eCall System with Improved Speech Intelligibility

Gonzalo Lucioni[1], Michael Pauls[2]

[1] *Novero GmbH, a Laird business, 44807 Bochum, E-Mail: Gonzalo.Lucioni@novero.com*

[2] *Novero GmbH, a Laird business, 44807 Bochum, E-Mail: Michael.Pauls@novero.com*

Abstract

Intelligibility is a quite abstract term. Main tele-communication measures are dealing rather with quality perception than with intelligibility. But there is no doubt, that an uninterrupted transmission of the speech signal and high echo attenuation are fundamental requirements to obtain low listening effort and good intelligibility.

This article deals with improving the echo canceller and the double-talk performance in vehicle-related eCall solutions while allowing at the same time higher loudness and distortion levels. The approach takes into account in particular limitations by energy budget through the use of eCall backup batteries, backup loudspeakers or by computing power for the digital signal processing.

1. Introduction / Motivation

eCall systems in motor vehicles are becoming increasingly a mass product. This is possible because the definition of the necessary standards, the regulation of jurisdiction and construction of the public eCall infrastructure is carried out by the legally responsible authorities to a large extend. Examples for this are ERA / GLONASS in Russia or the PAN European eCall in the EU.

It turns out, however, that the criteria for voice communication still need to be adjusted to the eCall specific requirements [1]:

- A sufficient acoustic sound level and good speech intelligibility in addition to a good double-talk behavior are extremely important for the eCall scenario after a crash.
- Other criteria that are relevant to speech quality such as the variation of the loudness rating, the insertion of comfort noise, or a low distortion may be considered as secondary to the eCall scenario. The latter might be accomplished by technologies like loudspeaker nonlinear control [9].

Echo cancellation solutions, which comply with all relevant criteria, in particular under non-linear conditions, are still the subject of intense research, see e.g. [2]. A major problem with such approaches is that the solution complexity is usually linked to the degree of non-linearity. This means, on the other hand, that the solution complexity might become huge even at acceptable audio distortion levels.

In this paper, an extension of prior art linear echo cancellers is examined with the aim of improving the double-talk performance when combined with higher acoustic sound levels and stronger distortion. The approach is particularly interesting for eCall systems, which have a limited energy budget when operated via backup battery. eCall systems are required, on the other hand, to guarantee a high sound level inside the vehicle, which might be subject to significant non-linear distortion, especially when the energy budget is getting low.

2. Echo cancellation with non-linear reference signals

In the proposed approach, the usual reference signal (R) for linear echo cancellation is replaced by the upstream, possibly nonlinear echo source. This may be for example the digitized signal of the electrical output from the audio amplifier PA (R1), which might be the output voltage, the output current measured via a shunt resistor, or both. Alternatively, it might be also a loudspeaker signal (R2) sampled via a suitable sensor.

If necessary, vibration-induced distortion can be detected in addition via a suitable sensor (R3) [3] and a multi-channel approach to echo cancellation, see e.g. [2], might be used to cancel the echo signal with diverse non-linear sources as reference signals. This possibility is particularly interesting for those scenarios where there is a significant mechanical-acoustic enclosure coupling between the speaker and microphone.

The following figure shows the target system and related system components

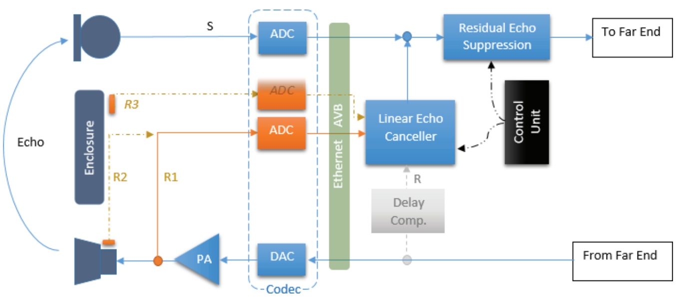

Figure 1: target system and related system components with proposed reference points.

It should be noted here that due to the proposed system structure, a time alignment between the reference signals and the echo signal might be omitted.

The concept may also operate in an environment where the sensors for detecting the reference signals and related Analog-Digital-Converters are spatially separated from the digital signal processing.

3. Test Setup for Simulation related Audio Recordings

Using different recordings in real vehicle environment, audio data was generated to further examine and validate the previously described method. It was concentrated to the electrical output signal of the power amplifier for this study. The recordings were done with an automotive OEM hands-free device using an integrated cellular engine. The used device was prepared to trace the digital audio signals at different points in the signal processing chain. The configuration parameters of the signal processing were freely adjustable.

Test Setup in the Vehicle

The acoustic measurement setup in the vehicle was realized based on the ITU-T P.1140 recommendation [4]. The head and torso simulator was positioned on the driver's seat and the incoming telephone audio signal was played via the two in the front doors of the vehicle integrated speakers. The Volume control for all recordings was adjusted to obtain an RLR value of -13dB at the position of the head and torso simulator in driver's position. The hands-free microphone was mounted, as prevalent in vehicle hands-free devices, in the roof module near the rear view mirror between driver and front passenger directing to the driver´s mouth.

The microphone path was tuned to obtain an SLR value of 13dB in the used setup and the cellular connection was realized using a 3G system simulator configured to use AMR Full-Rate speech codec with 12.2Kbit/s bit rate.

The recordings were done with signals based on test cases from ITU-T P.1140 [4] from the sections Echo and double-talk performance. The main focus was laid on the measurement methods using real speech measurement signals. These are "*Terminal Coupling Loss (TCLw)*" (8.8.1) with compressed speech according to ITU-T P501 [5] and "*Attenuation Range in Sending Direction during Double Talk*" (8.10.1) based on ITU-T P.502 Appendix III [6]. Additionally the measurements of the temporal echo behavior and Detection of Echo components during double talk were recorded in the same test setup for possible future analysis. They will not be considered further in this paper. Environmental noise scenarios were also not taken into account for this investigation for the time being.

Recording Points and Parameterization

For the simulation different recording points for the audio signals were defined at the hands-free device.

Within the signal processing chain the audio signals (R, Tx1 und Tx2) can be digitally traced and stored for further use. The audio signals at the power amplifier output (R1) were recorded in analog domain and digital converted in the measurement system.

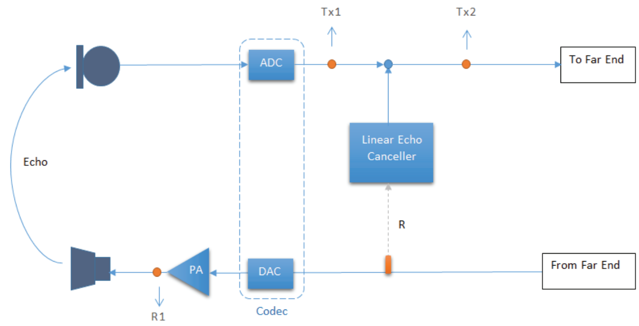

Figure 2: Definition of recording points in signal processing chain and at PA output.

In the parameterization of the signal processing in the receiving branch specifics of eCall application have been considered. Strong compression has been applied to receive audio signal to obtain sufficient signal energy also via possibly small backup speakers and under limited operating voltages. Moreover, the level was increased so far, allowing distortions up to 9% in the upper range of digital signal level (>0dBm0).

This already showed significantly impaired echo performance of the original signal processing algorithms used by the OEM hands-free device using the standard digital reference signal.

Additionally to the voltage recording at reference point (R1) a current measurement has been recorded via a shunt resistor. Using the current/voltage relation, containing even some loud-speaker feedback, is expected to improve echo cancellation further when used as reference signal.

The recordings were afterwards prepared for the simulation by time synchronization and alignment of sample rates.

It is noted that due to crystal drift of approximately 15.5 ppm between the measurement equipment and the OEM device a proper resampling of the (R1) signal was performed hereafter for simulation purposes.

4. ITU-T P.1140 Simulation

The previously obtained recordings have been evaluated according to ITU-T P.1140 by simulation using an AEC algorithm prior art, more specifically, here according to [2, 7, 8].

TCLw Simulation

The overall TCLw evaluated according to the ITU-T P.1140, 8.8.1 when using the compressed speech from reference points (R) and (R1), respectively, is shown below

	Reference R	Reference R1
TCLw	24.02 dB	50.14 dB

Note that the TCLw values already consider the gain of 6dB applied after echo cancellation to obtain appropriate SLR. The corresponding TCLw evaluation as a function of related frequency groups is depicted below

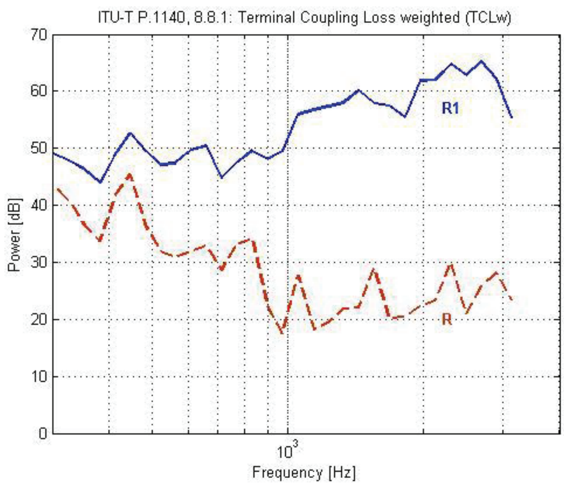

Figure 3: TCLw simulation results

The TCLw evaluation shows the expected improvement when using a pure linear echo cancellation approach in combination with the non-linear reference signal (R1). The result almost outperforms the TCLw of 46 dB required by ITU-T P.1140 for maximum setting of volume control.

When using the reference signal (R), the TCLw does not comply with the ITU-T P.1140 requirement. To solve this issue, the residual echo suppression part of the algorithm is extended by a non-linear term [8], where the level of nonlinearity is adjusted individually for the particular setup. The nonlinear setting is such that the required TCLw requirement is met, respectively. This algorithm is used hereafter for the double-talk simulation.

Double-Talk Simulation

The ITU-T P.1140, 8.10.1 test consists of each 12 speech sequences which interact as follows:

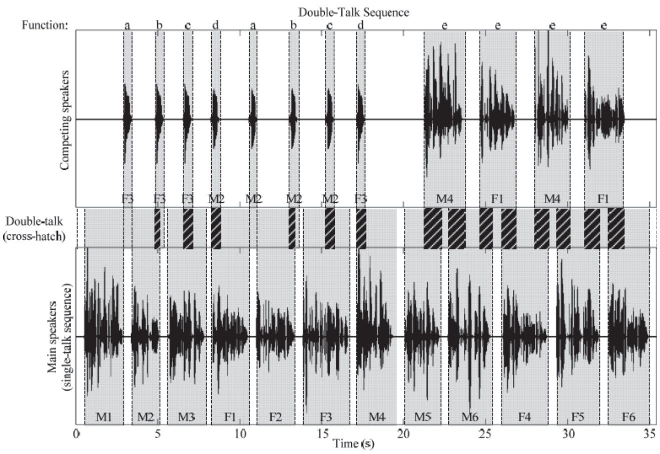

Figure 4: ITU-T P.502 real speech double-talk sequence

The evaluation of the individual sequences related to occurrence of level differences between the respective double-talk sequence after echo cancellation and corresponding single-talk sequence according to ITU-T P.1140 is shown hereafter:

Fct.	Reference R	Reference R1	Spk.
a	0.2 dB	0.1 dB	F3
b	6.0 dB	0.6 dB	F3
c	4.4 dB	2.0 dB	F3
d	11.7 dB	10.9 dB	M2
a	2.4 dB	2.3 dB	M2
b	3.9 dB	2.9 dB	M2
c	13.7 dB	3,5 dB	M2
d	6.5 dB	2.0 dB	F3
e	9.3 dB	1.1 dB	M4
e	5.3 dB	2.4 dB	F1
e	13.2 dB	1.8 dB	M4
e	15.4 dB	2.4 dB	F1

The median thereof is 6.2 dB (*class 2b*) and 2.2 dB (*class 1*) for (R) and (R1), respectively.

In the case of reference signal (R) a closer examination of the target signal after echo cancellation shows quite significant losses in the double-talk event, especially for the sequences of the function e: *'Sentence masking another sentence*'. Thus, an improvement is absolutely necessary, most probably at the expense of echo suppression capability during double-talk.

The figure below illustrates the level of the single-talk sequence, the double-talk sequence (dotted line) and the level difference between the two sequences versus time. For simplification, speech pauses are masked out.

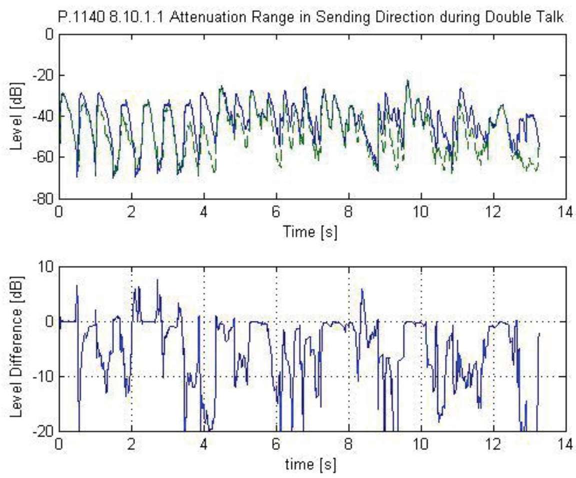

Figure 5: Double Talk performance with digital Reference signal R

In the case of the reference signal (R1), the evaluation of occurrence of level differences over all 12 speech sequences, with pauses in between being masked out, confirms with 1.8 dB the *class 1* categorization:

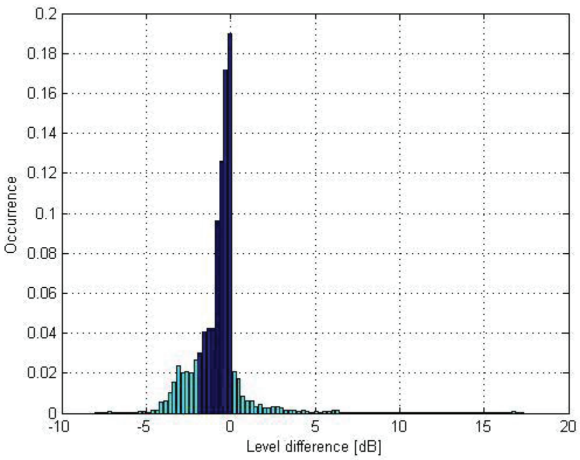

Figure 6: Histogram of level differences during double-talk

A closer look into the target signal after echo cancellation shows a very good double-talk performance with low level differences such that almost no further improvement is needed, especially for the eCall case.

The figure below illustrates again the level of the single-talk sequence, the double-talk sequence (dotted line) and the level difference between the two sequences versus time.

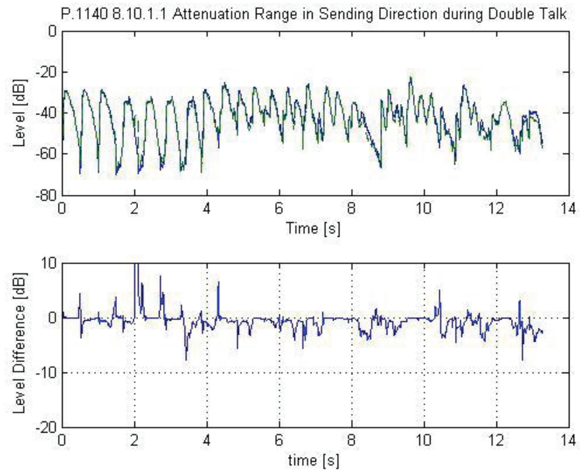

Figure 7: Double-Talk performance with digital converted reference signal R1

5. Summary and Outlook

In this paper, the combination of non-linear reference signals with linear echo canceller solutions is examined for hands-free devices in an environment prone to nonlinear distortion. As a result, significant improvements in echo cancellation performance and double-talk behavior are determined both particularly interesting for scenarios with limited energy budget and high acoustic sound levels as e.g. in eCall systems.

Moreover, the concept may be extended to a multi-channel approach where different, non-linear sources are considered.

As an outlook, the approach with additional current measurements will be investigated further, e.g. to cover the nonlinear behavior of an eCall backup loudspeaker.

References

[1] Gierlich H.W., Kettler F., "eCall and Voice Quality – The forgotten Dimension", in ITU Symposium on The Future Networked Car, Geneva Motor Show, 5-6 March 2014.

[2] Malik, S., "Bayesian learning of linear and nonlinear acoustic system models in hands-free communication", Dissertation, Institut für Kommunikationsakustik, Ruhr-Universität Bochum, Universitätsbibliothek der Ruhr-Universität Bochum, 2012.

[3] T. Gupta, S. Suppappola, and A. Spanias, "Nonlinear Acoustic Echo Control Using an Accelerometer", in Proceedings of the 2002 IEEE International Conference on Acoustics, Speech and Signal Processing, May 2009, pp. 1313-1316.

[4] ITU-T P.1140, Speech communication requirements for emergency calls originating from vehicles. (06/2015).

[5] ITU-T P.501, Test signals for use in telephonometry, (01/2012).

[6] ITU-T P.502, Objective test methods for speech communication systems using complex test signals, Amendment 2, Appendix III (09/14).

[7] G. Enzner and P. Vary, "Frequency-domain adaptive Kalman filter for acoustic echo control in hands-free telephones", Signal Processing, vol. 86, no. 6, pp.1140 - 1156, 2006.

[8] Enzner, G.: Model-Based Optimum Filtering Approach to Acoustic Echo Control: Theory and Practice, Dissertation, no. 22 in Aachener Beiträge zu Digitalen Nachrichtensystemen (ABDN) ed. by Vary, P., Aachen, Germany: Verlag Mainz, Apr. 2006, ISBN: 3-861-30648-4.

[9] W. Klippel, "Optimal Design of Loudspeakers with Nonlinear Control", AES 32nd International Conference, Copenhagen, Denmark, 2007