# A COMPRESSED ENCODING SCHEME FOR APPROXIMATE TDOA ESTIMATION

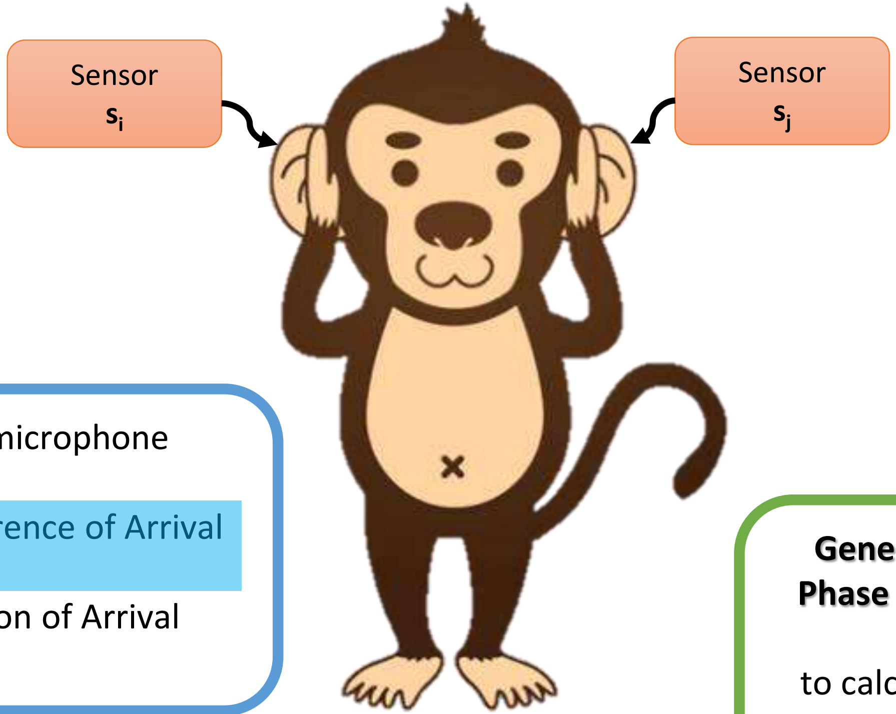
Elizabeth Vargas, James R. Hopgood, Keith Brown, Kartic Subr

*Heriot-Watt University*

*University of Edinburgh*

EUSIPCO 2018

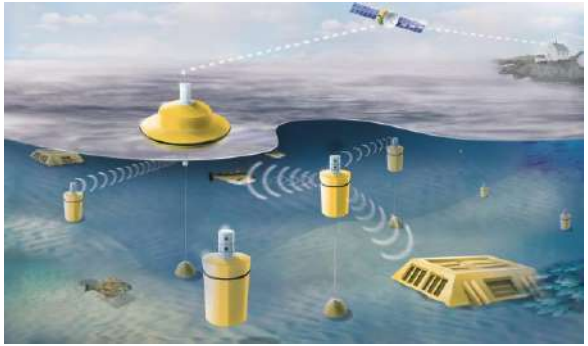
# Acoustic Source Localization




1. Record signal using microphone array
2. Calculate Time Difference of Arrival (TDOA)
3. Calculate the Direction of Arrival (DOA)

**Generalised Cross-correlation Phase Transform (GCC-PHAT)** is commonly used to calculate the Time Difference of Arrival (TDOA)

# Constraint Transmission



Underwater Sensor Networks

Inexpensive Mobile Networks
“Matrix Voice”*

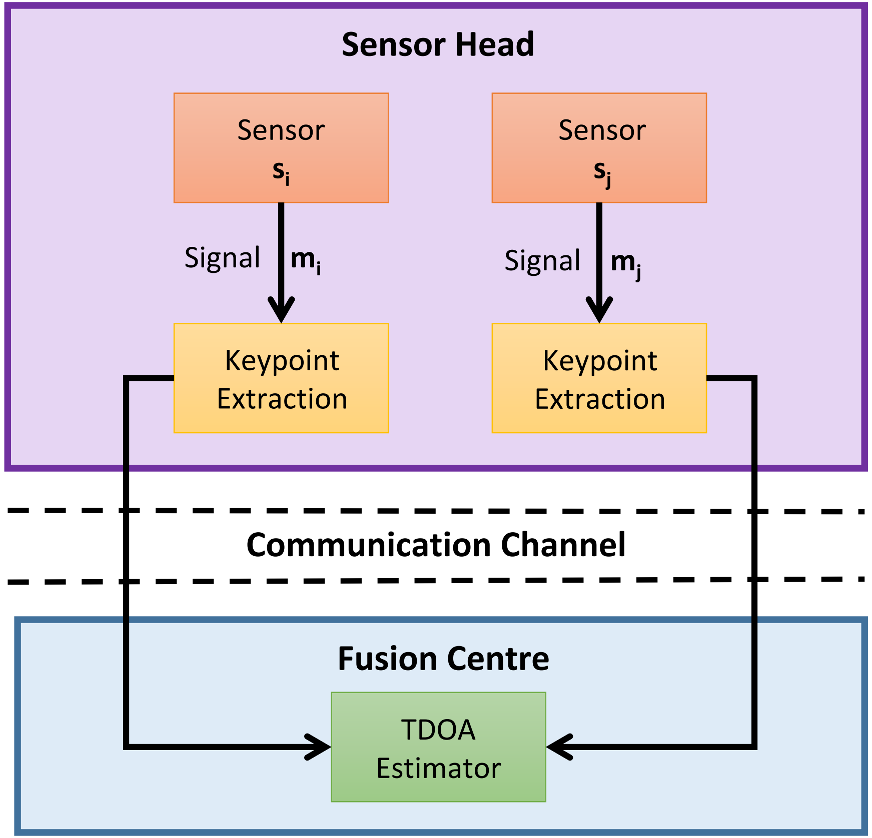
Disaster Zones

https://www.matrix.one/products/voice

# QUESTION

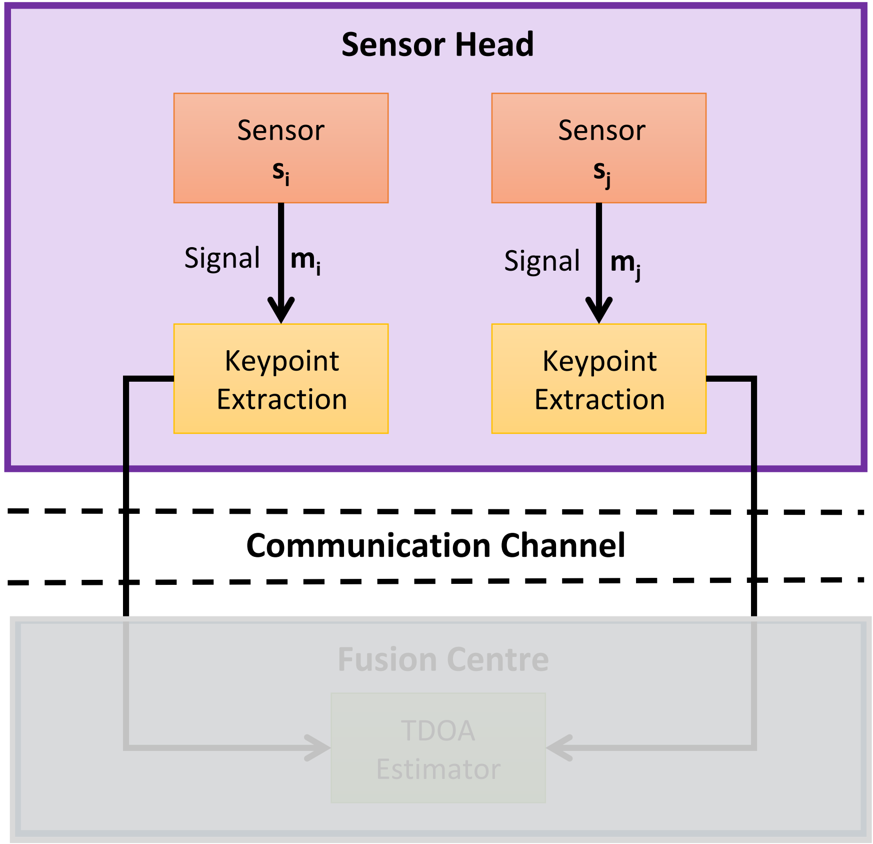
Is it necessary to use the entire signal to calculate TDOA accurately?

# Constraint Transmission



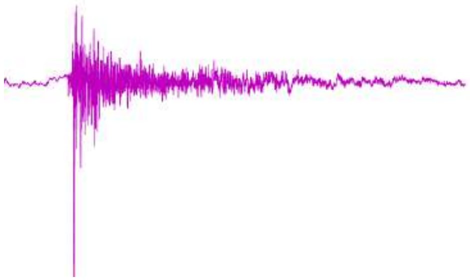
# PROPOSED METHOD

# Constraint Transmission

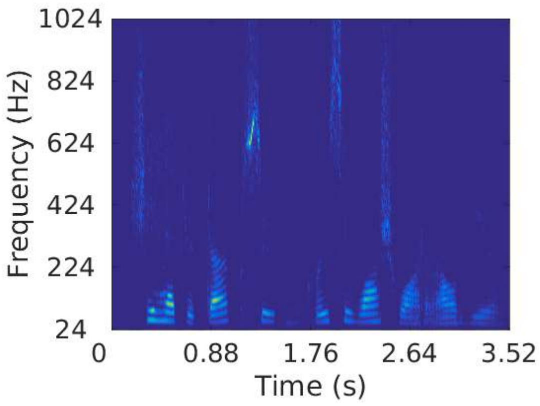


# Proposed Method

**Spectrogram** → Keypoints → Add Row Coefficients → High Energy Rows → Binary Mask → Cross Correlation



**Signal**




**Spectrogram**

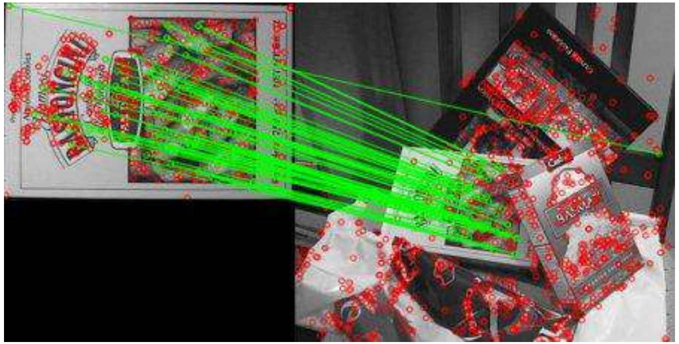
**Parameters**

```
window = 256
noverlap = 80%
nfft = 1024
```

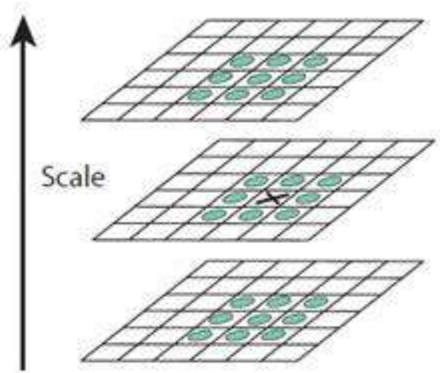
# PROPOSED METHOD

Spectrogram → **Keypoints** → Add Row Coefficients → High Energy Rows → Binary Mask → Cross Correlation

## Scale Invariant Feature Transform (SIFT)



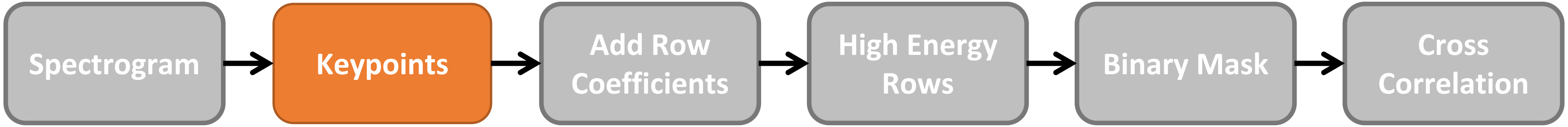
**Computer Vision Technique**
Keypoint detection for object recognition



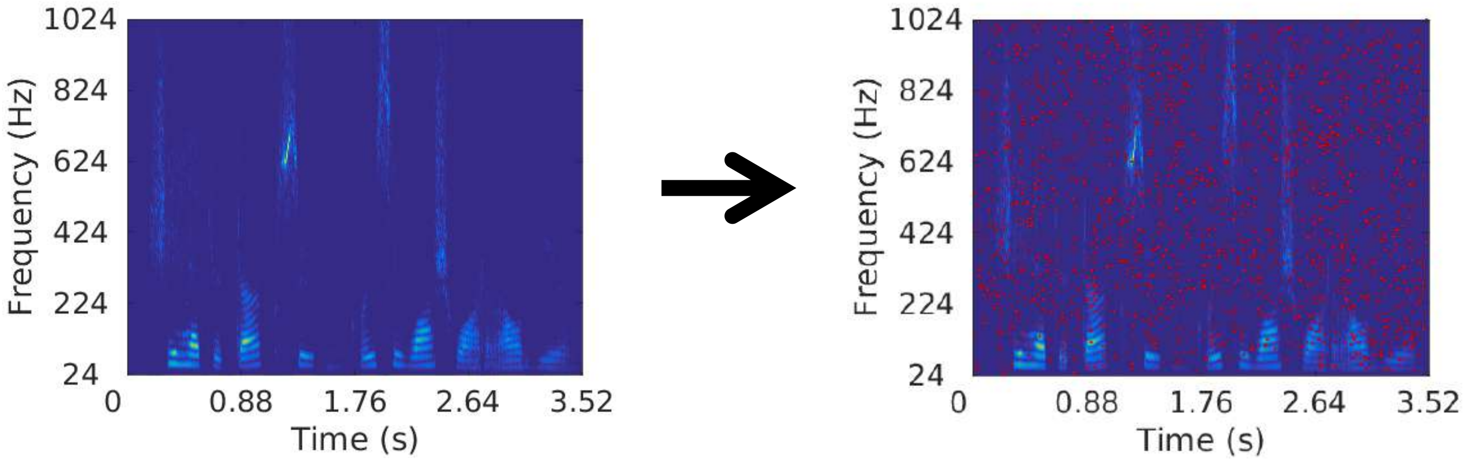

We applied the keypoint detector to the signal spectrogram

https://opencv-python-tutroals.readthedocs.io/en/latest/py_tutorials/py_feature2d/py_sift_intro/py_sift_intro.html
https://docs.opencv.org/3.4/dc/dc3/tutorial_py_matcher.html

# Proposed Method

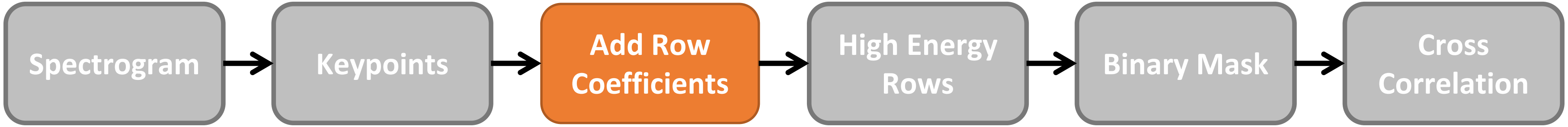


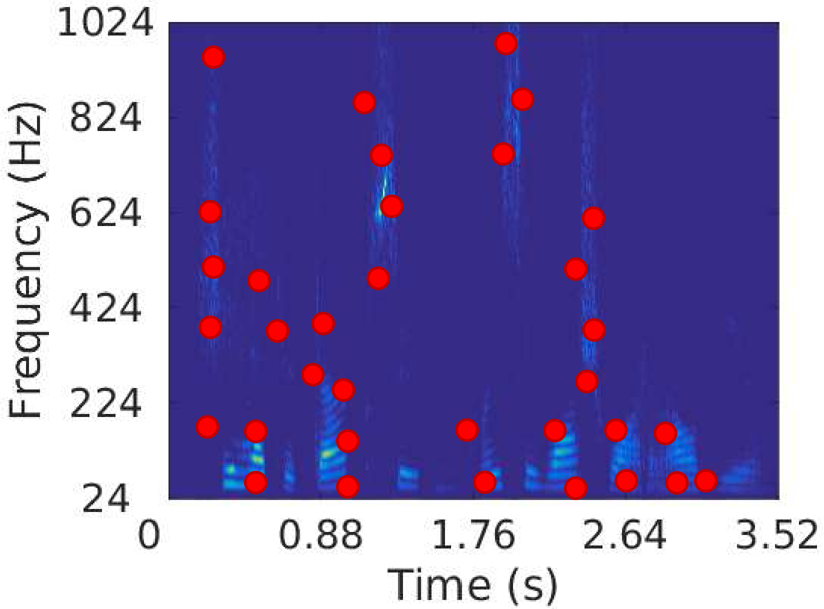

**Scale Invariant Feature Transform (SIFT)** performed on the spectrogram



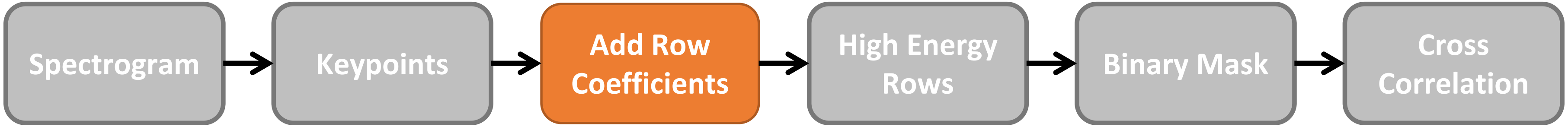

**spectrogram** **keypoints**

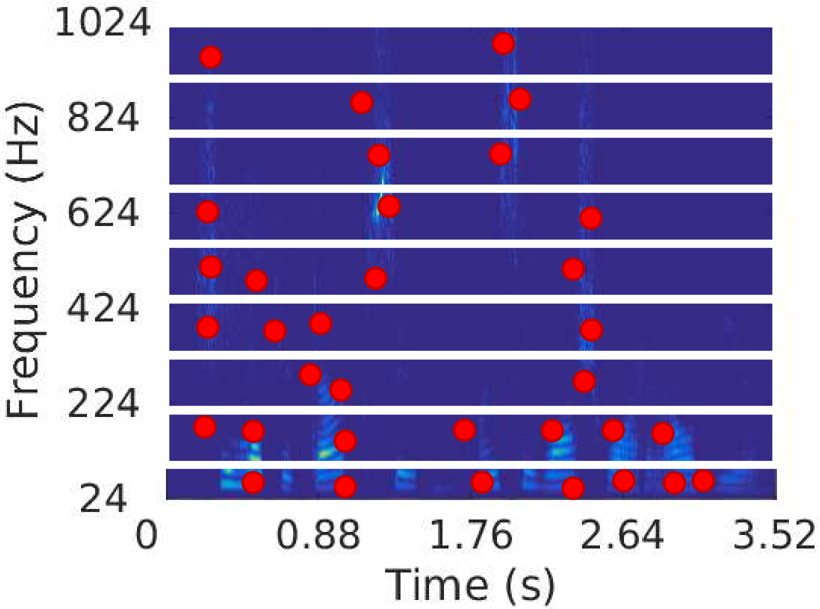
# PROPOSED METHOD



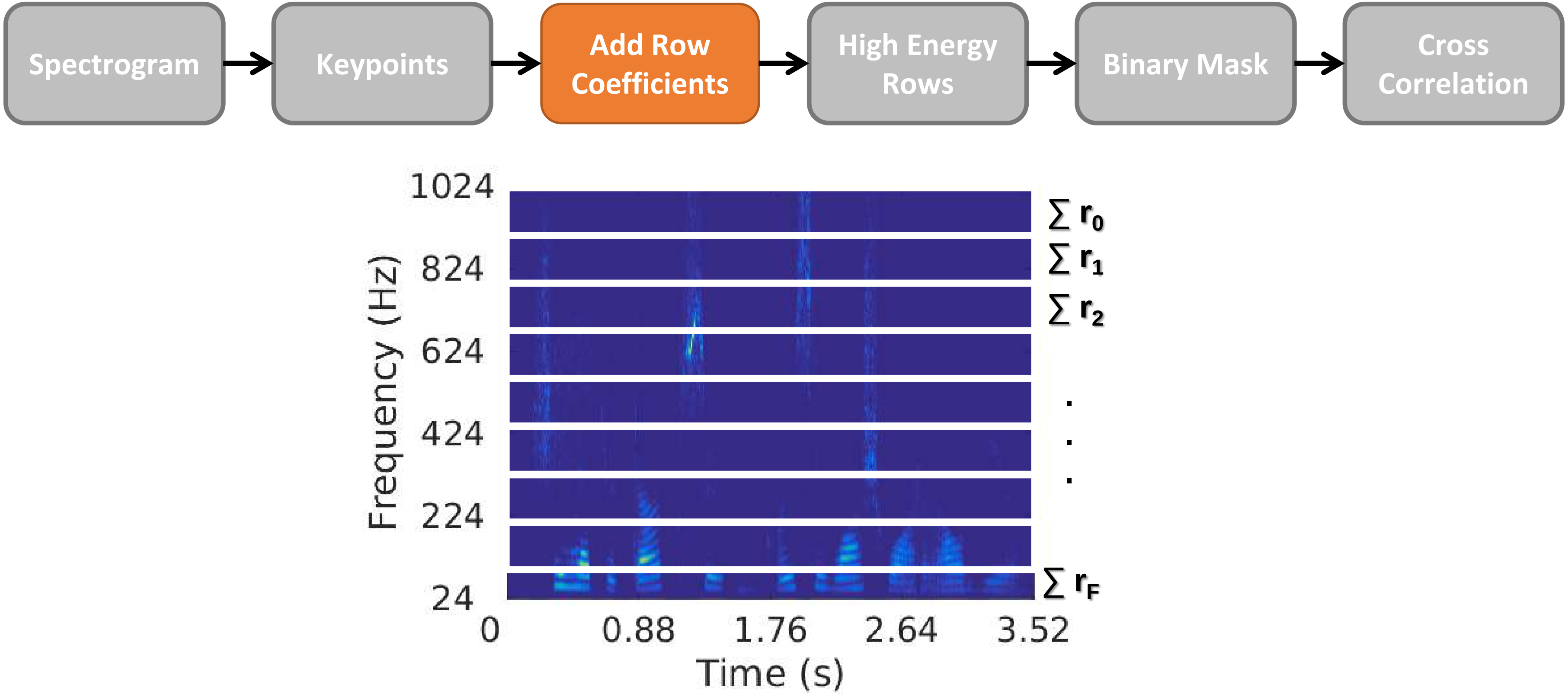



# PROPOSED METHOD

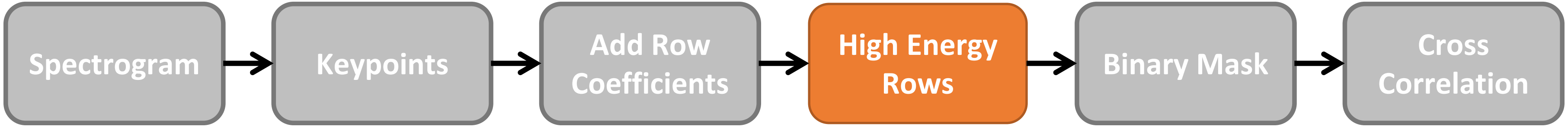


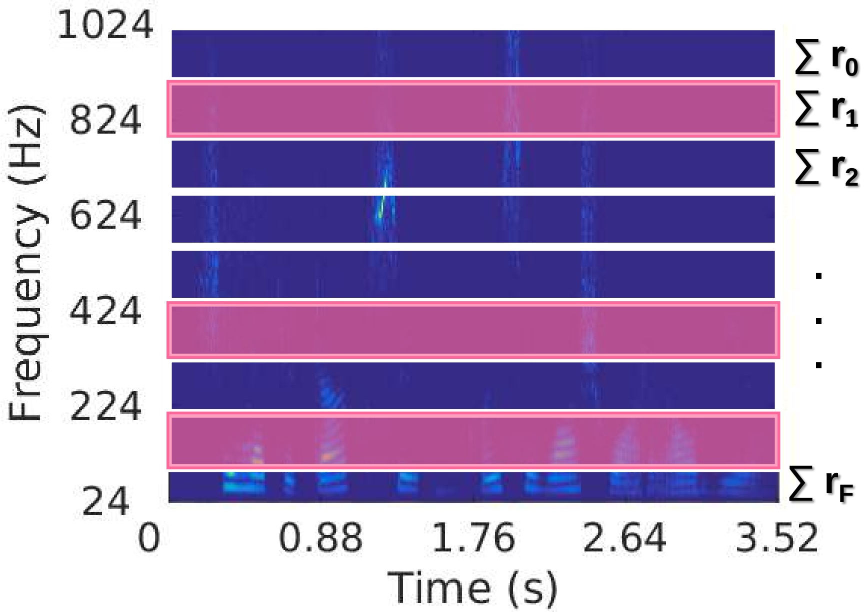



# PROPOSED METHOD



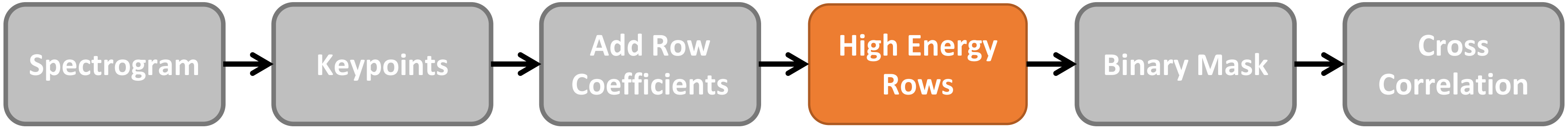
# Proposed Method




**Use rows with highest sums.**
**We used a random configuration initially, however it did not provide good results**

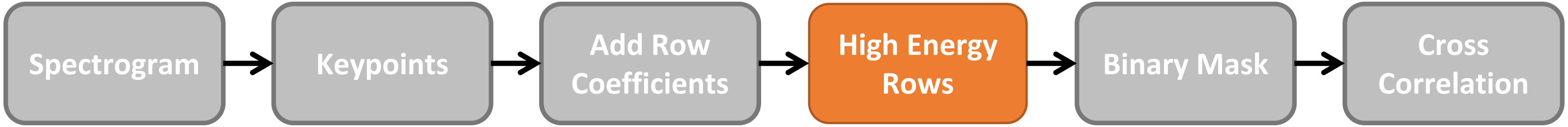
# PROPOSED METHOD



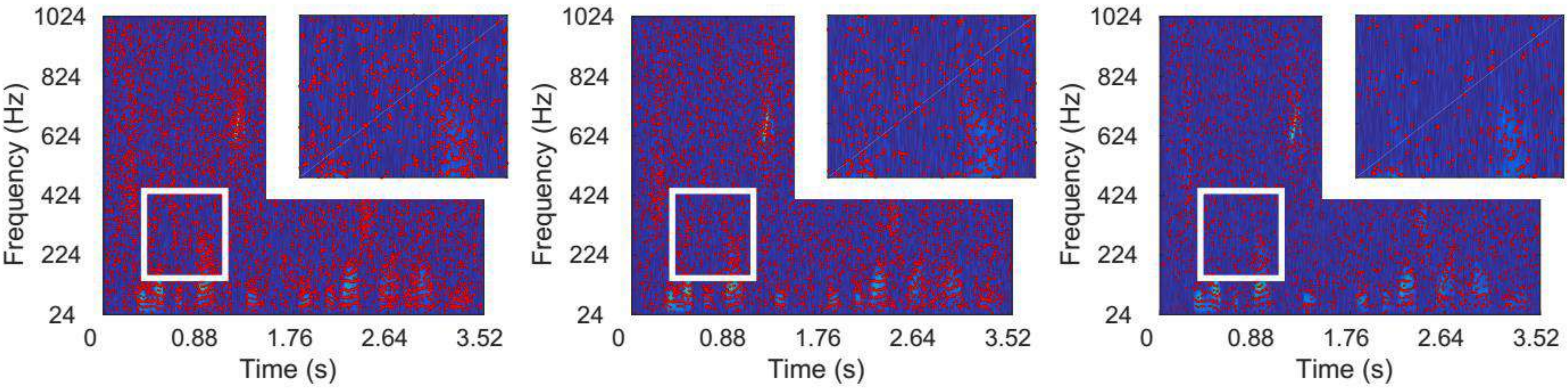

## Compression Ratio

$$compression\ ratio = \frac{uncompressed\ size}{compressed\ size}$$

# PROPOSED METHOD



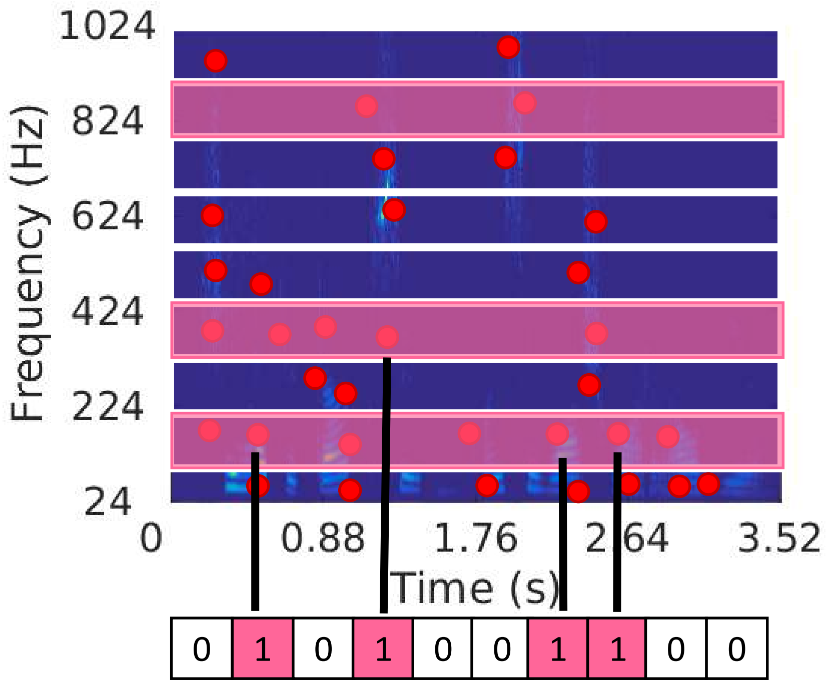

Various compression ratios represented as chosen rows (keypoints)



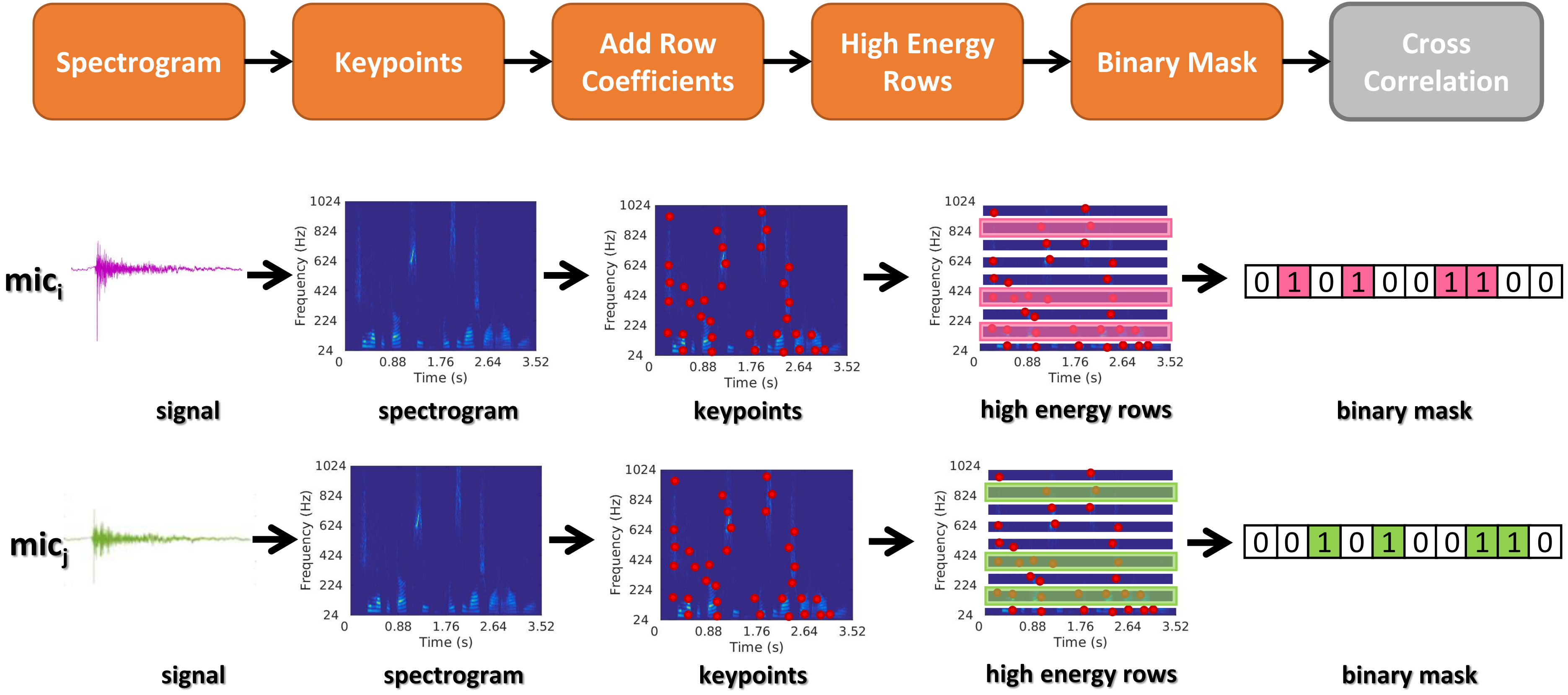

**40 : 1** **55 : 1** **90 : 1**

# PROPOSED METHOD

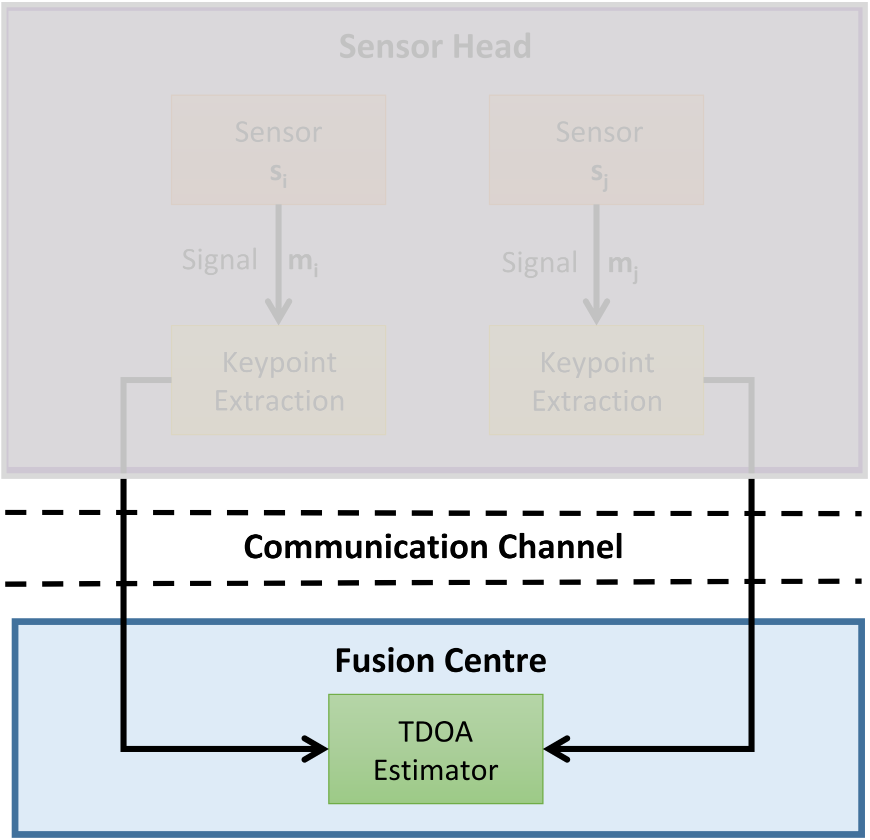
Spectrogram → Keypoints → Add Row Coefficients → High Energy Rows → **Binary Mask** → Cross Correlation



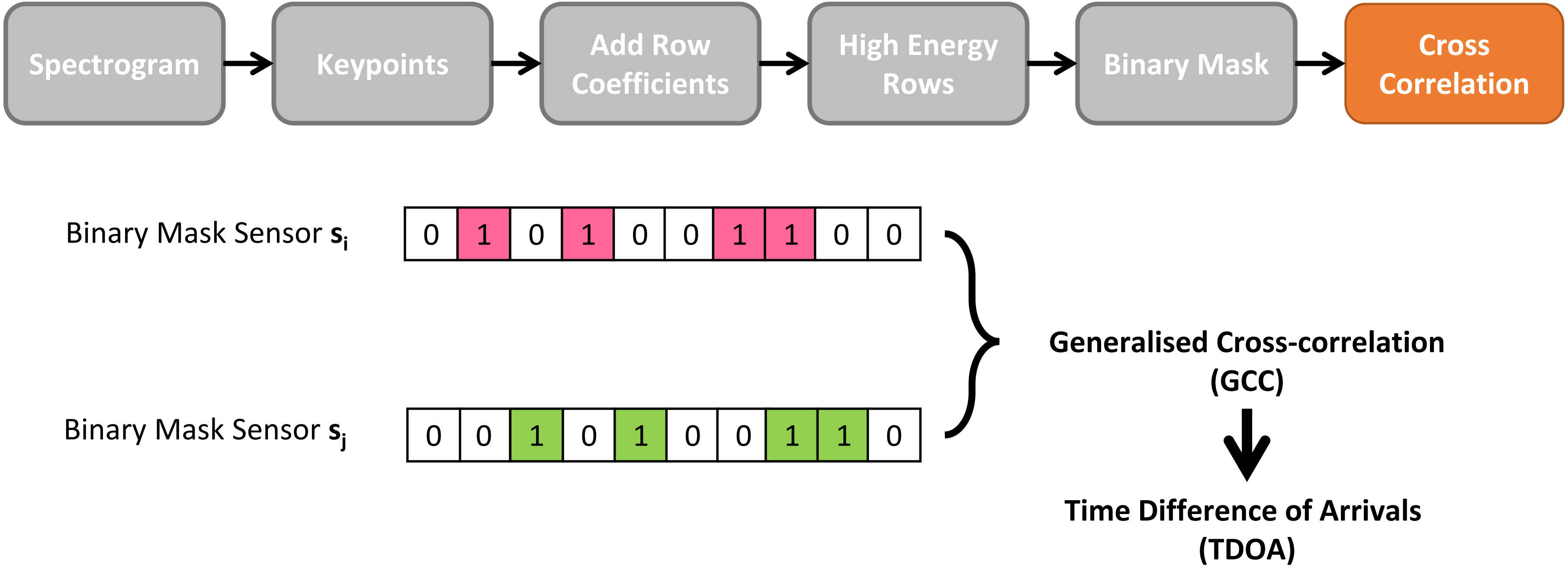
# PROPOSED METHOD



# Constraint Transmission



# Proposed Method




**Calculation is very fast because the masks are binary**

# RESULTS

# METRICS

- **Relative Error**

$$tdoa\ error\ (\%) = \frac{\|tdoa\ - gt\|}{\|gt\|} * 100$$

$$doa\ error\ (\%) = \frac{\|doa\ - gt\|}{\|gt\|} * 100$$

The TDOA depends on the distance between Microphones and the source location.

Relative error help us compare TDOA error for different source locations.

**What does it mean?**

- If the ground truth is 0.02 and I get 0.01 my relative error is **50%**
- If the ground truth is 0.02 and I get 0.001 my relative error is **95%**
- If the ground truth is 0.02 and I get 0.1 my relative error is **400%**

# METRICS

- **Compression**

$$compression\ ratio = \frac{uncompressed\ size}{compressed\ size}$$

$$compression = \frac{signal\ length\ * 64\ (bits)}{number\ of\ one's\ * 32\ (bits)}$$

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| 0 | 1 | 0 | 1 | 0 | 0 | 1 | 1 | 0 | 0 |

# DATA USED

- Simulated environments by means of the image-source method

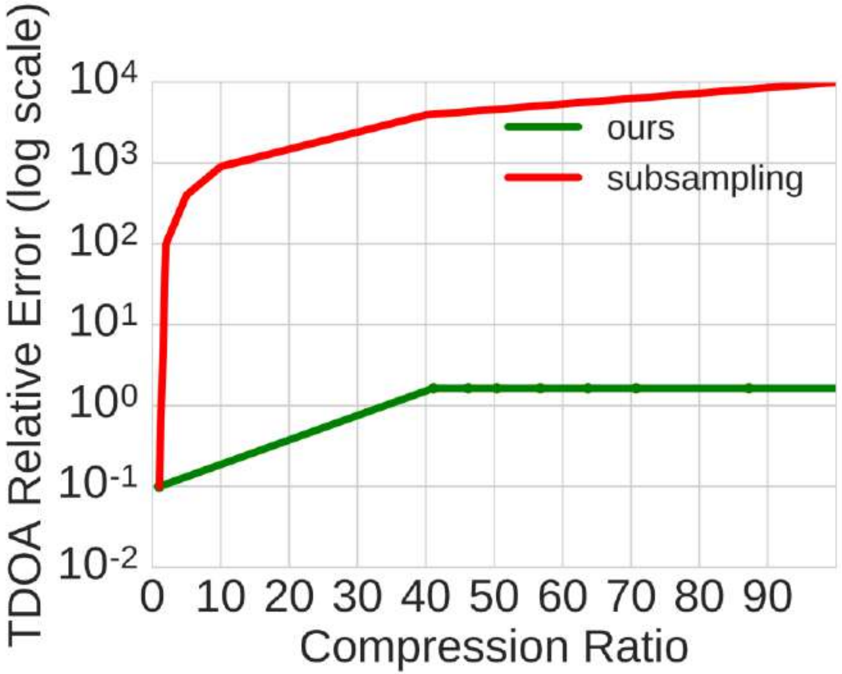
**http://www.eric-lehmann.com/ism_code.html**

- Experiments using speech signals from the TIMIT database

**https://catalog.ldc.upenn.edu/ldc93s1**

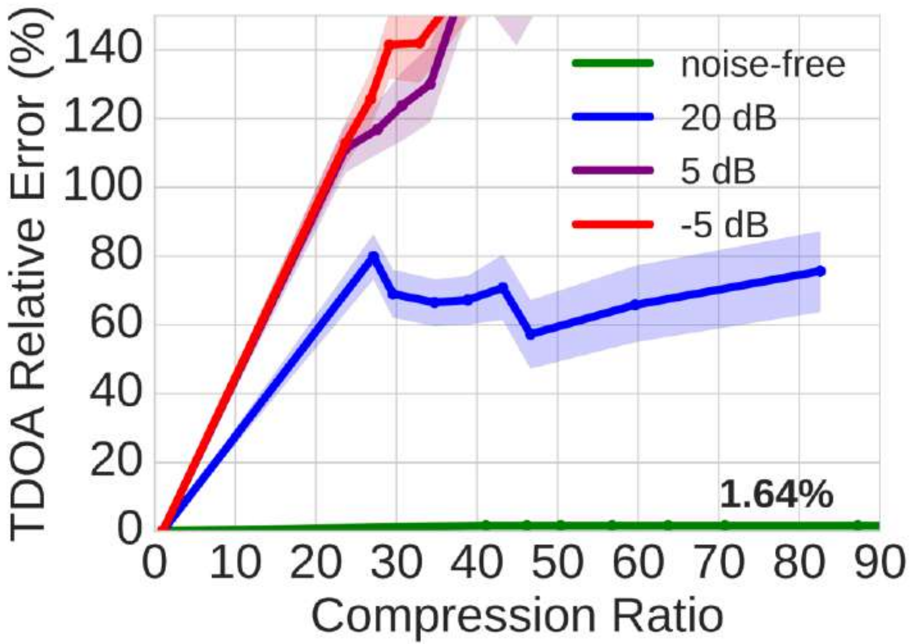
# PROPOSED METHOD VS NAÏVE APPROACH

The accuracy provided by our method is better than the accuracy when subsampling the signal



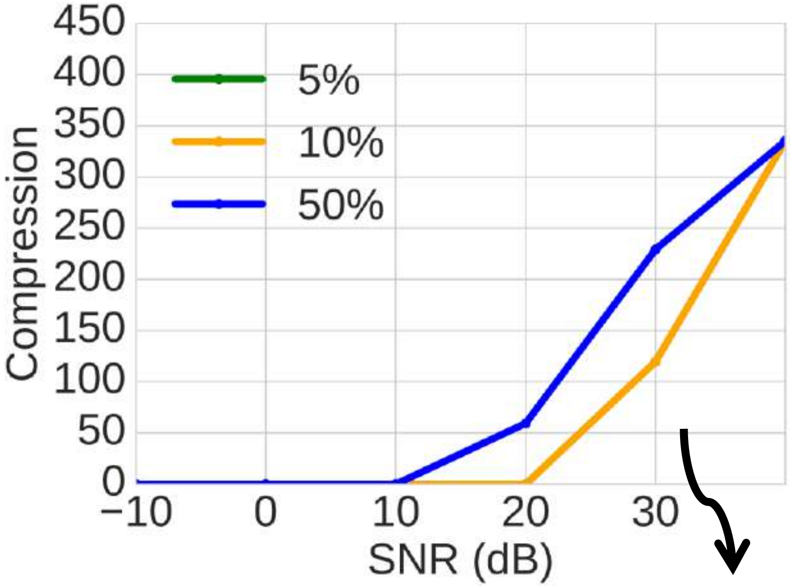
# PROPOSED METHOD ON VARIOUS NOISE LEVELS

Our approach produces reasonable accuracy for low noise levels



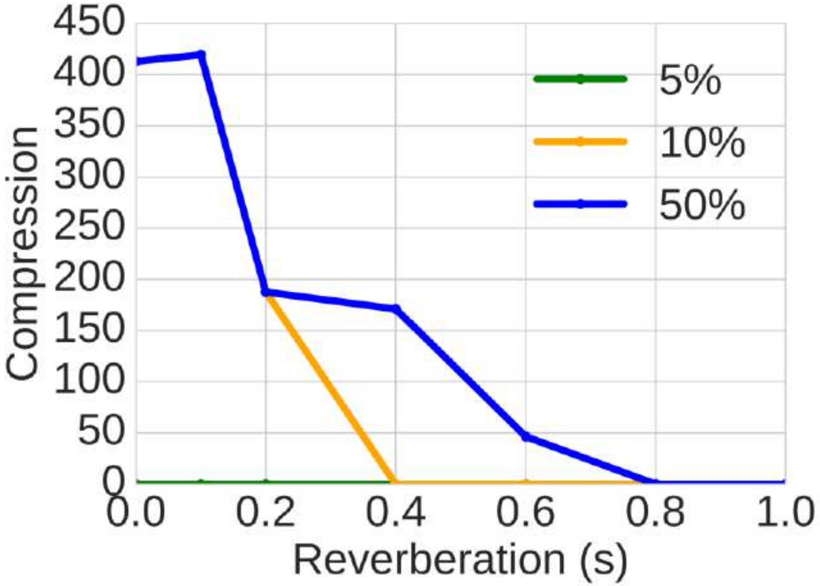
# ROBUSTNESS TO NOISE AND REVERBERATION

We calculated the maximum amount of compression achieved when obtaining an error below **50%, 10%** and **5%**




**Noise**

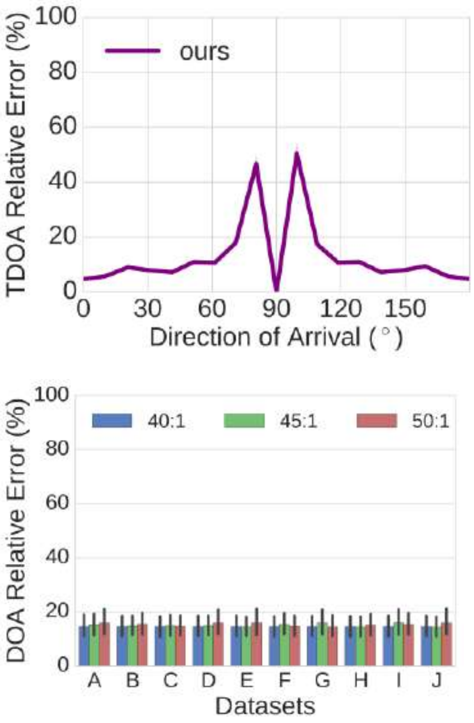
*Green and yellow line are the same*



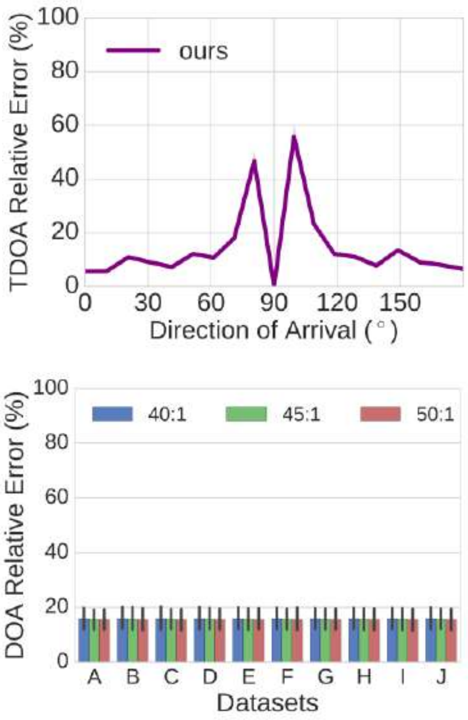

**Reverberation**

# DIRECTION OF ARRIVAL: 10 DIFFERENT SPEECH SIGNALS

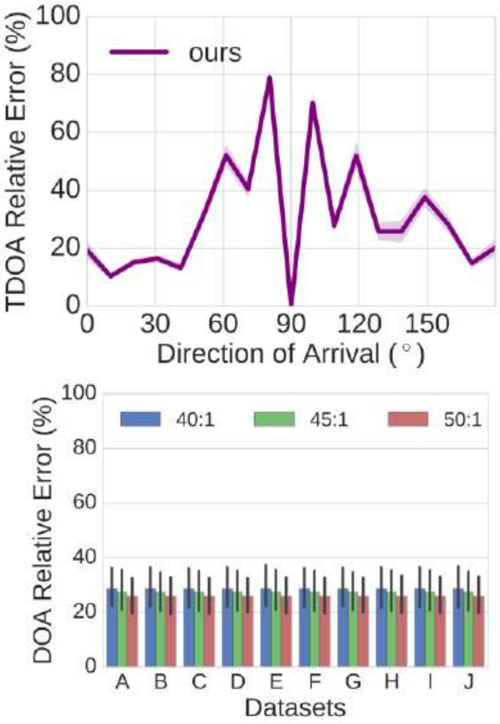
**reverberation $T_{60}$ = 0.1**




**reverberation $T_{60}$ = 0.2**




**reverberation $T_{60}$ = 0.3**



# CONCLUSIONS

## QUESTION

Is it necessary to use the entire signal to calculate TDOA accurately?

- It is not necessary to use the entire signal. It is possible to select signal samples to calculate TDOA.
- Computer Vision techniques such as **SIFT** (scaled-invariant feature transform) applied to the signal spectrogram is a possible way to calculate TDOA accurately.
- We tested our algorithm under low noise up to -10 dB and reverberation up to 0.8 seconds.
- Further work is needed on the keypoint selection and pruning, considering reverberation and noise, e.g. including information about neighbouring points.