**Feature Extraction:**

Feature extraction is the process of computing a compact numerical representation that can be used to characterize a segment of audio.

**Audio features**

->Volume

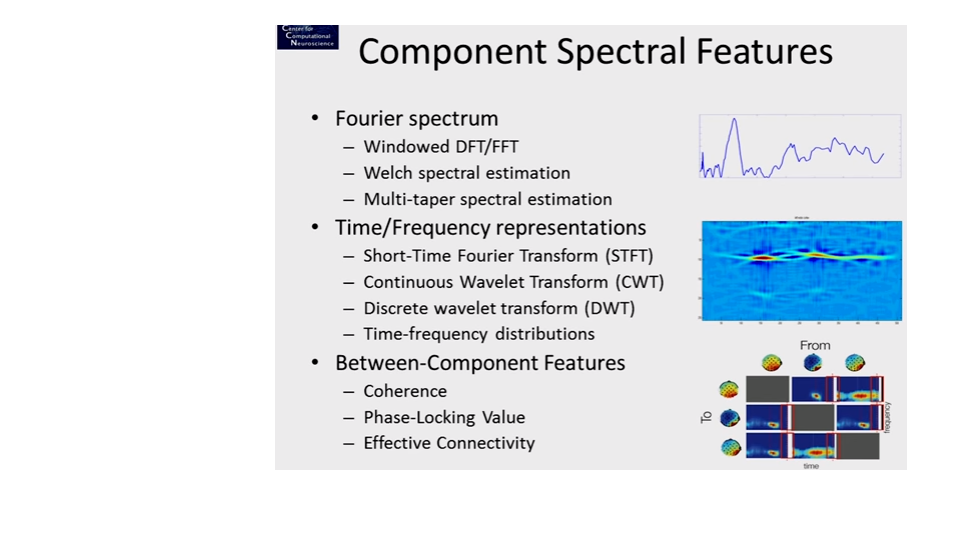
->Pitch

->ZCR(zero cross rate)

->Energy

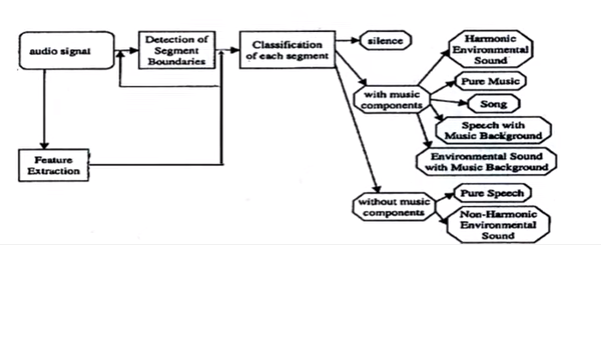
->Centroid

->bandwidth



**Classification:**

Each segment of audio is classified as into one of the basic audio types according to procedure.



**Details of features:**

There are two stages in the audio feature extraction methodology:

1. *Short-term feature extraction*: It splits the input signal into short-term widnows (frames) and computes a number of features for each frame. This process leads to a sequence of short-term feature vectors for the whole signal.
2. *Mid-term feature extraction:*In many cases, the signal is represented by statistics on the extracted short-term feature sequences described above.It extracts a number of statistcs (e.g. mean and standard deviation) over each short-term feature sequence.

The total short term features:

|  |  |  |
| --- | --- | --- |
| Feature Id | Feature name | Description |
| 1 | Zero Crossing Rate | The rate of sign-changes of the signal during the duration of a particular frame |
| 2. | Energy | The sum of squares of the signal values, normalized by the respective frame length. |
| 3 | Entropy of Energy | The entropy of sub-frames' normalized energies. It can be interpreted as a measure of abrupt changes. |
| 4 | Spectral Centroid | The center of gravity of the spectrum. |
| 5 | Spectral Spread | The second central moment of the spectrum. |
| 6 | Spectral Entropy | Entropy of the normalized spectral energies for a set of sub-frames. |
| 7 | Spectral Flux | The squared difference between the normalized magnitudes of the spectra of the two successive frames. |
| 8 | Spectral Rolloff | The frequency below which 90% of the magnitude distribution of the spectrum is concentrated. |
| 9-21 | MFCCs | Mel Frequency Cepstral Coefficients form a cepstral representation where the frequency bands are not linear but distributed according to the mel-scale. |
| 22-33 | Chroma Vector | A 12-element representation of the spectral energy where the bins represent the 12 equal-tempered pitch classes of western-type music (semitone spacing). |
| 34 | Chroma Deviation | The standard deviation of the 12 chroma coefficients. |

Mel-frequency cepstrum Coefficients

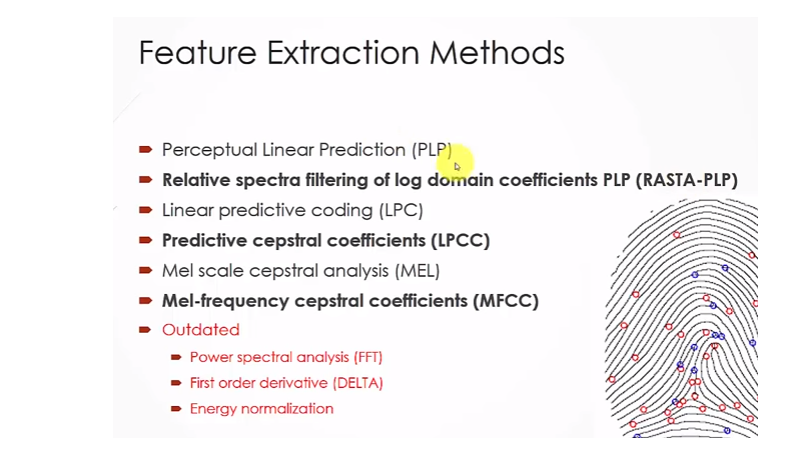
MFCCs are commonly derived as follows:

1. Take the [Fourier transform](https://en.wikipedia.org/wiki/Fourier_transform) of (a windowed excerpt of) a signal.
2. Map the powers of the spectrum obtained above onto the [mel scale](https://en.wikipedia.org/wiki/Mel_scale" \o "Mel scale), using [triangular overlapping windows](https://en.wikipedia.org/wiki/Window_function#Triangular_window).
3. Take the [logs](https://en.wikipedia.org/wiki/Logarithm) of the powers at each of the mel frequencies.
4. Take the [discrete cosine transform](https://en.wikipedia.org/wiki/Discrete_cosine_transform) of the list of mel log powers, as if it were a signal.
5. The MFCCs are the amplitudes of the resulting spectrum.

Feature extraction for emotion dataset:

***Rhythmic Features****:* The rhythmic features were derived by extracting periodic changes from a beat histogram. An algorithm that identifies peaks using autocorrelation was implemented. We selected the two highest peaks and computed their amplitudes, their BMPs (beats per minute) and the high-to-low ratio of their BPMs. In addition,features were calculated by summing the histogram bins between 40-90, 90-140 and 140-250 BPMs respectively. The whole process led to a total of 8 rhythmic features.

***Timbre Features****:* Mel Frequency Cepstral Coefficients (MFCCs) are used for speech recognition and music modelling [8]. To derive MFCCs features, the signal was divided into frames and the amplitude spectrum was calculated for each frame. Next, its logarithm was taken and converted to Mel scale. Finally, the discrete cosine transform was implemented. We selected the first 13 MFCCs. Another set of 3 features that relate to timbre textures were extracted from the Short-Term Fourier Transform (FFT): Spectral centroid, spectral rolloff and spectral flux.For each of the 16 aforementioned features (13 MFCCs, 3 FFT) we calculated the mean, standard deviation (std), mean standard deviation (mean std) and standard deviation of standard deviation (std std) over all frames. This led to a total of 64 timbre features



**Emotion Labelling: (of emotion dataset)**

The Tellegen-Watson-Clark model was employed for labelling the data with emotions. this particular model because the emotional space of music is abstract with many emotions and a music application based on mood should combine a series of moods and emotions. To achieve this goal without using an excessive number of labels, we reached a compromise retaining only 6 main emotional clusters from this model.

Ref:

>Github pyaudio python.

>MULTI-LABEL CLASSIFICATION OF MUSIC INTO EMOTIONS