Individual-Level Reconstruction of Vocal Behavior 1

In our lab we monitor the communication of songbirds that are housed together. To know which bird utters which vocalization, we attach miniature backpacks on these birds [1]. Each backpack carries an accelerometer (acc_i) that records vibrations from the body of bird i. From these recordings we want to extract the individual vocalizations. However, the accelerometer signals have properties that render them more difficult to analyze than microphones: they only monitor low-frequency vibrations, are affected by radio noise, and wing flaps of the bird mask all vocal signals. Segmentation and clustering of longitudinal data into recurring patterns is a universal theme in science and medicine [2]. You will develop and evaluate a customized solution standard problems in this field. To inform your algorithm you can use the recordings from a microphone (mic) attached to the wall, which records all vocalizations simultaneously.

Idea:

Radio noise is one of the main problems we face in using the acc_i recordings. However, if only one acc_iat the time experiences radio noise, while all other$acc_{j\neq i}$not, we can infer the vocalization of bird i, by using the microphone data. To do so, the first goal is to learn a reconstructed microphone signal $mic_i = f(acc_i)$, for the trivial case when only one bird vocalizes and there is no radio noise. In a second step, you will use this to learn the full reconstruction mapping $mic_i = g(acc_{1..n}, mic)$.

Data: Each group will use a different data set of mic and acc recordings, which consists of audio files.

Tasks:

1. Introduction: Review literature on the segmentation problem with focus on sound recordings and vocal signals. What is the cocktail party problem? What are denoising strategies? What are the basic characteristics of mapping functions, such as linear regression (simple) and neural networks (complex)?
2. Read and inspect the data provided by the supervisors (convert raw data to spectograms). Describe it in the methods sections. Describe your strategy of tackling the problem.
3. What is radio noise? What is silence? How can they be characterized (Methods)? How frequently does it occur (Results)?
    a. To define silence, look at RMS histograms and set a reasonable threshold.
    b. To define radio noise, look at the radio signal strength signal and try to find a threshold on it, below which the accelerometer signal is noise.
4. Learn $mic_i = f(acc_i)$, for the trivial subset (S_trivial) when only one bird vocalizes and there is no radio noise. You can use a function of your choice, such as linear regression.
5. Apply this function to bird i in the bigger subset (S_clean) when there is no radio noise. Notice that for S_trivial you can simply use $mic_i = mic$.
6. On S_clean, learn to infer the reconstructed signal stemming from one accelerometer, without using its data. In other words, learn the function $mic_i = g(acc_{j\neq i}, mic)$.
    a. as a first idea you can test $g(acc_{j\neq i}, mic) = mic - \sum_{j\neq i} f(acc^j)$
    b. use a neural network to learn $g(acc_{j\neq i}, mic)$

7. Evaluation (Results):

    a. For each bird and all times in S_trivial plot the mic spectrogram on top of g. There should be close correspondence.

    b. For each bird and all times in which bird i is silent but another bird is not, plot the mic spectrogram on top of g. The output g should be zero.

    c. For each bird i and all times at which two or more birds produce sounds, plot the mic spectrogram on top of g.

8. Put your results into perspective (Discussion). What works, what doesn't? Why? Improvements possible?

Requirements/ Organization

- This is a group project (group size: 5 students). Use the google spreadsheet on Moodle to sign up to a group.
- You are required to write a report together with the other group members + one page where you describe your contribution to the project.
- Hand-in deadline for the report: **30th June 2020**
- Grading: The project counts 25% of final grade. The project grade will be calculated from a group grade and an individual grade.

Supervision/ Questions:

Tomas Tomka (tomas@ini.ethz.ch), Linus Rüttimann (rlinus@ini.ethz.ch), Corinna Lorenz (corinna@ini.ethz.ch)

[1] Ter Maat, A., et al. (2014). Zebra finch mates use their forebrain song system in unlearned call communication. PloS one 9.10: e109334

[2] Coffey, K.R., Marx, R.G. & Neumaier, J.F. DeepSqueak: a deep learning-based system for detection and analysis of ultrasonic vocalizations. Neuropsychopharmacol. 44, 859–868 (2019).