Mel Frequency Cepstral Coefficients

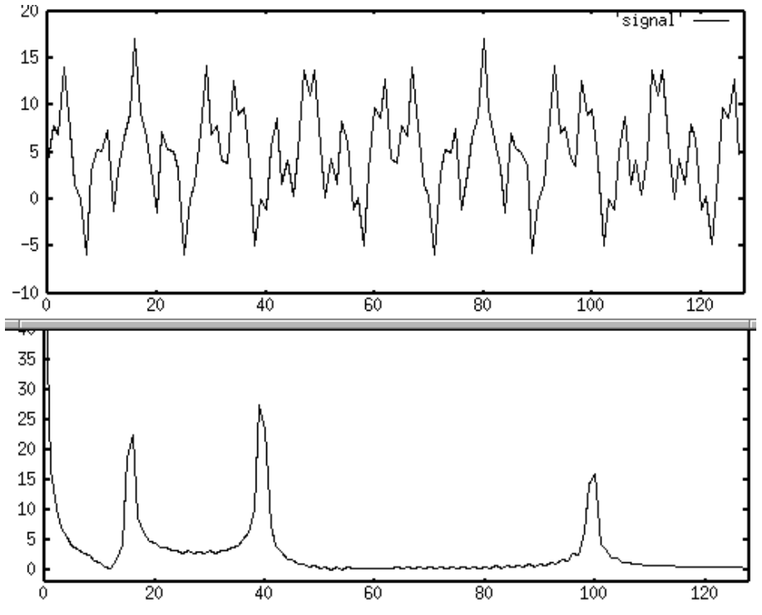
LOG10

The log function is applied to the Mel magnitudes. This can be motivated from two perspectives: Firstly, loudness perception of humans is logarithmic, and secondly, the logarithm transforms the product of the two parts of speech into an addition which is easier to separate apart.

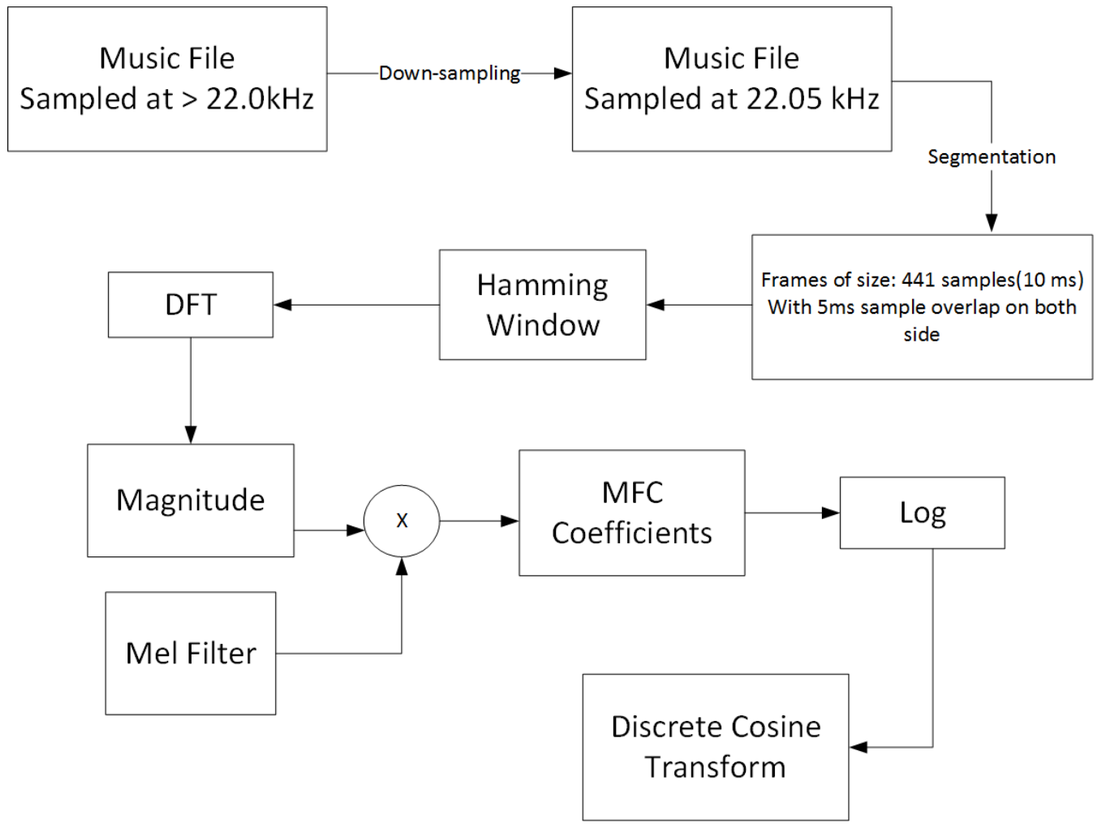
DISCRETE COSINE TRANSFORM

A discrete cosine transform is applied to each frame of Mel log magnitudes, of which the higher coefficients are discarded and only first 15 are kept.

The discrete cosine transform (DCT) helps separate the song into parts (or spectral sub-bands) of differing importance.



COMPLETE FLOW CHART



Therefore, in essence, we try to extract the information out of a music file which consists of hundreds of thousands of samples into some net total of 30 coefficients per song. This compression (or feature extraction), though results in some loss of information, is sufficient for implementing classification techniques for basic feature detection, like genre, of the song.