Mel Frequency Cepstral Coefficients

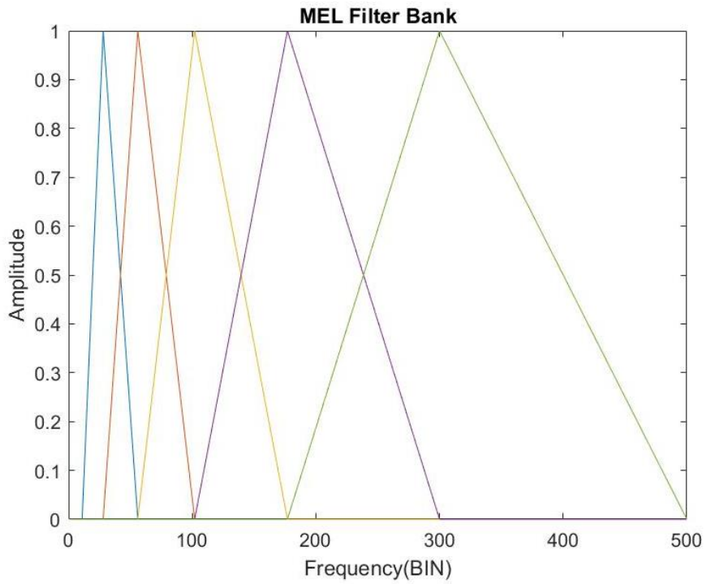
MEL SCALING

Frequencies are mapped to a perceptual pitch scale, the so-called *Mel scale*. Human's perception of pitch has been found to be linear in frequency up to about 1000 Hz, above which pitch perception is logarithmic. This relationship is empirically calculated as:

$$M(f) = 1125\ln(1 + \frac{f}{700})$$

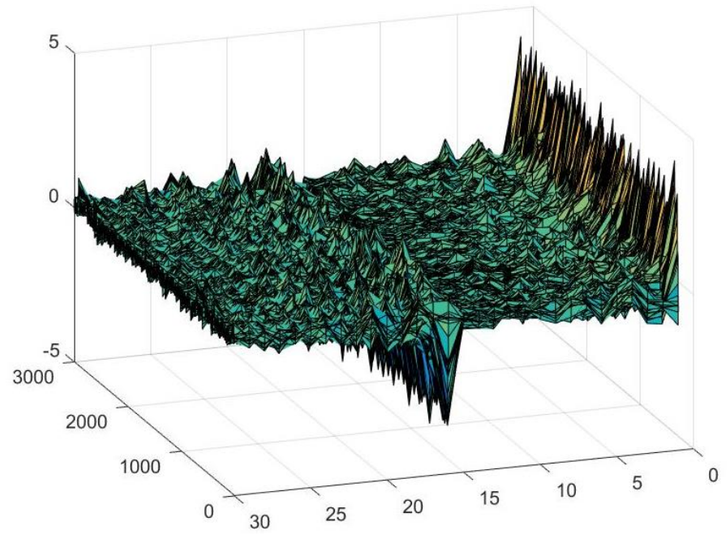
Mel scaling is also used for dimensionality reduction: the frequency bins of the magnitude spectrum (usually between 256 and 1024) are mapped to relatively few MEL bands (e.g., 40), justified by the fact that the human ear only distinguishes few so-called *critical bands* as well.

After the scaling, a MEL Filter bank is constructed as:

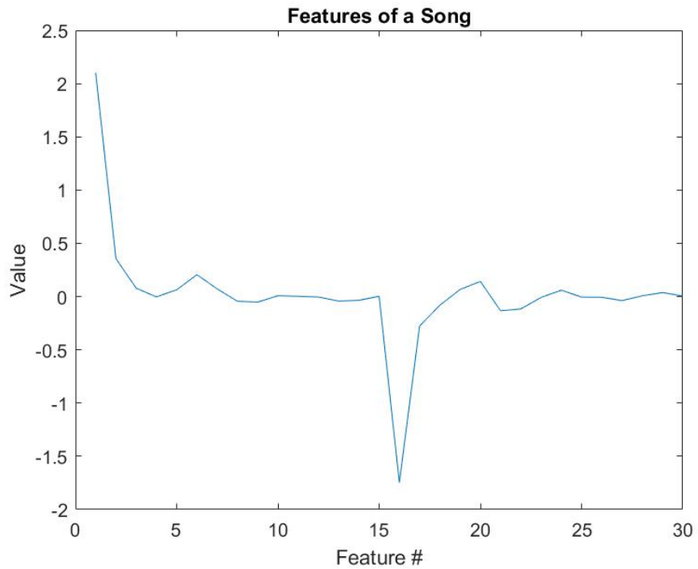


For simplicity, only 5 filters were created, which will give us 5 coefficients for each frame. In the demo, we used 40 such filters.

After choosing only the top 15 coefficients we get the following feature space for a song clip of 10s:



Here, along with MFCC coefficients we have also used delta MFCC coefficients which represents change in frequency power.



This leads to excessive useless computation due to a lot of correlation present among the features. Therefore, we average out the features to one frame as shown: