Speech Quality Evaluation in Vehicles

H.W. Gierlich

HEAD acoustics GmbH, 52134 Herzogenrath, h.w.Gierlich@head-acoustics.de

Abstract

Hands-free communication in cars is still a challenging topic. The present article describes the challenges when implementing, optimizing and testing systems for hands-free communication in vehicles. An overview is given about the parameters impacting speech quality in sending, receiving and in the conversational situation. Test setups as used today are described and methodologies used to test and optimize the various parameters impacting speech communication quality in cars are given.

Motivation

The deployment of hands-free systems in cars is and remains a challenging topic. The car environment is a quite adverse environment. Microphones are typically placed far away from the driver or co-driver, loudspeaker systems in cars are not placed in optimum positions with regard to optimum speech reception and acoustical echo cancellation. The driving noise caused by the car at various speeds is a factor which may affect the speech quality significantly. Finally, the complex structure of modern car hands-free systems is adding further to the complexity of the task.

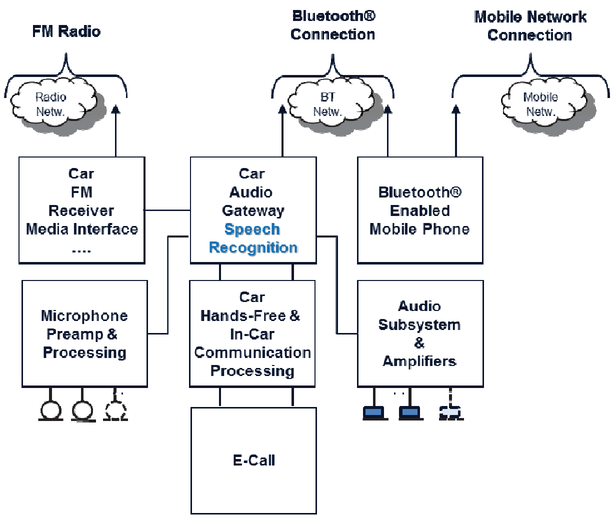

Figure 1: Block diagram of a typical architecture in a vehicle

Figure 1 gives an overview about the car hands-free configuration in modern cars. The car hands-free processing is typically part of a head unit which is interconnected to various subsystems such as microphones, audio subsystem, network transport subsystem (mobile phone connected via Bluetooth® to the car hands-free system and to the mobile network). This configuration may change from car to car but also depends very much on the individual audio package supplied in the various versions of one car. The system has to be "future proof" since the car owner expects his car to be a long term investment operating during its entire life time with all different mobile phones which the user may buy in the future. To adapt the long term development cycles of cars to the short term development cycles of mobile phones is challenging. The user however expects the car hands-free system to work seamlessly well without any degradation. In order to deal with the adverse environment in a car a lot of signal processing is integrated in modern car hands-free systems.

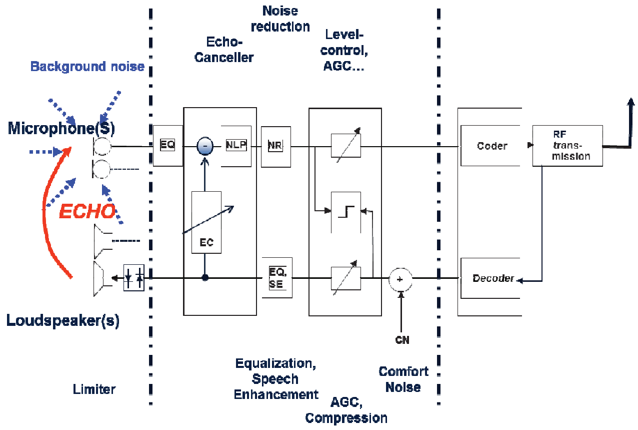

Figure 2: Block diagram of signal processing typically used in car hands-free

The block diagram in **figure 2** gives an overview about typical signal processing blocks used. Besides optimum acoustical placement of the microphones, microphone or microphone arrays are used in conjunction with noise cancellation algorithms to remove as much background noise as possible while keeping high speech quality in sending direction. In addition equalization, automatic gain control and different sorts of compression may be used. Speech coding has to be applied in order to transport the speech signal over the mobile network. In many designs Bluetooth® connections are used to connect the mobile phone to the car hands-free processing. This involves additional speech coding since the Bluetooth® speech coder is different from speech coding used in mobile networks.

In receiving a similar structure can be found. Different types of automatic gain control, compression and equalization algorithms are applied in order to provide a sufficient speech quality using the car loudspeakers.

The acoustical coupling between loudspeaker and microphone introduces an additional problem. Due to the delay in such connections the acoustical coupling will lead to echo problems for the far end user in the connection. Therefore echo cancellation is deployed. Different types of echo canceller solutions in conjunction with different types of noise cancellation algorithms form a complex system of signal processing which has to provide sufficient speech quality in all conversational situations. The conversational quality is mainly determined by the delay introduced by the complete system as well as by the performance of echo and noise cancellation in different driving situations in different background noise situations.

Test setup

Since each car provides a quite different acoustical environment all hands-free systems are typically tested in the individual target car. In this target car the different hands-free systems in conjunction with the different packaging are deployed and tested in a lab cab environment. A general test setup is shown in **figure 3**.

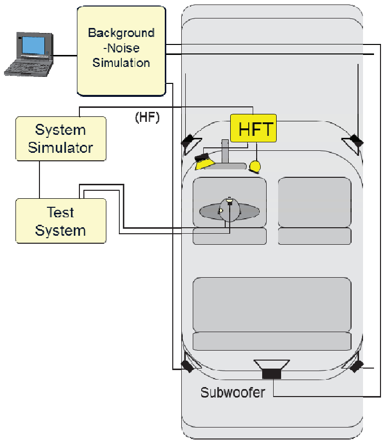

Figure 3: Test setup in a vehicle

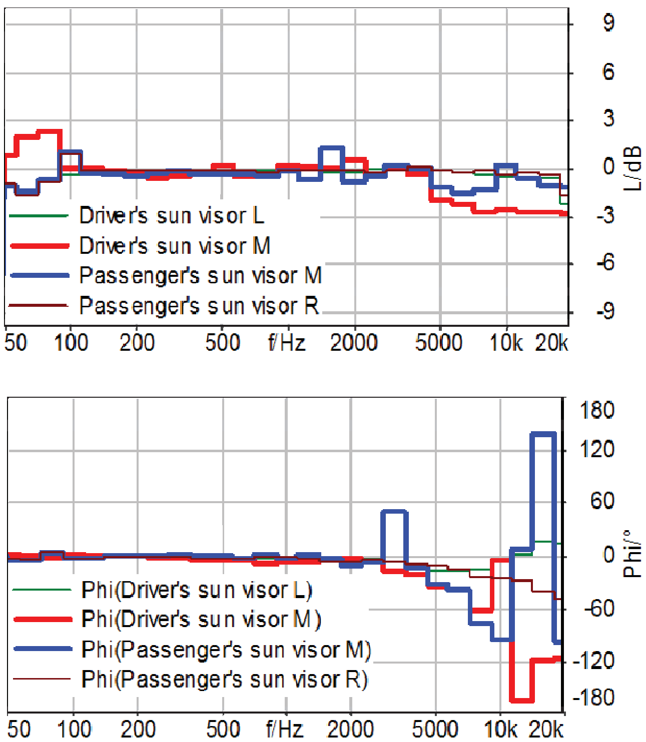

Figure 4: Background noise reproduction accuracy at two microphone locations in a vehicle (driver and co-driver seat)

In order to test under most realistic conditions driving noise is pre-recorded and simulated in the car. Most recent technologies allow simulating the background noise realistically (with correct magnitude and phase across the relevant frequency range) at various locations inside the car. The background noise is recorded at the different microphone positions in the car and equalized to the different microphone positions in the car using a multi-channel loudspeaker setup which in general is described in [1]. The equalization is performed by equalizing each loudspeaker to each microphone position using inverse matrix calculation for the sound field. A typical result for such equalization at two microphone positions (one above the driver, a second one above the co-driver) is shown in **figure 4**. **Figure 4 shows** that magnitude and phase of the background noise are well reproduced across the entire frequency range. This is important in order to take care of modern types of microphone arrangements such as beamforming microphones. It should be noted that the tests are mostly carried out under error free network conditions by simulating the radio network using a radio network simulator.

Tests in sending

The major target of all tests in sending direction is to ensure a good speech sound quality, to preserve good speech intelligibility and avoid any additional artifacts or additional components such as background noise. Obviously the performance in background noise is of vital importance in the sending direction. In order to achieve a good background noise performance the positioning of microphones and the microphone construction itself should be chosen such that an optimum of signal to noise ratio can be achieved already by the microphone placement and construction. The position found here is individual to each car. Nevertheless noise suppression has to be applied. Noise suppression has to be targeted to the different driving situations and to the different background noises produced at different vehicle speeds. This task is more challenging when moving from narrowband to wideband or even super wideband transmission of speech. While in narrowband a simple high pass can result in a sufficient signal to noise ratio improvement in combination with a moderately aggressive noise cancellation algorithm, in super-wideband and fullband the low frequency content of the speech has to be preserved in order to keep a good balanced sound quality impression. Since most of the noise energy is found in the low frequency domain noise cancelling has to be more advanced in order to remove background noise in the low frequency domain. In addition high quality noise cancellation is expected in the high frequency domain where mostly no speech signal energy is present. Furthermore the requirement on natural sounding speech is increasing when moving from narrowband to wideband or even super wideband speech. A good measure allowing estimating the speech sound quality separately from the intrusiveness of the noise components is found in [9]. The speech quality determined in S-MOS and the noise intrusiveness determined in N-MOS is determined in a range of 1 to 5 (1 = poor … 5 = excellent). Obviously background noise simulation is present during these tests. The tests are targeted to different vehicle driving speeds and different background noises individually pre-recorded for each vehicle and played back as described previously. Furthermore there are requirements and frequency response characteristics which are getting more demanding when moving from narrowband to wideband or even super wideband and fullband. Besides these very important parameters also the perceived loudness currently determined as "loudness ratings" [[2]) and the variation of loudness during the call are determined. In order to ensure a mostly undistorted transmission of the speech signal distortion measurements are available mainly targeted to distortions introduced by acoustical components or artifacts such as overload in the signal processing. A detailed description of these tests can be found for narrowband, wideband, super wideband and fullband systems in [3], [4], [5].

Tests in receiving

Naturally in a car environment speech quality in receiving is less demanding than in sending. The main reason is the generally high quality audio systems providing fullband loudspeaker and amplifiers which are normally used for audio playback but can be used for speech communication in the same way. However, it is important to verify that any audio signal processing is deactivated. Additional audio processing would introduce additional delay and potentially non-linearities. Any additional delay leads to reduced conversational speech quality. Any additional non-linearity will lead to a degradation of the echo canceller performance since echo cancellers have to rely on a linear echo-path. Except measurements in background noise which are currently not covered (except ensuring a certain speech level reproduced in the car with background noise) the measurements in receiving concentrate on the determination of loudness (based on loudness ratings) listening speech quality (without background noise) and noise and distortion components which might be produced by loudspeakers and other acoustic equipment in the car.

Tests targeted to conversational quality

Three main parameters describe conversational speech quality in general:

- Round trip delay
- Echo performance
- Double talk capability

Delay is introduced by radio transmission and all types of signal processing in the car hands-free system. In order to provide good conversational quality an end-to-end delay of 150 ms should not be exceeded [5]. This value will ensure seamless interaction between conversational partners without noticeable interruptions in the conversational flow. In car hands-free systems delay consists of the implementation dependent delay (signal processing in the car hands-free implementation) and access dependent delay. The access dependent delay is mainly the delay introduced by speech coding and radio transmission to the mobile radio network. This part of the delay cannot be changed. In addition a short range wireless access delay e.g. if a device is connected by Bluetooth® to the car hands-free system may be introduced. The current requirements in [3], [4], [5] suggest a round trip delay of 170 ms which is already beyond the 150 ms required for seamless interaction between two users. The test of delay is made in a rather simple way. The tests are conducted separately in sending and receiving. The round trip delay consist of the send and receive delay. In the latest version of the ITU-T Recommendations the access specific delay is included in the delay measurement which leaves about 70 ms for the hands-free signal processing to provide all required functionalities such as noise and echo cancellation, gain control, equalization etc.

In order not to hamper the conversation by echo a sufficiently well designed echo cancellation has to be implemented. The main features to be tested are Echo Loss (steady state convergence of the echo canceller), convergence and re-convergence of the echo canceller, , the occurrence of short echo artifacts, sufficient echo loss in the high frequency domain as well as sufficient echo performance in different background noise situations. A variety of tests are targeted to the performance of echo cancellation.

Double talk capability describes the ability of a hands-free implementation to simultaneously transmit sending and receiving without cutting off speech segments or producing echo signals. In general these double talk tests are targeted to the proper implementation of echo cancellers in conjunction with non-linear processes. For testing speech signals are inserted in sending and receiving simultaneously. The focus of the test is to verify that the send signal is not interrupted in case of a double talk signal being present. The measurement is realized by first measuring the send direction without any double talk signal present and conducting a second measurement with the double talk signal present. The speech signals measured in send direction with and without double talk are compared to each other (after time alignment) and attenuation changes are analyzed. If no attenuation is identified a double talk capability of class 1 (see [7]) can be achieved which corresponds to no noticeable speech clipping during double talk. In a similar way echo performance is analyzed. For testing the echo performance in double talk again signals are inserted in send and receive simultaneously. Here a voiced sound is used which provides a combfilter structure with different fundamental frequencies in sending and receiving. The signal is constructed in such a way that the combfilter structure does not overlap and the signals can easily be separated in the frequency domain. More detailed information can be found in [7]. A classification scheme ranging from full duplex systems to no duplex systems is standardized in [7] and used in car hands-free testing.

Special issues with short range wireless devices

A common technology to connect the car hands-free system to the mobile network is the use of the personal smartphone as a gateway. The underlying technology is typically Bluetooth®. The user can bring his own device into the car, it is connected to the car hands-free Bluetooth® interface and the smartphone itself is connected to the mobile network. In this configuration the mobile phone acts as a gateway. The gateway functionality must be restricted to just the transport of speech data from one interface to the other. Any signal processing other than speech coding must be deactivated in order to ensure the signal transport without any additional signal manipulation. Due to the lack of protocol and performance specifications in the existing Bluetooth® specifications this functionality is mostly not or only partially implemented in today's mobile phones. Although instructed to switch off signal processing many phones still leave parts of the signal processing active such as additional volume control, filtering, noise and/or echo cancellation and others. The setup used to test and verify the transparency of speech communication over the Bluetooth® link of mobile phones in gateway mode is shown in **figure 5**. The mobile phone is connected at the one hand side to a Bluetooth® reference interface at the other side to a network reference interface. A series of tests as specified in [3], [4], [5] is used to verify the intended performance of mobile phones used as gateways.

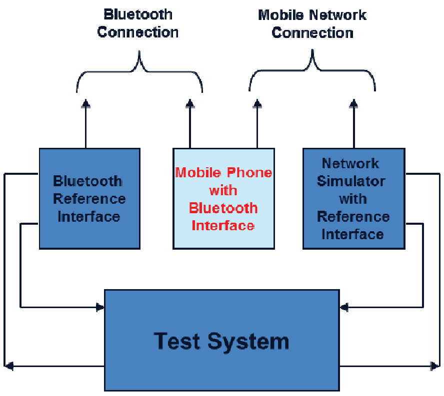

Figure 5: Test setup for Bluetooth® testing

The configuration of car hands-free systems for emergency calls

Though at first glance a hands-free system might be suitable for emergency call communication from the vehicle to the infrastructure special requirements in an emergency call apply. The major difference between hands-free systems intended to be used in emergency call is the handling of background noise. In case an emergency call is generated automatically very likely the car is crashed and windows most likely are damaged. There will be no driving noise. Instead of driving noise environmental noise will be present. There may be even a situation that driver or passengers are not able to talk themselves. In such situation the information contained in the environmental noise is of vital importance for the public safety operator (PSAP) in order to get information about the crash situation. Any noise cancelling must be deactivated in the situation even more it would be desirable if background noise would be amplified to make it easier for the PSAP operator to analyze the environment.

The main focus of such a system is not so much on the speech quality but rather on speech intelligibility. Therefore response characteristics are more emphasized to transmit high frequency components rather than low frequency components since it is known that frequency components in the area between 2 and 4 kHz certainly contribute to increased speech intelligibility. In double talk situations the main focus is on the complete transmission of the talker's voice in the car. Double talk situations may occur more likely in emergency calls compared to ordinary conversations. In such situations conversations are less disciplined leading to increased double talk. Finally for emergency call systems it is important that a minimum of intelligibility must be guaranteed throughout the whole car. From each position not just from the driver's position hands-free communication must be possible. All these aspects are reflected in special eCall requirements which can be [8].

Conclusions

Testing of car hands-free systems requires a target car and an almost realistic reproduction of the driving conditions by simulating background noises in lab type environments. These types of setups are available and are combined with advanced types of testing in sending, receiving and for the conversational situations. Internationally agreed test procedures are available and manifested in a variety of ITU-T Recommendations. The connection of car hands-free systems to the network using Bluetooth® technology is still a challenging topic. This topic is mainly not a technical topic but is rather a topic of being not recognized by the mobile phone makers since gateway tests are not specified in Bluetooth® nor are they specified in any Global Certification Forum (GCF) requirement. These types of testing are out of the focus of many mobile phone makers. Additional effort such as whitelisting of mobile phones (see http://www.itu.int/en/ITU-T/C-I/Pages/HFT-mobile-tests/test_event_3.aspx) and improved specifications are under way. When developing hands-free communication for emergency call systems special requirements have to be taken into account. These are specified for narrowband and wideband systems in ITU Recommendation [8].

References

[1] ETSI TS 103 224: Speech and multimedia Transmission Quality (STQ); A sound field reproduction method for terminal testing including a background noise database

[2] Recommendation ITU-T P.79: Calculation of loudness ratings for telephone sets

[3] Recommendation ITU-T P.1100: Narrow-band hands-free communication in motor vehicles

[4] Recommendation ITU-T P.1110: Wideband hands-free communication in motor vehicles

[5] Recommendation ITU-T P.1120: Super-WideBand (SWB) and FullBand (FB) hands-free communication in motor vehicles

[6] Recommendation ITU-T G.114: One-way transmission time

[7] Recommendation ITU-T P.502: Objective test methods for speech communication systems using complex test signals

[8] Recommendation ITU-T P.1140: Speech communication requirements for emergency calls originating from vehicles

[9] ETSI EG 202 396-3: Speech Processing, Transmission and Quality Aspects (STQ); Speech Quality performance in the presence of background noise Part 3: Background noise transmission - Objective test methods