**Acoustic Event Detection, Segmentation, and Score-Performance Alignment in Opera Recordings**

**Possible Topics for Bachelor and Master Theses and Praktika**

**Context and Goals**

In the context of a multi-national ***European Research Project*** (<https://mip-frontiers.eu>), we collaborate with the ***Vienna State Opera*** (<https://www.wiener-staatsoper.at/en>) on the topic of ***real-time opera tracking*** (based on machine listening). The goal is to develop algorithms that are given the score (sheet music) of an opera, in machine-readable form, and a live audio stream of an actual opera performance, and can reliably track the current position in the score (for purposes of synchronisation, automatic subtitling, etc.). Opera recordings are complex sound mixtures of instrumental music, singing (soloists, several soloists together, or a whole chorus), interspersed with speech and other sounds and noises, breaks, applause, etc.

In the context of this project, we offer several topics for bachelor's theses, master theses, and/or "Praktika". The topics revolve around ***segmenting opera recordings according to certain sound classes*** (which involves ***building machine learning classifiers***) and trying to ***align recordings and scores*** (or references) based on this segmentation.

In cases where substantial annotation efforts are involved (see below), we can also offer some financial remuneration.

**Possible topics for projects (examples):**

1. Singing voice detection & male-female classification
2. Music, applause, speech detection
3. Melody/F0 extraction for audio-to-score alignment
4. Audio-to-lyrics alignment via singing voice detection and speech understanding

Each of the tasks also involves, as a first step, listening to the operas and peforming a ***manual annotation*** (e.g., marking all segments that contain singing). We will focus on two operas by W.A. Mozart: *Don Giovanni* (recorded at the Vienna State Opera) and *"Die Zauberflöte"* (recorded at the Bruckner University of Music, Linz).

**Prerequisites:**

Knowledge of machine learning, minimal interest in music, good listening capabilities.

To give a more concrete idea of what these projects could look like, we decribe two them in a bit more detail:

**Project 1 - Singing Voice Detection and Male-female Classification**

For Project 1, the first goal is to detect in the recordings the parts where singing voice is present. First the student will investigate the current singing voice detection state-of-the art to extract relevant features with signal processing algorithms. Then machine learning experiments [1] [2] will be performed on already available opera datasets (detailed in [2]) to learn singing voice detection. The models will be extended to learn gender classification. The two opera recordings mentioned above (Don Giovanni and Zauberflöte) will be manually annotated regarding the occurrence of singing voice, which will make it possible to test the learned models on a new dataset. Finally the classifiers will be combined with current score follower systems (with our help) to evaluate improvements on tracking accuracy and robustness.

**Project 3 - Melody/F0 Extraction for Audio-to-score Alignment**

For Project 3, the first goal is to isolate audio segments where the voice is predominantly present, in order to extract the fundamental frequency and/or the melody as a function of the time. Pre existing melody/f0 extraction algorithms from the current literature [3] [4] will be tested and compared on opera recordings. Annotations related to onset times of each singing note present in the segments will allow to evaluate the quality of the extracted melody. Various alignment methods will then be experimentally tested regarding the possibility of matching the audio to the score, based only on the extracted melody lines.

**References**

[1] Schlüter, J., & Lehner, B. (2018, September). Zero-Mean Convolutions for Level-Invariant Singing Voice Detection. In ISMIR (pp. 321-326).

[2] Lehner, B., Schlüter, J., & Widmer, G. (2018). Online, loudness-invariant vocal detection in mixed music signals. IEEE/ACM Transactions on Audio, Speech and Language Processing (TASLP), 26(8), 1369-1380.

[3] Salamon, J., & Gómez, E. (2012). Melody extraction from polyphonic music signals using pitch contour characteristics. IEEE Transactions on Audio, Speech, and Language Processing, 20(6), 1759-1770.

[4] Kim, J. W., Salamon, J., Li, P., & Bello, J. P. (2018, April). Crepe: A convolutional representation for pitch estimation. In 2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) (pp. 161-165). IEEE.

**Contact**

Gerhard Widmer [www.cp.jku.at/people/widmer](http://www.cp.jku.at/people/widmer) [gerhard.widmer@jku.at](mailto:gerhard.widmer@jku.at)