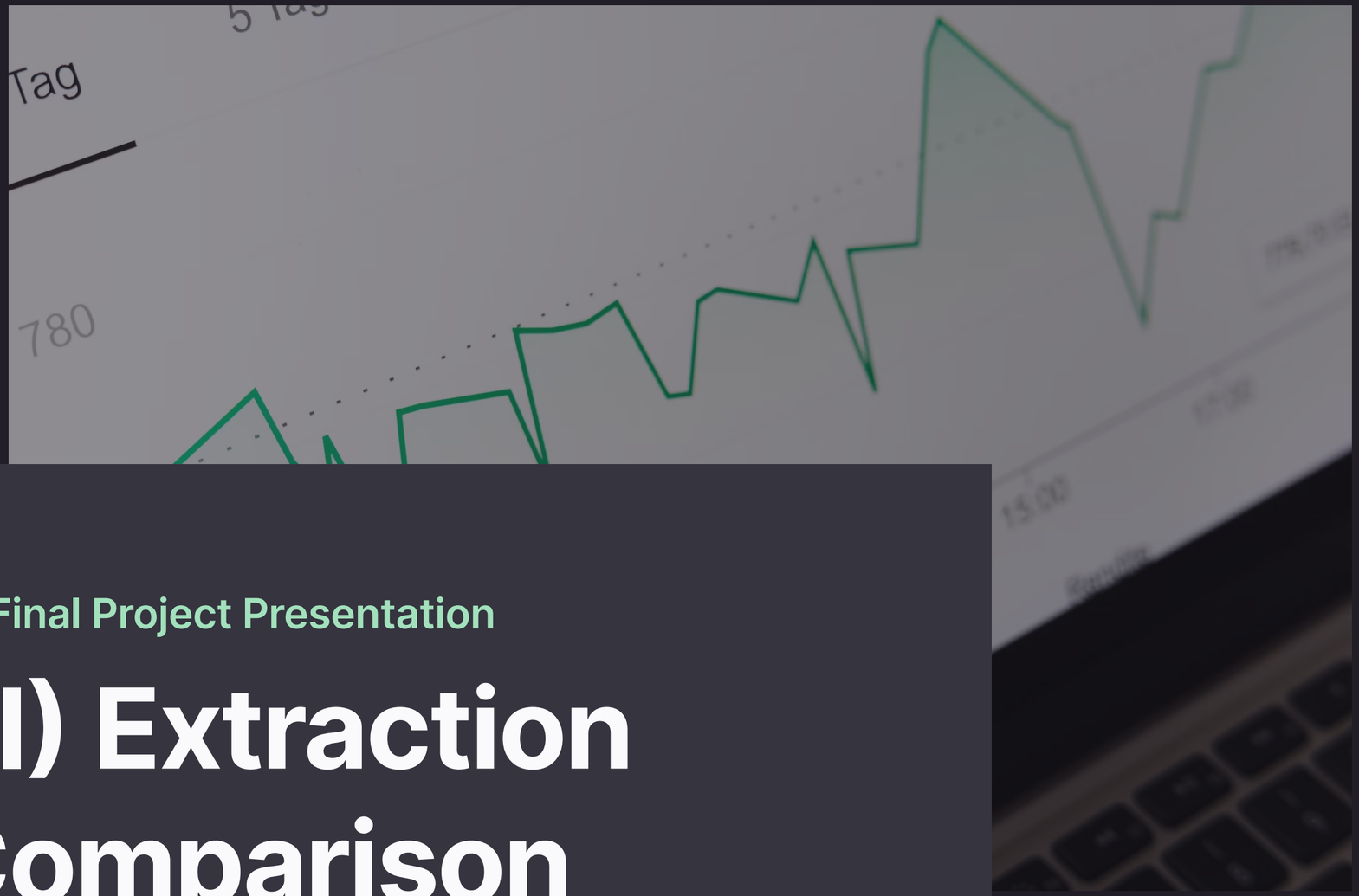


Speech Signal Processing - Final Project Presentation

Epoch (GCI) Extraction Methods Comparison

Abhinav Marri
Harinie Sivaramasethu

Table of Contents

Objectives

1. Analyse different extraction methods

2. Implement 3-4 methods

3. Get Metrics

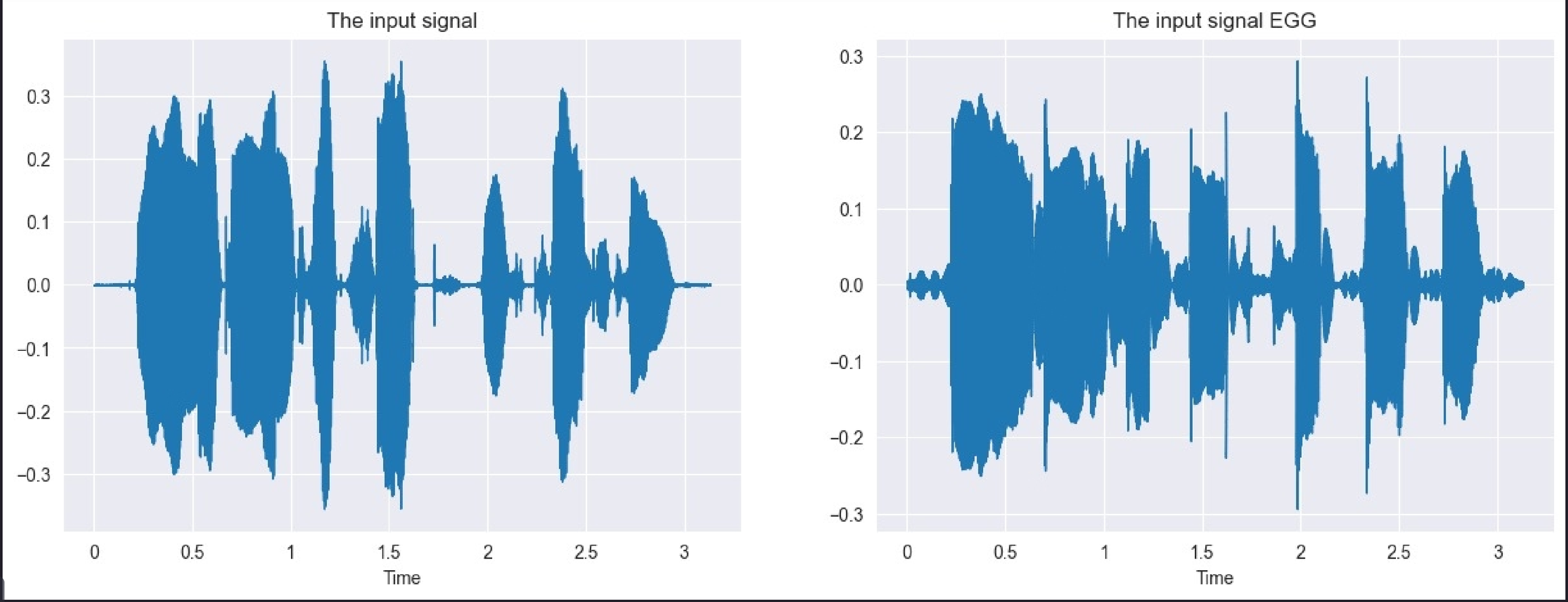
4. Comprehensive study of complexity and quality of results

Methods Analysed:

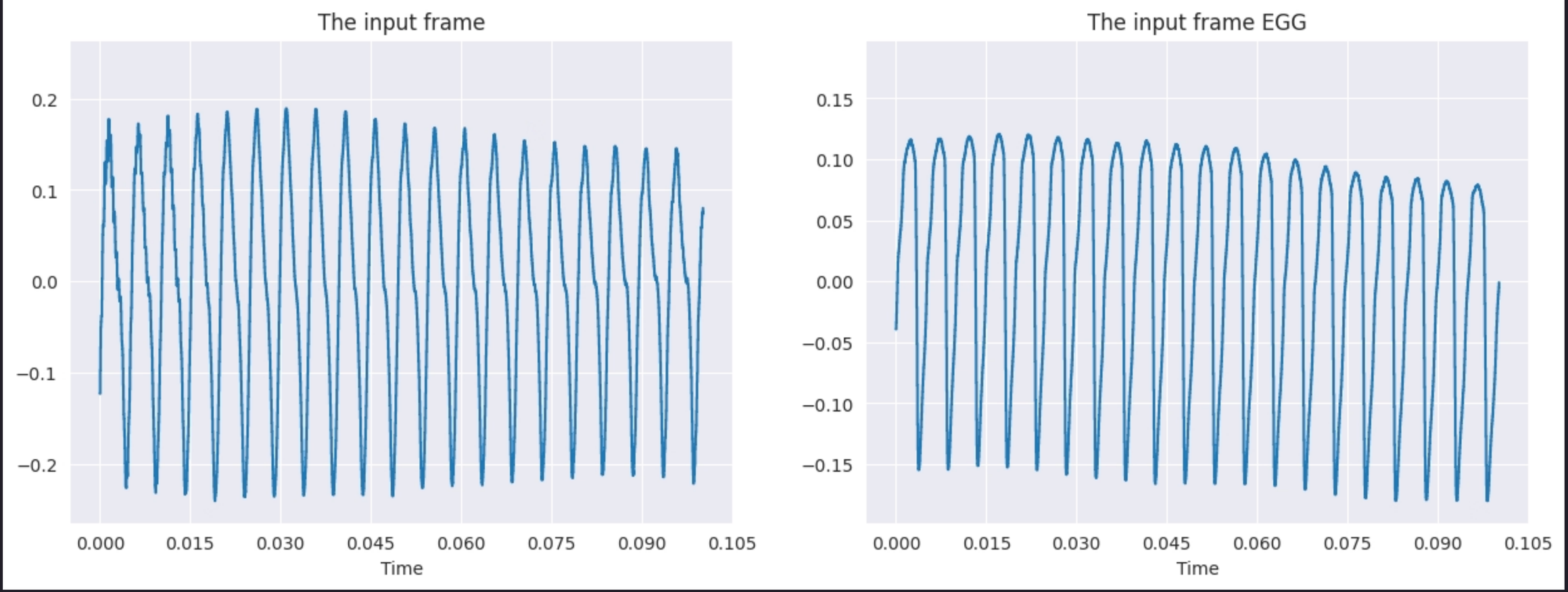
- LP Error
- MLED (Maximum Likelihood Epoch Estimation)
- ZFF (Zero Frequency Filtering)
- ZP-ZFR (Zero-Phase Zero Frequency Resonator)
- DYPSA (Dynamic Programming Phase Slope Algorithm)

Input

- We used CMU_ARCTIC database which contains sterio signals (Speech and corresponding EGG)
- Following signal was used for all the methods:

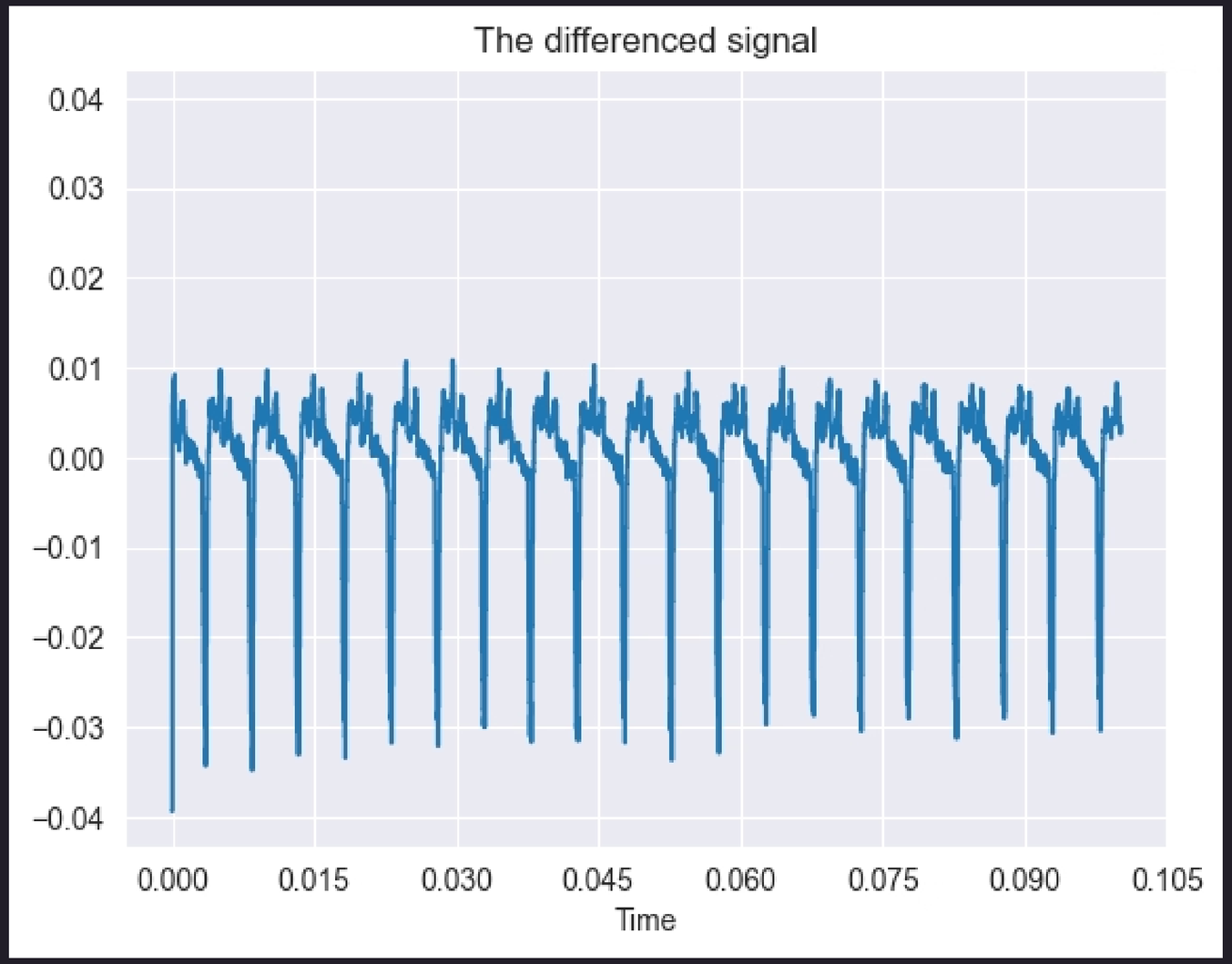


The following voiced frame was chosen and has a duration of 100 ms.



Try Pitch

Differenced EGG (electroglottograph)



The negative peaks of the differenced EGG signify epoch locations.

1] LP ERROR

Reference Paper:

TIRUPATTUR V. ANANTHAPADMANABHA AND B. YEGNANARAYANA, MEMBER, IEEE

"Epoch Extraction from Linear Prediction Residual for Identification of Closed Glottis Interval"

Steps:

- Find Linear Predicted Signal
- Find LP Error
- The LP Error peaks coincides with the second derivative of glottal pulses.
- Hilbert Envelop is added to smoothen out the signal

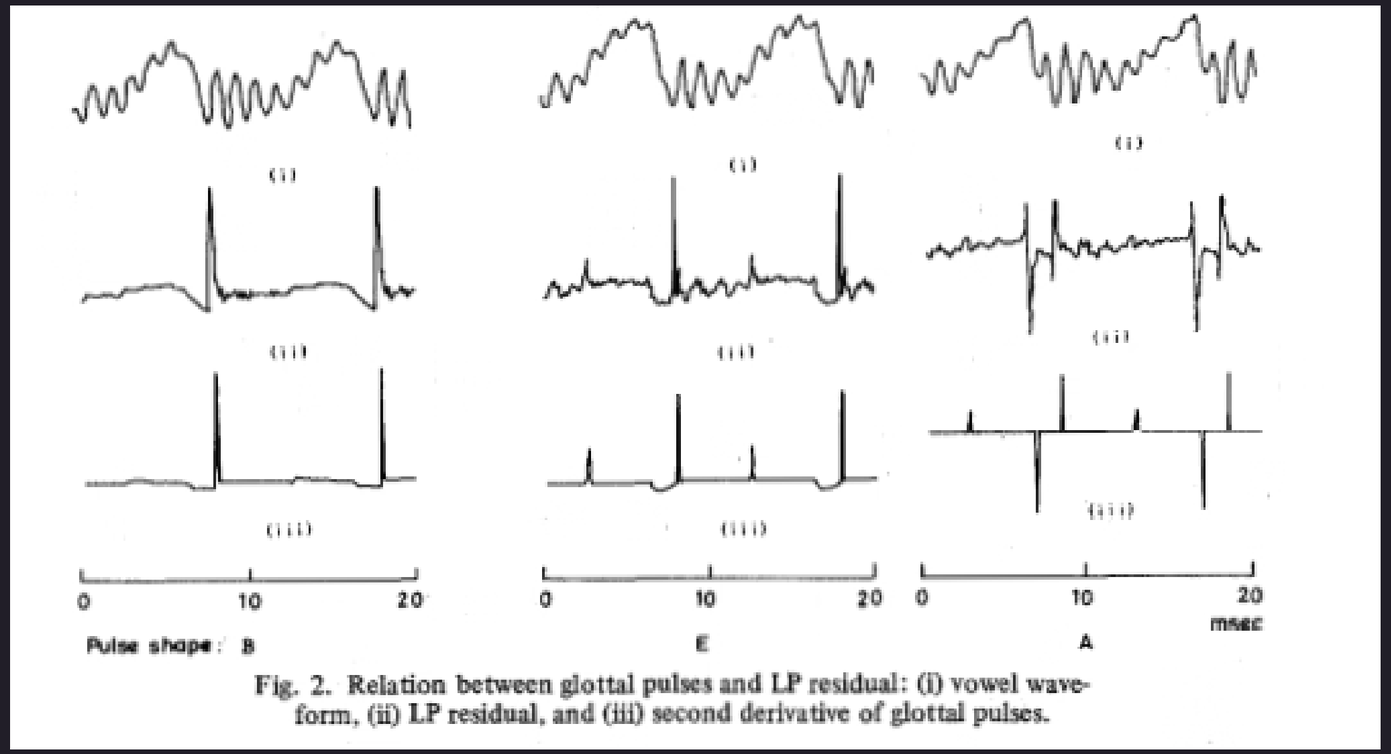
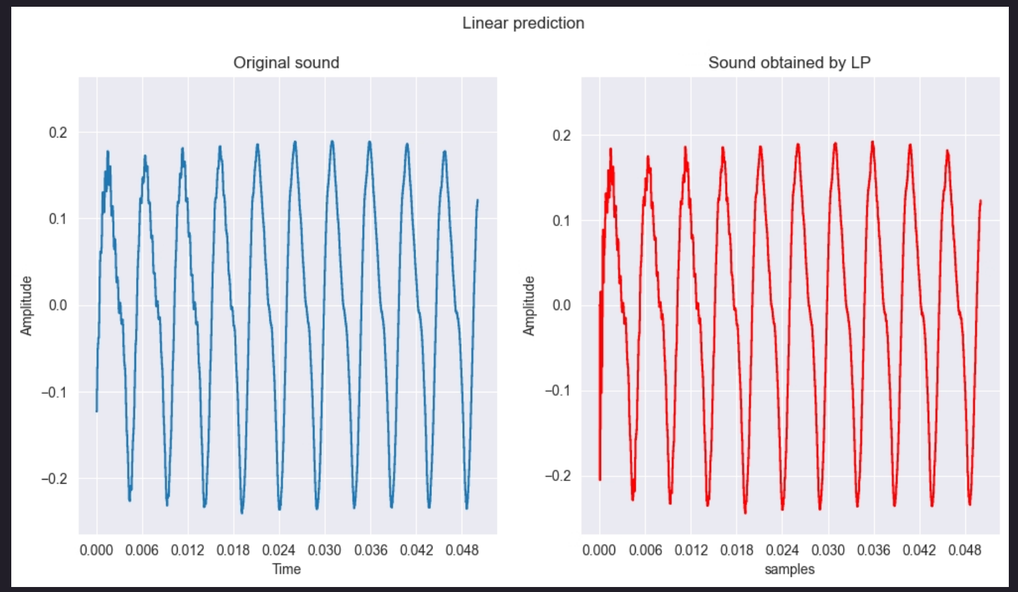
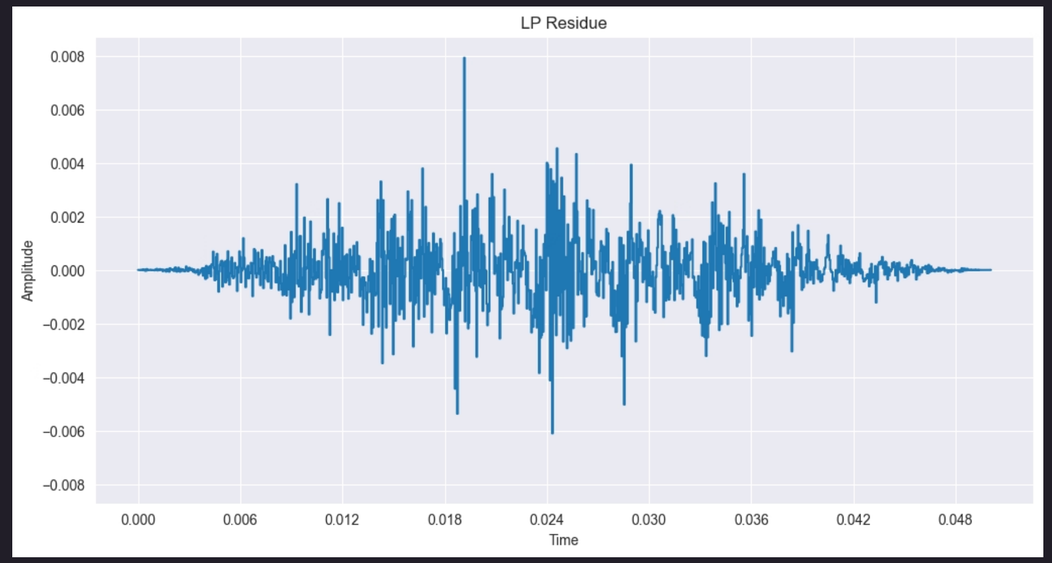


Fig. 2. Relation between glottal pulses and LP residual: (i) vowel waveform, (ii) LP residual, and (iii) second derivative of glottal pulses.

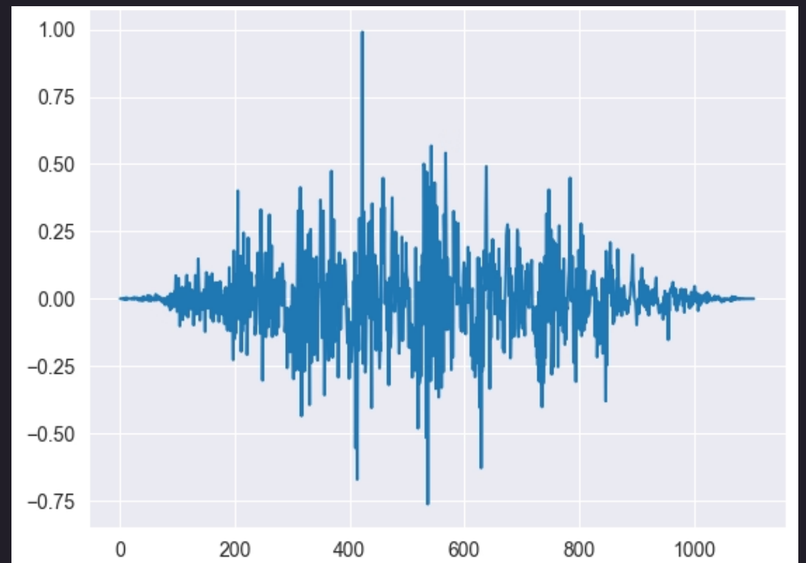


Differencing

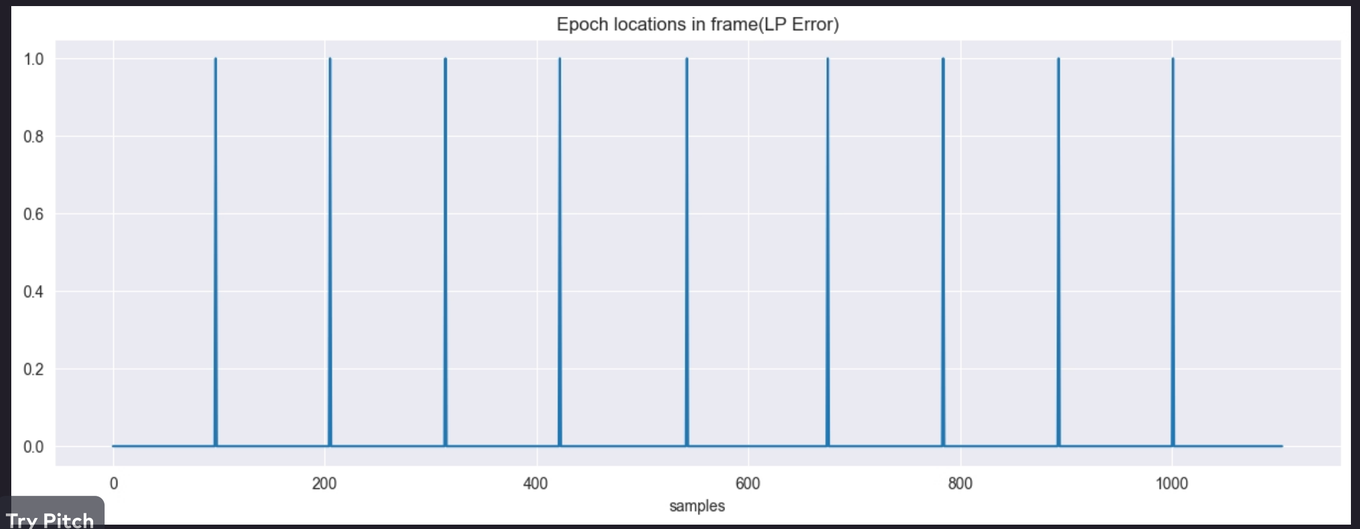
Lp residual and windowing



Applying Hilbert Envelope

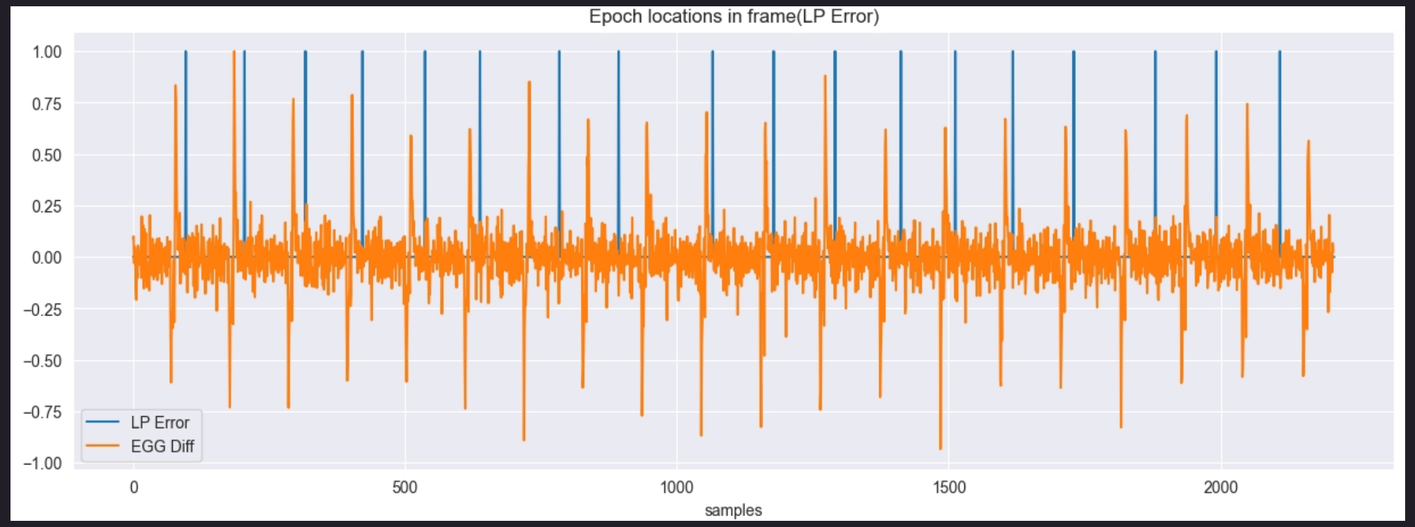


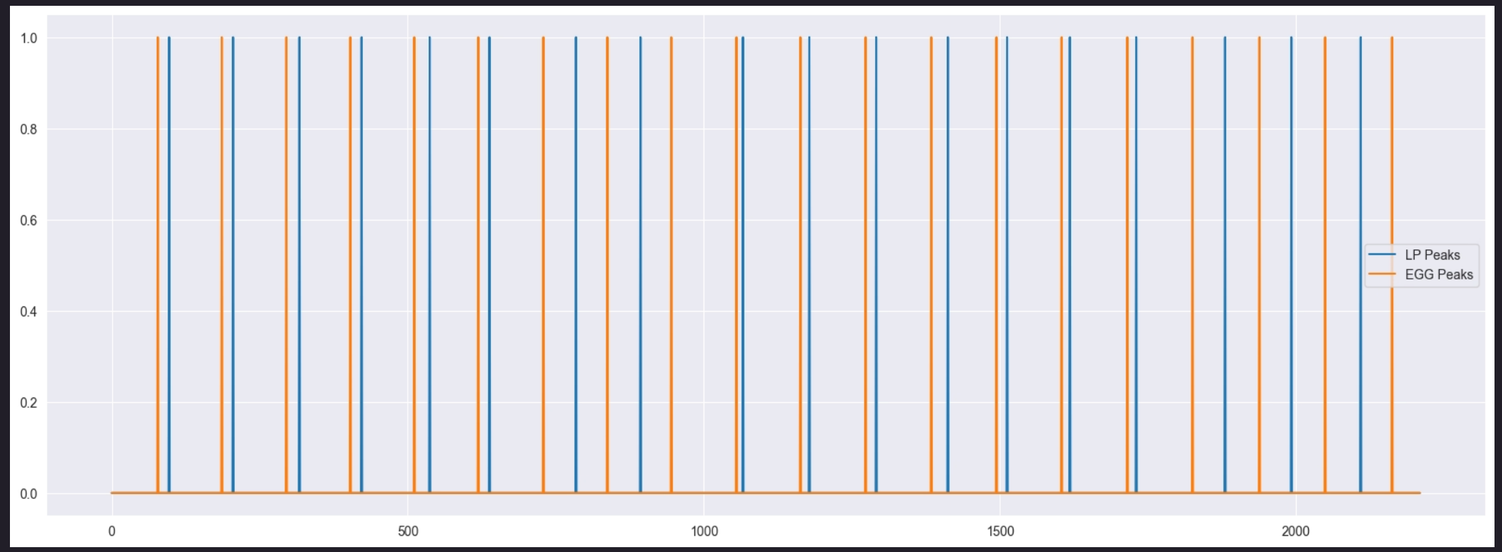
Extracting Epochs



Try Pitch

Epoch locations coincide with double differenced EGG signal.



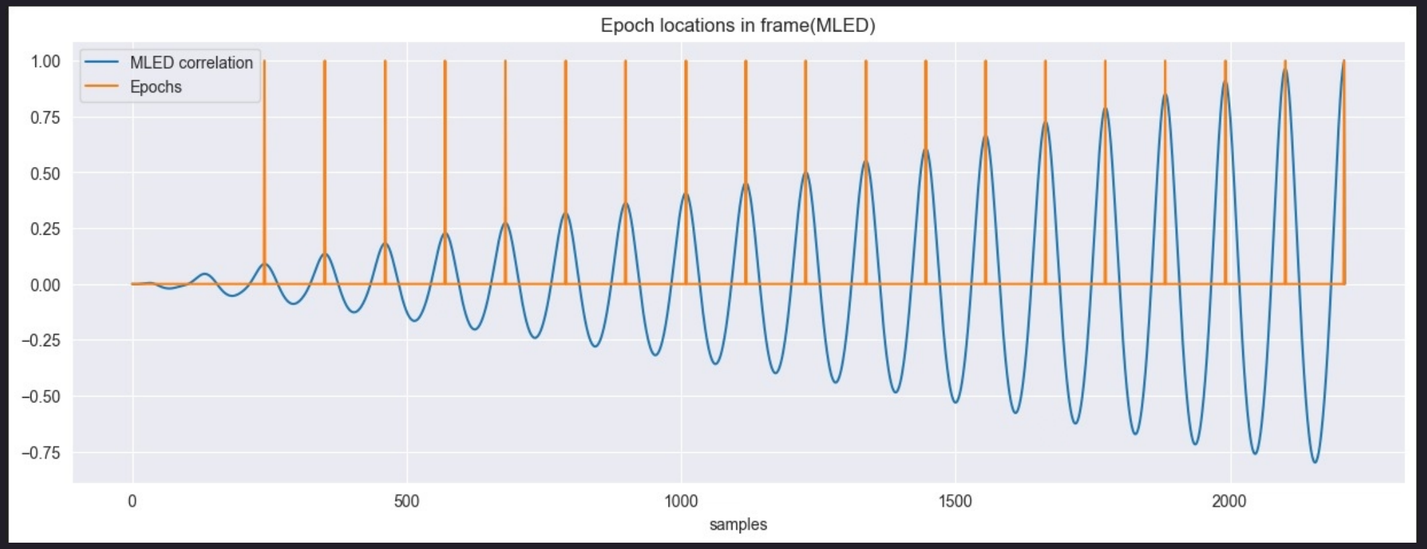


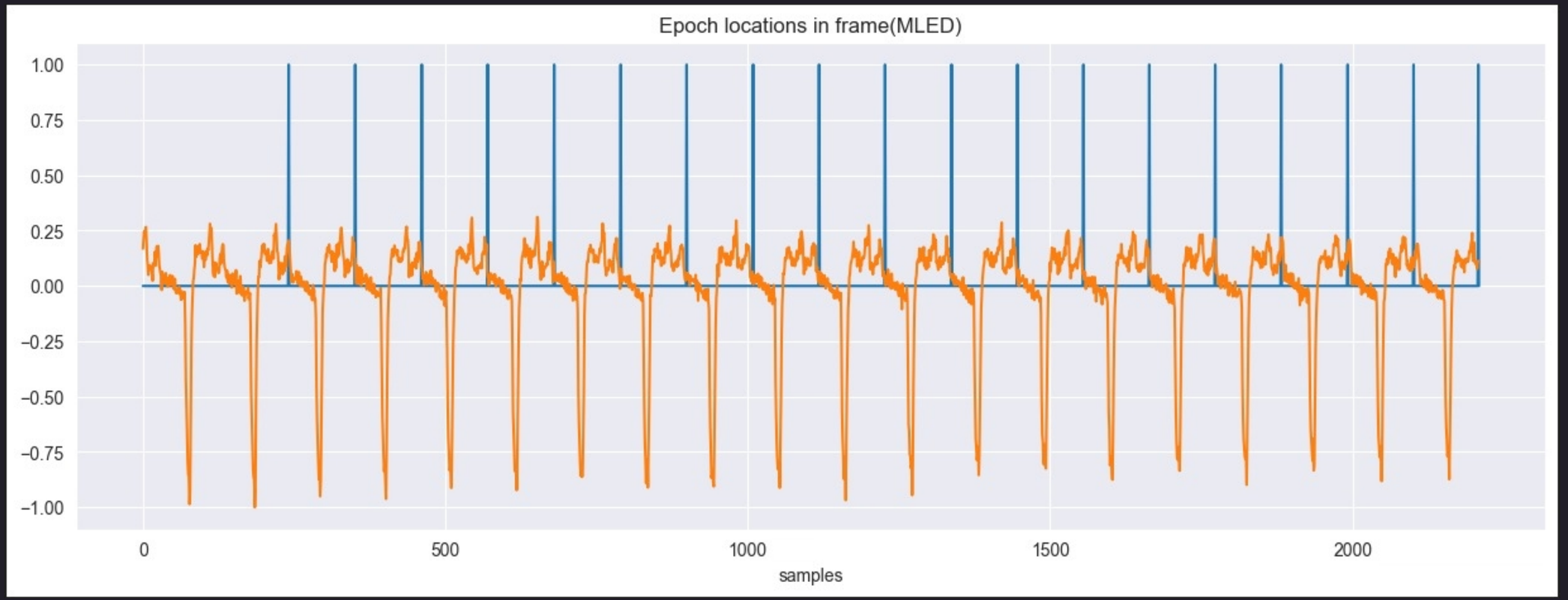
2] MLED

Reference Paper:
Yan Ming Cheng, member, IEE, and Douglas O'Shaugnessy, member, IEEE
"Automatic and Reliable Estimation of Glottal Closure Instant and Period"

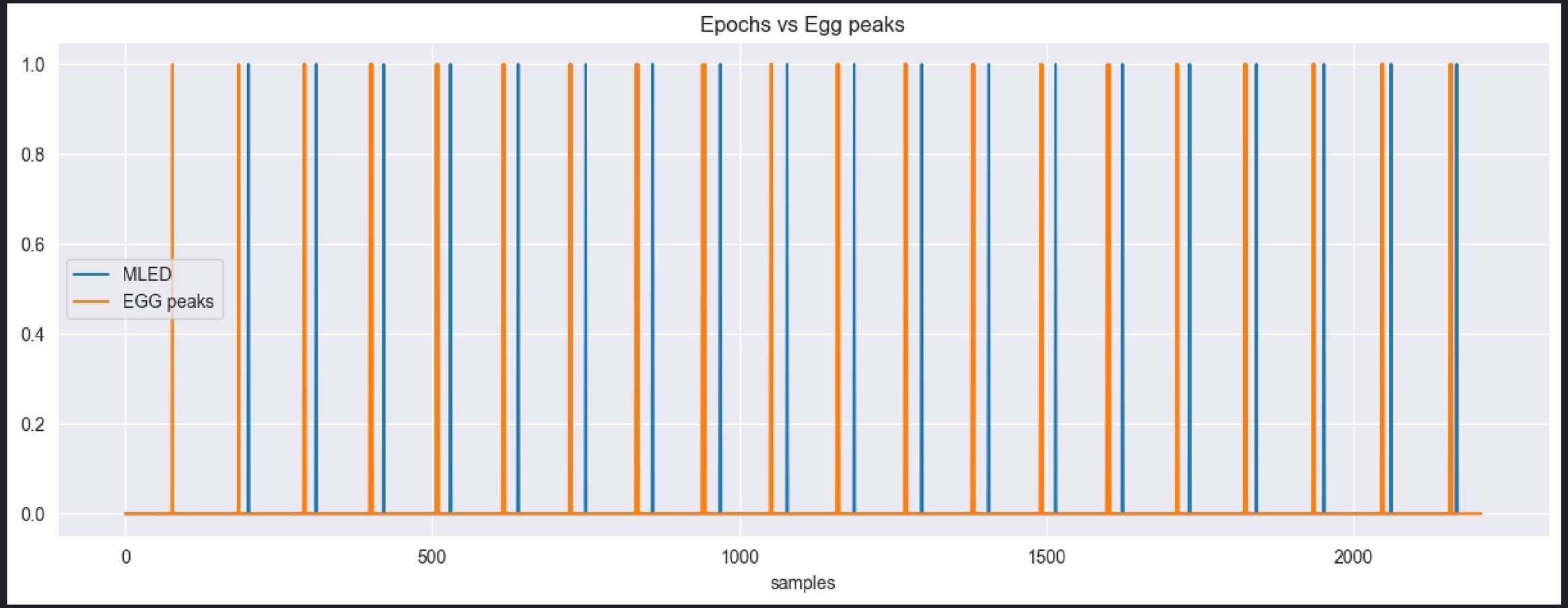
Steps:

- Compute LPC Coefficients to make LP filter.
- We correlated the predicted signal with the actual signal.
  This is the method of using the partial maximum-likelihood function.


Epoch locations in frame(MLED)
MLED correlation
Epochs
1.00
0.75
0.50
0.25
0.00
−0.25
−0.50
−0.75
0
500
1000
1500
2000
samples


Epoch locations in frame(MLED)
1.00
0.75
0.50
0.25
0.00
−0.25
−0.50
−0.75
−1.00
0
500
1000
1500
2000
samples

Epoch locations and egg peaks :



3] ZFF

Reference Paper:
K. Sri Rama Murty, B. Yegnanarayana, Senior Member, IEEE, and M. Anand Joseph
"**Epoch Extraction From Speech Signals**"

Steps:

- 1. Difference the speech signal s[n] (to remove any time varying low frequency bias in the signal).

  $x[n] = s[n] - s[n-1]$

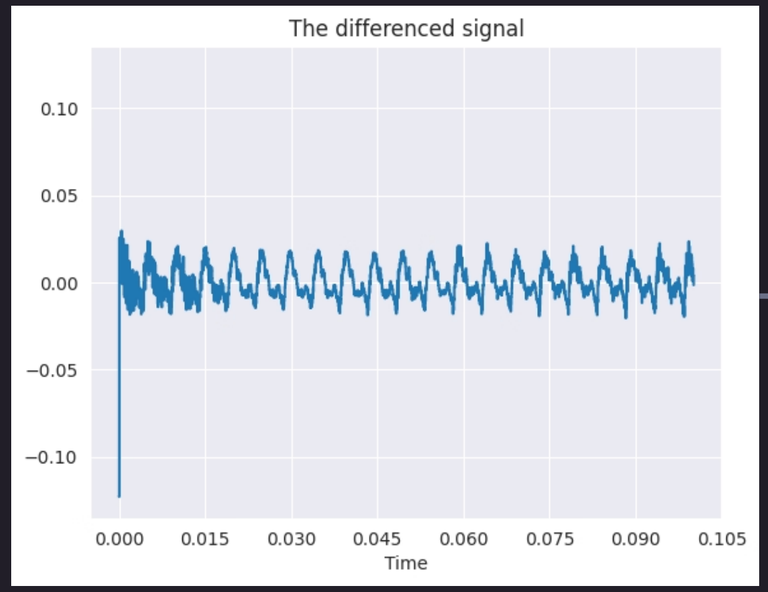
- 2. Pass the differenced speech signal twice through an ideal resonator at zero frequency.

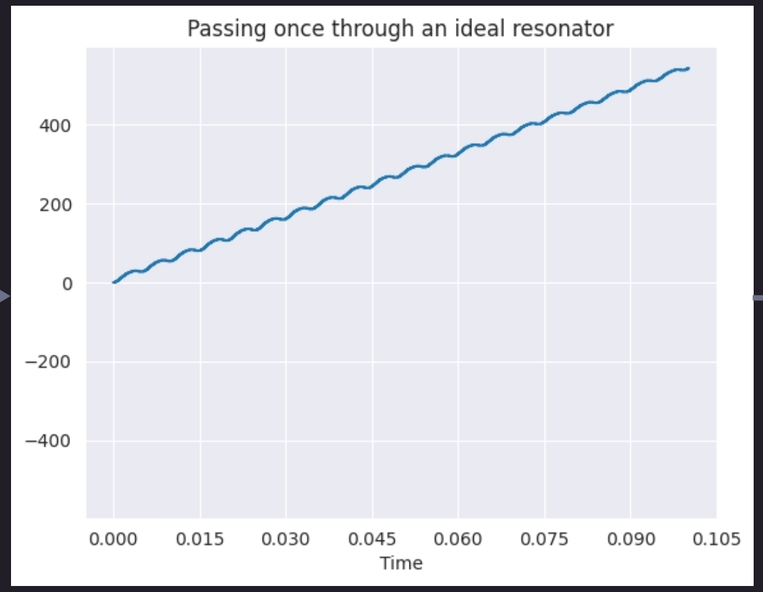
  $y1[n] = -(\Sigma[ak\ y1[n-k]\ \text{for}\ k\ \text{in}\ [1, 2]]) + x[n]$

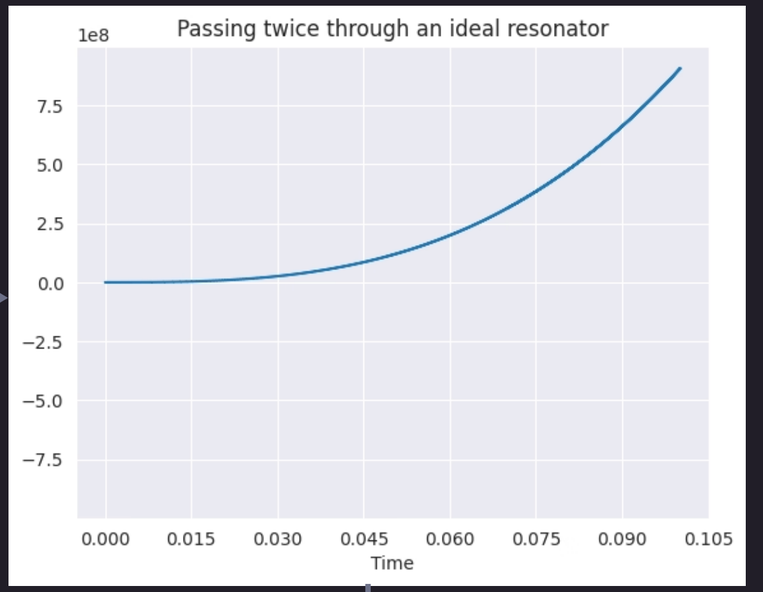
  $y2[n] = -(\Sigma[ak\ y2[n-k]\ \text{for}\ k\ \text{in}\ [1, 2]]) + y1[n]$

- 3. Remove the trend in y2[n] by subtracting the average over 10 ms at each sample. The resulting signal is called the zero-frequency filtered signal, or simply the filtered signal.

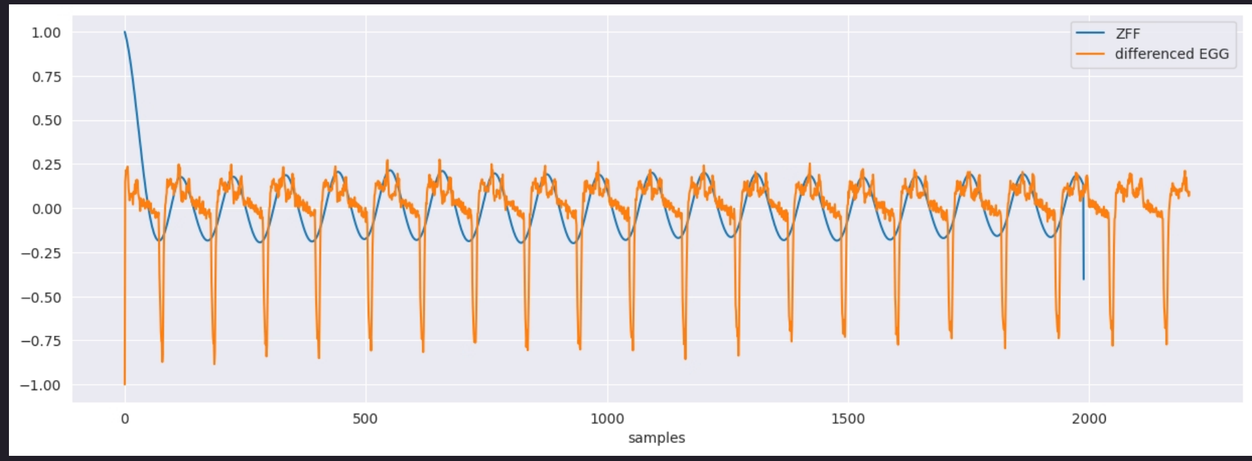
  $y[n] = y2[n] - (1 / (2*N + 1)) * \Sigma[\ y2[n+m]\ \text{for}\ m\ \text{in}\ [-N, N]\ ]$



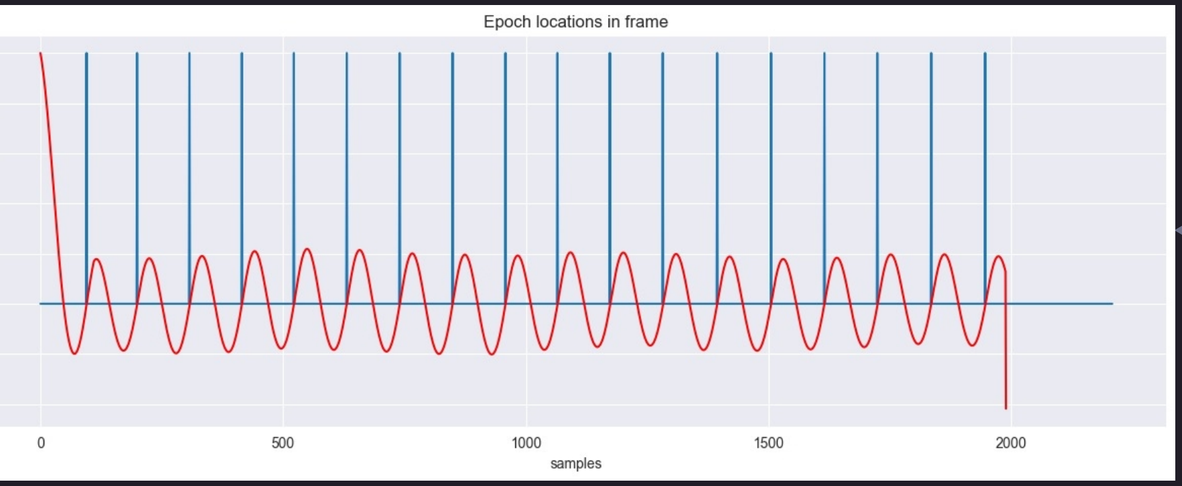




After Twice Mean Subtraction

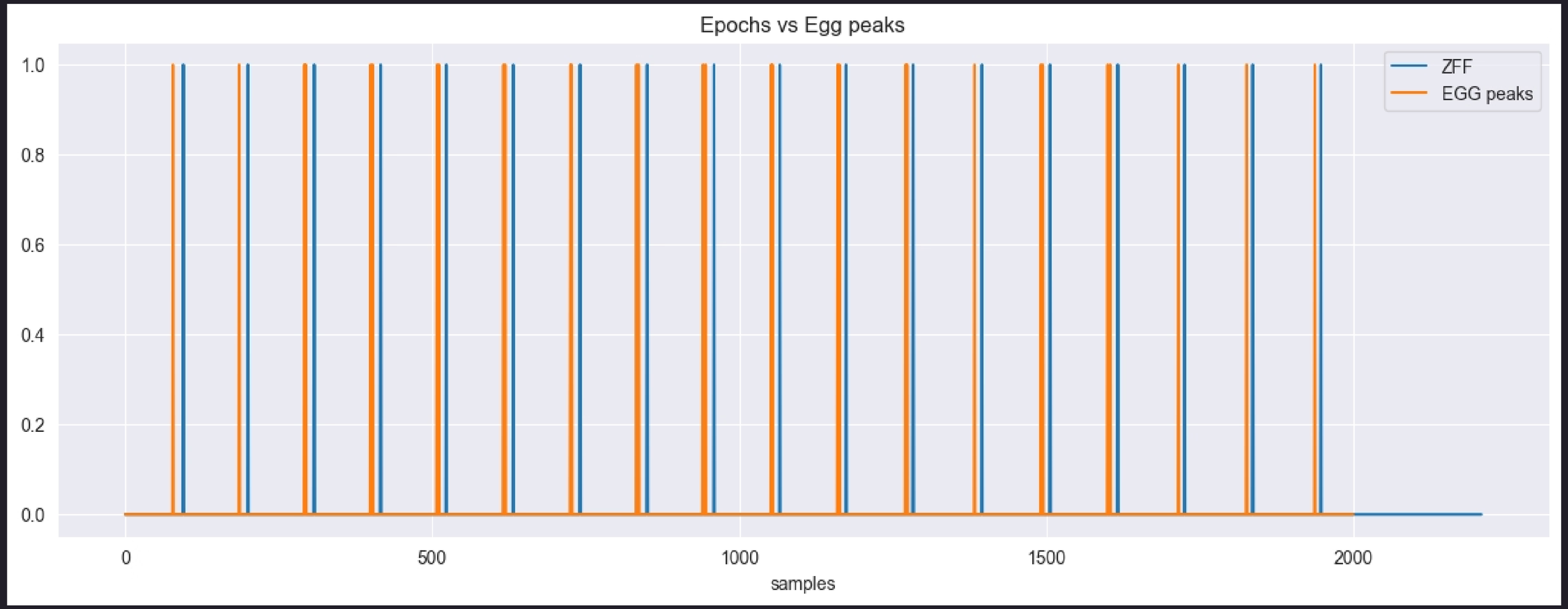


Epochs Extracted



Try Pitch

Epoch locations and egg peaks :



4] ZP-ZFR

Reference Paper:
Krishna Gurugubelli and Anil Kumar Vuppala , Member, IEEE
"Stable Implementation of Zero Frequency Filtering of Speech Signals for Efficient Epoch Extraction"

The ZFR system has a non-linear phase, which can mess up the detection of glottal closure instances. This is why we look to use an algorithm that has zero phase and causes zero phase distortion.

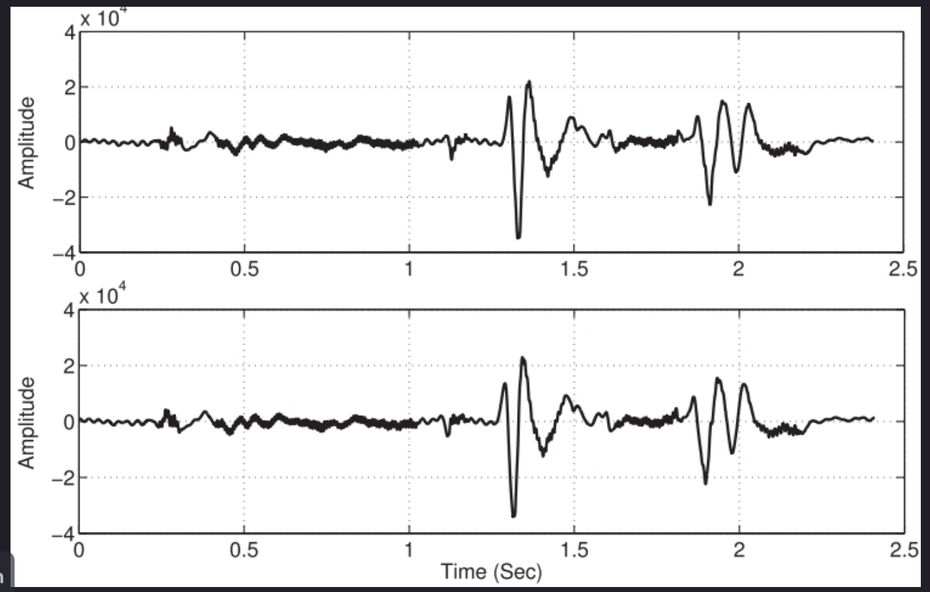
Steps:

- Pass through difference filter : x[n] = s[n] - s [n-1] for pre-emphasis.
- We then pass pre-emphasized speech signal to zero-phase zero frequency resonator block
- We perform trend removal.
- The negative peaks are identified and these correspond to the epoch locations in the speech signal.

The Zero Phase Zero Frequency Resonator :

$$H_{ZP-ZFR}(z) = H(z)H(z^{-1})$$
$$= \frac{1}{(1-rz^{-1})^2}\,\frac{1}{(1-rz)^2}$$
$$= \frac{-z^{-2}}{(1-rz^{-1})^2(r-z^{-1})^2}.$$

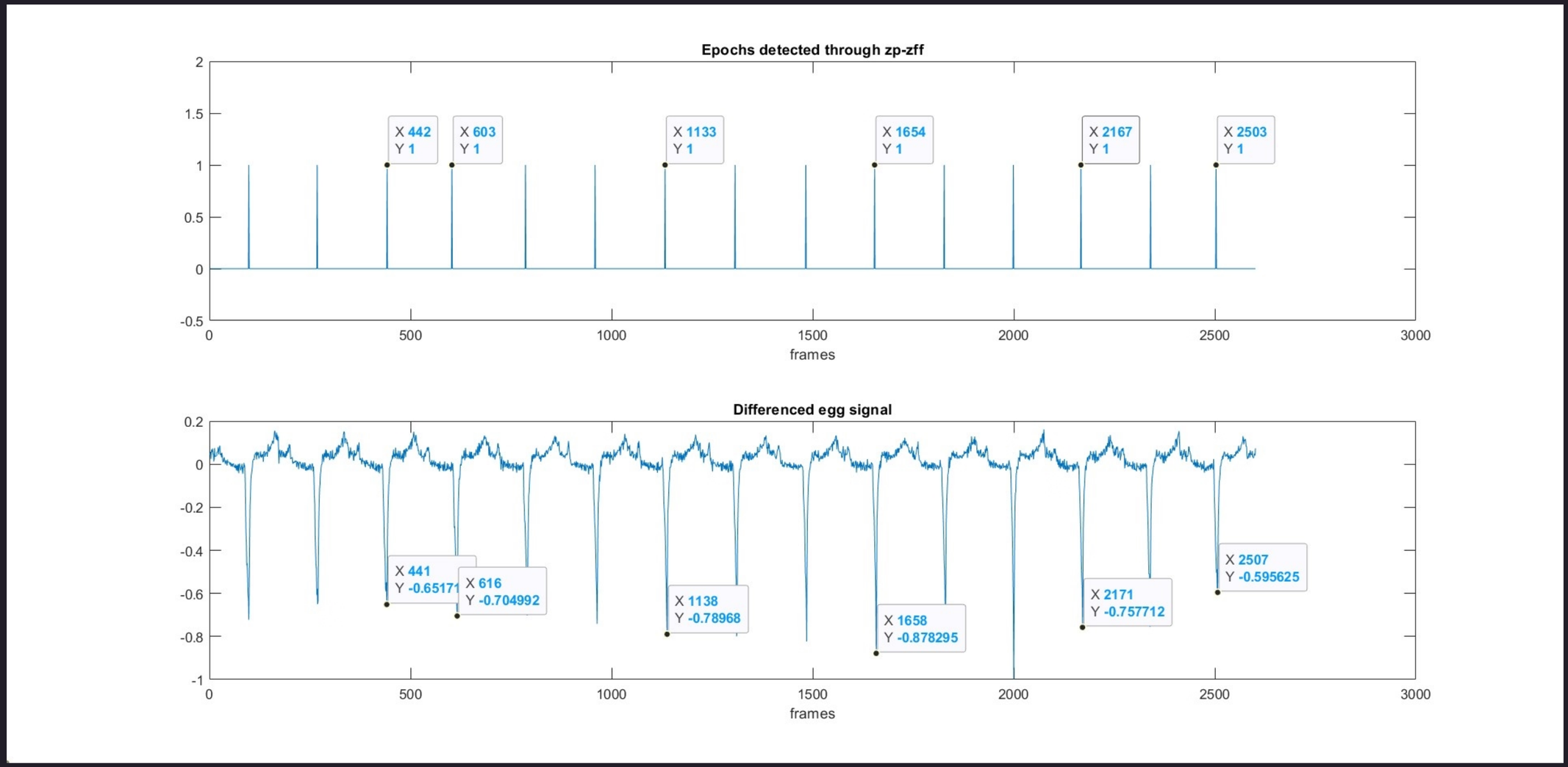
- r is a parameter that determines the bandwidth of the filter.
- The choice of r in between 0.9 to 1 gives us the closest approximation to the frequency response of a ZFR with zero phase.



frequency response of ZFR

frequency response of ZP ZFR at r = 0.98

Epochs obtained :



5] DYPSA

Reference Paper:

Patrick A. Naylor, Anastasis Kounoudes, Jon Gudnason, Member, and Mike Brookes

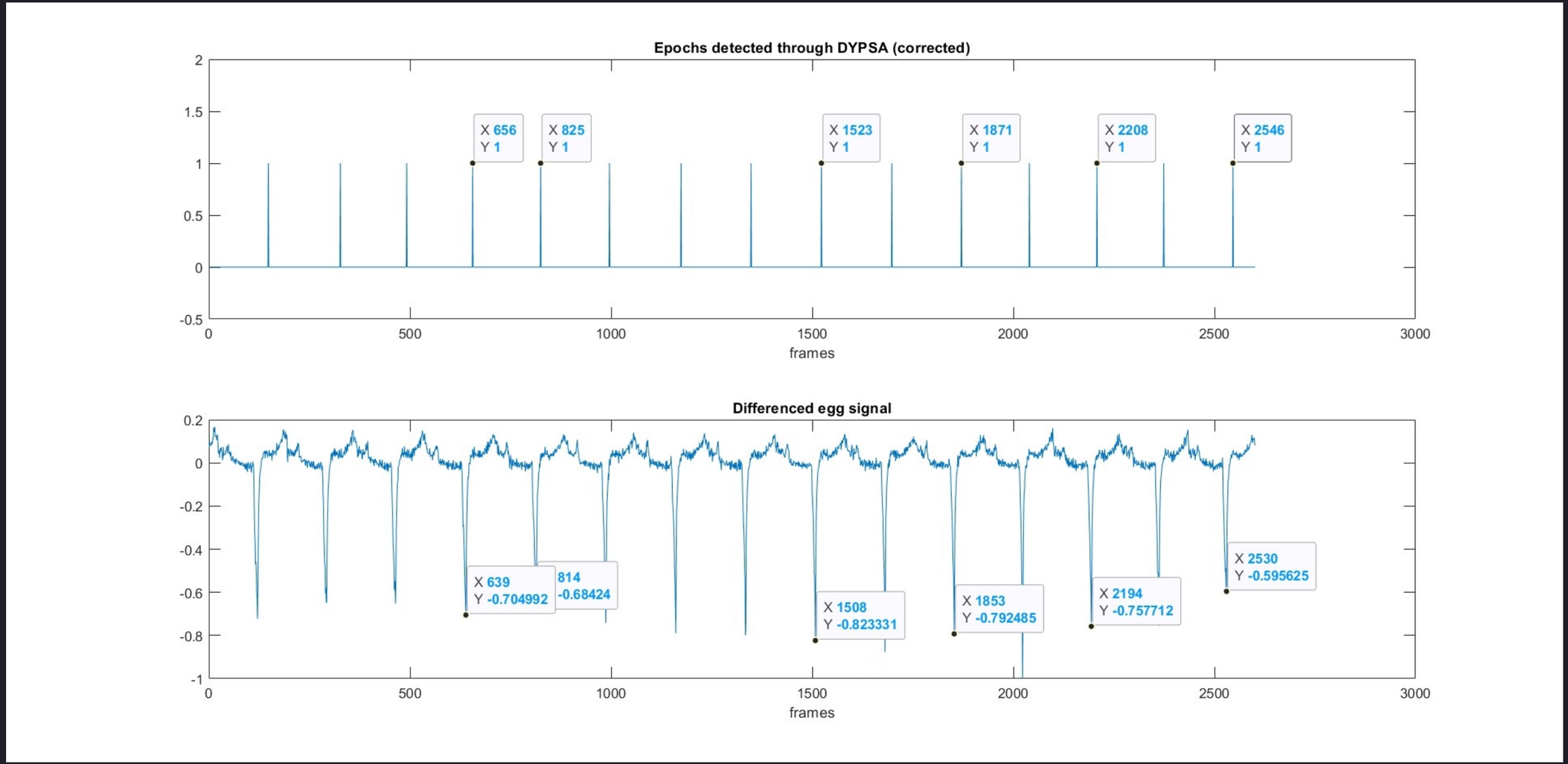
"Estimation of Glottal Closure Instants in Voiced Speech Using the DYPSA Algorithm"

Steps:

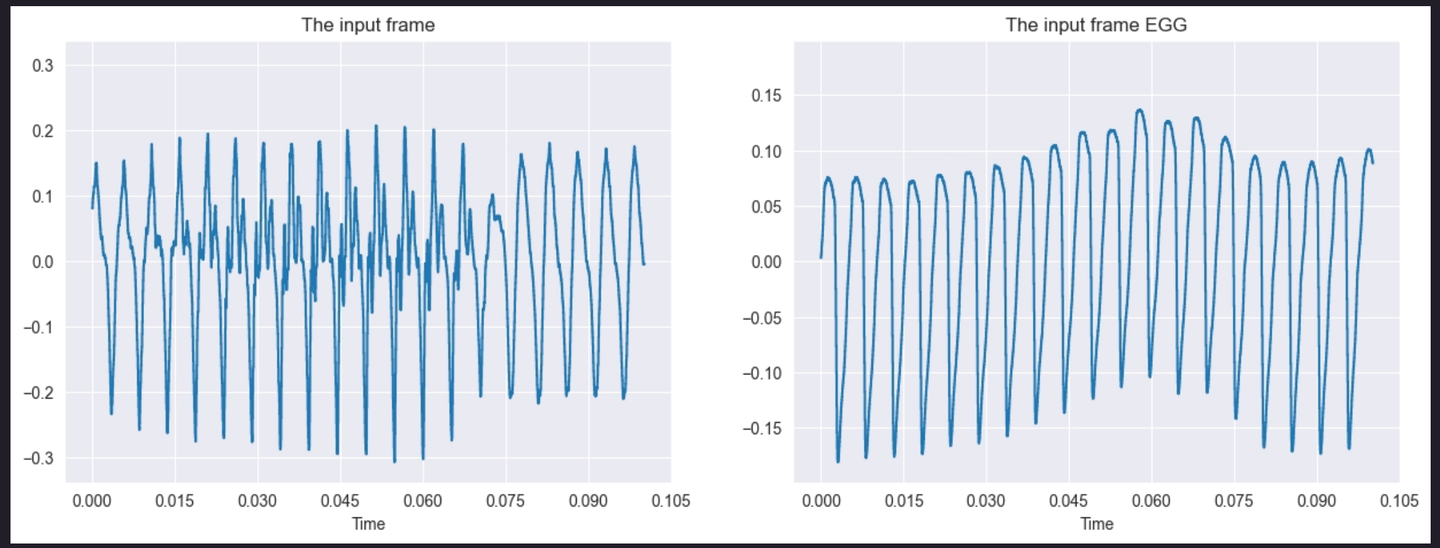
$$\tilde{\tau}_n(\omega) = \frac{d \arg\left(\tilde{X}_n(\omega)\right)}{d\omega}$$

- Generate candidate GCIs using zero crossings of the phase-slope function.
- Employ a phase-slope projection technique to recover candidates for which the phase-slope function does not include a zero crossing. (Here, we get almost all true GCI + no. of false GCI
- We then use DP (Dynamic Prog.) to identify the true GCIs from the set of candidates by minimizing a cost function.

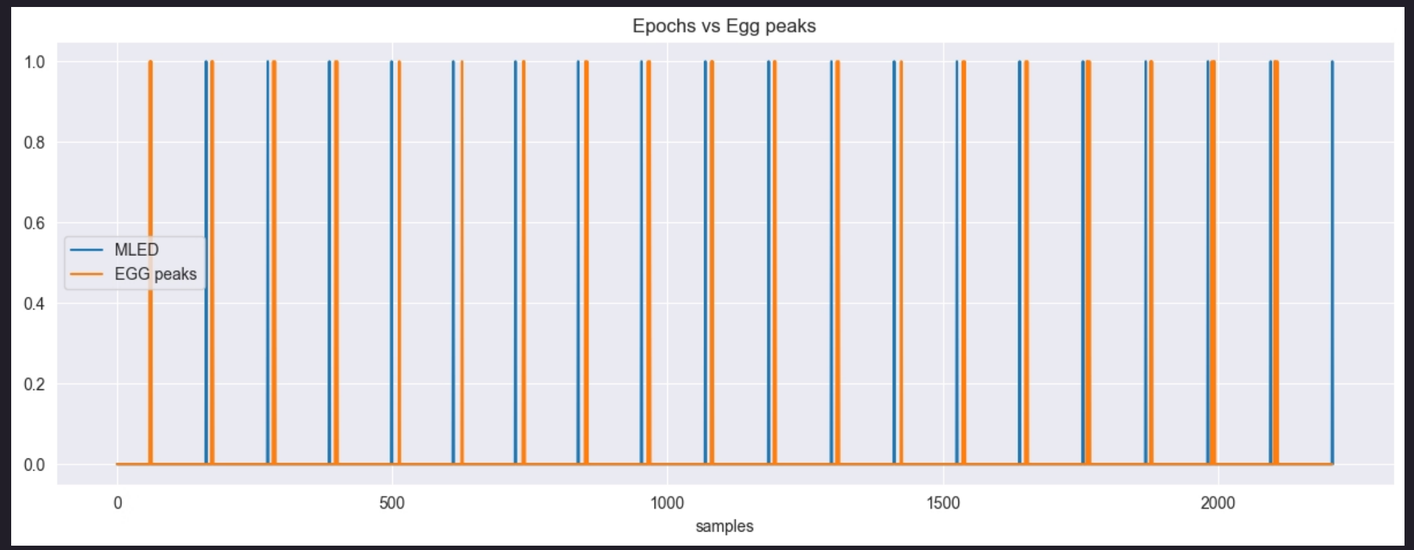
Epochs obtained :



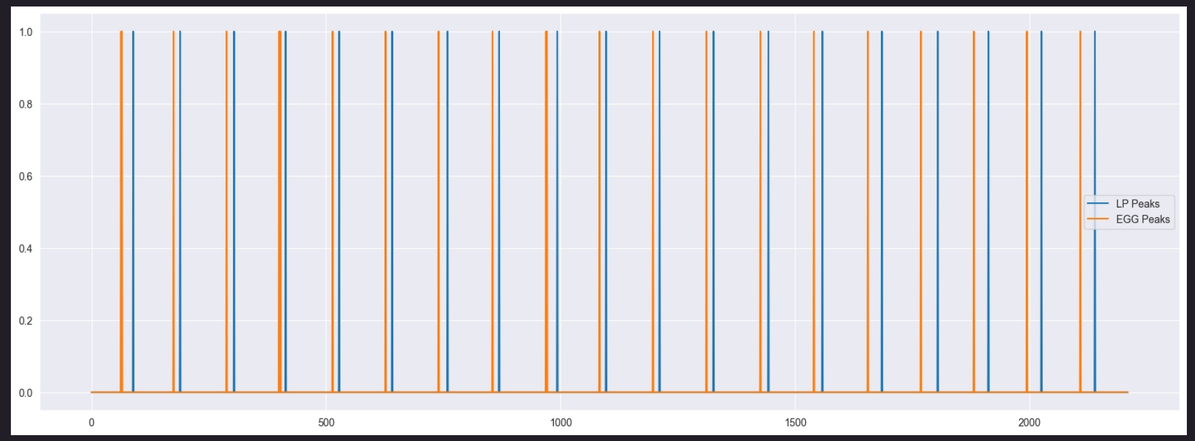
Epoch extraction on different voiced frames [0.85s to 0.95s] :



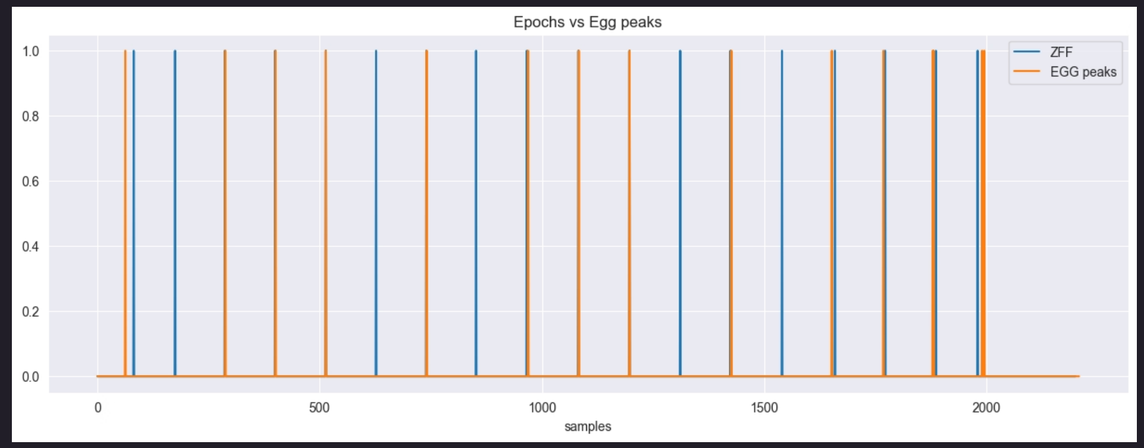
1) MLED

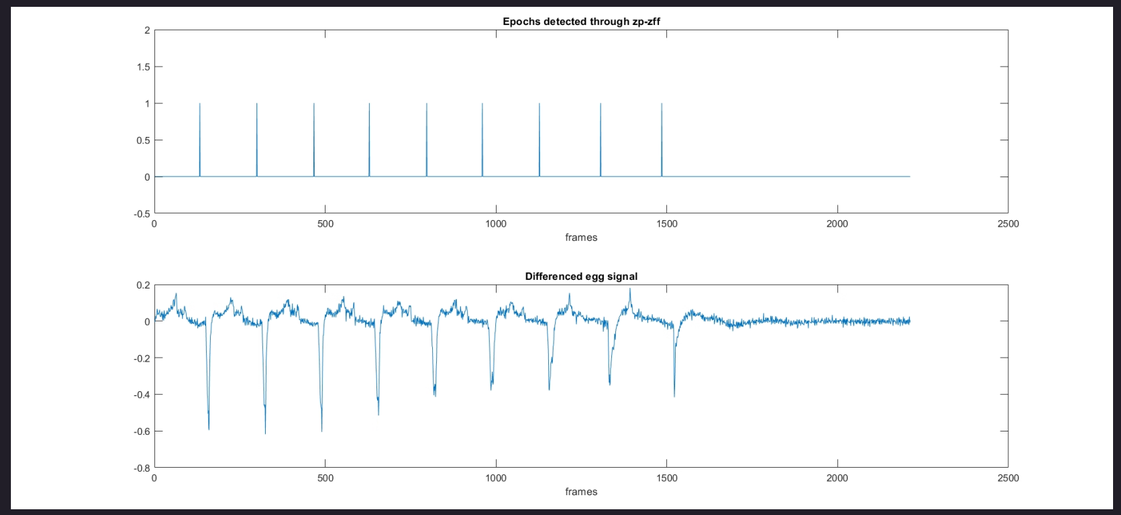


2) LP ERROR

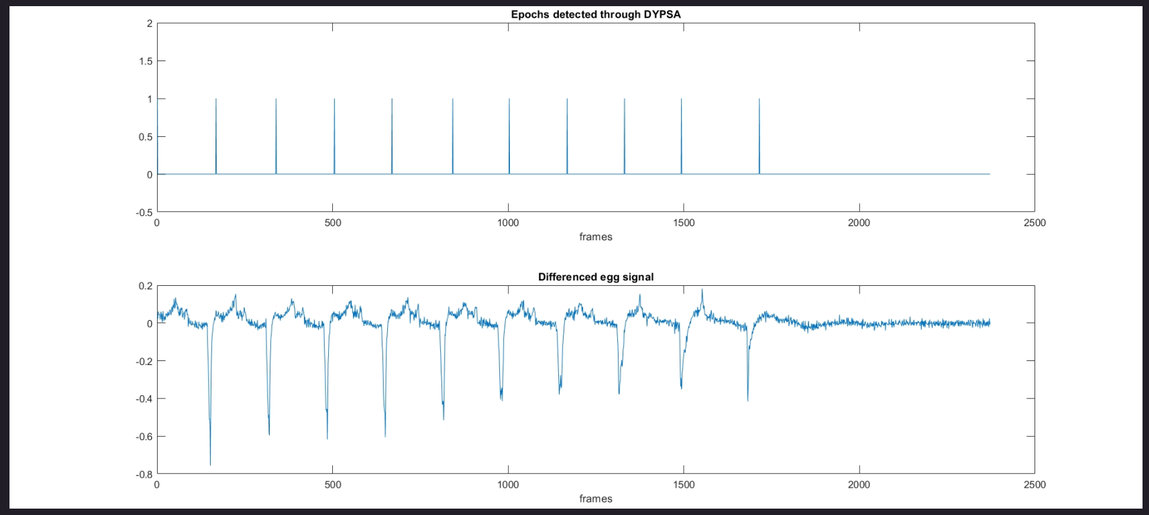


3) ZFF



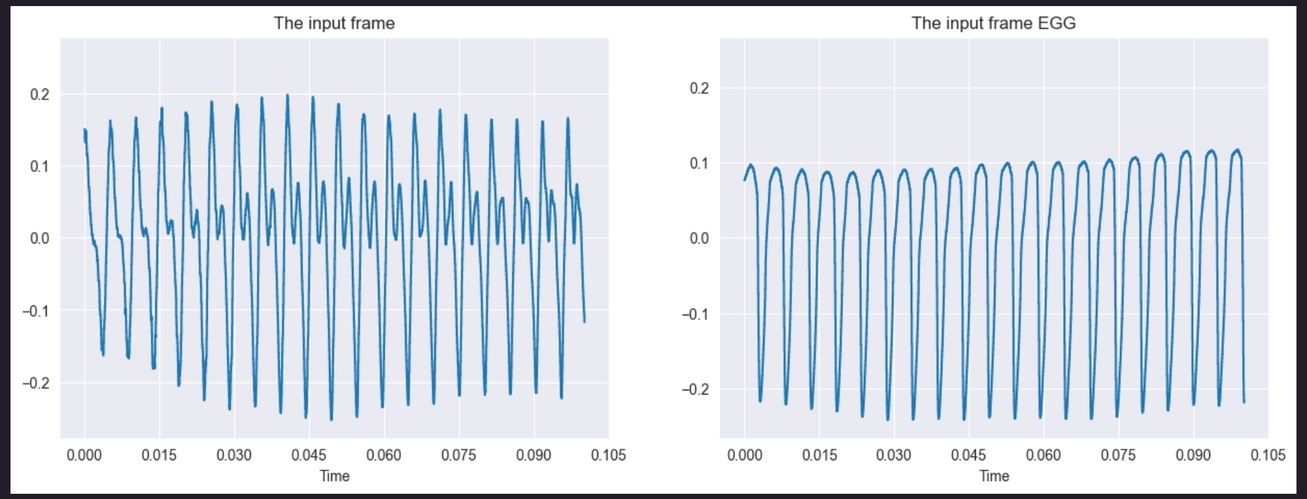


4) ZP-ZFR

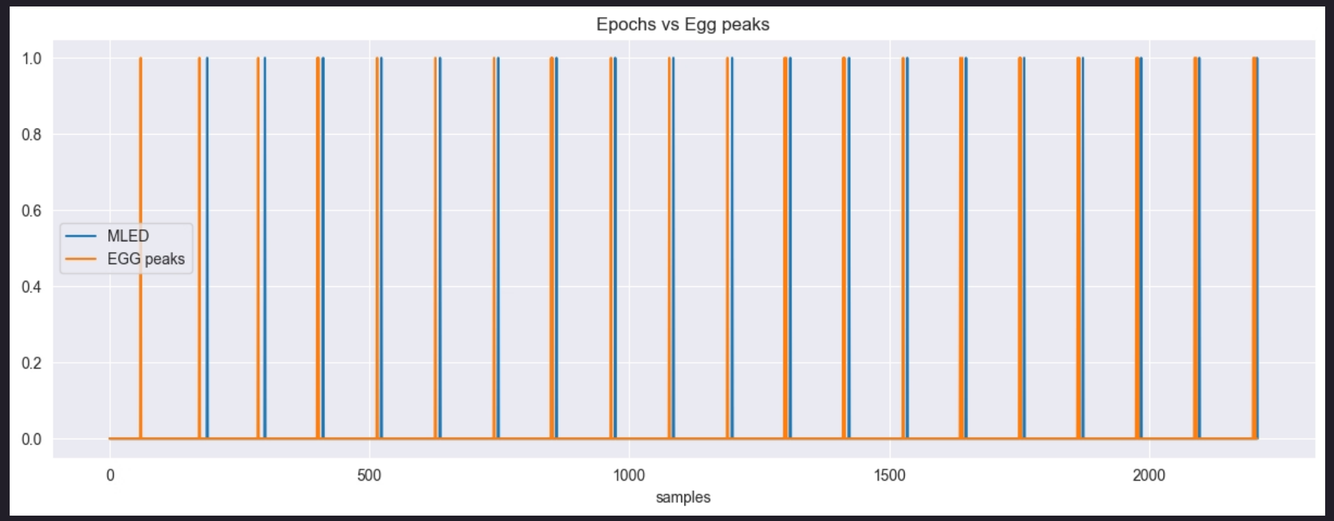


5) DYPSA

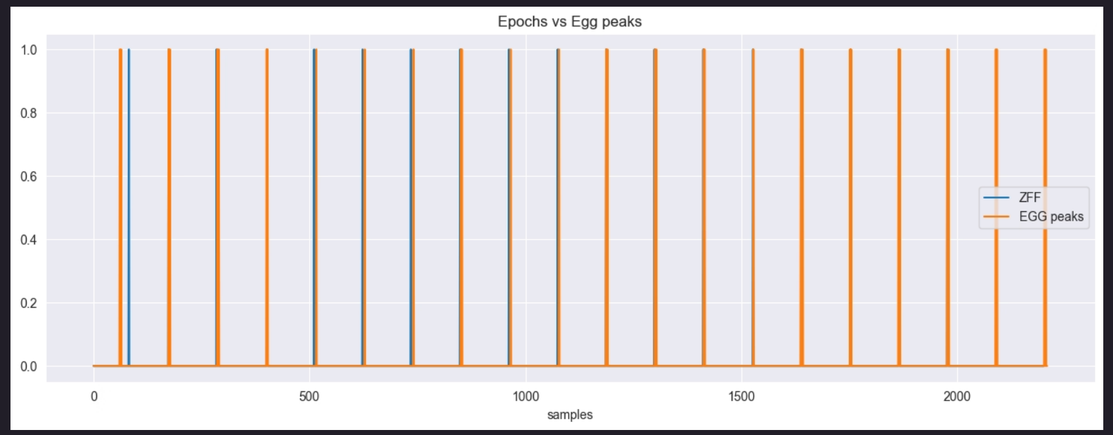
Epoch extraction on different voiced frames [0.25s to 0.35s] :



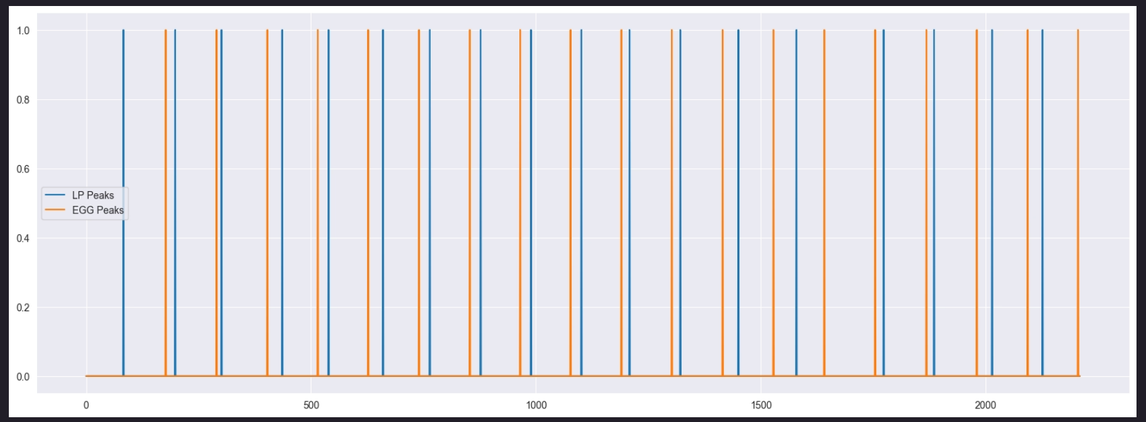
1) MLED

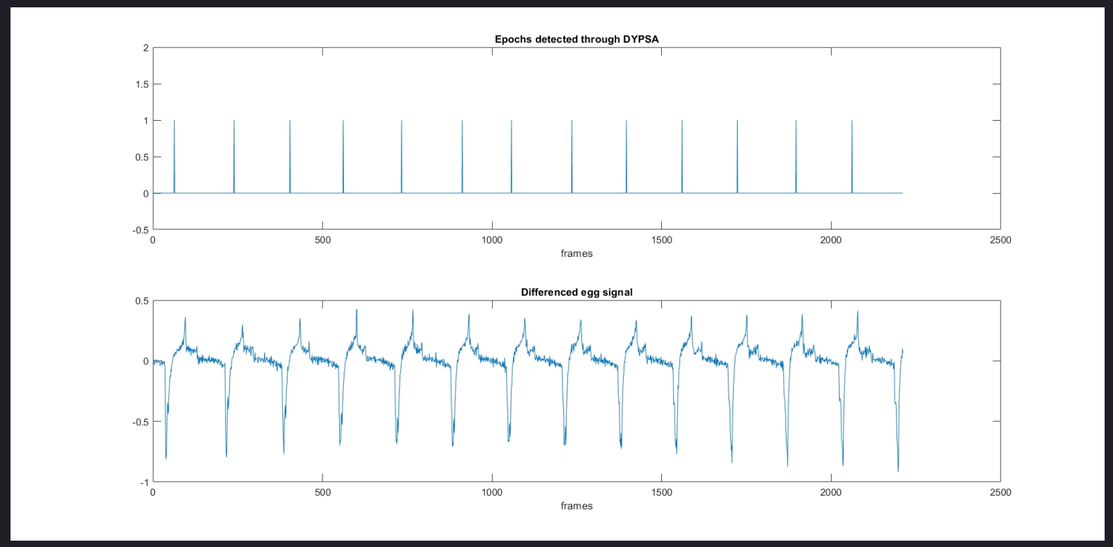


2) LP ERROR

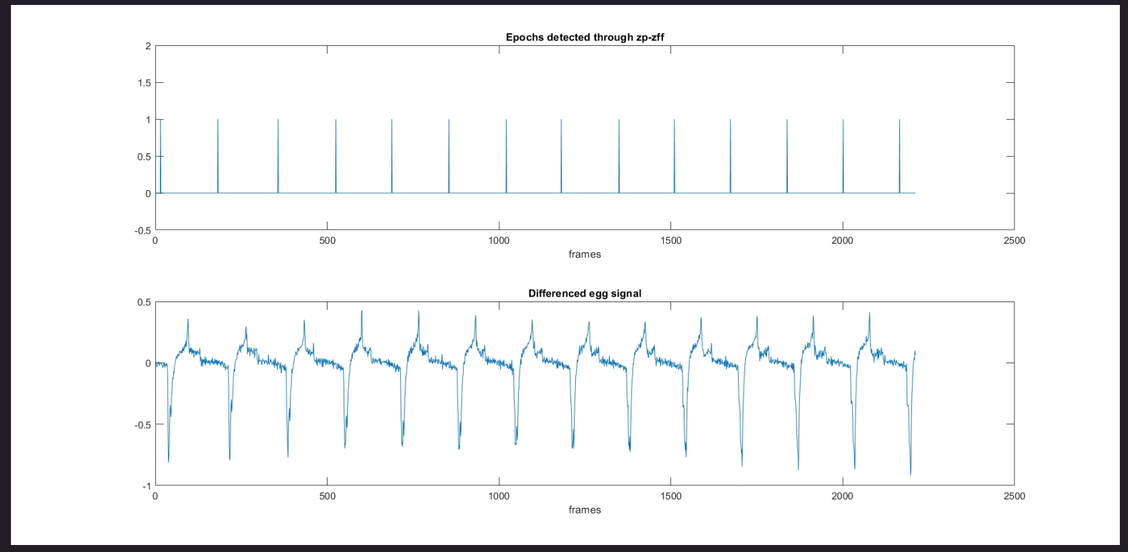


3) ZFF





4) ZP-ZFR



5) DYPSA

METRICS :

Reliability Measures:

1. Identification rate (IDR)

2. Miss rate (MR)

3. False alarm rate (FAR)

Accuracy Measures:

1. Identification accuracy (IDA)

2. Identification rate within ± 0.25 ms (IDR within ± 0.25 ms)

Reliability Measures:

EGG - Original
Test - Our test case where we derive epochs

1. Identification rate (IDR)

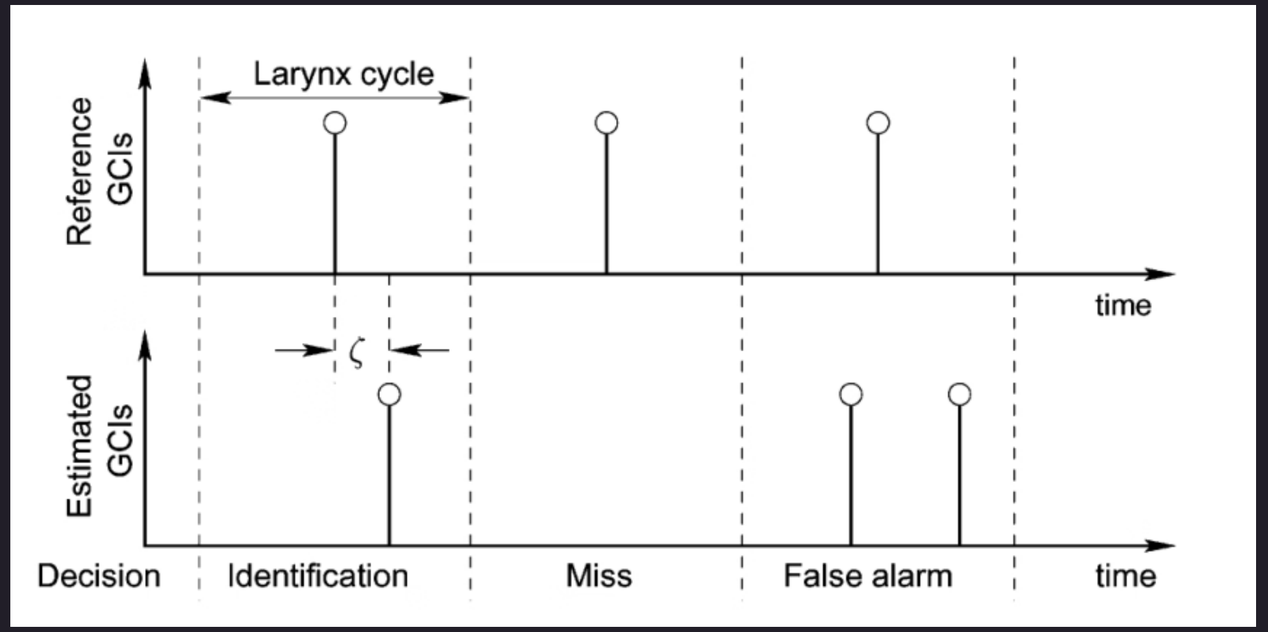
Formula = Number of Epochs in test / Number of epochs in EGG

2. Miss rate (MR)

Formula = Number of misses by test / Number of epochs in EGG

3. False alarm rate (FAR)

Formula = Number of false alarms in test / Number of epochs in EGG



Accuracy Measures:

1. Identification accuracy (IDA)

Formula = Distance between closest epochs in test and epochs in EGG in ms

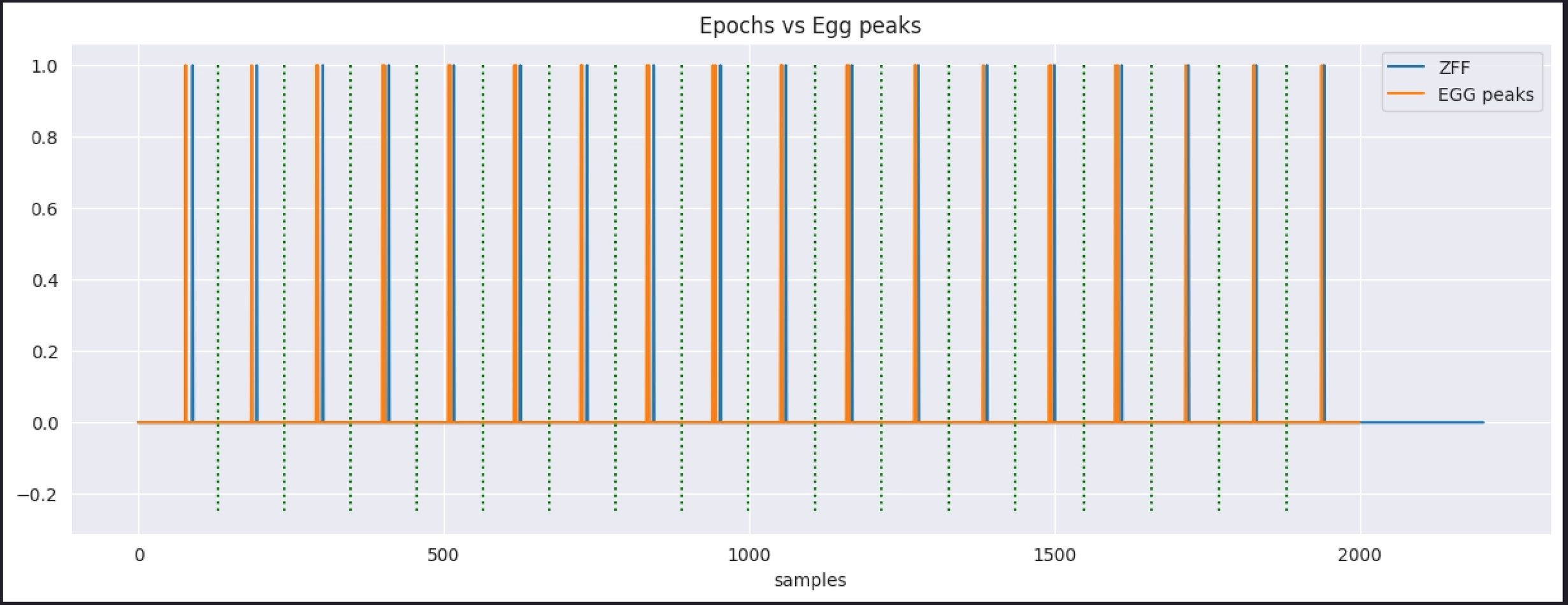
2. Identification rate within ± 0.25 ms (IDR within ± 0.25 ms)

Formula = Number of epochs in test / number of epochs in EGG

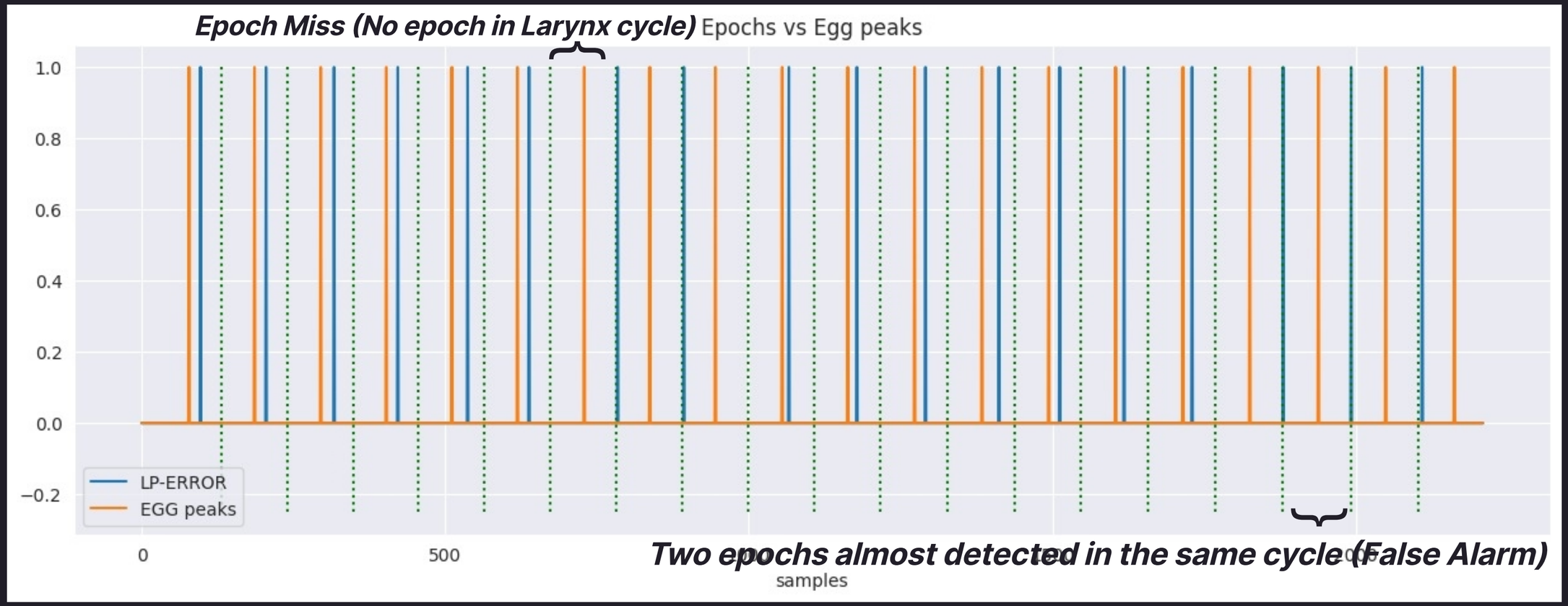
***The epochs identified in test must be within 0.25 ms of actual EGG epoch**

Larynx cycles plotted :

- GREEN Lines depict Larynx cycles



Examples of False Alarm and Miss :



Epoch Miss (No epoch in Larynx cycle)

Two epochs almost detected in the same cycle (False Alarm)

Identification accuracy measured in our frame under observation :

ZFF

Epoch	Distance to EGG
Epoch 1	0.45 ms
Epoch 2	0.32 ms
Epoch 3	0.41 ms
Epoch 4	0.41 ms
Epoch 5	0.36 ms
Epoch 6	0.41 ms
Epoch 7	0.41 ms
Epoch 8	0.45 ms
Epoch 9	0.32 ms
Epoch 10	0.32 ms
Epoch 11	0.32 ms
Epoch 12	0.23 ms
Epoch 13	0.36 ms
Epoch 14	0.36 ms
Epoch 15	0.41 ms
Epoch 16	0.23 ms
Epoch 17	0.23 ms

0.352140856342537 ms

ZP-ZFR

Epoch	Distance to EGG
Epoch 1	0.50 ms
Epoch 2	0.36 ms
Epoch 3	0.45 ms
Epoch 4	0.45 ms
Epoch 5	0.41 ms
Epoch 6	0.45 ms
Epoch 7	0.45 ms
Epoch 8	0.50 ms
Epoch 9	0.36 ms
Epoch 10	0.36 ms
Epoch 11	0.36 ms
Epoch 12	0.27 ms
Epoch 13	0.41 ms
Epoch 14	0.41 ms
Epoch 15	0.45 ms

0.414210128495842 77 ms

DYPSA

Epoch	Distance to EGG
Epoch 1	0.63 ms
Epoch 2	0.50 ms
Epoch 3	0.59 ms
Epoch 4	0.59 ms
Epoch 5	0.54 ms
Epoch 6	0.59 ms
Epoch 7	0.59 ms
Epoch 8	0.63 ms
Epoch 9	0.50 ms
Epoch 10	0.50 ms
Epoch 11	0.50 ms
Epoch 12	0.41 ms
Epoch 13	0.54 ms
Epoch 14	0.54 ms
Epoch 15	0.59 ms

0.5502645502645502 ms

MLED

Epoch	Distance to EGG
Epoch 1	0.77 ms
Epoch 2	0.63 ms
Epoch 3	0.73 ms
Epoch 4	0.73 ms
Epoch 5	0.68 ms
Epoch 6	0.73 ms
Epoch 7	0.73 ms
Epoch 8	0.77 ms
Epoch 9	0.63 ms
Epoch 10	0.63 ms
Epoch 11	0.63 ms
Epoch 12	0.54 ms
Epoch 13	0.68 ms
Epoch 14	0.68 ms
Epoch 15	0.73 ms
Epoch 16	0.54 ms
Epoch 17	0.54 ms

0.66960117380285 46 ms

LP ERROR

Epoch	Distance to EGG
Epoch 1	1.02 ms
Epoch 2	0.88 ms
Epoch 3	0.98 ms
Epoch 4	0.98 ms
Epoch 5	0.93 ms
Epoch 6	0.98 ms
Epoch 7	0.98 ms
Epoch 8	1.02 ms
Epoch 9	0.88 ms
Epoch 10	0.88 ms
Epoch 11	0.88 ms
Epoch 12	0.79 ms
Epoch 13	0.93 ms
Epoch 14	0.93 ms
Epoch 15	0.98 ms
Epoch 16	0.79 ms
Epoch 17	0.79 ms

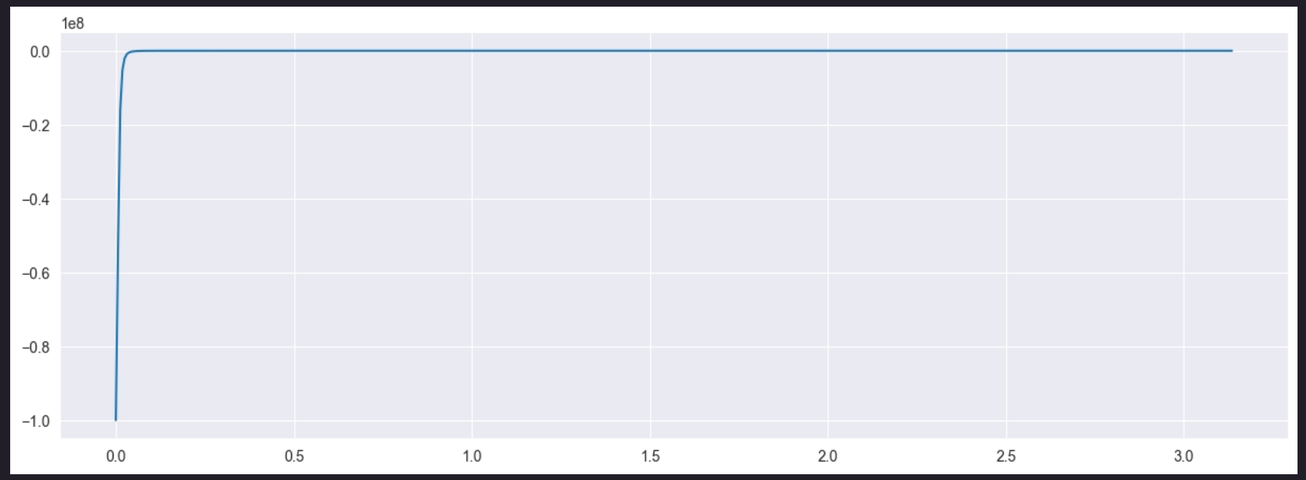
0.91960117380285 46 ms

		IDR (%)	MR (%)	FAR (%)	AVG IDA (in ms)
1	LP-ERROR	87.71	13.23	7.3529	0.9196011738028546
2	MLED	90.23	10.29	4.4117	0.6696011738028546
3	ZFF	95.8	5.88	1.4705	0.352140856342537
4	ZP-ZFR	96.7	2.94	-	0.4142101284958427 7
5	DYPSA	92.1	2.94	-	0.5502645502645502

Above values have been calculated and averaged over few voiced frames.

Challenges Faced:

- Trying to comprehend MLED mathematically.
- Trying to recreate the zero - phase zero frequency resonator in python. Recreating the filter gave us the following frequency response.



- As we can see this is not the expected frequency response. therefore, we have resorted to using a matlab function for ZP-ZFF.

Pitch

Want to make a presentation like this one?

Start with a fully customizable template, create a beautiful deck in minutes, then easily share it with anyone.

Create a presentation (It's free)