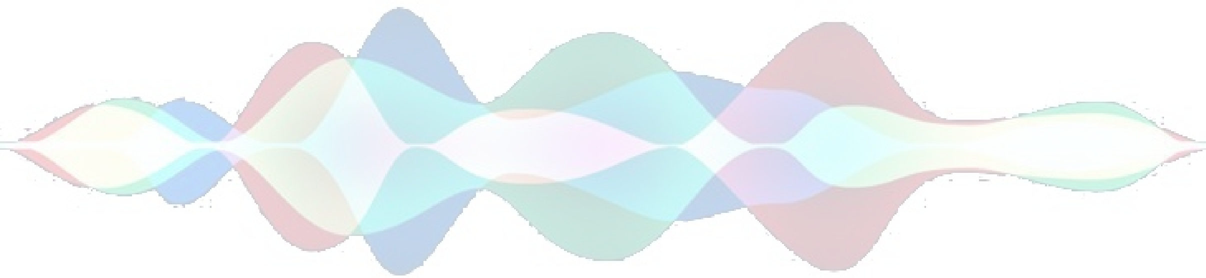
EduBrain 清帆科技

语音信号处理技术

Voice Activity Detection (VAD) 介绍

Weekly Tech Salon

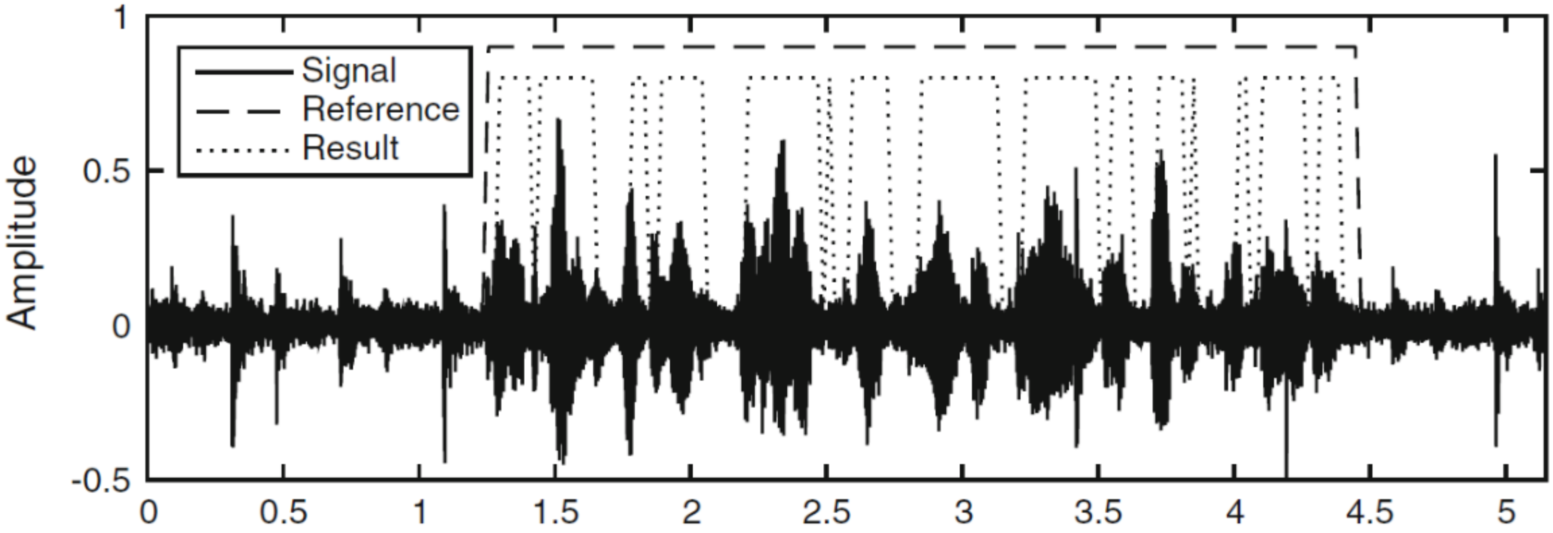


概览

- A brief introduction to Voice Activity Detection (VAD)
  - What is VAD?
  - Application
- How VAD works
  - Requirements
  - Features used for VAD
  - Algorithm
- Resources

What is VAD?

Voice Activity Detection (VAD) refers to the analysis of an audio signal to determine whether speech is present or not. It' s a binary classification problem.



Applications of VAD

- Loudness measurement and control
- Dialog enhancement
- Perceptual audio coding
- Broadcast monitoring
- Silence compression
- Blind upmixing
- Speaker diarization
- ….

VAD 通常被视为一种驱动技术，我们主要对应用 VAD 的系统的表现颇为关注，而对于 VAD 的输出结果并没有直接的兴趣．

而 VAD 的效果通常对于各种系统的表现有重要影响．

Requirements

正是由于 VAD 的效果通常对于各种系统的表现有重要影响．这就对 VAD 相关算法提出了很多性能需求．

- Accuracy
- Robustness
- Latency
- Computational Load
- Memory Requirement

Features uesd for VAD

Feature Extraction

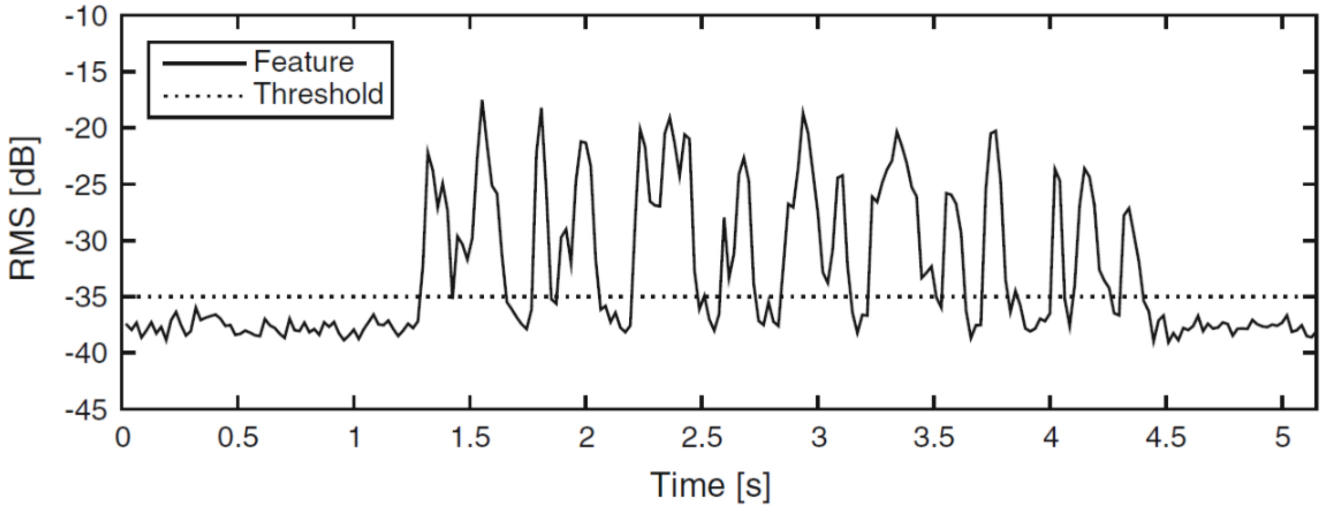
- Pre-processing of the audio signal, e.g. re-sampling, filtering, or noise reduction,
- Feature computation in the time domain or frequency domain,
- Feature selection,
- Centering and variance normalisation,
- Projection and dimensionality reduction,
- Analysis or Linear Discriminant Analysis,
- Filtering (smoothing or differentiating).

Features uesd for VAD

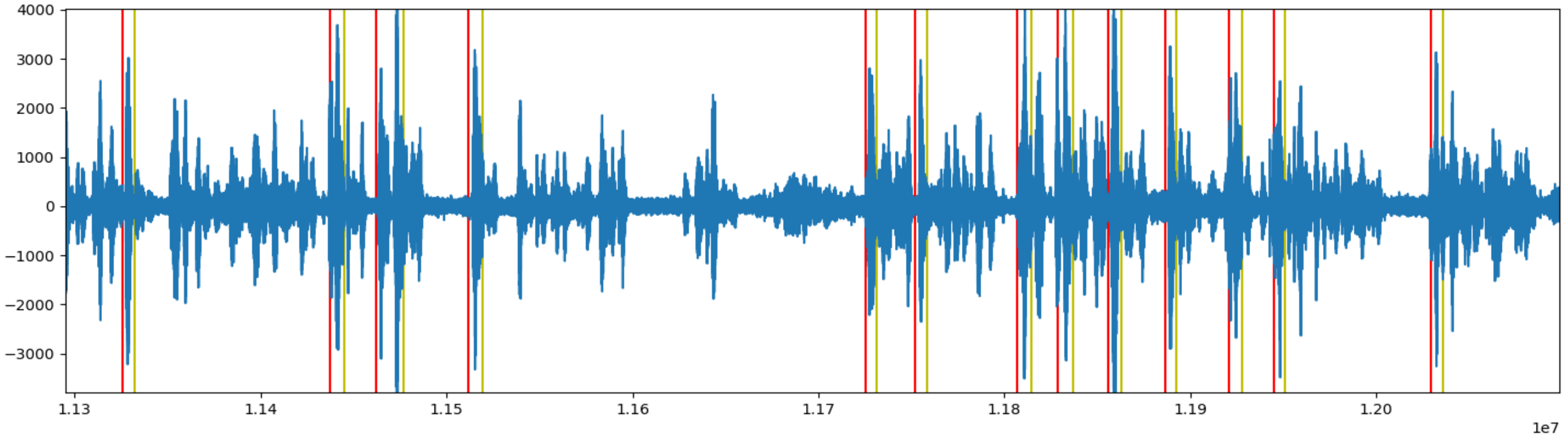
- **Intensity Features(强特征)**

e.g. Short-term Energy

Defect: sensitive to background noise when the noise is not stationary or if the SNR is low.



实例 1



测试音频长度 40 min （这里是局部放大展示的切分效果）
红线为切分开始点， 黄线为切分终止点

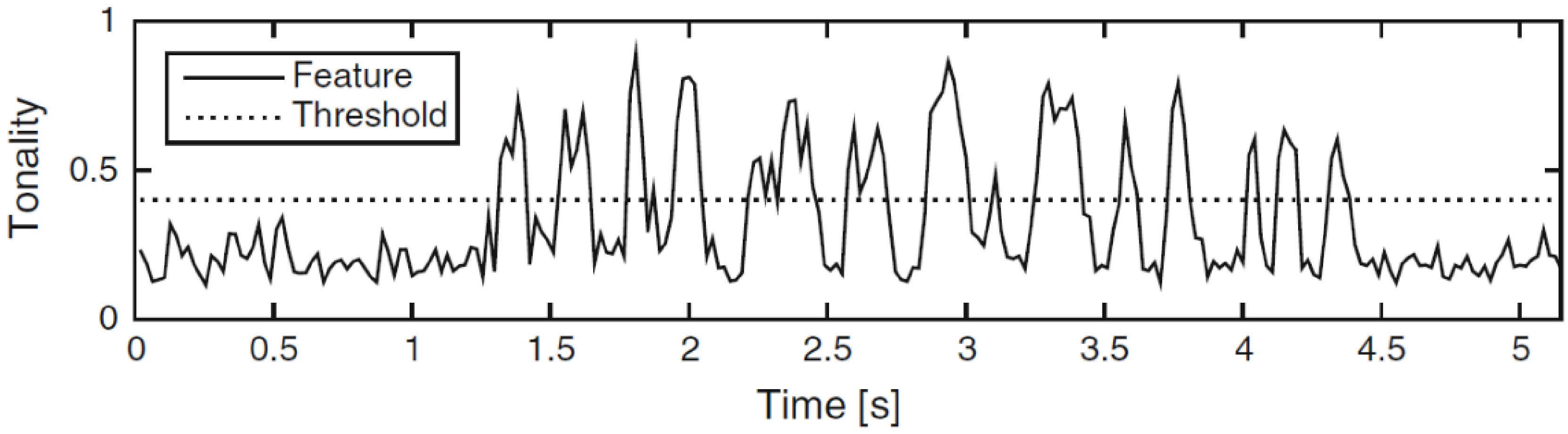
很显然，单纯基于强特征的 VAD 效果很难保证。

Features uesd for VAD

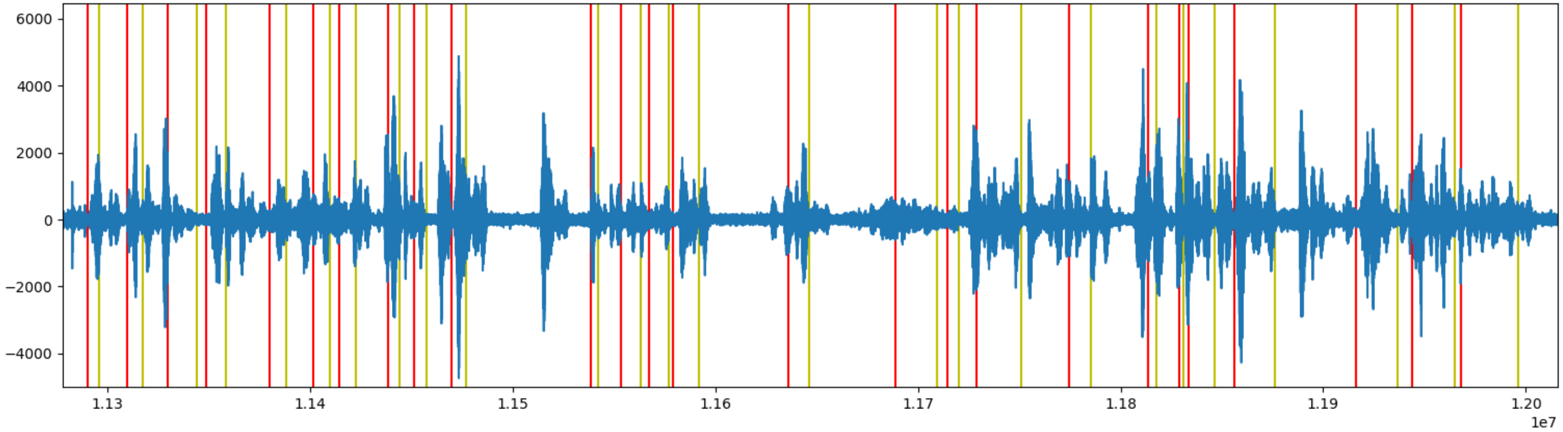
- **Tonality features(音调特征)**

e.g. Spectral flatness measure, Spectral crest factor

Defect: Since the fundamental frequency features strong modulations and the vocal tract filtering also varies with time, voiced speech is stationary only over short periods of time.



实例 2



测试音频长度 40 min （这里是局部放大展示的切分效果）
红线为切分开始点， 黄线为切分终止点

基于音调特征的 VAD 效果显然更佳，但是切分过碎， 容易受声音节奏变化影响

Features uesd for VAD

- Spectral shape features(谱形特征)

e.g. MFCC, PLPC, RASTA-PLPC

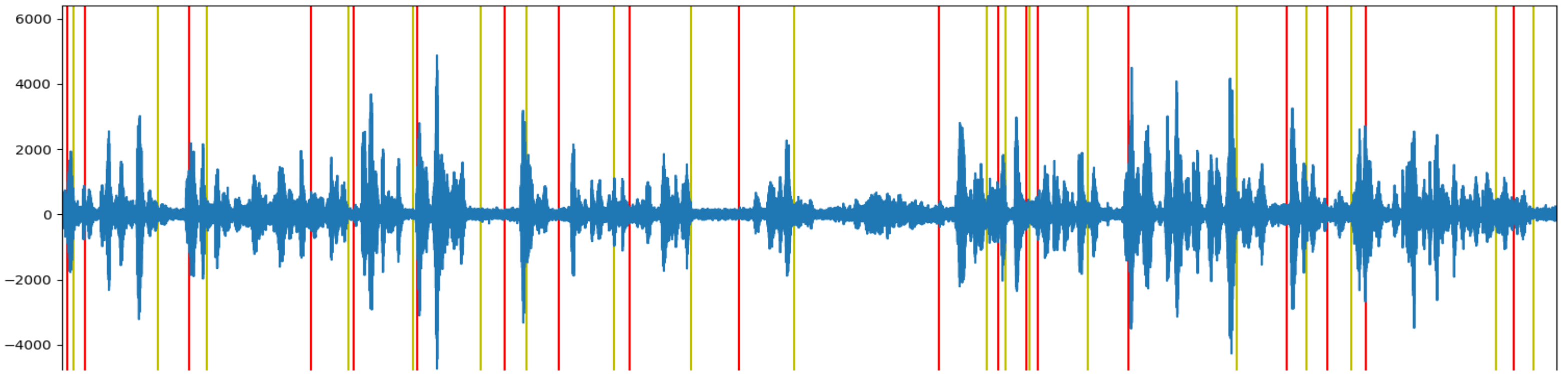
Defect: high dimensionality

- Other features

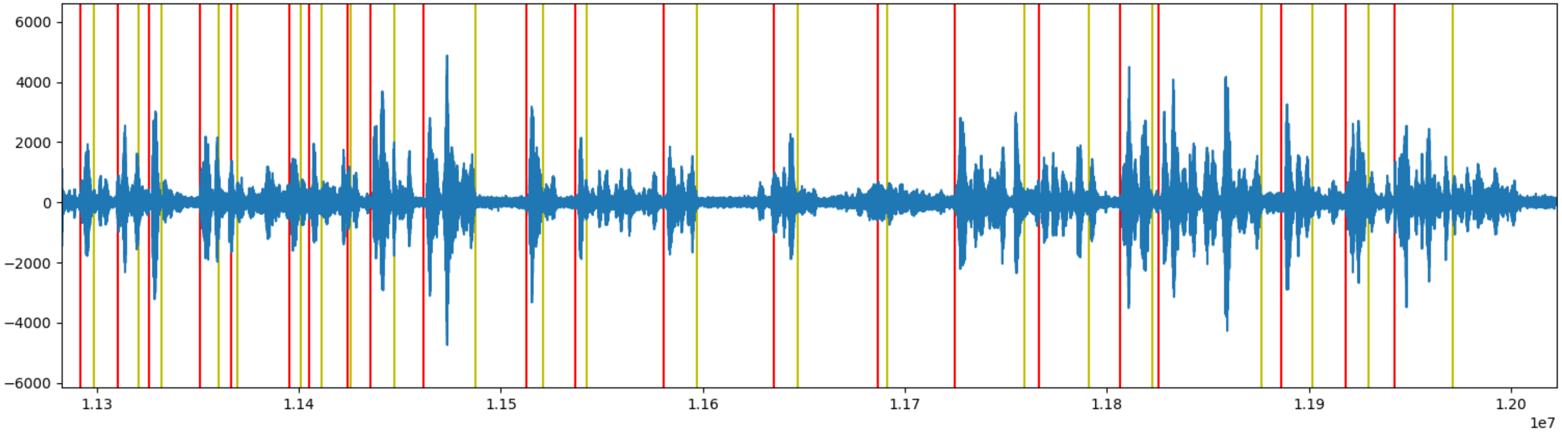
e.g. Line-spectral frequencies, Zero-crossing Rate, Entropy-Based...

实例 3

Entropy-Based VAD



基于多特征的 VAD



测试音频长度 40 min （这里是局部放大展示的切分效果）

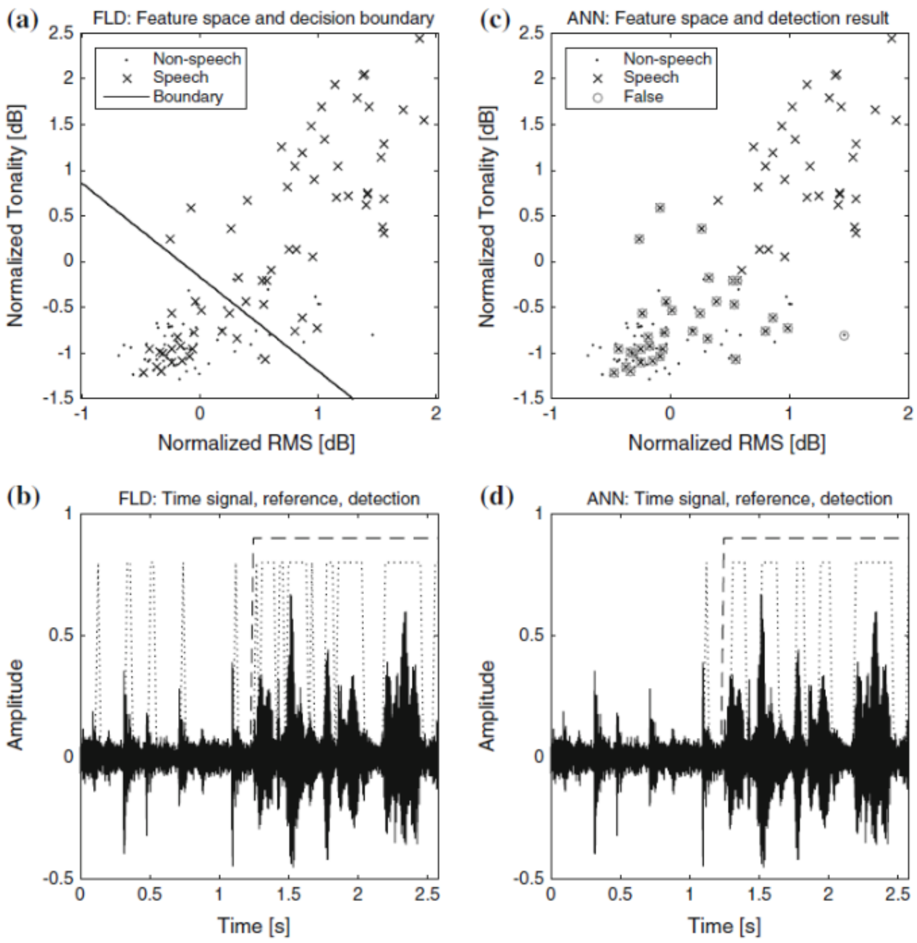
红线为切分开始点， 黄线为切分终止点

Algorithm

本质上，VAD 就是一个 2 分类问题

Basic Form: a scalar feature compared with a threshold which has been determined heuristically.

In Practice: multiple features are evaluated, then we may use FLD, SVM or ANN to do the classification.



Challenges

- **Low Latency (** 降低延迟 **)**
- **Reduce the effect of Background Noise (** 降低背景噪音干扰 **)**
- **Generalization (** 多场景的泛化能力 **)**
- **Ambiguous Ground Truth (** 真值标记困难 **)**

Resources

- **Voice Activity Detection Toolkit**

该Toolkit 包含有4种基于python 和tensorflow的分类器：

Adaptive context attention model (ACAM)

Boosted deep neural network (bDNN)

Deep neural network (DNN)

Long short term memory recurrent neural network

- **[pdf]Recurrent neural network for Voice Activity Detection**

来自Google research 的一篇论文，建立了一个多层的RNN进行VAD，并应用于他们的语音识别程序中，有效减少了17%的运算时间，并使得识别准确率相对提高了1%.