

Automatic Extraction of the Main Melody in Music Signals

Jean-Louis Durrieu

GET ENST Paris,
France
durrieu@enst.fr

Abstract. In order to mine a multimedia database in an efficient way, one needs more than the meta data that are usually available. Extracting relevant features directly from the source file is necessary and our research aims at providing some of these features. Extracting the melody from a song might benefit to several applications such as Query-By-Humming, Cover-Version Identification and any other application that involves a symbolic melody similarity step. In particular we might be able to build a database of monophonic melodies which can be useful to the above applications. We want to experiment source separation approaches for the task of automatic melody extraction. At first, we focus on songs with a predominant singer voice over a background music, which is the case for most of the popular pop songs. We use a Non-negative Matrix Factorization (NMF) approach to model the background music along with a specific modeling of the voice signal, with spectral Gaussian Mixture Models (GMM). We combine them by considering that the mixed signal is the sum of the voice signal and the music signal. The parameters of the models are estimated thanks to the Expectation-Maximization (EM) algorithm. Thanks to these parameters, we build the sequence of frequencies that most likely generated the voice signal, thus obtaining the wanted melody.