8th International Electronic Conference on Sensors and Applications

1–15 Nov, 21

Finding Earthquake Victims by Voice Detection Techniques

Ruchi Jha* ,Walter Lang, Reiner Jedermann

Institute For Microsensors, Actuators, And Systems (IMSAS)

University Of Bremen, Bremen, Germany

Background And Motivation

1. Most earthquakes above 5.5 on Richter scale, can cause large-scale destruction.

2. Many of the casualties stuck in damaged structures can live only for a few hours after the initial disaster.

3. Approximately 80-percent of victims can be successfully rescued alive if they are detected by help teams within 48 hours of disaster occurrence.

4. Despite of technological excellence, rescue is challenging when the victim cannot be found through a direct line of sight.

5. A method is needed to detect people through physical attributes like motion, odour, speech, heat signature or gaseous emissions unique to human body for better rescue operations.

Introduction

Voice Activity Detection (VAD)

Detecting
presence or absence
of
human speech.

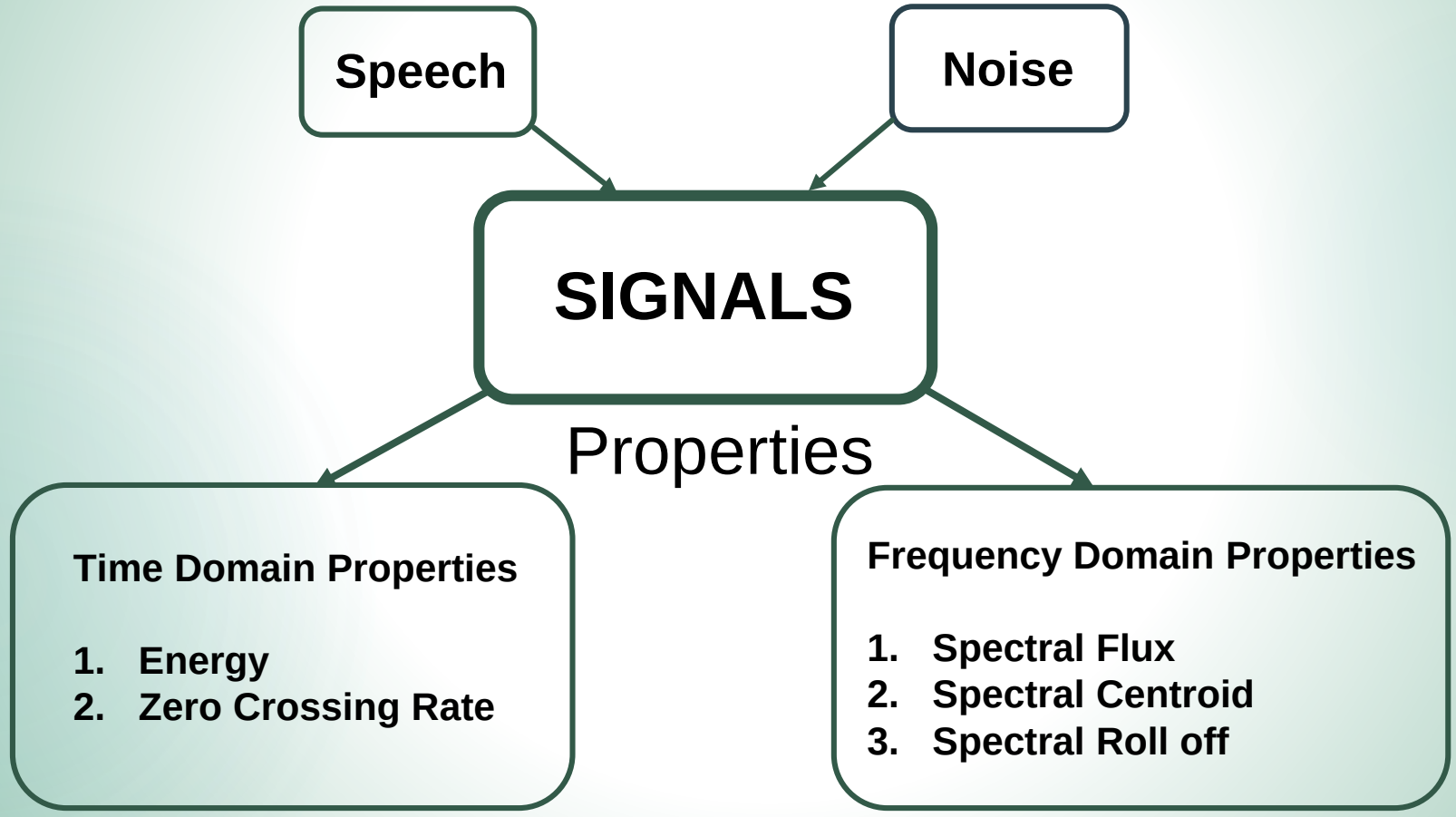
Applications

1. Audio Conferencing
2. Echo Cancellation
3. Speech Recognition
4. Speech Encoding
5. Hands-free telephony

Methods

1. Energy Thresholds
2. Pitch and Harmonic detection
3. Zero-crossing rates
4. Waveform and spectral analysis

Scientific Approach



System Description

1. Collection of Noise and Voice Samples

Count	Group	Sources	Name	Examples
11	Noise Studio	Audio CD [14]	N1 to 11	Traffic, touring cars, motorcars, cleaning, airplane, buzzer, river, applause, industry, chattering.
9	Voice Samples	CD, TV, Studio recording	VF1 to 4 (female) VM1 to 5 (male)	Female and Male sound recordings in English and German.
7	Noise Street	Own Outside recording	SN1 to 7	Street noises with birds, cars, tram, glasses, music, river and wind.
5	Voice Mix	Own recording Outside	MIX 1 to 5	Mix sounds of people speaking with background noise.
6	Voice Studio	Studio recording	VF… (female) VM… (male)	Speech recorded in Spanish(S), German (D), Hindi (H), English (E), and Latvian (L).

Total Samples = 38; Sampling frequency = 44100 Hz (mono) ; Length = 2 seconds

2. Feature Extraction

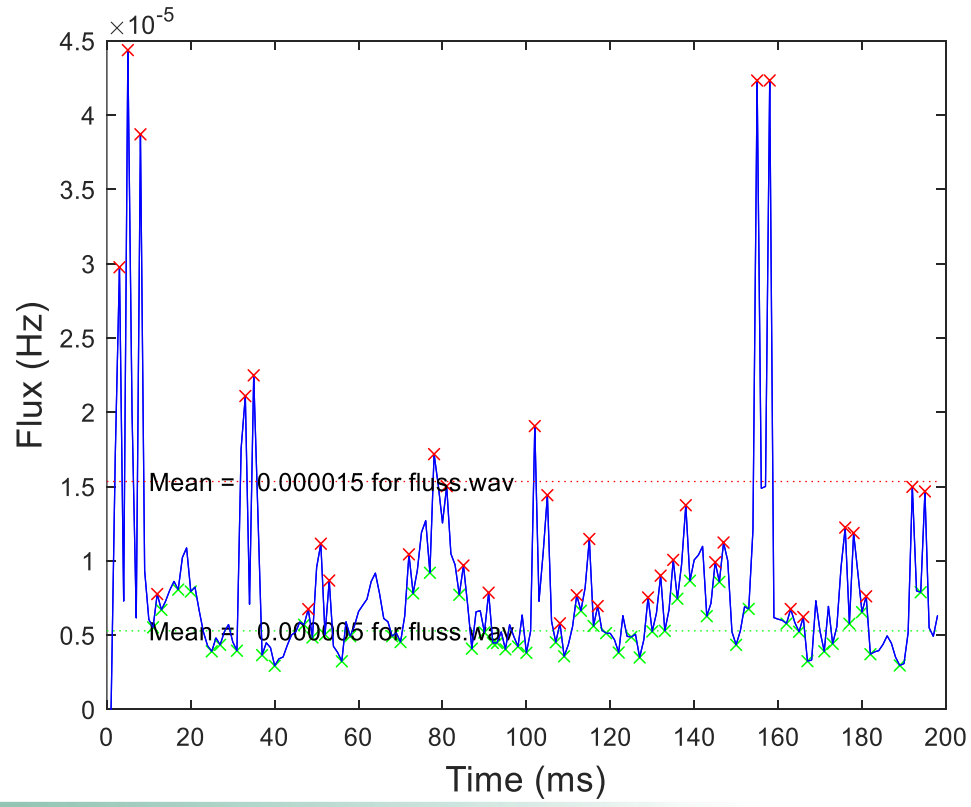
1. Trails with Average values of Spectral features.

Many Buffer range of values overlapping for speech and Noise. (Not opted)

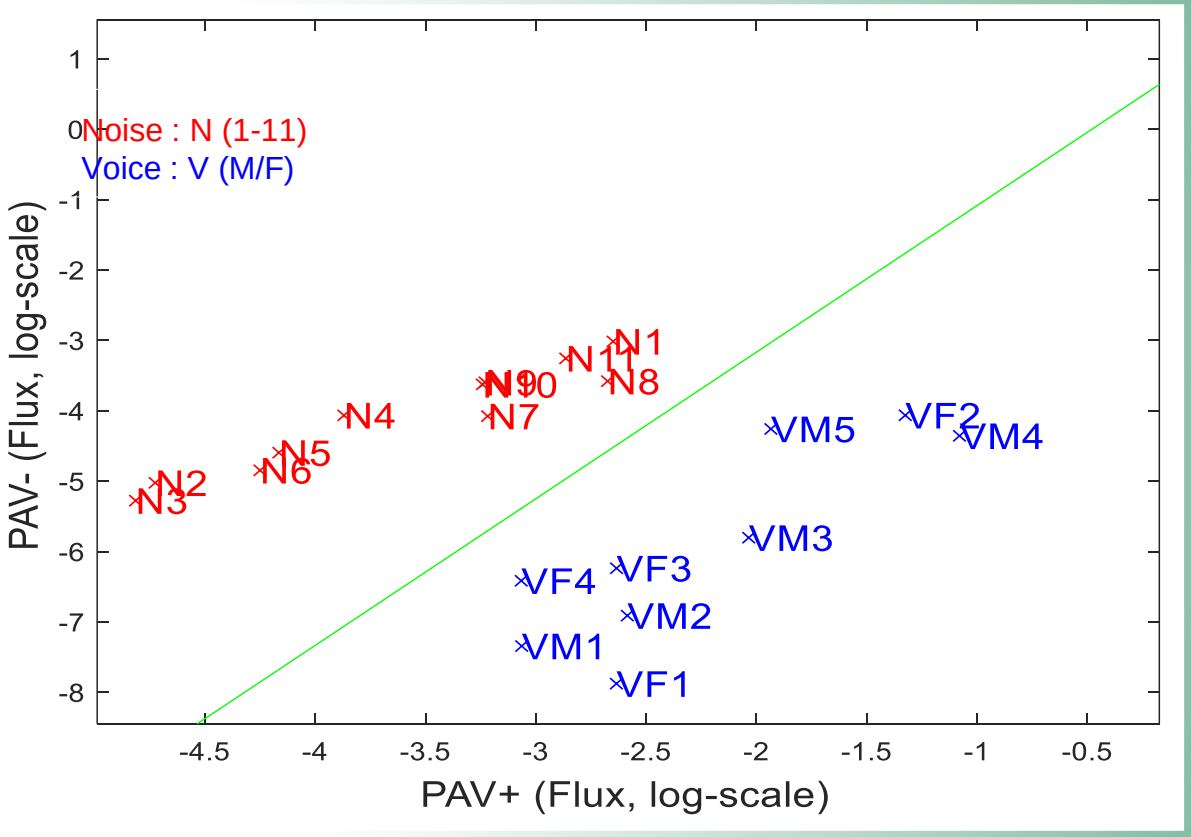
2. Calculation of Average of Local Minima and Maxima (ALMM) of spectral features.

- MATLAB script to find peak values.
- Drawing a conclusion for separating speech and noise.

..continued.



Flux Peaks in a River sound sample



Discriminating Speech (in blue) and Noise (in red) based on the ALMM of Flux values

3. Methodology

Cross Validation : Estimates the performance of a model.

1. **Manual method**

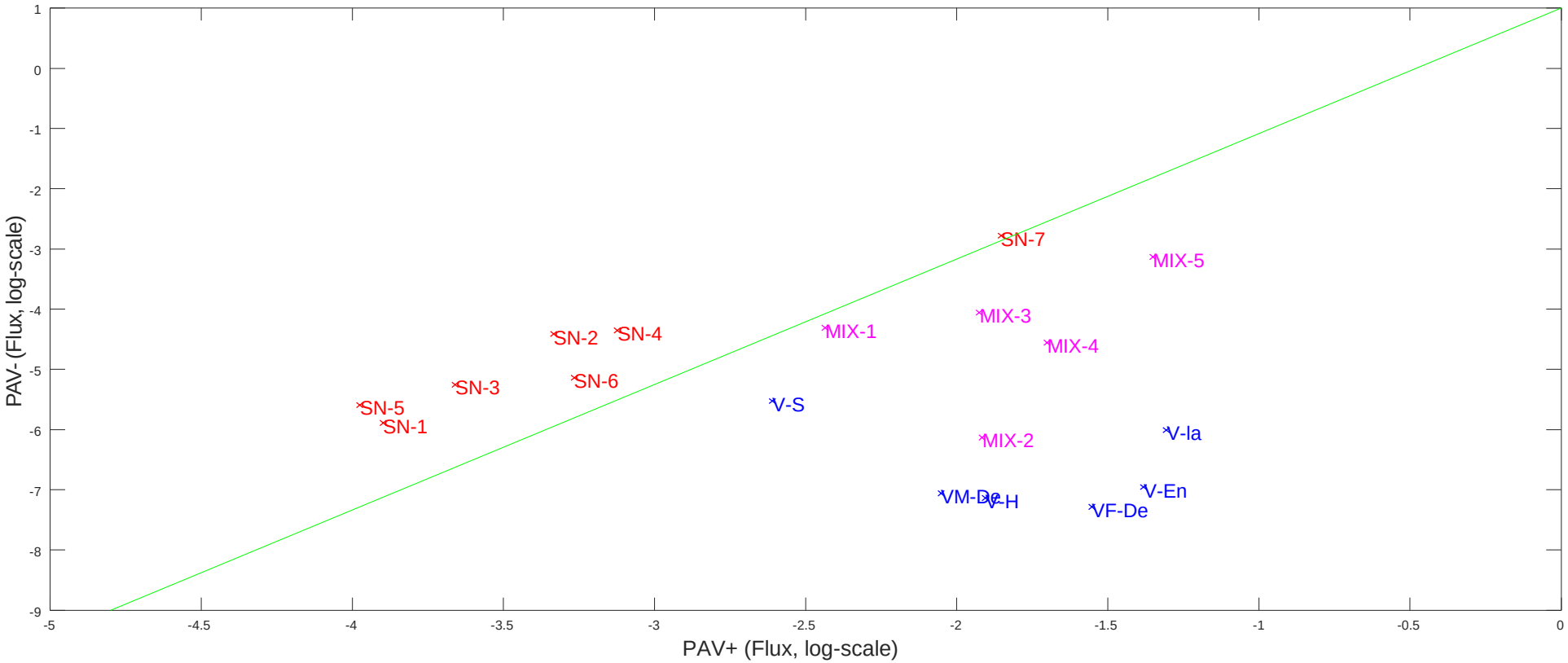
- Linear boundary between the sample.

2. **Machine learning**

- Selection of a predictive model : LDA (Linear Discriminant Analysis)
- Training the LDA Model : To obtain a linear boundary.
- Testing the model : To get a predicted result.

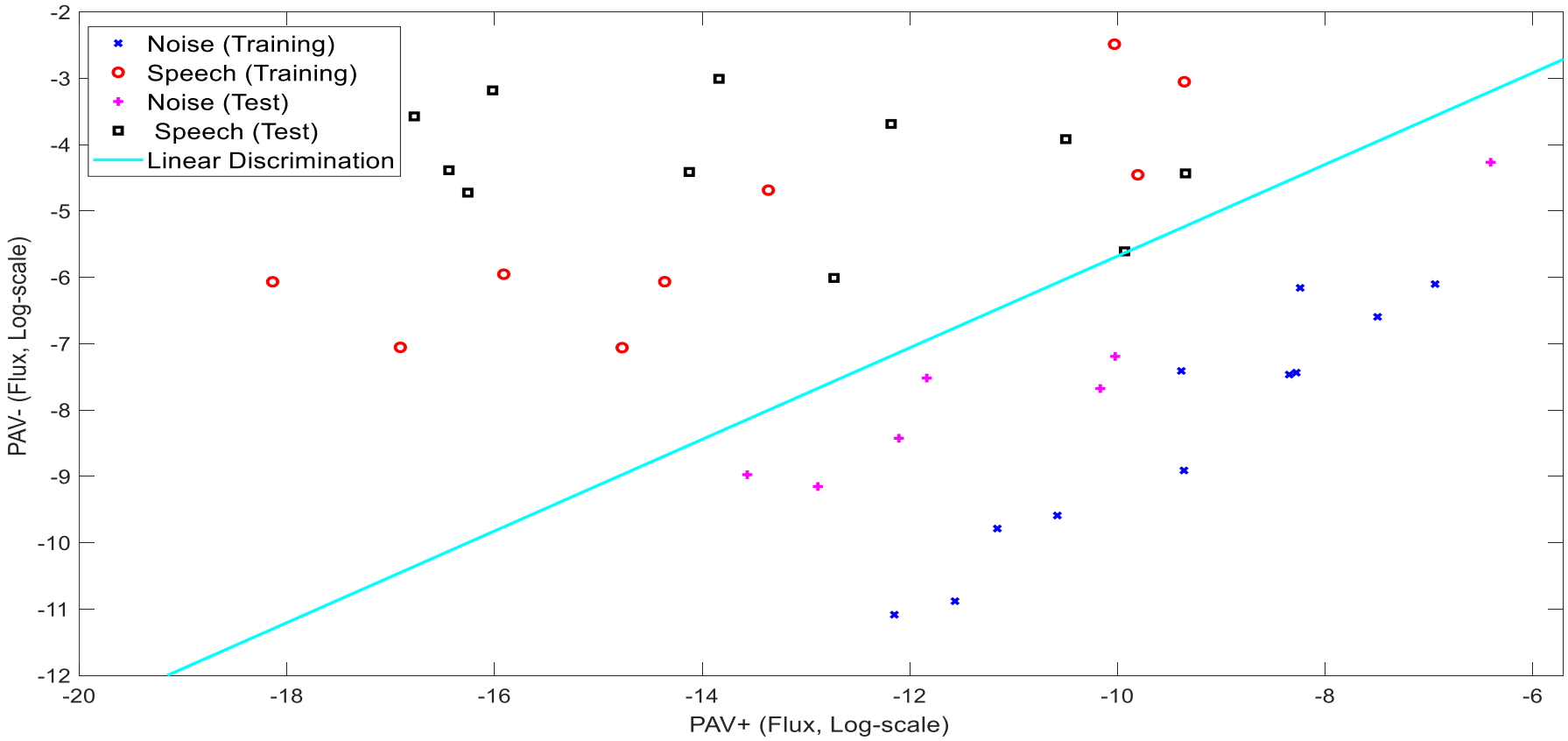
Results

1. Cross validation Results for manual boundary selection



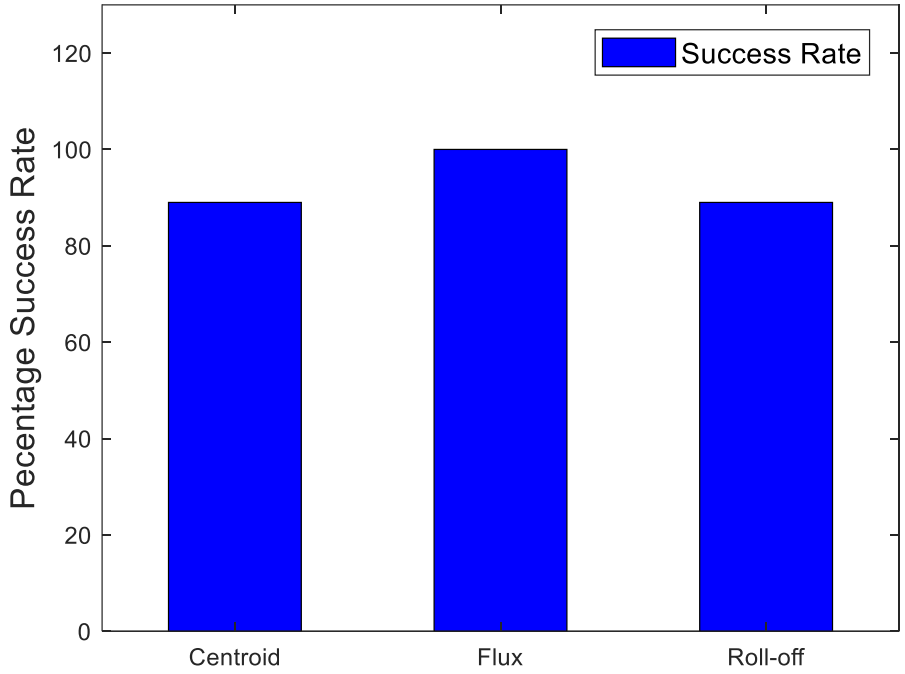
A distinction of noise and speech using Manual method with Flux values

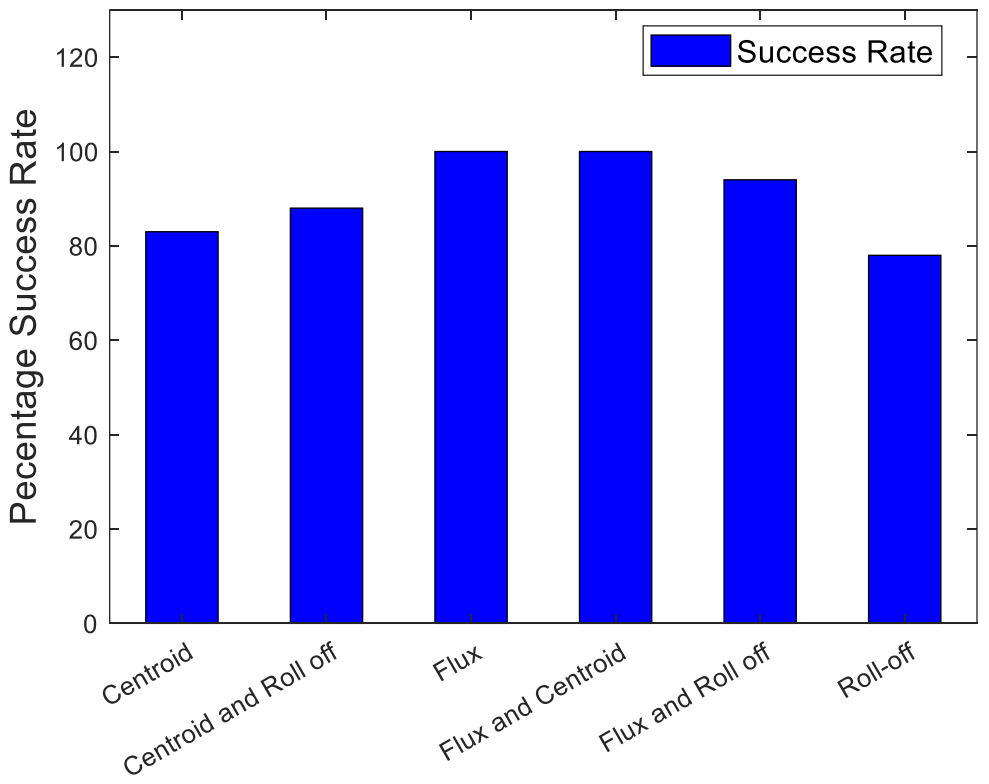
2. Cross validation Results for Machine learning



A distinction of noise and speech using ML linear classification with Flux values

3. Overall success rate for both methods and combination of parameters.



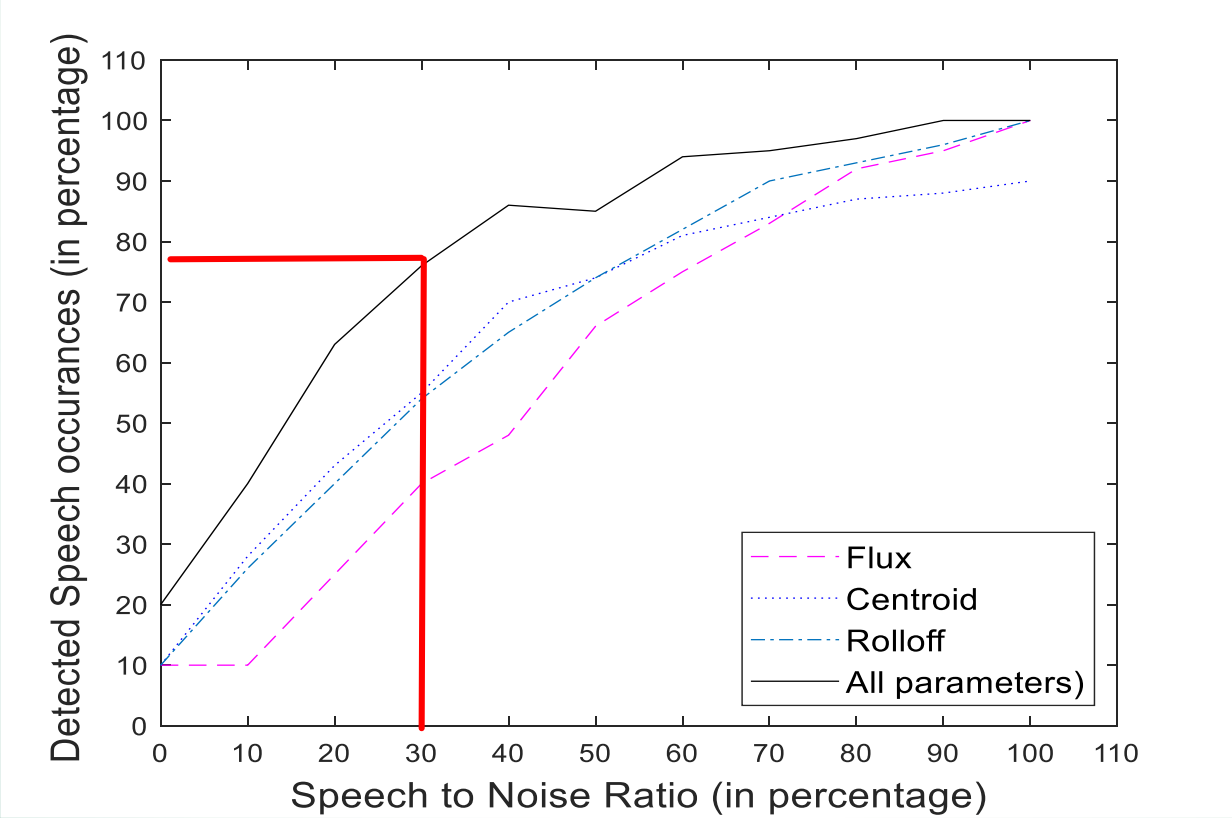


Results of cross-validation : Manual methods

Results of cross-validation : Machine learning

Additional Observation with mixed Sample type for Training and Testing

- Models tested from low to high amplitude of varying speech.



Prediction result obtained by varying speech content in the samples

Conclusion

1. Flux, Centroid, Roll-off And Their ALMM Values Can Be Used For VAD.
2. 2. Machine Learning Is More Effective Than Manual Technique.
3. 3. LDA Method Is Better Than QDA.
4. 4. Combination Of Parameters Should Be Preferred.
5. 5. Requires More Research To Detect Low Percentage Of Speech(< 30 %).

Challenges

- Insufficient literature on the methods of using spectral values.
- Recording and procuring correct samples.
- Difficult to detect low amplitude of speech in a sample.

Future Possibilities

- More data set, with diverse recordings.
- More signal features for speech identification.
- More machine learning models can be tested.

Thank you