Project Summary:
Recognition and Analysis of Melodies from Polyphonic Musical Audio Data

The transcription of the melody out of polyphonic audio data is recognized as an unsolved problem among researchers. While humans easily spot the melody line from various musical pieces, there is no reliable technical method for the automatic estimation of the melody.

Anyway over the last 10 years there has been a remarkable progress in the area of transcription of polyphonic music – a topic that is closely related to melody recognition. The latest publications show promising results in the transcription of polyphonic audio pieces, though in most cases the audio test data is subject to several restrictions: be it that only one special instrument is allowed (eg. piano), be it the limitation of the maximum number of instruments playing simultaneously or the exclusion of percussive instruments.

The aim of this PhD thesis is to develop an algorithm for the automatic estimation of the melody (predominant pitch) from contemporary western music. Most scientists comply with the statement, that this difficult task cannot be solved using signal processing methods alone. In fact one also has to employ knowledge about music theory or even knowledge about the processes in the auditory cortex of humans. Hereby the combination and further development of techniques from audio signal processing on the one hand and computational musicology on the other hand is the main focus of this work.

Basically the concept for the algorithm can be divided into two sequential steps: At first all those sound events should be extracted from the complex polyphonic music signal which could be part of the melody. Simple decision rules based on psycho acoustics and music theory will help to make a preselection among simultaneously occurring pitches. Secondly a melody will be estimated from the remaining pitch candidates. The model to find the most plausible succession of scores could either be deterministic or statistical – most likely it will be a combination of both methods. In any case the model will be due to voice-leading and harmonic progression principles, which have already been the subject of extensive theoretical attention in the scope of musicology.

Research work in an interdisciplinary context depends even more than usual on the cooperation of experts from various research fields. That is why I want to point out particularly that this thesis project benefits from guidance and support of two scientific institutes: The Fraunhofer Institute for Digital Media Technology in Ilmenau and the university of Konstanz. The Fraunhofer IDMT enjoys an excellent international reputation for its research on digital audio processing whilst Prof. Jan Beran (department of mathematics and statistics of Konstanz University) actively participates in the research on computational musicology.