M2 Data Science
DS-telecom-15 "Audio and Music Information Retrieval"

What you need to know

- Transformée de Fourier à Court Terme
  - formule, explication
  - influence des paramètres types de fenêtre et longueur de fenêtre
  - paradoxe temps/fréquence
  - interprétation passe-bande et passe-bas de la TFCT

- Reconstruction du signal à partir de la TFCT
  - sans modifications: addition/ recouvrement (OLA): principe et formule
  - *(avec modifications: algorithme de Griffin & Lim: principe et formule) -> NON*

- Applications du traitement par TFCT
  - débruitage par soustraction spectrale: principe et formule
  - vocodeur de phase: principe et formule
    - fréquence instantanée
    - comment déterminer le paramètre du unwrapping de la phase

Lecture 2 What you need to know

- What is a pitch
  - what are the Pythagorean and Equal temperaments (tuning-systems), and their pro and cons

- What is the Constant-Q-Transform
  - what problem does it solve?
  - how does it solve it ?

- What are the Chroma/ Pitch-Class-Profile
  - what they represent, how they are computed
  - what are there advantages (easiness of computation) / drawbacks (timbre sensitivity)

- What is a Hidden Markov Model
  - what are the variables that define it
  - what are the three main problems it allows to solve
  - what is the Viterbi decoding algorithm

- How are chords estimated using an HMM approach, a Deep Learning approach ?

- How are the performances of such a system measured ?

Lecture 3 What you need to know

- Audio features
    - what are the Cepstrum, real-Cepstrum, MFCCs
    - what is the relationship between the cepstrum and the source-filter model

- Representations
    - What is a time-time Self-Similarity-MatrixWhat are the homogeneity, repetition assumptions

- Methods
    - What is the checker-board/ novelty curve segmentation method
    - What is the summary score ?

- Deep learning method
    - How to apply a ConvNet for the structure estimation problem
    - How to apply Self-Supervised-Learning for structure estimation (McCallum method)

- Evaluation
    - How to evaluate the performance of a Music Structure Discovery system

Lecture 4 What you need to know

- Music auto-tagging
    - what is the cold-start problem and how to solve this using content-based recommendation

- Deep Learning
    - What are the most common architectures and meta-architectures
    - What is a TCN
    - What is metric learning ? especially what is the contrastive loss, the triplet loss
    - What is few-shot learning ? especially what is a prototypical network
    - What is self-supervised learning ? being able to explain one of the paradigms presented during the lecture (Neural-Autoregressive models, ContrastivePredicting Coding, simCLR, PACE, SPICE, Audio/Visual Synchronization)

- Music Classification
    - being able to describe one of the deep neural network used for Music Auto-tagging

Lecture 5 What you need to know

- Deep Learning
  - Why we say that T/F representations are not like natural images and understand why we need to adapt ConvNet architectures
  - What is the Time-Translation Invariance property and how we can achieve it in the case of Conv1D
  - What is SincNet, Wavenet, Harmonic-CQT and why are they used

- Audio Source Separation
  - What are the SDR, SIR, SAR and how they are computed
  - What is deep clustering
  - What is a masking network for source separation and be able to provide the details of one deep systems that allows to estimate it
    - in the T/F domain (such as U-Net)
    - in the Time domain (such as ConvTasNet)
  - What is an Ideal Binary Mask, an Ideal Ratio/Soft Mask