

## **Existing System:**

In existing voice recorder once we start recording it records everything from the start to end point. It records silent speech, human voice, non-human voice and pretty much everything which generates sound.

This behaviour of voice recorder leads to keeping unnecessary data, increases size of audio files, needs more storage, consuming large bandwidth, decreasing speed of data transfer and at the end reduces the overall performance.

In order to overcome above problems we come up with the idea of intelligent voice recorder which will help to overcome above mention problems.

## **Proposed system:**

This is same as regular voice recorder but with little more features. It records only human voice and totally ignores other voice. Once you run the application it keeps processing voice with incoming signals and when it comes to know that incoming signals are of human voice it store it in a file of specific format. As the application is continuous in nature it would have to terminate externally by human intervention.

## **Applications:**

- 1) Military applications [any kind of activity with spying purpose]
- 2) Streaming of voice on tv channels

## Intelligent Voice Recorder

---

- 3) Messaging app (working on speech mode, no need to click voice symbol for each message as it is in continuous manner)

### **References:**

#### *1. Working with audio files in python*

<https://realpython.com/playing-and-recording-sound-python/#playing-audio-files>

#### *2. input through microphone and speech recognition*

<https://realpython.com/python-speech-recognition/>

#### *3. human voice detection hint*

<https://stackoverflow.com/questions/18355448/detect-human-voice-from-audio-file-input>

### **Libraries that can become useful for human voice detection:**

#### *a) Speex :*

<https://speex.org/docs/manual/speex-manual/node4.html>

It contains VAD too.

#### *b) LibVAD*

#### *c) FTT algorithm*

#### *d) Sphinx4*

#### *e) PocketSphinx*

#### *f) Freeswitch*

### **VAD (Voice Detection Activity):**

[https://www.etsi.org/deliver/etsi\\_i\\_ets/300700\\_300799/300730/01\\_20\\_103/ets\\_300730e01c.pdf](https://www.etsi.org/deliver/etsi_i_ets/300700_300799/300730/01_20_103/ets_300730e01c.pdf)

## Intelligent Voice Recorder

---

The function of the VAD is to distinguish between noise with speech present and noise without speech present. This is achieved by comparing the energy of a filtered version of the input signal with a threshold. The presence of speech is indicated whenever the threshold is exceeded

The function of the VAD is to indicate whether each 20 ms frame produced by the speech encoder contains speech or not.

From speex:

*When enabled, voice activity detection detects whether the audio being encoded is speech or silence/background noise. VAD is always implicitly activated when encoding in VBR, so the option is only useful in non-VBR operation. In this case, Speex detects non-speech periods and encode them with just enough bits to reproduce the background noise. This is called ``comfort noise generation'' (CNG).*