Speech Quality and Tuning Aspects for eCall Implementations

Frank Kettler, Radi Serafimov
HEAD acoustics GmbH, 52134 Herzogenrath, E-Mail: frank.kettler@head-acoustics.de

Introduction

A high effort is spent by the automotive industry in order to reduce the number of killed or seriously injured people in road traffic. In parallel, emergency call (eCall) systems will become mandatory in 2015 to build up the PAN European eCall system. In an emergency case, these eCall systems send a minimum set of data (MSD) and establish a phone call between the public safety answering point (PSAP) operator and the vehicle. Up to now, only very rudimental tests are defined to verify the quality or intelligibility of transmitted speech in a conversation over these hands-free systems. Moreover, the existing tests are based on current hands-free specifications that focus on speech quality rather than speech intelligibility.

This paper discusses transmission aspects that could be considered to optimize eCall systems and to cover the more important intelligibility aspect in such emergency situations.

Current Audio Test Specifications

For narrowband and wideband scenarios, speech quality test specifications are available for hands-free communication, such as the VDA Specification for car hands-free systems [1], ITU-T Recommendation P.1100 [2] and P.1110 [3]. These documents and tests are designed to guarantee a high or at least sufficient speech quality in all conversation situations. They are not targeted at speech intelligibility aspects.

The GOST ERA-Glonass Specification [4] focusses on eCall systems, however, the basis for this specification is again ITU-T Recommendation P.1100 and P.1110. None of these specifications addresses eCall specific needs, such as: good intelligibility at all positions in the car (front and backseat passengers), double talk capability with emphasis on intelligibility rather than quality, eCall specific background noise scenarios for testing, such as open (i.e. broken) windows, etc. Furthermore, eCall systems are typically designed as backup systems. A system optimization in the same manner as "regular" hands-free communication systems (HFT) may not always be possible. Therefore, different requirements should be applied. No eCall specific test specification is available today. This is non-satisfying for a communication system which is mandatory in the very near future and may consequently lead to misunderstandings and disputes between automotive industry and suppliers.

Echo Attenuation vs. Double Talk Performance

The balance between sufficient echo attenuation under single and double talk conditions and a good double talk performance (both subscribers talk at the same time) is generally one of the most challenging aspects in the design of HFT systems. Due to the limited capability of echo cancellation algorithms (EC), a higher portion of echo attenuation needs to be provided by implemented echo suppression (ES). However, this typically hampers double talk performance due to the inserted attenuation in the microphone path. The principle is shown in **figure 1**.

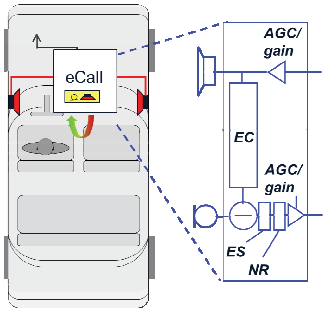

Figure 1: Signal processing components in an eCall system

An HFT may provide a rather good double talk performance today assuming high performing algorithms. This is typically tested for the driver's position using nominal playback volume [1], [2], [3]. Vice versa, eCall systems need to provide a good double talk capability for playback volume which is sufficient for all positions in the car (including passengers on the backseat). Consequently, the playback volume needs to be increased compared to the HFT, this again leads to a higher demand for the ES component.

Vice versa, it is not necessary to tune eCall systems in order to provide a high quality in terms of natural speech under double talk conditions but to keep both channels (the receiving and sending direction) open to guarantee high intelligibility. For this purpose, it may even be possible to distribute the necessary attenuation (echo suppression) in sending and receiving direction using comb filters that do not overlap in the frequency domain as e.g. already proposed in [6]. An example of two original test signal spectra and the filtered signals is given in **figure 2** in order to demonstrate the principle.

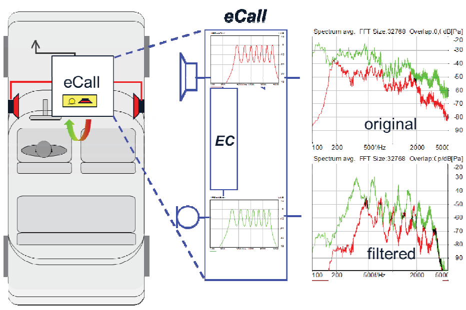

Figure 2: Principle of comb filtered transmission paths during double talk (see also [6])

Background Noise Scenarios

Hands-free implementations are typically tested in constant driving conditions (see e.g. [1]). In eCall scenarios it must be considered that the vehicle may not be moving, but standing e.g. on a noisy street with broken windows. Recordings have been carried out on a highway parking place with four windows open respectively closed. The noise spectra are given in **figure 3** by the blue curve (closed windows) and black curve (open windows). For comparison, a constant driving situation (red curve) shows especially a dominant low frequency content of the driving noise and a lower energy above approximately 1 kHz.

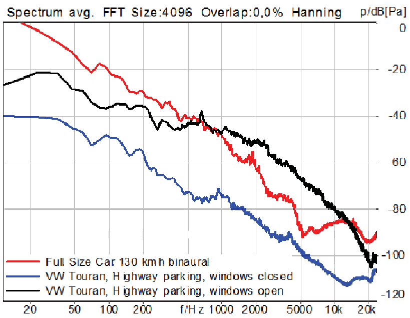

Figure 3: Comparison of noise spectra

The noise levels for these scenarios (130 km/h constant driving condition, 69 dB(A)) and the standing car with open windows (72 dB(A)) is comparable, but the spectral content is significantly different.

Speech Intelligibility

eCall loudspeakers may not be mounted in acoustically optimized locations In consequence level differences between the front seat passenger's and the back seat passenger's position may be of up to 10 dB which needs to be compensated by higher playback volume. Furthermore, the mounting position may lead to strong resonances. One example is shown in **figure 4**. The red and green curves show the spectrum of a speech signal played back via a loudspeaker below the dashboard in the driver's footwall and recorded using a HATS (left and right ear, green/red curve) positioned on the driver's seat.

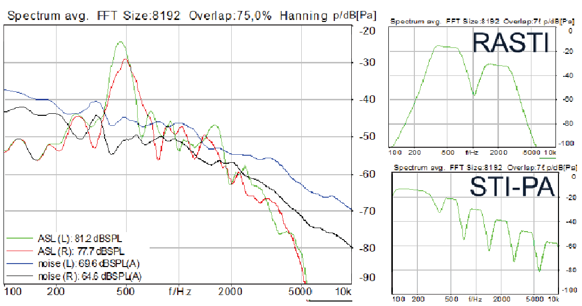

Figure 4: Spectra of speech (left) and RASTI / STI-PA test signals (right)

A strong resonance around 500 Hz can be observed. Speech intelligibility analysis methods such as RASTI or STI-PA make use of modulated noise sequences. The spectra of these noise sequences are given on the right hand side of **figure 4**. The combination of the strong resonances with the spectral characteristics of the RASTI respectively STI-PA test signal may lead to very different intelligibility scores. The STI-PA result in this example was determined to 35 % (left ear) vs. 82 % when using the RASTI method. This indicates that these analysis methods are not applicable for such acoustical scenarios in receiving direction. Furthermore, any noise reduction in sending direction of an HFT implementation may attenuate the RASTI respectively STI-PA test signal and make these measurements even impossible.

Use of NELE Algorithms

eCall scenarios may also be a practical use case for near end listening enhancement algorithms as today often implemented in mobile phones [5]. Today these algorithms typically use the microphone signal for noise estimation and adapt the loudspeaker signal (downlink signal) according to an optimization rule typically targeting at improved speech intelligibility. However, these algorithms require the noise estimation at the user's ear in the car which may be the driver's or passenger's position. If the noise is measured with the eCall module microphone, a correction filter may need to be applied in order to properly estimate the noise spectrum in the user's ear in the vehicle.

Conclusions

It is recommended to consider eCall specific tests, tolerances and requirements instead of adapting HFT specifications. Receiving and sending frequency response characteristics should be optimized for speech intelligibility rather than speech quality. eCall specific loudness requirements in different positions in the vehicle need to be considered. The same applies to the integration of speech intelligibility tests in both directions, an aspect which is completely out of consideration in current hands-free specifications. A new focus should also be put on background noise scenarios to consider realistic eCall conditions. Tests should be designed to preserve the naturalness and information content of transmitted background noise in sending direction as this is often the only available information at the PSAP side to judge the severity of the emergency case. Furthermore, double talk performance should be kept as good as possible, perhaps well balanced between lower speech quality and high intelligibility.

Literature

[1] VDA Specification for Car Hands-free Terminals, Draft Version 1.6, 2006

[2] ITU-T Recommendation P.1100, Narrow-band hands-free communication in motor vehicles (03/11)

[3] ITU-T Recommendation P.1110 Wideband hands-free communication in motor vehicles (12/09)

[4] GOST ERA-Glonass Specification, In-vehicle emergency call system compliant test methods for quality of speaker phone in a vehicle, 2012

[5] B.Sauert, 6th Workshop on Speech in Noise: Intelligibility and Quality, Marseille, France, January 10, 2014

[6] Gierlich, H.W., Ein Verfahren zur akustischen Entkopplung von Freisprecheinrichtungen, DAGA '87 Aachen, Berichtsband S. 485-488