AUDIO DATA AND SOFTWARE END USER LICENSE AGREEMENT

SCOPE

This End User License Agreement ("License") granted by and between Tampere University of Technology ("TUT") to the End User.

By accessing or using the Database and/or Software Tools, the End User acknowledges that the End User has read, understood, and agrees to be bound by this License.

DEFINITIONS

”Database” means the TUT database described in the Appendix 1 composed of recordings of everyday audio scenes and their annotations, disclosed on a voluntary basis by the concerned persons. All data contained within Database have been collected and processed in accordance with the laws applicable in Finland.

“End User” means an individual or legal person using the Database and/or the Software Tools as a single user on an individual computer or as multi–user on several individual computers or workstations.

“Software Tools” means the software described in Appendix 1.

“Licensed Materials” means the Database and Software Tools.

COPYRIGHT AND LICENSE

Title to, ownership of and all intellectual property rights and other interests in the Licensed Materials belong to and remain the sole and exclusive property of TUT or its licensors, as the case may be. TUT and its licensors retain all rights in the Licensed Materials, including but not limited to distribution rights, not expressly granted to End User in this Agreement.

Subject to the terms and conditions of this License, TUT grants to the End User a perpetual, free of charge, non-exclusive, non-transferable, non-assignable, non-sub-licensable license under TUT’s intellectual property rights in the Licensed Materials to use the Licensed Materials for its own internal and non-commercial use and for the purpose of scientific research. The License does not include the right to use the Licensed Materials either alone or in the connection with creation of commercial products. A separate special agreement may be concluded thereon. However, small portions may be incorporated scientific publications as long as the publication complies with terms and conditions in this License (Publication).

ACCESS

The Database is accessible at the following address: http://www.cs.tut.fi/sgn/arg/dcase2017/ .

DISTRIBUTION

No restrictions are made on disclosure of the Licensed Materials to others. However, End User will not transfer nor permit access to the Licensed Materials over a remote connection to any third party.

TECHNICAL SUPPORT

TUT has no obligation to furnish End User with any technical support whatsoever. Any such support is subject to separate agreement between the parties.

FEEDBACK

Upon using the Licensed Materials, on a voluntary basis End User may provide feedback on the Licensed Materials to TUT and mark such report, concept, finding or other idea as “Feedback to TUT” (collectively, “Feedback”). Anything not marked as “Feedback to TUT” is not Feedback. If End User provides TUT with any Feedback, End User grants TUT, under End User’s intellectual property rights in such Feedback, a non-exclusive, perpetual worldwide license to use such Feedback for TUT’s internal research, promotion, and business.

PUBLICATION

If End User publishes results obtained by using the Licensed Materials and/or derivate works (as described above), including, but are not limited to, research papers, articles and presentations for conferences or educational purposes, End User will acknowledge the Licensed Materials by citing to one or more reference available at http://www.cs.tut.fi/sgn/arg/dcase2017/, or as may be provided from time to time by TUT.

ILLEGAL OR CRIMINAL USE OF THE DATABASE

Any illegal or criminal use of the Licensed Materials by End User is strictly prohibited.

THIRD PARTY LIMITATIONS

The Licensed Materials may include third party materials, including, for example, ambient music and speech from radio or TV (“Third Party Content”). TUT will notify End User via respective webpage if TUT becomes aware of any third party requirements for use of such Third Party Content. The Licensed Material is the scientific research work and any Third Party Content is incorporated based on the “fair use” and “citation right” –concepts.

LEGAL DISCLAIMER

The Licensed Materials are provided without any warranty. TUT will not be held responsible for any damage (physical, financial or otherwise) caused by the use of the Licensed Materials. TUT will not be held responsible of any criminal or illegal use of the Licensed Materials by End User.

AMENDMENTS AND MODIFICATIONS

TUT is allowed to change this License at any time. TUT will inform about any changes on its website..

EXPORT CONTROL

End User will follow all export control laws and regulations relating to the Licensed Materials.

Appendix 1

LICENSED MATERIALS

Description of the Database

The Database includes the following and any other data that TUT makes available to End User under this License:

The data consists of 15 hours of binaural audio, recorded in 15 different everyday environments: lakeside beach, inside bus, city center, cafe/restaurant, inside car, forest path, grocery store, home, library, metro station, office, park, residential area, train and tram. There is approximately one hour recorded in each of these. There are a total of 270 recordings of average length 3.5 minutes. All locations for recordings are in Finland. The recordings contain sounds that are usually occurring in these environments: natural sounds (birds, wind, etc), traffic sounds and sounds related to human presence (footsteps, conversations, laughter, etc). In private areas (home, office), the recording was made with the approval of the recorded person(s).

The Database includes annotations of all audible events in street (residential area and city center) environments: the start and end times of each sound event, and a description of the type of event.

Description of the Software Tools

The Software Tools include the following and any other software that TUT makes available to End User under this License:

A baseline classification and detection systems and evaluation is provided along with the Database, as well as software for creating mixtures of sound events and backgrounds. The baseline system implements three applications using shared code base: 1) baseline acoustic scene classification subsystem 2) binary sound event detection subsystem, and 3) sound event detection subsystem. The subsystems include training stage and testing stage with evaluation using cross-validation or development data.

Two versions are provided for each subsystem: 1) multilayer perceptron (MLP) neural network based, and 2) Gaussian Mixture model (GMM) based.

The three subsystems are based on the same acoustic feature extraction algorithm and classification algorithms. Mel-frequency cepstral coefficients (MFCC) are used to represent the coarse shape of the power spectrum of the acoustic signals. Gaussian Mixture models (GMM) and multiayer perceptron (MLP) are provided as options for modeling class-conditional densities of these features. The MLP-based solutions are based on the same network architecture.

In the MLP-based acoustic scene classification subsystem, a network containing two dense layers of 50 hidden units per layer is implemented. Classification decision is based on the network output layer which is of softmax type. In the GMM-based acoustic scene classification subsystem, the acoustic scene-conditional densities of MFCCs are modeled with GMMs. In the classification stage, the likelihood of test signal coming from each modeled scene class is evaluated and classification is done by maximum-likelihood classifier.

In the MLP-based binary sound event detection subsystem,a network with two dense layers of 50 hidden units per layer is implemented, and the detection is based on an output layer containing a single sigmoid unit. In the GMM-bases binary sound event detection subsystem, presence of absence of the target event class is modeled using two GMM models: one trained with acoustic material where target sound event is active and one trained with acoustic material where the target sound event is inactive. In the detection stage, likelihood ratio between these two models is used to get active region for the target sound event frame by frame.

In the MLP-based sound event detection subsystem, a network with two dense layers of 50 hidden units per layer is implemented, and the detection is based on anutput layer containing sigmoid units that can be active at the same. In the GMM-based sound event detection subsystem, for each sound event class two GMM models are trained: one trained with acoustic material where sound event is active and one trained with acoustic material where sound event is inactive. In the detection stage, likelihood ratio between these two models is used to get active sound event for each processing frame, for each event class.

The implementation is provided for Python platform. The implementation is dependent on standard Python libraries and external libraries: Keras, Theano, tensorflow, numpy, scipy, scikit-learn, pyyaml, librosa, soundfile and sed_eval.

The mixture generation software performs creation of mixture recipes, which include information on the presence of the target event in the mixture, its temporal position, the amplitude scaling factor of the event audio signal, the temporal position of the event in the original recording. Based on the created mixture recipes, the software then performs synthesis of audio files of these mixtures. The Python implementation is dependent on standard Python libraries and external libraries: yaml, librosa, soundfile, numpy, pandas, tqdm, hashlib.