**Voice Activity Detection:**

Also known as VAD in other terms actually this phenomenon is used to detect human speech whether it is present or not.

Any VAD algorithm consist of:

* First of all there is a noise removal model (recommended: spectral subtraction)
* Features and quantities are calculated from a section of input.
* Then we define a classification rule to take decision is it speech or not.

That classification rule finds out when the value exceeds threshold value

VAD that google developed for ‘webRTC’ is one of the best available currently.

**VAD ALGORITHMS:**

1. <http://practicalcryptography.com/miscellaneous/machine-learning/voice-activity-detection-vad-tutorial/>

This is a basic algorithm which output 1 when speech is detected else 0.

1. Break audio signal into frames.
2. Extract features from each frame.
3. Train a classifier on a known set of speech and silence frames.
4. Classify unseen frames as speech or silence.
5. <https://github.com/marsbroshok/VAD-python>

Input audio data treated as following:

1. Convert stereo to mono
2. Move a window of 20ms along the audio data
3. Calculate ration between energy of speech band and total energy for window
4. If ratio is more than threshold (0.6 by default) label windows as speech
5. Apply median filter with length of 0.5s to smooth detected speech regions
6. Represent speech regions as intervals of time
7. <http://iopscience.iop.org/article/10.1088/1742-6596/705/1/012037/pdf>

Actually it is a research paper in which a simple but efficient voice activity detector based on the Hilbert transform and a dynamic threshold is presented to be used on the pre-processing of audio signals

**BASIC IDEA:**

* They are breaking sound into three sections
  + - VOICED, where speech is detected
    - UNVOICED, where vocal chords vibrated and sound produced
    - SILENCE, where vocal chords produced nothing.

This is known as VAD.

* They are using Zero crossing rate, Mean square of the energy and Signal Covering using Hilbert-transform.