Tampere University

Project work

SGN 14007

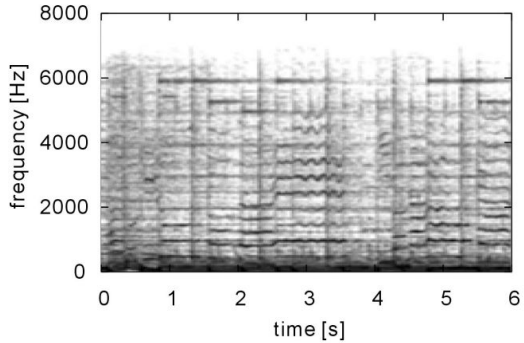
Annamaria Mesaros

Practical arrangements

- Few different topics
  - Implementing an audio signal processing algorithm and writing a report about it
  - Done in two-person groups (doing the project alone is OK too)
  - Marked as Pass / Fail
  - Required for passing the course
  - Select one of the given topics (or propose your own topic to be approved)

- Return the script and report by **Wednesday 18 Dec 2019** (midnight) to Moodle

Topic 1
Separation of drums from music signals



- Separation of harmonic and percussive components from music spectrogram (time-frequency domain)
  - Anisotropy: dependence on direction
  - Temporal continuity of harmonic sounds
  - Temporal localization of percussive sounds (continuity along frequency)
- Iterative search for such spectrograms that maximize the anisotropy (continuity
- over time vs. over frequency) [1]
- Together, the two spectrograms (harmonic / percussive) equal the original spectrogram: $H_{h,i} + P_{h,i} = W_{h,i}$

N. Ono, K. Miyamoto, J. L. Roux, H. Kameoka and S. Sagayama, “Separation of a monaural audio signal into harmonic/percussive components by complementary diffusion on spectrogram,” in Proc. EUSIPCO, 2008

Topic 1
Separation of drums from music signals

Hints to the implementation

- Equations 24-30 of reference [1] describe the algorithm implementation, you do not need to care about the derivation of the iterative algorithm
- Framewise processing (short-time Fourier transform) has been discussed at the lectures and at the exercises
- Test material: drum and harmonic instruments (+ singing) separately, from which the total signal is obtained by mixing (drums + the rest)
  - Note that the algorithm works only for mono signals. If you want to process stereo signals, you may process both channels separately.
  - You may use audio material of your own or signals linked to the course web pages
- Evaluation: signal-to-noise ratio
  - s(t) = original, e(t) = original minus separated

$$\mathrm{SNR} = 10 \log_{10} \left(\frac{\sum_t s(t)^2}{\sum_t e(t)^2} \right)$$

Report:

- What kind of audio material is the algorithm suitable for and why?
- How should the separation quality be measured and assessed?

Topic 2
Pitch detection of monophonic music

- Pitch detection of a segment of monophonic music (single instrument)
    - Use (preferably) YIN method for pitch detection
    - Transcribe the phrase to musical notes
    - Estimate note lengths
    - Create a synthesized signal composed of pure notes with the same pitch and lengths

A. de Cheveigné and H. Kawahara. YIN, a fundamental frequency estimator for speech and music. The Journal of the Acoustical Society of America, 111:1917, 2002. doi:10.1121/1.1458024

Topic 2
Pitch transcription of monophonic music

Hints to the implementation

- Based on the reference paper, use the cumulative mean normalized difference function

$$d'_t(\tau) = \begin{cases} 1, & \text{if } \tau = 0, \\ d_t(\tau) \Big/ \left[(1/\tau) \sum_{j=1}^{\tau} d_t(j) \right] & \text{otherwise.} \end{cases}$$

- Frame-wise processing to calculate pitch in each frame
- Use pitch contour to estimate length of each note
  - Alternatively, you can do onset detection (look up more information on this!)
- Test material: short excerpt of solo musical instrument (not drums!)
  - You may use audio material of your own or signals linked to the course web pages
  - If you want to process stereo signals, you may create the mono as the channels average
- Evaluation:
  - Listen to the synthesized signal and rate the outcome
  - Visually compare spectrograms of the original and synthesized audio

Report:

- What kind of audio material is the algorithm suitable for and why?
- How should the transcription quality be measured and assessed?

Topic 3
Music synchronization

- Synchronization of different performances (A, B, C, ..) of the same musical piece
  - Given piece A (solo singing), annotate start time of each word
  - Infer the annotation to the other ones through synchronization with A
  - Do synchronization using chroma features and dynamic time warping
    - You can also try with MFCC
  - Time stretch all other versions to match A, and run synchronization again
  - Compare warping path between A and original vs time-stretched versions

Chroma, DTW, MFCC and time stretching presented in the course
See also the tutorial Music Synchronization with Dynamic Time Warping in librosa
https://librosa.github.io/librosa/auto_examples/plot_music_sync.html

Topic 3
Music synchronization

Hints to the implementation

- You can use the librosa tutorial for the alignment
- Determine time-stretching factor based on the length of singing
  - You may cut the audio files to eliminate silence at the beginning and at the end
- You can use your own time stretching implementation, or a library version
- Test material: provided short excerpts of singing
  - You may use audio material of your own, for example two versions of the same song, in which case you annotate and infer song structure (intro, verse, chorus, bridge)
- Evaluation:
  - Listen few examples and rate the correctness of the inferred annotations
  - Visually compare the aligned signals

Report:

- What kind of audio material is the algorithm suitable for and why?
- How should the alignment quality be measured and assessed?

Topic 4
Spatial audio: cross-talk cancellation

Playback of binaural recordings on loudspeakers, by canceling crosstalk; also creates the impression that sounds are externalized

- Typically binaural recordings are meant to be played such that sound originating from the left ear is only played back to the left ear; played back with stereo loudspeakers, the sound from the left ear also travels to the right ear (cross-talk), ruining the spatial audio quality
- Method to cancel as much as possible this cross talk:
  - Create rough HRTF filters that spatialize a sound on headphones, based on the head model as a sphere
  - Create a point-like virtual source using the provided audio example
  - Process the binaurally recorded audio to allow listening on loudspeakers

Zölzer (ed.) (2002) DAFX – Digital Audio Effects. John Wiley & Sons, Chapter 5. Binaural Techniques in Spatial Audio; Section 5.4. Matlab code available

Topic 4
Spatial audio: cross-talk cancellation

Hints to the implementation

- You can use the Matlab code as a guide for your Python implementation
- Create HRTF filters based on the head model as a sphere
- Create a point-like virtual source from the provided audio example
- Process the binaural signal to cancel cross-talk

Evaluation:

- Vary the azimuth angle of virtual source; listen examples
- Describe the perceived effects when listening original and processed signals on loudspeakers; comment on differences between spatial audio quality with headphones (binaural) and loudspeakers (processed)

Report:

- What kind of audio material is suitable for obtaining the best effect? Why?
- How should the spatial quality be assessed or improved?

Topic of your choice

- Audio classification (classical or more recent machine learning methods)
  - Advanced audio processing course project!
- Something else?

Email to annamaria.mesaros@tuni.fi describing the planned project, including:

- Problem to be solved
- Data to be used
- Method to be used
- Evaluation procedure

Instructions

1. Read the scientific paper/provided materials
2. Implement the algorithm using Python
3. Evaluate and interpret the results (using some test data)
4. Write a report about the results and your observations

Return

1. Python Script
   - **project_family_names.py**
   - Return only a part of the test material to check the functionality of the algorithm (one signal) even though the evaluation results should be computed using more data
2. Report
   - **project_family_names.pdf**
   - What problem is being solved in the project work?
   - How is the problem solved?
   - What assumptions were made?
   - Short description of the implementation. What stages does the algorithm consist of?
   - What is evaluated in your experiments, and how?
   - Include figures and diagram whenever descriptive
   - Mention in the report how the work was divided in your group
   - Length **maximum 4 pages** (depending on the figures)

General

- Questions regarding the project work in Moodle to course assistants
  - We make an effort to read and answer the questions promptly
  - If the questions are related to Python code, the code has to be very well commented!

- We will try to arrange a support session for answering questions and advising related to the project work (after the exam?).
  - More information about that on the course web page.
  - You may also ask questions at the end of the weekly exercise sessions