Description of the VS10 dataset (audio+annotations)

General about the VS10 dataset

The VS10 dataset contains video recordings of ten vehicles (*Citroen C4 Picasso*, *Mazda 3 Skyactive*, *Mercedes AMG 550*, *Nissan Qashqai*, *Opel Insignia*, *Peugeot 3008*, *Peugeot 307*, *Renault Captur*, *Renault Scenic* and *VW Passat B7*) passing by the camera at a known constant speed. Specification of vehicles (engine type, power, transmission type and production year) is given in Table 1 in the paper. The dataset comprises **304** video recordings, each containing a single drive of a single vehicle. The speed of vehicles ranges from 30 to 105 km/h, with the exact values presented in the rightmost column in Table 1 in the paper. The speed is maintained stable by the on-board cruise control, all vehicles were equipped with.

Annotation text files contain the speed of a vehicle and its pass-by-camera instant. Relative time from the beginning of the file, given in seconds with a two-decimal precision, was measured. Precise annotations were obtained by visual screening, i.e., by identifying a video frame when a vehicle starts to exit the camera view, which approximately corresponds to the closest point of approach.

Folder audio+annotations

Folder **audio+annotations** contains 12 zip archives with audio files (44100 Hz sampling rate, WAV format, 32-bit float PCM), extracted from video files, and corresponding annotations. In addition to 10 archives with audio files containing the sound of vehicles passing by the camera, we provide archives **NoCar.zip** and **NoCarTest.zip** with additional audio files (without corresponding video) containing only environmental noise (no vehicles passing by the camera). The audio files were used in the experiment described in the paper (Section 4) to train and test the proposed audio-based vehicle speed estimation method.

Our estimation method was evaluated using 10-fold cross validation. To that end, we split files to training and validation files, 80%-20% split. The split procedure is as follows: i) sort all speeds of a vehicle into ascending order, ii) divide the sorted speeds into batches of 5 speeds, iii) randomly select one speed in each batch to be used for validation, the other ones for training. This strategy ensures that low-, medium- and high-speed audio of each vehicle are used in both training and validation. Each archive contains a file **Train_valid_split.txt**, which associates labels *train* or *valid* with each audio file.

Naming convention for dataset files includes the vehicle name and the speed. For example, Peugeot307_79.wav and Peugeot307_79.txt represent the names of audio and annotation files, respectively, of Peugeot 307 driven at 79 km/h.

Total size of 11 archives is 1.01 GB.