

COMP.SGN.220 Advanced Audio Processing

Exam 1.3.2024

Exam prepared by Tuomas Virtanen

Use of a non-programmable calculator is allowed (but not necessary for the completion). Indicate if you have completed the project work or exercises of the course some previous year.

1. For each of the three tasks (a-c) below that will be solved by a deep neural network based acoustic model, different kind of acoustic features are used in each task. Name an acoustic feature that would be good for each task. Describe why it would be good. Use maximum *one sentence* per each answer. (6 p.)
 - a) Estimating which animal species are present in a 10-second audio clip.
 - b) Estimating the audio signals of individual animals in a 10-second audio clip.
 - c) Estimating the spatial positions of different animals in a 10-second audio clip consisting of multiple audio channels recorded with closely-spaced microphones.

2. A CNN-based deep neural network is used for an animal monitoring application to estimate which of the target sound classes "dog barking", "cat meowing", or "bird chirping" are present in a 10-second audio clip. Describe the most important processing steps of the method at the inference stage. (6 p.)

3. You are developing a smartphone app to implement the monitoring system of question 2. Describe what kind of data you will need to train the system, so that it will work well on smartphones with a wide range of users. Describe also how you can obtain such data. (6 p.)

4. You are developing an RNN-based audio enhancement system for bird watchers. that will take a 10-second audio recording containing bird sounds as an input and will output a corresponding audio signal where background noises are suppressed. Describe the most important processing steps that the system uses at the usage stage. (6 p.)

5. Your task is to design a source localization system that will estimate the spatial coordinates of a sound source with respect to a microphone array, in a 10-second audio clip, as the function of time. The audio is captured with a mobile phone sized microphone array with three microphones which coordinates are known. Describe how the system will produce the estimates of coordinates, given the input audio. In case you are using a machine learning based approach, describe also how you train the system. You can decide yourself what kind of representation of spatial coordinates is used. (6 p.)