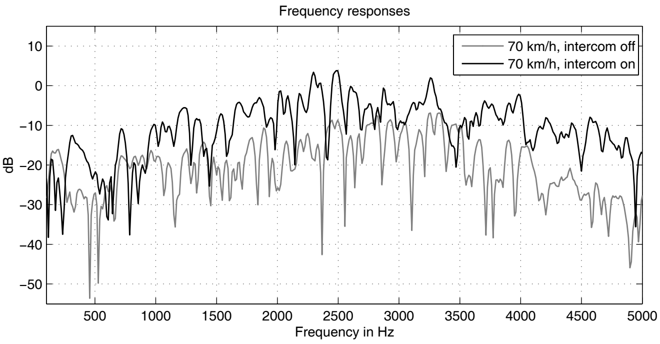

Figure 1: Measured frequency responses.

With a real system installed in a sedan-type car ratios of about 5 to 15 dB were measured. In all cases the background noise level was not affected by the intercom system. The amount of improvement depends on many influences such as the size of the car, the position of the microphones and loudspeakers and of course on the adjusted output gain of the system. Usually more improvement is achieved at higher frequencies since several of the processing units, such as beamforming, are not very effective at low frequencies and also the directivity pattern of a human mouth is not as distinctive at low frequencies as it is at high ones.

Distortions for the Speaking Passenger

A large ratio $R(e^{j\Omega})$ indicates an improvement of the communication quality for the listening passenger. The communication quality from the point-of-view of the speaking passenger, on the other hand, cannot be measured by $R(e^{j\Omega})$. If the gain of the intercom system is rather large the speaking passenger might be disturbed by getting aware of his own echo. The longer the delay of the system is the more disturbing is a certain amount of echo. At a system delay of about 5 ms and a coupling of about -10 dB from the mouth of a speaker to his ears, for example, most people do not realize any echo due to the self masking effect of the human auditory system. If the delay exceeds 30 ms (again with a coupling of about -10 dB) nearly everyone gets aware of the echo.

These disturbing echo effects can be detected – at least approximately – by measuring the impulse response between the mouth of a speaker and his ears. Again a head-and-torso simulator can be utilized for this purpose. By comparing the absolute value of the impulse response with a masking envelope curve, audible echoes can be detected. The problem with this kind of analysis is the determination of the margin envelopes. These envelopes depend on many boundary conditions. For example, different slopes are necessary for different types and levels of the background noise. For this reason, a large amount of research is required until an appropriate echo masking model will be set up for this special purpose.

References

[1] G. Schmidt, T. Haulick: Signal Processing for In-Car Communication Systems, in E. Hänsler, G. Schmidt (eds.), Topics in Acoustic Echo and Noise Control, Springer, 2006.