**RL4Audio Project: Acoustic Event Detection**

This project is a feasibility study that investigates possible *Reinforcement Learning* applications in the audio domain. The targeted task for this project is *acoustic event detection.* An agent has to select the onset, offset, and a label for an event and based on these actions, a reward will be calculated and is further used to lead the agent to achieve the goal of detecting acoustic events.

**Hardware requirements:**

PC equipped with GPU.

**Time requirement:**

Period of at least 3 months (minimum 20-30 hours/week).  
Starting from July 2019 or later.

**Candidate requirements:**

* Python programming
* Pytorch
* Familiarity with reinforcement learning
* Familiarity with acoustic event detection
* Willingness to learn new topics

**Study material:**

1-Reinforcement learning (RL):

* Basics of DL ([Book](https://www.deeplearningbook.org/) by Ian Goodfellow)
* Basics of RL ([Book](http://incompleteideas.net/book/RLbook2018trimmed.pdf) by Sutton and Barto)
* RL Lectures by [David Silver](http://www0.cs.ucl.ac.uk/staff/d.silver/web/Teaching.html) or from [Berkley](http://rail.eecs.berkeley.edu/deeprlcourse/)
* [Overview](https://spinningup.openai.com/en/latest/spinningup/keypapers.html) on state-of-the-art RL literature (no in depth knowledge required)
* [Object Detection](http://openaccess.thecvf.com/content_cvpr_2018/papers/Pirinen_Deep_Reinforcement_Learning_CVPR_2018_paper.pdf) and [Object Tracking](http://openaccess.thecvf.com/content_cvpr_2017/papers/Yun_Action-Decision_Networks_for_CVPR_2017_paper.pdf) with RL  
  (The first steps of the project will be to implement and solve a small toy problem for detecting objects in images with RL, before we continue with Audio)

2-Acoustic event detection:

* Reading the following articles:
  1. Polyphonic sound event detection using multi label deep neural networks, E Cakir, T Heittola, H Huttunen, T Virtanen, 2015 international joint conference on neural networks (IJCNN)
  2. Metrics for Polyphonic Sound Event Detection, A Mesaros, T Heittola, T Virtanen, Applied Sciences
* Familiarity with the tasks below and being able to reproduce the baseline results:
  1. <http://www.cs.tut.fi/sgn/arg/dcase2016/task-sound-event-detection-in-synthetic-audio>
  2. <http://www.cs.tut.fi/sgn/arg/dcase2016/task-sound-event-detection-in-real-life-audio>
  3. <http://dcase.community/challenge2019/task-sound-event-detection-in-domestic-environments>

**IL4Audio Project: Audio Synthesis**

This project is a feasibility study that investigates possible *Imitation Learning* applications in the audio domain. The targeted task for this project is *acoustic event/scene audio synthesis.* An agent has to choose a set of parameters in an audio synthesizer to synthesize an audio piece based on a recoding example. Based on these actions, a reward will be calculated and is further used to lead the agent to achieve the goal of synthesizing audio pieces from real recordings.

**Hardware requirements:**

PC equipped with GPU.

**Time requirement:**

Period of at least 3 months up to 6 months (minimum 20-30 hours/week).  
Starting from July 2019 or later.

**Candidate requirements:**

* Python programming
* Pytorch
* Familiarity with imitation learning
* Familiarity with reinforcement learning
* Familiarity with acoustic event detection/acoustic scene classification
* Willingness to learn new topics

**Study material:**

1-Imitation Learning (IL):

* Basics of IL ([ICML Tutorial](https://www.youtube.com/watch?v=WjFdD7PDGw0), [MSR Talk](https://www.youtube.com/watch?v=4PnNlvPGbUQ))
* Familiarity with the following papers and their implementations:
  1. [One-Shot Imitation Learning](https://arxiv.org/abs/1703.07326)
  2. [One-Shot Visual Imitation Learning via Meta-Learning](https://arxiv.org/abs/1709.04905)
  3. [Generative Adversarial Imitation Learning](https://arxiv.org/abs/1606.03476)

2-Reinforcement learning (RL):

* Basics of DL ([Book](https://www.deeplearningbook.org/) by Ian Goodfellow)
* Basics of RL ([Book](http://incompleteideas.net/book/RLbook2018trimmed.pdf) by Sutton and Barto)
* RL Lectures by [David Silver](http://www0.cs.ucl.ac.uk/staff/d.silver/web/Teaching.html) or from [Berkley](http://rail.eecs.berkeley.edu/deeprlcourse/)
* [Overview](https://spinningup.openai.com/en/latest/spinningup/keypapers.html) on state-of-the-art RL literature (no in depth knowledge required)

3-Acoustic event/scene analysis:

* Reading the following articles:
  1. Polyphonic sound event detection using multi label deep neural networks, E Cakir, T Heittola, H Huttunen, T Virtanen, 2015 international joint conference on neural networks (IJCNN)
  2. Metrics for Polyphonic Sound Event Detection, A Mesaros, T Heittola, T Virtanen, Applied Sciences
  3. Acoustic scene classification with fully convolutional neural networks and I-vectors, M Dorfer, B Lehner, H Eghbal-zadeh, H Christop, P Fabian, G, Widmer, Tech. Rep., DCASE2018 Challenge
* Familiarity with the tasks below and being able to reproduce the baseline results:
  1. <http://www.cs.tut.fi/sgn/arg/dcase2016/task-sound-event-detection-in-synthetic-audio>
  2. <http://www.cs.tut.fi/sgn/arg/dcase2016/task-sound-event-detection-in-real-life-audio>
  3. <http://dcase.community/challenge2019/task-sound-event-detection-in-domestic-environments>
  4. <http://www.cs.tut.fi/sgn/arg/dcase2016/task-acoustic-scene-classification>

**GAN4Audio Project: Audio Synthesis**

This project is a feasibility study that investigates possible *Generative Adversarial Networks (GAN)* applications in the audio domain. The targeted task for this project is *acoustic event/scene audio synthesis* given a specific event/acoustic scene label*.* A generator has to learn to generate audio pieces containing specific events/from specific acoustic scenes while a discriminator is trying to distinguish them from real recordings. The two models compete with each other until the generated samples are indistinguishable from real ones.

**Hardware requirements:**

PC equipped with GPU.

**Time requirement:**

Period of at least 3 months up to 6 months (minimum 20-30 hours/week).  
Starting from July 2019 or later.

**Candidate requirements:**

* Python programming
* Pytorch
* Familiarity with GANs
* Familiarity with Raw audio models
* Familiarity with acoustic event detection/acoustic scene classification
* Willingness to learn new topics

**Study material:**

1-Generative Adversarial Networks (GAN):

* Basics of DL ([Book](https://www.deeplearningbook.org/) by Ian Goodfellow)
* GAN [Tutorial](https://arxiv.org/abs/1701.00160) by Ian Goodfellow
* Familiarity with following papers and their implementations:
  1. [GAN](http://papers.nips.cc/paper/5423-generative-adversarial-nets)
  2. [WGAN-GP](https://arxiv.org/abs/1704.00028)
  3. [FID](https://arxiv.org/abs/1706.08500)
  4. [WaveGAN](https://github.com/chrisdonahue/wavegan)
* Additional papers:
  1. [GANSynth](https://arxiv.org/abs/1902.08710)
  2. [Auxiliary Classifier Generative Adversarial Network for Acoustic Event Detection](https://ieeexplore.ieee.org/document/8523637)
  3. [Spectral Norm GAN](https://openreview.net/forum?id=B1QRgziT-)
  4. [BigGANs](https://arxiv.org/abs/1809.11096)
  5. [StyleGAN](https://github.com/NVlabs/stylegan)

2-Raw Audio Models :

* Familiarity with following papers and their implementations:
  1. [Wavenet](https://deepmind.com/blog/wavenet-generative-model-raw-audio/)
  2. [Realistic music generation](https://arxiv.org/abs/1806.10474)
  3. [Efficient Neural Audio Synthesis](https://arxiv.org/abs/1802.08435)

3-Acoustic event/scene analysis:

* Reading the following articles:
  1. Polyphonic sound event detection using multi label deep neural networks, E Cakir, T Heittola, H Huttunen, T Virtanen, 2015 international joint conference on neural networks (IJCNN)
  2. Metrics for Polyphonic Sound Event Detection, A Mesaros, T Heittola, T Virtanen, Applied Sciences
  3. Acoustic scene classification with fully convolutional neural networks and I-vectors, M Dorfer, B Lehner, H Eghbal-zadeh, H Christop, P Fabian, G, Widmer, Tech. Rep., DCASE2018 Challenge
* Familiarity with the tasks below and being able to reproduce the baseline results:
  1. <http://www.cs.tut.fi/sgn/arg/dcase2016/task-sound-event-detection-in-synthetic-audio>
  2. <http://www.cs.tut.fi/sgn/arg/dcase2016/task-sound-event-detection-in-real-life-audio>
  3. <http://dcase.community/challenge2019/task-sound-event-detection-in-domestic-environments>
  4. <http://www.cs.tut.fi/sgn/arg/dcase2016/task-acoustic-scene-classification>