A Background Noise and Impulse Response Corpus for Research in Automotive Speech and Audio Processing

Tobias Hübschen, Marco Gimm and Gerhard Schmidt
Digital Signal Processing and System Theory, Kiel University, Germany, Email: {thu,mgi,gus}@tf.uni-kiel.de

Introduction

Hands-free communication systems have been part of automotive vehicles for many years. Nevertheless, such systems are still in the focus of researchers and developers as systems are to be improved or optimized for certain use cases, or specific capabilities/functionalities are to be added to existing systems. More recently, in-car communication systems were integrated into select vehicle models in addition to the conventional hands-free communication systems. However, the research and development of such systems remains an ongoing task.

An integral part of developing above-mentioned speech communication systems is to continuously ensure the desired system performance, whether this be with focus on system stability or on speech/communication quality. Generally at the end of the development process, requirements given by, e.g, ITU-T Rec. P.1120 or P.1150 [1, 2] have to be met in addition if a system is to be implemented in commercially available vehicles. For both development and evaluation, specific vehicle-dependent data is required.

With the intention of providing a large amount of such data, this work introduces the Automotive Noise and Impulse Response (ANIR) corpus. This corpus offers both background noise recordings and impulse response measurements from the interior of automotive vehicles and, thus, allows for the simulation or reproduction of the acoustic environments inside the vehicles. This corpus, currently limited to recordings/measurements from a Daimler V Class, is freely available at [3]. By example of the data obtained from the Daimler V class, this work presents the measurement setup used for the creation of the corpus and the specific properties of the obtained background noise recordings and impulse responses. This work closes with a discussion on the applicability of the corpus in its current state and an outlook.

Measurement Setup

The Daimler V Class was equipped with both microphones and loudspeakers for the measurements. The rough location of a microphone/loudspeaker is indicated by the passenger zone in which it is located. According to [2], the vehicle has a total of eight passenger zones in three rows; two zones in front and three zones each in the middle and back row. In the following, passenger zones are referred to by their assigned number as given in Fig. 1. Precise locations of microphones/loudspeaker in a 3D-space are, additionally, available via [3].

The vehicle itself was retrofitted to accompany a multitude of potential microphone/loudspeaker positions for hands-free or in-car communication system usage. In

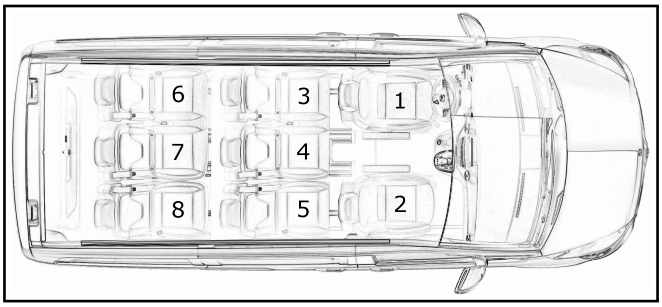
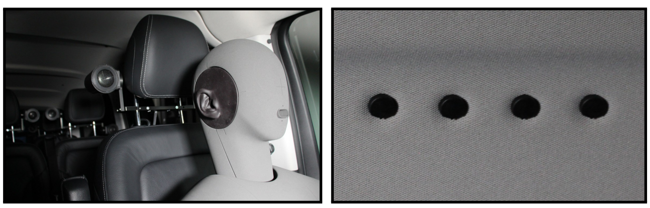


Figure 1: Indication of passenger zones in a Daimler V Class according to [2].

all outer zones, namely zones 1, 2, 3, 5, 6, and 8, a Mechakustik EBM1 [4] microphone was mounted to the inside of the vehicle's roof. In each zone, these microphones were placed laterally centered but a few centimeters in front of a passenger's expected head position. In zones 4 and 7, microphone arrays were installed instead of the single microphones. Such a microphone array (Fig. 2) consists of four microphones with an equidistant spacing of 4 cm and is in alignment with the other microphones in the same passenger row. An identical array was installed in the first row between zones 1 and 2 and an additional single microphone was placed on the B-pillar in zone 1. Consequently, a total of 19 microphones were installed as part of the vehicle's retrofit.

Furthermore, each passenger zone was equipped with a pair of headrest loudspeakers (Fig. 2), where the left and right loudspeaker can emit signals individually. These loudspeakers are a custom development with a 3D-printed housing, which also contains an amplifier. Consequently, a total of 16 loudspeakers were installed as part of the vehicle's retrofit, while only the 12 loudspeakers installed in the outer zones were used for the creation of the corpus. The four loudspeakers built in by the manufacturer were, additionally, made available for third-party signals, while maintaining full functionality of the vehicles original audio system, by inserting a digital signal



(a) (b)

Figure 2: Measurement setup in Daimler V Class. Headrest loudspeaker and head-and-torso-simulator in (a) and microphone array in (b).

Figure 3: View from the trunk of the retrofitted Daimler V Class showing headrest loudspeakers and the 18 microphones installed in the roof.

processor [5] into the system. Fig. 3 gives an impression of the vehicle after completion of the retrofit.
In order to simulate human listeners and speakers, all outer zones were equipped for the measurements with either a head-and-torso-simulator (HATS) according to [6] as depicted in Fig. 2 or with a binaural headset [7]. Consequently, there were 12 ear-microphone and six mouth-loudspeaker positions available for the measurements.
During the measurements, signals were fed to loudspeakers and were collected from microphones using synchronized high-resolution sound cards. All recordings and measurements were performed using the Kiel Real-Time Application Toolkit [8]. This toolkit also ensured the calibration of the microphone channels. The microphones were calibrated using a 1 kHz calibrator in such a way that a sound pressure of 114 dBSPL corresponds to an unweighted digital power level of -10 dBov.

Background Noise

Background noise was recorded synchronously at 48 kHz for all 31 available microphones during a number of driving sessions where the current driving conditions were logged. The entire recordings were post-processed offline to include ear-microphone equalization in the corresponding channels. For this purpose, linear-phase versions of the equalization filters were generated and applied. Due to the linear-phase property of the filters, it was possible to compensate the (group) delay of the filters by an additional time shift. Thus, the 12 equalized channels maintained their synchronicity with the other 19 microphone channels.
After equalization, the recordings were cut into 30 s long segments according to the logged driving conditions. Regarding driving speed, the database contains conditions with 0, 30, 50, 60, 70, 80, 100, 120, 150, 180, and 200 km/h. The smoothed power spectral densities for some of these conditions are depicted in Fig. 4. For each of the driving speeds, the air conditioning system was either set to low, medium, or high intensity (level 1, 4, and 7, respectively, in a Daimler V Class) while the vents were directed towards the windows, where possible. The smoothed power spectral densities of conditions between which only the AC intensity was varied are depicted in Fig. 5. By combining the 11 driving speeds with the three air intensities, a total of 33 database conditions was created. The conditions are available as wave files in 32-bit IEEE float format.

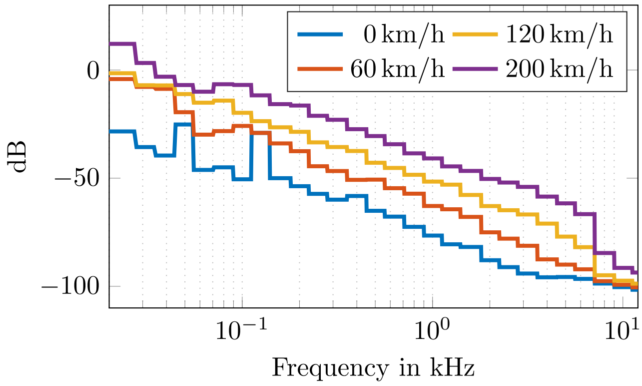


Figure 4: Terz-smoothed power spectral densities at the roof microphone in zone 1 (driver) for low AC setting with various speeds.

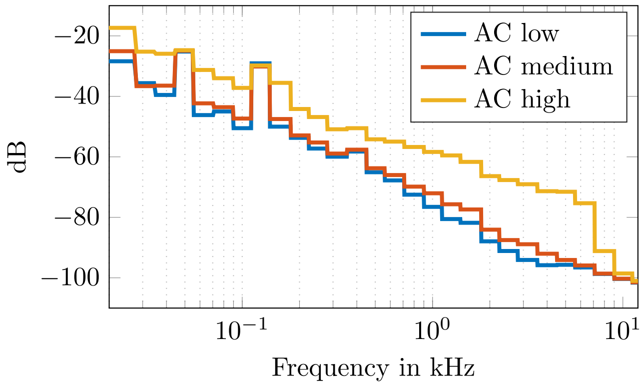


Figure 5: Terz-smoothed power spectral densities at the roof microphone in zone 1 (driver) for 0 km/h and with various AC intensities.

Impulse Responses

Impulse responses were measured between all loudspeaker/microphone combinations, yielding a total of 682 impulse response files. They were estimated using the Normalizes-Least-Mean-Square (NLMS) algorithm at 48 kHz sample rate. After obtaining the original estimations, a number of post-processing steps were applied.
For all measurements, the delay induced by the measurement equipment was compensated. For measurements where ear microphones were involved, the linear-phase equalization filters were applied and the filter delay was compensated analogously to the background noise recordings. For measurements where mouth loudspeakers were involved, the delay of the mouth-equalization filters was compensated. Mouth-equalization filters were, however, not applied as post-processing as they were placed in the digital signal path during the impulse response measurements. The linear-phase mouth-equalization filters had been estimated beforehand as part of the preparation for the measurements. An additional post-processing of impulse responses where mouth-loudspeakers were involved was a calibration in such a way that a digital speech excitation of -26 dBov corresponds to an average sound pressure of -1.7 dBPa at the mouth reference point (MRP). To achieve this, calibration factors had been determined beforehand alongside the mouth-equalization filters. The calibration of the microphones as described in the measurement setup section

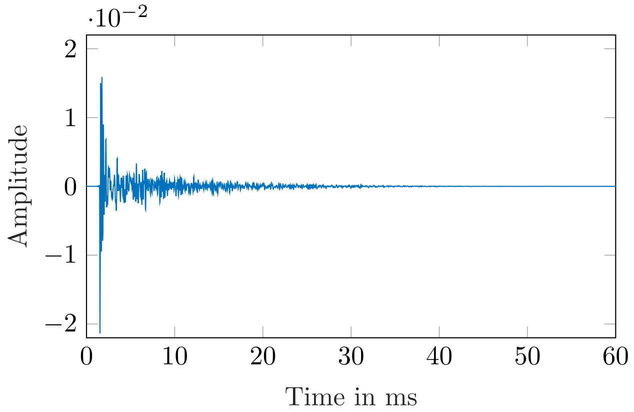


Figure 6: First 60 ms of the impulse response between the left headrest louspeaker and the roof microphone in zone 1.

also holds for the impulse response measurements. The measured impulse response of a common echo/feedback path in the Daimler V Class is depicted in Fig. 6.
Impulse responses are available as binary files in float format. Example scripts reading in impulse response files are provided for both MATLAB© and Python.

Discussion

The previous sections presented how the background noise and impulse responses were recorded/measured and what data is currently available in the ANIR corpus. The following non-exhaustively outlines use cases for the described data.
Generally, the corpus may be used to digitally simulate the acoustic environment inside a Daimler V Class. Acoustic sources in such simulations are either loudspeakers which can be fed by, e.g., a communication/audio system, or HATS loudspeakers which qualify to simulate human speakers according to [6]. Accordingly, acoustic sinks in such simulations are either microphones which can feed, e.g., communication systems, or HATS microphones which qualify for receiving signals as human listeners according to [6]. Thus, simulated microphone signals may be rated according to [9, 10] if desired. If simulated signals are presented to human subjects in this context, an additional equalization of the playback system should be implemented. Other subjective or instrumental analyses of the signals, are, of course, also possible.
If instead of a full simulation a physical sound field reproduction is desired, methods similar to ETSI TS 103 224 [11] may be used to obtain the sound field inside the vehicle when situated in a laboratory environment. The length of the background noise conditions in the corpus is derived from [11] and the length may be extended by applying a suitable looping method as suggested in [2].
Regarding speech communication systems, pretests for certification according to ITU-T Rec. P.1120 or P.1150 [1, 2] are simplified by form of simulations, which can accompany an entire development process. Users must simply select the correct impulse responses for their applicable acoustic signal paths (e.g. input, output, echo, or feedback) and the desired noise files at the corresponding microphones. Due to the multitude of microphone positions in the corpus, especially new microphone positions and array applications can be investigated. Regarding in-car communication system testing, the corpus is designed to enable many of the testing procedures recommended in [2]: All background noise user scenarios required by the recommendation are contained in the corpus. The impulse responses were calibrated in such a way that the required speech level at the HATS MRP can be ensured. Furthermore, the mandatory inter-zone testing configurations can be achieved as HATS impulse responses exist between the mouth of the driver and the ears of the passengers in all outer zones.
Consequently, the corpus in its current state provides valuable data for research and development of speech communication and audio systems in the Daimler V Class. The applicability of the corpus will be improved in the future by repeating the described measurements for other vehicles of interest.

References

[1] ITU-T Rec. P.1120: Super-wideband and fullband stereo hands-free communication in motor vehicles, 2017

[2] ITU-T Rec. P.1150: In-car communication audio specification, 2020

[3] Automotive Noise and Impulse Response (ANIR) Corpus, URL (last access on 24.03.2022): `https://dss-kiel.de/index.php/media-center/data-bases/anir-corpus`

[4] Mechakustik EBM1 datasheet, URL (last access on 24.03.2022): `https://dss-kiel.de/images/media_center/data_bases/anir/downloads/v_class/mechakustik_ebm_1_a.pdf`

[5] Audiotec Fischer: HELIX VEIGHT DSP MK2 datasheet, URL (last access on 24.03.2022): `https://dss-kiel.de/images/media_center/data_bases/anir/downloads/v_class/HELIX-V-EIGHT-DSP-MK2_Web.pdf`

[6] ITU-T Rec. P.58: Head and torso simulator for telephonometry, 2021

[7] Head acoustics GmbH: BHS II, mobile headset for aurally accurate sound recordings, URL (last access on 24.03.2022): `https://www.head-acoustics.com/products/artificial-head-binaural-recording/bhs-ii`

[8] Kiel Real-Time Application Toolkit (KiRAT), URL (last access on 24.03.2022): `https://dss-kiel.de/index.php/research/realtime-framework`

[9] ITU-T Rec. P.800: Methods for subjective determination of transmission quality, 1996

[10] ITU-T Rec. P.863: Perceptual objective listening quality prediction, 2018

[11] ETSI TS 103 224: Speech and multimedia Transmission Quality (STQ); A sound field reproduction method for terminal testing including a background noise database, 2014