Challenges in Mobile Hands-free System Optimization

Frank Kettler, Hans Wilhelm Gierlich, Frank Rosenberger
HEAD acoustics, 52134 Herzogenrath, Germany, Email: Frank.Kettler@head-acoustic.de

Introduction

Built-in car hands-free telephones may consist of different components consequently leading to a heterogeneous system design. Each component like the microphone (or microphone array respectively), the hands-free unit itself and sometimes the mobile integrated additionally introduce various kinds of signal processing. This may even lead to cascaded algorithms providing the same kind of, or at least comparable signal processing in order to improve speech quality. A typical example is the noise reduction which can be found in the microphone array (beam forming, algorithmic noise reduction), additionally in the hands-free unit and in the integrated mobile. In order to guarantee a good speech quality for the complete system it is absolutely necessary to choose the parameter settings of these algorithms according to the characteristics of the cascaded systems. Typically these cascaded systems have been optimized independently of each other and the combination of the systems do not necessarily lead to a sufficient speech quality of the whole system. Guidelines are discussed how these complete systems can be optimized.

Complex cascaded systems

The hands-free telephones (HFT) are completely integrated in vehicles and consequently use components like the audio playback system in the car. Figure 1 shows the principal configuration of an in-car hands-free telephone system.

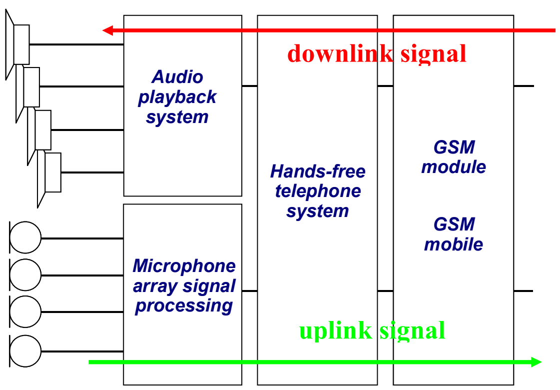

Figure 1: Components of integrated HFT systems in vehicles (example for GSM mobile network)

The heterogeneous, cascaded systems can typically be found in cars because the single components (GSM modules or mobiles, HFT algorithms, audio playback systems, microphones) are provided by different manufactures. Moreover car manufactures typically offer a market-dependent variety of implementations for each model, like different audio playback systems, the integration of different mobiles, features like speech recognition, etc.

The HFT signal processing block in Figure 1 represents all relevant components like acoustic echo cancellation (AEC), post processing, level switching and attenuation control, noise reduction (NR) or comfort noise injection (CN). Typically the receive direction is less critical in terms of signal processing influencing the transmitted speech. Although the relevant algorithms are found in sending direction (uplink), cascaded components in receive direction may also lead to potential problems if not properly considered.

The downlink direction

The signal in receive direction (downlink) is picked up and decoded by a GSM module or a GSM mobile. This signal is then transmitted via the HFT algorithms, the integrated audio play back system and the loudspeakers in the car cabin to the driver’s ear. It can be assumed that the loudspeakers themselves are of high quality because they are typically also used for radio or CD playback in the car.

The audio system is part of the echo path for the acoustic echo canceller. Additionally this echo path is represented by the loudspeakers, the microphone, the impulse response in the car cabin between loudspeaker and microphone and control mechanisms e.g. to automatically adjust the playback volume to the driving noise (automatic gain control, AGC). This leads to two consequences:

- The audio system may introduce time-invariant characteristics (e.g. delay), which needs to be considered by the AEC algorithm.
- The adaptive filter in the AEC does not have the appropriate, “correct” reference signal to estimate the impulse response between loudspeaker and microphone.

The audio system may introduce additional non-linearities which significantly impairs the performance of the acoustic echo cancellation.

Any kind of signal processing introduced by the GSM module or GSM mobile should not degrade the performance of the AEC. It influences the reference signal but not the echo signal. But differences in sensitivity of different types of mobiles will lead to different playback volumes in the car, if not considered e.g. by loading “mobile-specific” settings in the audio system or the HFT algorithm. This requires a communication between these system components.

The uplink direction

The signal in sending direction (uplink) is picked up by a microphone or a microphone array. This signal is then passed to the HFT algorithms and transmitted via the GSM module or a mobile into the mobile network. The sending direction is extremely critical because all kind of signal processing implemented to improve the quality of the transmitted speech can be found here (AEC and post processing, NR, CN injection, …). Note that the same kind of signal processing as implemented in the HFT part may also be found in the GSM mobile – if not deactivated. This configuration leads to AECs in tandem and post processing, cascaded NR algorithms or different stages of level adjustments. GSM modules - as indicated in figure 1 - typically represent only the GSM coding and decoding without additional signal processing.

Depending on the system design different optimization strategies are necessary:

- selective optimization of one algorithm part on the acoustical conditions while disabling all other algorithms providing the same kind of signal processing
- choice of parameters in parts of the algorithms in order to reduce undesired artifacts of the cascaded algorithms (providing the same kind of signal processing)

Cascaded noise reduction algorithms

Typically these algorithms are implemented to reduce the noise picked up by the microphone or the microphone array. In principal it is recommended to implement and optimize one noise reduction algorithm in the signal processing chain. The combination of a microphone array and one noise reduction algorithm in the HFT module is useful, recommended and -in most cases- sufficient. In cascaded systems today, the signal in sending direction is transmitted through -at least- two noise reduction algorithms, one provided by the HFT part and one provided by the GSM module.

This will lead to undesired artifacts like a distorted speech (driver's voice) or even a too low noise signal from the driving car. The driver's voice sounds unnatural because typically the Lombard effect influences the characteristics of this voice. In order to optimize the overall performance it may be necessary

- to adjust the HFT NR algorithms in a way that the residual noise level is lower than an implemented threshold of the NR algorithm in the GSM mobile or
- to minimize the influence of the HFT NR algorithm and rely on the NR in the GSM mobile.

It should be noticed that microphone suppliers announce already microphones with implemented noise reduction algorithms. This would result in a third cascaded algorithm.

Optimizing double talk performance

Implemented signal processing like the AEC combined with additional post processing (level switching, attenuation control, echo suppression) influence double talk performance. The importance of this quality aspect is considered e.g. in ITU-T Recommendation P.340 [1] and the VDA Specification for mobile hands-free telephones [2]. The test signal is a periodical repetition and combination of two uncorrelated CS signals [2]. One signal is applied in receiving direction typically via a GSM system simulator during laboratory tests. The other signal is played back by the artificial mouth of a HATS (Head And Torso Simulator according to ITU-T Recommendation P.58 [3], P.581[4]) positioned on the driver's seat. The signal recorded in sending direction at the output of the GSM simulator is then analyzed in terms of level variation versus time. The following figures represent four typical measurement results. The level variation is also given in each figure.

- One HFT algorithm combined with GSM module, external loudspeaker and microphone (no audio system)
- An after-market HFT implementation combined with a GSM mobile and external loudspeaker. The built-in audio system in the car is not used for telephone play back.
- A complex system with cascaded signal processing in the HFT module, the audio system and the GSM mobile.
- A HFT implementation combined with a GSM mobile but deactivated signal processing in this mobile. The built-in audio system is used for play back in the car cabin. The HFT algorithms are optimized on car-specific characteristics.

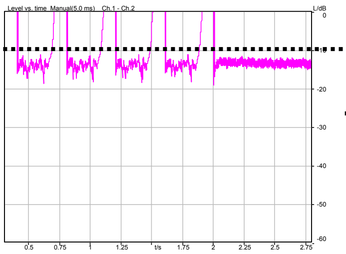

Figure 2: HFT algorithm in development state (with GSM module)

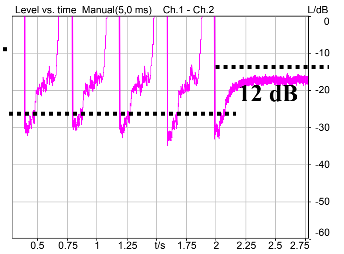

Figure 3: After-market HFT implementation with external loudspeaker

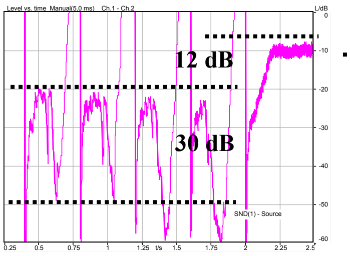

Figure 4: HFT implementation in a complex cascaded system

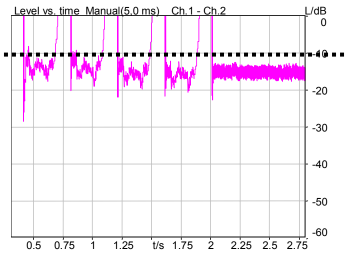

Figure 5: HFT implementation in a complex system with GSM mobile (after optimization)

In the complex cascaded system (figure 4), the near end speech is nearly completely switched off. The measured curve indicates, that two gain control stages are active attenuating the drivers voice by more than 40 dB. These two level variations are introduced by cascaded algorithms. The double talk performance is worse than the performance of a relatively simple after-market solution (figure 3).

Summary

Complex system architectures are critical in terms of the resulting speech quality, even if high quality components are used. General optimization criteria are hard to give. They highly depend on the individual characteristics of the cascaded algorithms.

References

[1] ITU-T Rec. P.340, Transmission Characteristics of Hands-free Telephones

[2] VDA-Specification for Car Hands-free Terminals, V. 1.4

[3] ITU-T Rec. P.58, Head and Torso Simulators for Telephonometry

[4] ITU-T Rec. P.581, Use of Head and Torso Simulators for Hands-free Terminal Testing