Third Party Listening Test in Emergency Call Scenarios using Different Languages

Frank Kettler, Silvia Poschen, Radi Serafimov

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

Introduction

ECall systems in vehicles (IVS) need to be tested in terms of voice transmission quality for homologation in Russia. The test procedures in GOST/MGS R55531 [1] cover objective (instrumental) and subjective (auditory) parts. Both tests are mandatory. The MGS R55531 specification describes very complex conversational tests, which are hardly feasible during certification of vehicles. An alternative method based on Third Party Listening Tests (TPLT) as described in ITU-T P.832 [2] can be applied. This method uses pre-recorded conversations in native language, i.e. Russian language for the MGS certification test and a limited number of parameters [3]. This motivates the design of conversations in other languages enabling test labs to run such tests outside the Russian language area for system optimization [3] and certification preparation. Besides comparable time structure and content of such conversations, which is already challenging, the TPLT for each language shall lead to comparable results. The design and recordings of such conversations in three languages Russian, English and German are discussed together with first comparison results.

Motivation

Figure 1 of ITU-T P.832 describes a setup for TPLT based on binaural artificial head recordings in a setup between one hands-free implementation and a handset terminal. The application of this setup on eCall scenarios is shown in **figure 1**. The hands-free system is represented by the IVS in the vehicle under test (left hand side).

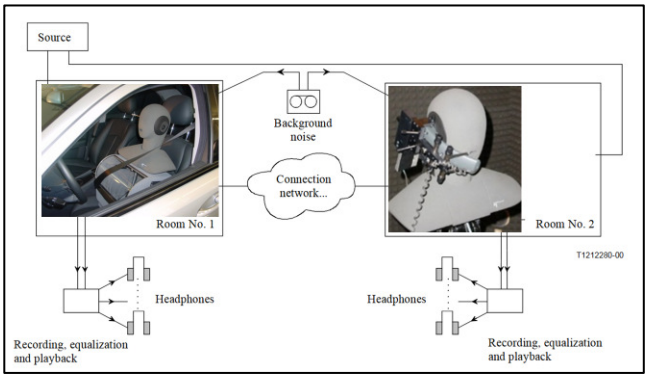

Figure 1: Principle of Third Party Listening Test acc. to [2] using pre-recorded conversations applied to eCall scenarios

The handset terminal represents the public safety answering point (PSAP, right hand side). The aim of the current work was the design of typical conversations between a PSAP operator and a driver in a vehicle calling the emergency services in order to report an accident. This represents a manually generated eCall rather than an automatically generated eCall, as conversations between persons being involved in an accident are much more complex to simulate compared to the situation of a manually generated eCall.

Design of Pre-recorded Conversations

For the design of realistic conversations in eCall scenarios, consultations were necessary with PSAP operators. Furthermore, conversations need to cover single and double talk periods. The latter typically represents the critical aspect in communications over hands-free systems. Furthermore double talk situations may appear in both directions, i.e. the driver interrupting the PSAP side and vice versa.

Although the minimum set of data (MSD, incl. GPS or GLONASS coordinates) is transmitted from the vehicle to the PSAP, incoming calls are typically answered by the PSAP side requesting information about the exact location (e.g. “Emergency control center, where is your emergency”). The driver, witnessing an accident, will then reply with the requested information (“An accident has happened on the…”), the PSAP requests further information, which is again provided by the calling side (see also [3]). The typical time structure of such a conversation is shown in **figure 2**. The red signal represents the PSAP signal, the green signal driver’s voice.

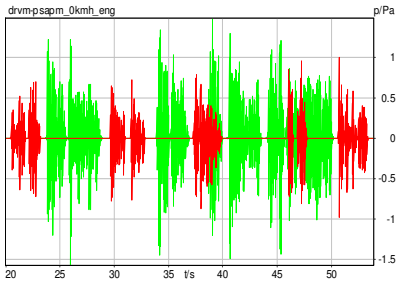

Figure 2: Structure of conversation, time signal (red: PSAP operators’ voice, green: drivers’ voice)

The speech of a simulated PSAP operator was recorded in a quiet, semi anechoic test room. Vice versa driver’s voice was recorded in a driving simulator, where the test persons were exposed to the driving noise at different speeds via closed headphones in order to initiate the Lombard effect. Recordings were made in Russian, English and German language with three speakers each, one male voice representing the PSAP operator and a male and female voice for the driver. The context of the conversations is nearly identical for the three languages except language specific expressions. The conversations were reviewed by native Russian, English and German partners and colleagues.

The active speech level of PSAP and driver’s voices was adjusted to -4.7 dB_{Pa} at the mouth reference point on

sentence basis. This eliminates the level offset caused by the Lombard effect for driver's voice while still maintaining other Lombard speech characteristics (pitch frequency shift, temporal structure, …). The exact speech level for driver's voice was adjusted during the tests using a 3 dB level offset for hands-free use according to ITU-T P.340 [4] and based on the vehicle specific noise level N at different speeds using formula (1) according to MGS R55531 and ITU-T P.1140 [5]:

$$I(N)=\begin{cases}0 & for \quad N<50\\ 0.3(N-50) & for \quad 50\le N<77\\ 8.0 & for \quad N\ge 77\end{cases} \qquad (1)$$

I(N): speech level increment
N: A-weighted noise level in test vehicle

The test setup for the speech recordings is shown in **figure 3**. The vehicle with the installed IVS is equipped with a background noise (BGN) simulation system according to ETSI EG 202 396-1 [6]. The IVS is connected to a network system simulator (2G connection according to [1]). An artificial head measurement system is positioned on the driver's seat in order to playback the pre-recorded utterances from driver's side. The measurement system and the noise simulation system are time synchronized in order to apply further noise processing (Time-synchronized Noise Compensation TNC, see [7].

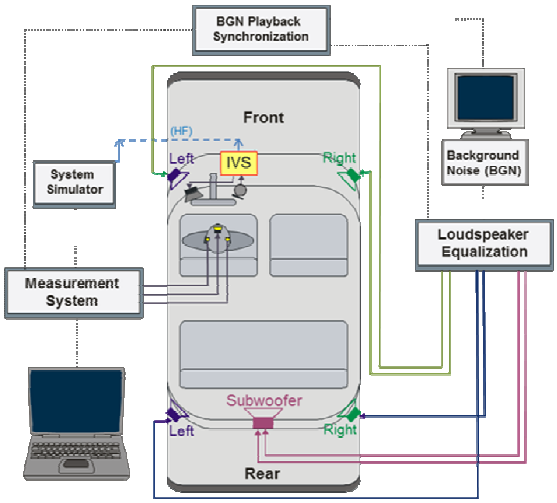

Figure 3: Test setup for speech recordings

The downlink signal inserted by the measurement system represents the PSAP speech (see **fig. 2**), which is bandlimited in order to simulate a typical handset in sending direction. Recordings were carried out binaurally in the vehicle with the HATS on the driver's seat. The uplink signal is recorded at the output of the network simulator. To simulate the PSAP side a binaural recording is generated using this uplink signal by applying a filter to simulate the frequency response of handset in receiving direction and by adding the sidetone of the PSAP speaker.

The course of the conversation highly depends on the uplink and downlink delay of the IVS under test which needs to be considered, in particular during changes of speech transmission direction. The time-synchronization of the test signals is shown in **figure 4**.

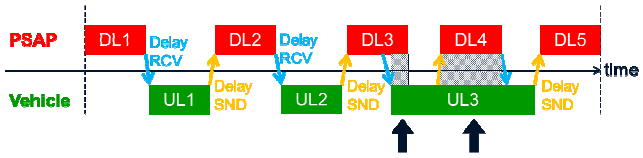

Figure 4: Time synchronization during recording process (DL: downlink signal, UL: Uplink signal)

In the test setup shown in **fig. 3**, the first driver's utterance (uplink signal 1, "UL1") needs to be delayed by the receiving delay ("Delay RCV") after the PSAP signal ends ("DL1"). Vice versa, the second PSAP utterance ("DL2") needs to be additionally delayed by the uplink delay ("Delay SND") and so forth. The time-synchronization is in particular important in order to ensure exactly the same double talk signals for each IVS under test (black arrow in **fig. 4**).

Listening Test

Recordings with the pre-recorded conversations in Russian, English and German language were carried out over 4 different devices. For simplification the tests were first carried out only for male voices and driver position, passenger position and female voices were skipped for most of the test cases. One aftermarket hands-free device providing good double talk performance was used (designated as "HFT" in the following), a one-Box solution with microphone and loudspeaker integrated in the same housing ("OneBox") and two different in-vehicle systems (IVS1, IVS2) were used. These two IVS are installed in two different vehicle, the "HFT" and "OneBox" solution are installed in another, third test car.

The recordings at the driver and PSAP position were assessed by native listeners. Six native German experts judge the German and English recordings (all experts are used to communicate in English), one expert and five naïve Russian test persons judge the Russian recordings. The intention of this first test was the verification of the recordings and the detection of obvious language dependent differences rather than the exact comparison of results. This is planned in a more intensive test run with a higher number of test subjects.

The following parameters were judged [1]:

- Listening effort in receiving, single talk, 0 and 120 km/h
- Listening effort in receiving, double talk, 0 and 120 km/h
- Listening effort in sending, single talk, 0 and 120 km/h
- Speech level variation during double talk, sending, 0 and 120 km/h
- Echo perception during single talk, 0 and 120 km/h
- Echo perception during double talk, 0 and 120 km/h

The following subset of results is discussed with focus on the comparison between languages. It should be considered

that these results only represent a tendency, as they were derived from a limited number of test persons.

Discussion of Results

Receiving direction (driver)

In the following analyses, the mean opinion scores (MOS) averaged over the six test persons are represented by the red (Russian-), grey (German-) and blue bars (English conversation). For information purpose the confidence interval is also indicated, although it must be clearly stated that this parameter does not provide any significance due to the very limited number of test persons without proved Gaussian result distribution.

Figure 5 shows the results for the listening effort in receiving direction under quiet, single talk conditions (0 km/h). The results are all above 4.2 MOS, the listening effort is low. No significant differences can be observed; neither for the four different implementations, nor for the different languages.

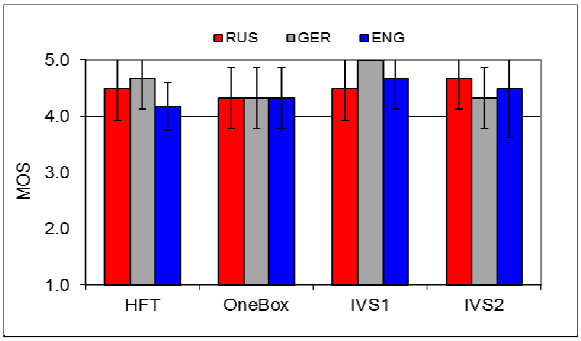

Figure 5: Listening effort in receiving direction (driver, single talk, 0 km/h)

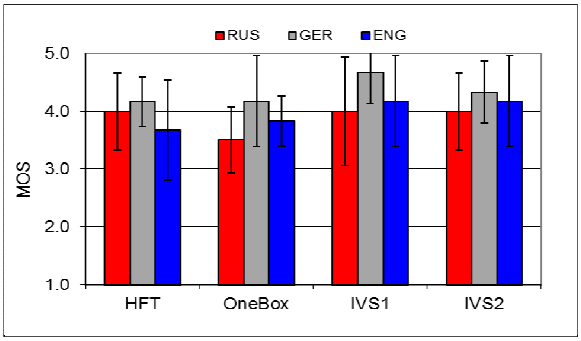

Figure 6: Listening effort in receiving direction (driver, single talk, 120 km/h)

As expected the listening effort decreases for the 120 km/h background noise scenario (**fig. 6**). The results are still high for all implementations, no significant differences can be observed for the three languages. Both IVS implementations seem to provide slightly higher results as these two devices provide a higher overall playback level (IVS2), respectively use an implemented automatic volume control (IVS1).

Listening effort significantly decreases under double talk conditions as shown in **figure 7**. The near end signal, audible as sidetone in the binaural recordings (drivers voice), masks the loudspeaker signal. However, the differences between the languages are not significant, in particular considering the high uncertainty within each test group.

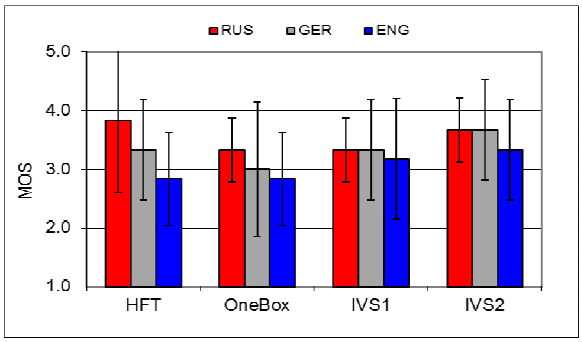

Figure 7: Listening effort in receiving direction (driver, double talk, 0 km/h)

Sending direction (PSAP)

Figure 8 analyses the listening effort results in sending direction (judged at the simulated PSAP side) under silent conditions (0 km/h). The results are again very high for all implementations and all languages. Note, that the confidence interval is now averaged over double the amount of listening examples compared to the receiving direction as for this test case a male and female driver was used.

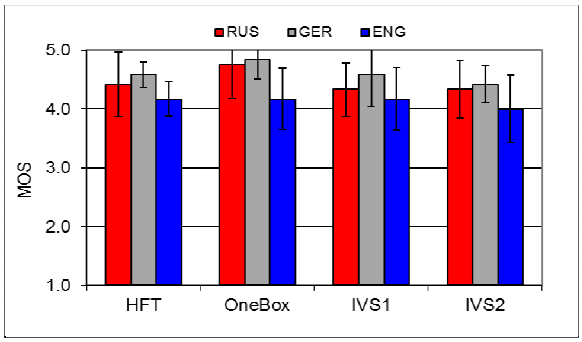

Figure 8: Listening effort in sending direction (PSAP, single talk, 0 km/h)

The test results for the 120 km/h test condition (**fig. 9**) indicate very comparable results for the different languages. The German speech samples seem to be rated slightly higher compared to the English samples for the "HFT" implementation. Furthermore, the three languages consistently indicate, that the listening effort is slightly higher for the IVS1 implementation compared to the other three test devices. This is explainable by the higher residual noise level in the transmitted uplink signal for this device.

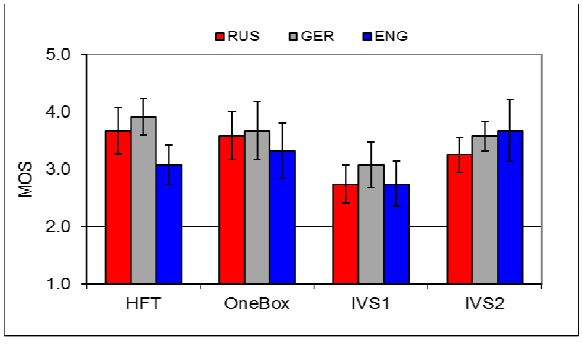

Figure 9: Listening effort in sending direction (PSAP, single talk, 120 km/h)

Significant differences can be detected under double talk conditions. **Fig. 10** clearly shows the limited double talk capability of the "OneBox" implementation. The results for the Russian, German and English conversations are consistent and detect these shortcomings. Both implementations "HFT" and "IVS1" provide good double talk capability (MOS 4: "level fluctuations are audible but not annoying").

Ambiguous results, in particular differences between the Russian and German conversations, appear for "IVS2". The uplink signal in the German conversation is significantly more impaired under double talk conditions than the Russian recording.

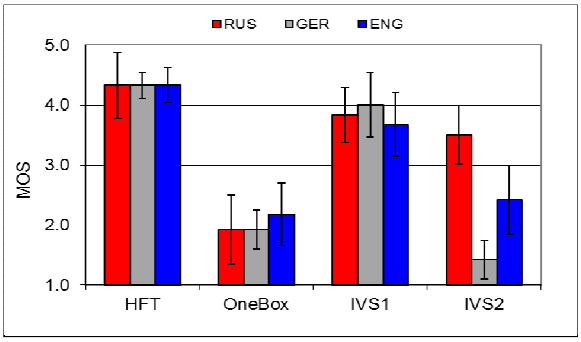

Figure 10: Speech level variation in sending direction during double talk (PSAP, 0 km/h)

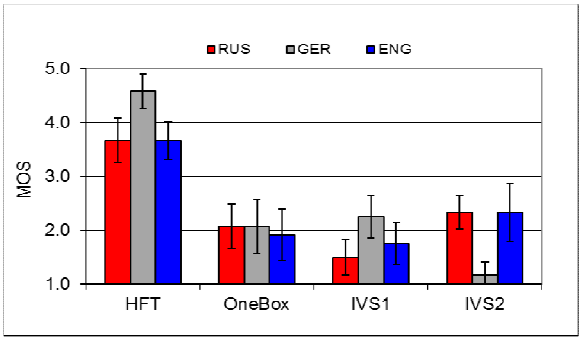

Figure 11: Speech level variation in sending direction during double talk (PSAP, 120 km/h)

The same tendency can also be analyzed with background noise playback simulating a 120 km/h driving noise in the vehicles. The residual echo suppression unit in this implementation is controlled by very short time constants, thus causes the insertion of uplink attenuation depending on the current short term level distribution in the uplink and downlink signal.

The level distribution during the second double talk sequence (DL4 overlapping UL3 in **fig. 4**) is analyzed in **fig. 12** for the three conversations (level vs time curves). It clearly shows, that the speech activity in the downlink signal is significantly higher in the German sequence.

Thus, the near end signal is more often attenuated and suppressed by the implemented echo suppression, which is triggered by the speech activity in the downlink path. This can be harmonized even without new recordings, adapting the German sequence by removing one irrelevant word, thus introducing a similar pause as for the two other languages.

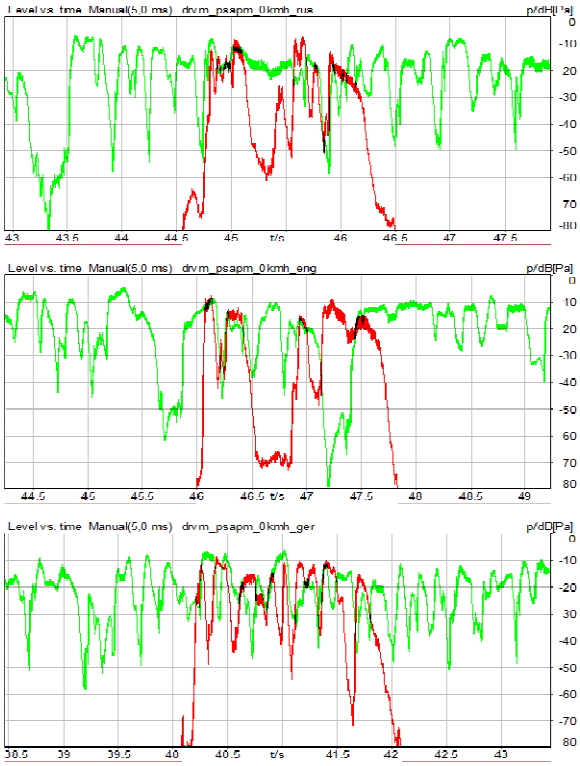

Figure 12: Comparison of double talk sequences, level vs. time (green: UL, driver's voice; red: DL, PSAP voice; Russian (up), English (mid), German (low))

Conclusion

A remarkable effort was spent to design and record conversations in three different languages suitable for TPLT of eCall systems. These recordings shall support IVS optimization and finally also ease the certification process. A first test run indicated already remarkable high consistency of the results for Russian, German and English conversations. The need for slight adaptations of the double talk sequence is identified in order to further improve result consistency for particular IVS implementations. Another test is planned with a higher number of test subjects.

References

[1] ERA-GLONASS Specification MGS R55531, In-vehicle emergency call system compliant test methods for quality of speaker phone in a vehicle, 2015

[2] ITU-T P.832, Subjective performance evaluation of hands-free terminals (05/2000)

[3] F. Kettler, R. Serafimov; Subjective IVS Testing Procedure using Different Languages; ETSI TC STQ Workshop on Telecommunication Quality beyond 2015, Vienna, October 2015

[4] ITU-T P.340, Transmission characteristics and speech quality parameters of hands-free terminals (05/ 2000)

[5] ITU-T P.1140, Speech communication requirements for emergency calls originating from vehicles (07/ 2015)

[6] ETSI EG 202 396-1, Speech and multimedia Transmission Quality (STQ); Speech quality performance in the presence of background noise; Part 1: Background noise simulation technique and background noise database (02/11)

[7] U. Müsch, F. Kettler, S. Bleiholder; Applications for Time–synchronized Noise Compensation (TNC); DAGA 2016, Aachen, March 2016