POS → (HR)

Positional Orthogonal - to - Skin

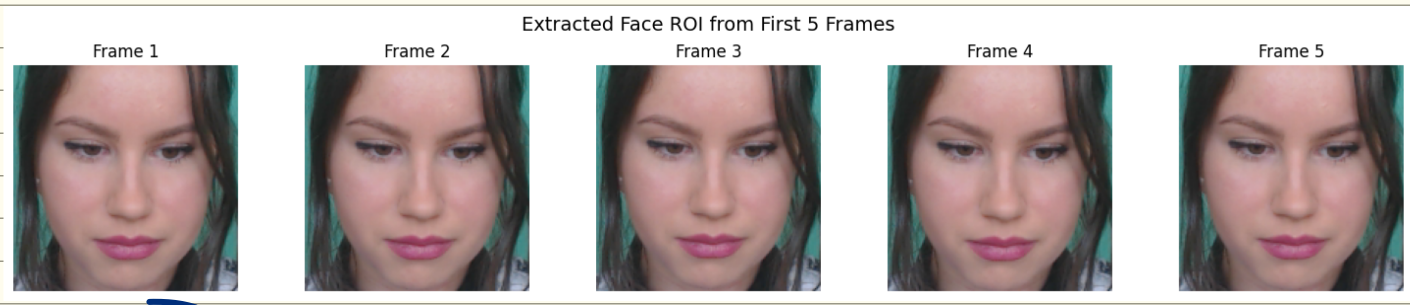
Description : POS transforms raw color into a clean rPPG pulse signal by projecting them onto an optimal plane that enhences pulsation and suppresses noise.

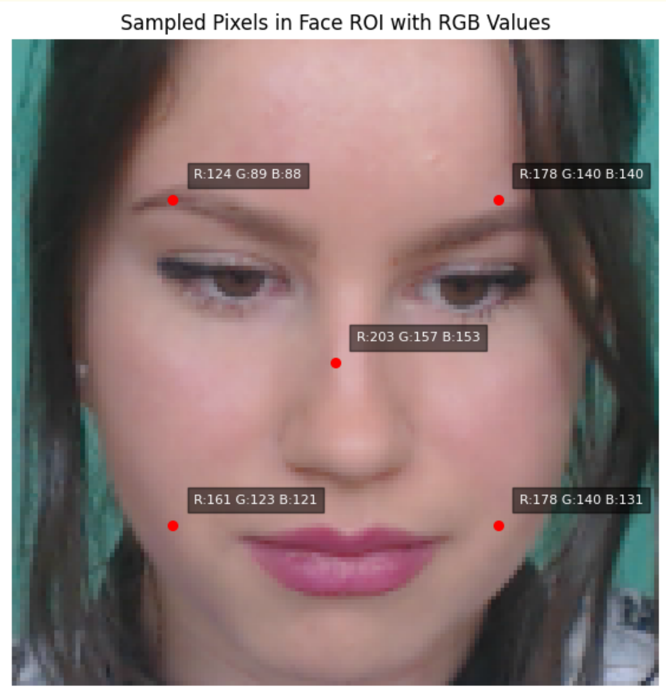
Input
- → Video Frames
- → ROI → Face Skin Area
- → RGB Signals → Avg. color values from ROI

Output
- → rPPG Signal
- → HR (FFT)

# Step- Zero

Assume we have 5 frames : (using haarcascade)



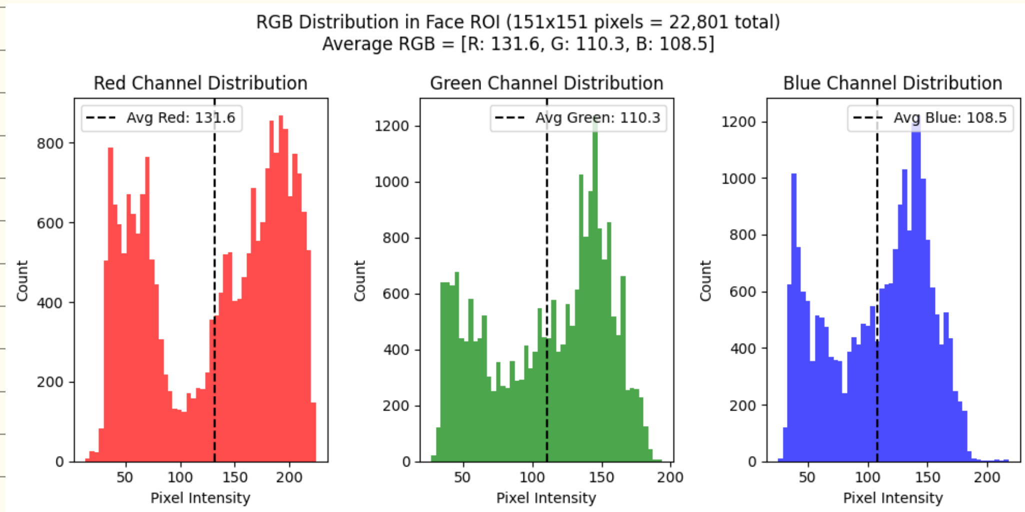




For A Single Fram:

Detected face ROI Size : 151 × 151

Total Pixels : 22801

Each Pixels has R, G, B values



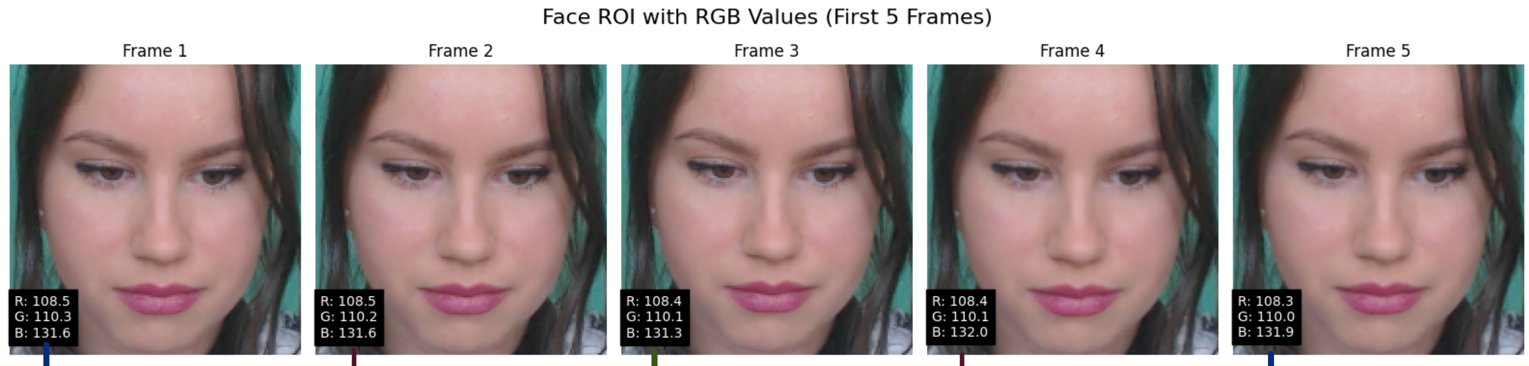

$$S(t) = \left[ \frac{1}{N} \sum_{p=1}^{N} R_p \, , \; \frac{1}{N} \sum_{p=1}^{N} G_p \, , \; \frac{1}{N} \sum_{p=1}^{N} B_p \right]$$

Here,

$N = 22801$

color (R, G, B)

For each fram we will have a 3D vector.




Avg RGB For each Fram

5 R, G, B vectors for 5 frams:

From-1 $\longrightarrow$ [array([ 108.46243586, 110.29025043, 131.5733959 ]),

From-2 $\longrightarrow$ array([ 108.46936538, 110.21608701, 131.55383536]),

From-3 $\longrightarrow$ array([ 108.38814087, 110.05201526, 131.32660848]),

From-4 $\longrightarrow$ array([ 108.36408889, 110.12666667, 132.03613333]),

From-5 $\longrightarrow$ array([ 108.25617778, 109.9708 , 131.94168889])]

mean

mean_R_G_B : [108.38804175 110.13116387 131.68633239]

$\mu_S : [\mu_R, \mu_G, \mu_B]$

# Step- One

## Normalization

$$\bar{S}(t) = \frac{S(t)}{\mu_s} - 1$$

**FRAM-1**

$R, G, B \rightarrow S(t)$ :

array([ 108.46243586,    110.29025043,     131.5733959 ])

mean_R_G_B $\rightarrow \mu_s$ :

[108.38804175     110.13116387    131.68633239]

Normalized_R_G_B $\rightarrow \bar{S}(t)$ :

[ 6.86368182e-04.   1.44451895e-03     -8.57617396e-04]

$$\frac{108.46243586}{108.38804175} - 1 \qquad \frac{110.29025043}{110.13116387} - 1 \qquad \frac{131.5733959}{131.68633239} - 1$$

Same For

| | | | |
|---|---|---|---|
| Fram-2 | [ 7.50300702e-04 | 7.71109073e-04. | -1.00615632e-03] |
| Fram-3 | [ 9.14456236e-07 | -7.18675885e-04 | -2.73167234e-03] |
| Fram-4 | [-2.20991777e-04 | -4.08350123e-05 | 2.65631925e-03] |
| Fram-5 | [-1.21659156e-03 | -1.45611713e-03 | 1.93912679e-03] |

# Step-TWO

## POS Projection (Color Signal Decomposition)

Color Projection Matrix:

$$C = \begin{bmatrix} 0 & 1 & -1 \\ -2 & 1 & 1 \end{bmatrix}$$

→ Used to Separate color components in a way that:

- Emphasizes pulsatile (HR) color variation
- Minimize illumination or motion artificats.

Ref: Wang et al.; "Algorithmic Principles of remote PPG"
IEEE TBME, 2017

They mathematically and experimentally showed that this projection:

→ Improves signals-to-noise ratio for HR

→ Fillters out common-mode illumination

Applying this to the Normalized R G B signal $\bar{S}(t)$:

$$\begin{bmatrix} S_1(t) \\ S_2(t) \end{bmatrix} = C \cdot \bar{S}(t)^T$$

which gives:

$$S_1(t) = \bar{G}(t) - \bar{B}(t)$$

$$C = \begin{bmatrix} 0 & 1 & -1 \\ -2 & 1 & 1 \end{bmatrix}$$

$$S_2(t) = -2 \cdot \bar{R}(t) + \bar{G}(t) + \bar{B}(t)$$

These capture different orthogonal color changes on the skin:

= $S_1(t)$ : Green-Blue variation $(G - B)$

→ HR Signal is strongest in the green channel.

→ Blue changes similarly with lighting, so subtracting B helps reduce illumination effects.

thus:

$$S_1(t) = \bar{G}(t) - \bar{B}(t)$$

≡ $S_2(t)$ : A mix R, G, B designed to capture motion-invariant color changes.

$$(-2R + G + B)$$

→ Red reflects deeper tissue and is more affected by lighting and motion.

→ This formula balances the three channles to cancel motion artifacts:

$$S_2(t) = -2\bar{R}(t) + \bar{G}(t) + \bar{B}(t)$$

# Example Calculation:

## FRAM-1

Normalized R, G, B $= [0.000686,\ 0.001445,\ -0.000858]$

Let,

$$\bar{R} = 0.000686$$

$$\bar{G} = 0.001445$$

$$\bar{B} = -0.000858$$

$$S_1 = \bar{G} - \bar{B} = 0.001445 - (-0.000858)$$

$$= 0.002303$$

$$\bar{S_2} = -2\bar{R} + \bar{G} + \bar{B} = -2(0.000686) + 0.001445 + (-0.000858)$$

$$= -0.000785$$

Same AS:

| | Fram-1 | Fram-2 | Fram-3 | Fram-4 | Fram-5 |
|---|---|---|---|---|---|
| $S_1 \rightarrow$ | [ 0.00230214 | 0.00177727 | 0.002013 | -0.00269715 | -0.00339524] |
| $S_2 \rightarrow$ | [-0.00078583 | -0.00173565 | -0.00345218 | 0.00305747 | 0.00291619] |

# Step - Three

## Alpha Scaling and Final Signal

Formula :

Motion Strength Factor = $\alpha = \dfrac{\sigma(S_1)}{\sigma(S_2)}$

$\alpha > 1$
strong signal

$\alpha < 1$
More motion

$$h(t) = S_1(t) - \alpha \cdot S_2(t)$$

Why? → Each Fram Contains

→ Useful Pulsatile signal

→ Unwanted noise (motion artifacts eg. head movement, expressions)

Let,

$$S_1 = [\text{Pulse} + \text{Motion}]$$

$$S_2 = [\text{Mostly Motion}]$$

$\Big\}$ From Step 2

$$\alpha = \text{Strength Motion}$$

Assume,

$$\alpha = \frac{\sigma(S_1)}{\sigma(S_2)} = 1$$

Therefore,

$$h(t) = S_1(t) - \alpha S_2(t)$$

$$= [\text{Pulse} + \text{Motion}] - 1.[\text{Motion}]$$

$$= \text{Pulse}.$$

# Example Calculation:

| | Fram-1 | Fram-2 | Fram-3 | Fram-4 | Fram-5 |
|---|---|---|---|---|---|
| $S_1 \rightarrow$ | [ 0.00230214 | 0.00177727 | 0.002013 | -0.00269715 | -0.00339524] |
| $S_2 \rightarrow$ | [-0.00078583 | -0.00173565 | -0.00345218 | 0.00305747 | 0.00291619] |

$$\sigma(S_1) = 0{\cdot}00256$$

$$\sigma(S_2) = 0{\cdot}00257$$

$$\alpha = \frac{0{\cdot}00256}{0{\cdot}00257} = 0{\cdot}997$$

$$h = S1 - 0{\cdot}997 \cdot S_2$$

| Fram-1 | Fram-2 | Fram-3 | Fram-4 | Fram-5 |
|---|---|---|---|---|
| [ 0.00306303, | 0.00345782, | 0.00535559, | -0.00565757, | -0.00621886] |

# Step Four

## Bandpass Filter

* Only HR component : 0.7 – 4.0 Hz

Which represents 42–240 bpm

Bandpass filter will remove everything below 0.7 Hz and above 4.0 Hz.

$h(t)$:

| Fram-1 | Fram-2 | Fram-3 | Fram-4 | Fram-5 |
|---|---|---|---|---|
| [ 0.00306303, | 0.00345782, | 0.00535559, | -0.00565757, | -0.00621886] |
| $h[0]$ | $h[1]$ | $h[2]$ | $h[3]$ | $h[4]$ |

sampling rate $f_s = 30$ Hz

Each sample is $\frac{1}{30}$ second part.

For Example we are using 3 point filter karnel.

Let,

High-pass karnel : $[-1, 2, -1]$

Low-pass karnel : $[0.25, 0.5, 0.25]$

Padding Signal :

$$[\ \underline{h[1]}\ ,\ h[0]\ ,\ h[1]\ ,\ h[2]\ ,\ h[3]\ ,\ h[4]\ ,\ \underline{h[3]}\ ]$$

| | | | | | | |
|---|---|---|---|---|---|---|
| Fram-2 | Fram-1 | Fram-2 | Fram-3 | Fram-4 | Fram-5 | Fram-4 |
| [ 0.00345782, | 0.00306303, | 0.00345782, | 0.00535559, | -0.00565757, | -0.00621886, | -0.00565757] |
| h[1] | h[0] | h[1] | h[2] | h[3] | h[4] | h[3] |

# Steps-Four

$h(t)$:

| Fram-1 | Fram-2 | Fram-3 | Fram-4 | Fram-5 |
|---|---|---|---|---|
| [ 0.00306303, | 0.00345782, | 0.00535559, | -0.00565757, | -0.00621886] |

Assume,

$$f_s = 30\ Hz$$

$$N = 5 \text{ Frames}$$

$$T = \frac{1}{f_s} = \frac{1}{30} \approx 0.0333 \text{ seconds}$$

## Step 4.1:

Time Axis : 5 frams at 30 Hz

Therefore, $t[n] = \left[0, \frac{1}{30}, \frac{2}{30}, \frac{3}{30}, \frac{4}{30}\right]$

$t[n] = [0.000, 0.033, 0.067, 0.100, 0.133]$

4.2 : Apply Discrete Fourier Transform (DFT)

$$H[K] = \sum_{n=0}^{N-1} h[n] \cdot e^{-j\frac{2\pi K n}{N}}$$

for $K = 0, 1, \ldots\ldots N-1$

From Euler's Formula,

$$e^{-j\theta} = \cos(\theta) - j\sin(\theta)$$

Thus,

$$W_N^n = e^{-j\frac{2\pi n}{N}}$$

$$= \cos\left(\frac{2\pi n}{N}\right) - j\sin\left(\frac{2\pi n}{N}\right)$$

$$\Rightarrow W_5^n = e^{-j\frac{2\pi n}{5}}$$

Now We have to Calculate

H[0] H[1] H[2] H[3] H[4]

---

⊛ Compute H[0]:

$$H[0] = \sum_{n=0}^{5-1} h[n] . e^{-j.0}$$

$$= \sum h[n]$$

Therefore $H[0] = 0.00306303 + 0.00345782 + 0.00535559 - 0.00565757 - 0.00621886$

$$= 1.0001 \times 10^{-8}$$

⊛ Compute H[1]:

For $N = 5$

$$H[1] = \sum_{n=0}^{5-1} h[n] \cdot e^{-j\frac{2\pi n}{5}}$$

Therefore $W_5^n = e^{-j\frac{2\pi n}{5}}$

$$= \cos\left(\frac{2\pi n}{5}\right) - j\sin\left(\frac{2\pi n}{5}\right)$$

- $W_5^0 = e^{-j\frac{2\pi 0}{5}} = e^{-j0^\circ}$

- $W_5^1 = e^{-j\frac{2\pi\cdot 1}{5}} = e^{-j72^\circ}$

- $W_5^2 = e^{-j\frac{2\pi\cdot 2}{5}} = e^{-j144^\circ}$

- $W_5^3 = e^{-j216^\circ}$

- $W_5^4 = e^{-j288^\circ}$

# Now,

Using Euler's Formula,

- $W_5^0 = e^{-j0^\circ} = \cos(0^\circ) - j\sin(0^\circ) = \boxed{1 - 0j}$

AND

$$h[0] = 0.00306303$$

Therefore,

$$h[0] * W_5^0$$

$$= 0.00306303 - j.0$$

$$\boxed{= 0.00306303}$$

h[n]

| n | h[n] | |
|---|---|---|
| 0 | [ 0.00306303, | Fram-1 |
| 1 | 0.00345782, | Fram-2 |
| 2 | 0.00535559, | Fram-3 |
| 3 | -0.00565757, | Fram-4 |
| 4 | -0.00621886] | Fram-5 |

$$\cdot\; W_5^1 = e^{-j72^\circ} = \cos(72^\circ) - j\sin(72^\circ)$$

$$\boxed{= 0.3090 - j\,0.9511}$$

AND

$$h[1] = 0.00345782$$

Therefore,

$$h[1] * W_5^1$$

$$= 0.00345782 * (0.3090 - j\,0.9511)$$

$$\boxed{= 0.001069 - 0.003289\,j}$$

h[n]

| | | |
|---|---|---|
| 0 | [0.00306303, | Fram-1 |
| 1 | 0.00345782, | Fram-2 |
| 2 | 0.00535559, | Fram-3 |
| 3 | -0.00565757, | Fram-4 |
| 4 | -0.00621886] | Fram-5 |

$$\cdot\ W_5^2 = e^{-j144^\circ} = \cos(144^\circ) - j\sin(144^\circ)$$

$$\boxed{= -0.809 - j\,0.588}$$

| n | h[n] | |
|---|---|---|
| 0 | [0.00306303, | Fram-1 |
| 1 | 0.00345782, | Fram-2 |
| 2 | 0.00535559, | Fram-3 |
| 3 | -0.00565757, | Fram-4 |
| 4 | -0.00621886] | Fram-5 |

And

$$h[2] = 0.00535559$$

Therefore,

$$h[2] * W_5^2$$

$$= 0.00535559 * (-0.809 - 0.588j)$$

$$\boxed{= -0.004333 - 0.003148j}$$

$\cdot \; W_5^3 = e^{-j216^\circ} = \cos(216^\circ) - j\sin(216^\circ)$

$$\boxed{= -0.809 + j\,0.588}$$

And

$h[3] = -0.00565757$

Therefore,

$h[3] * W_5^3$

$= -0.00565757 * (-0.809 + 0.588j)$

$$\boxed{= 0.004577 - 0.003325\,j}$$

| | $h[n]$ | |
|---|---|---|
| 0 | [0.00306303, | Fram-1 |
| 1 | 0.00345782, | Fram-2 |
| 2 | 0.00535559, | Fram-3 |
| $\boxed{3}$ | -0.00565757, | Fram-4 |
| 4 | -0.00621886] | Fram-5 |

$$\cdot W_5^4 = e^{-j288^\circ} = \cos(288^\circ) - j\sin(216^\circ)$$

$$\boxed{= 0.309 + j\,0.951}$$

And

$$h[4] = -0.00621886$$

Therefore,

$$h[4] * W_5^4$$

$$= -0.00621886 * (0.309 + 0.951\,j)$$

$$\boxed{= -0.001922 - 0.005914\,j}$$

| $n$ | $h[n]$ | |
|---|---|---|
| 0 | [0.00306303, | Fram-1 |
| 1 | 0.00345782, | Fram-2 |
| 2 | 0.00535559, | Fram-3 |
| 3 | -0.00565757, | Fram-4 |
| 4 | -0.00621886] | Fram-5 |

Now:

$$
\begin{array}{rr}
0.003063 & +\ 0.00000\, j \\
0.001069 & -\ 0.003289\, j \\
-0.004333 & -\ 0.003148\, j \\
0.004577 & -\ 0.003325\, j \\
-0.001922 & -\ 0.005914\, j \\
\hline
0.00245 & -\ 0.01568\, j
\end{array}
$$

Therefore, $H[1] = \sum h[n] \cdot W_5^{n}$

$$\boxed{= 0.00245 - 0.01568\, j}$$

Amplitude :

$$|H[1]| = \sqrt{(0.00245)^2 + (-0.01568)^2} = \boxed{0.01587}$$

Now,

$H[k]$ is the strength of Frequency $f_k$

Therefore, $f_k = \frac{k}{N} \cdot f_s$

$$= \frac{1}{5} \cdot 30$$

$$= 6 \text{ Hz}$$

Therefore, Heart Rate = $6 \times 60s$

$$\boxed{= 360 \text{ bpm}}$$

Note: $H[k]$ is being calculated only for finding the strength of $f_k$.

Following these steps we have to calculate:

$$\cdot H[2] = \sum_{n=0}^{5-1} h[n] \cdot W_5^{2n} = 0.0071$$

Therefore, $f_2 = \frac{2}{5} \times 30$

$= 2 \times 6$

$= 12$ Hz

Therefore Heart Rate $= 12 \times 60$s

$\boxed{= 720 \text{ bpm.}}$

• $H[3]$ = Conjugate of $H[2]$

• $H[4]$ = Conjugate of $H[1]$

| $K$ | $H[K]$ | $f_K$ | $H_K$ |
|---|---|---|---|
| 0 | 0·00. 06303 | 0 | 0 |
| 1 | $\boxed{0{\cdot}01587}$ | 6Hz | 360 bpm |
| 2 | 0·0071 | 12Hz | 720 bpm |
| 3 | −0·0071 | −12Hz | −720 bpm |
| 4 | −0·01587 | −6Hz | −360 bpm |

$\boxed{\text{Note:}}$ For FFT

- Only $\frac{N}{2}$ bins are unique
- Rest are mirrored.
- The second half of the FFT
  → Complex conjugate of the first half.

- No new information.

Bin: A frequency bin in FFT output refers to a slot that holds information about a specific frequency range present in the signal

Caution: For example only 5 Frames had been selected. Thats why the value is not realistic.

# Applying on A Full length Vedio

Total frames used: 1547

Normalized RGB shape: (1547, 3)

Dominant frequency: 1.7841 Hz

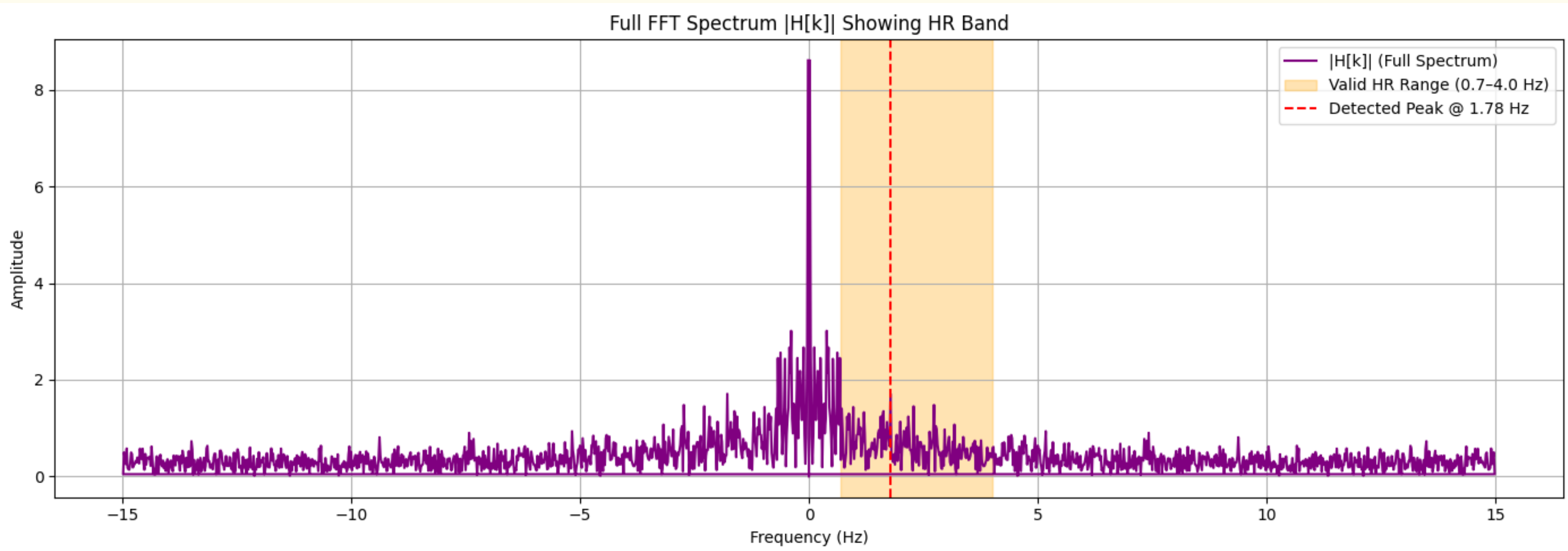
Estimated Heart Rate: 107.05 bpm

First 10 values of h(t):

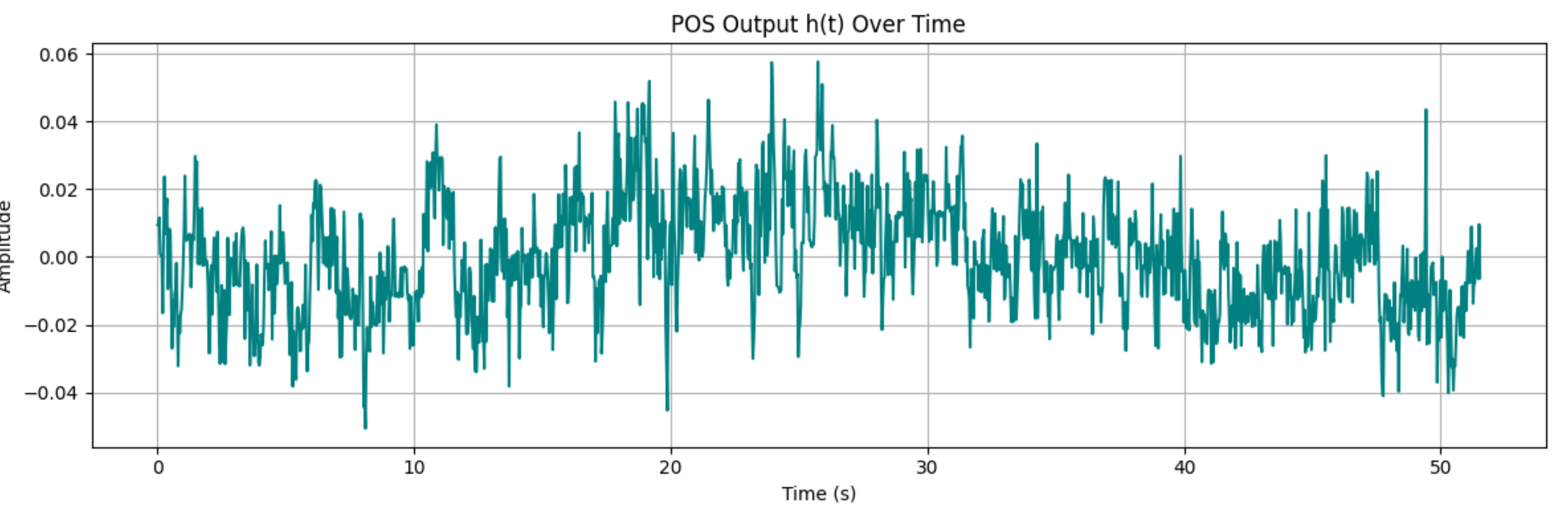
| | Frame | h(t) |
|---|---|---|
| 0 | 0 | 0.009481 |
| 1 | 1 | 0.009826 |
| 2 | 2 | 0.011632 |
| 3 | 3 | 0.000940 |
| 4 | 4 | 0.000363 |
| 5 | 5 | 0.000578 |
| 6 | 6 | -0.016535 |
| 7 | 7 | 0.006974 |
| 8 | 8 | 0.023692 |
| 9 | 9 | 0.006829 |

First 10 FFT bins:

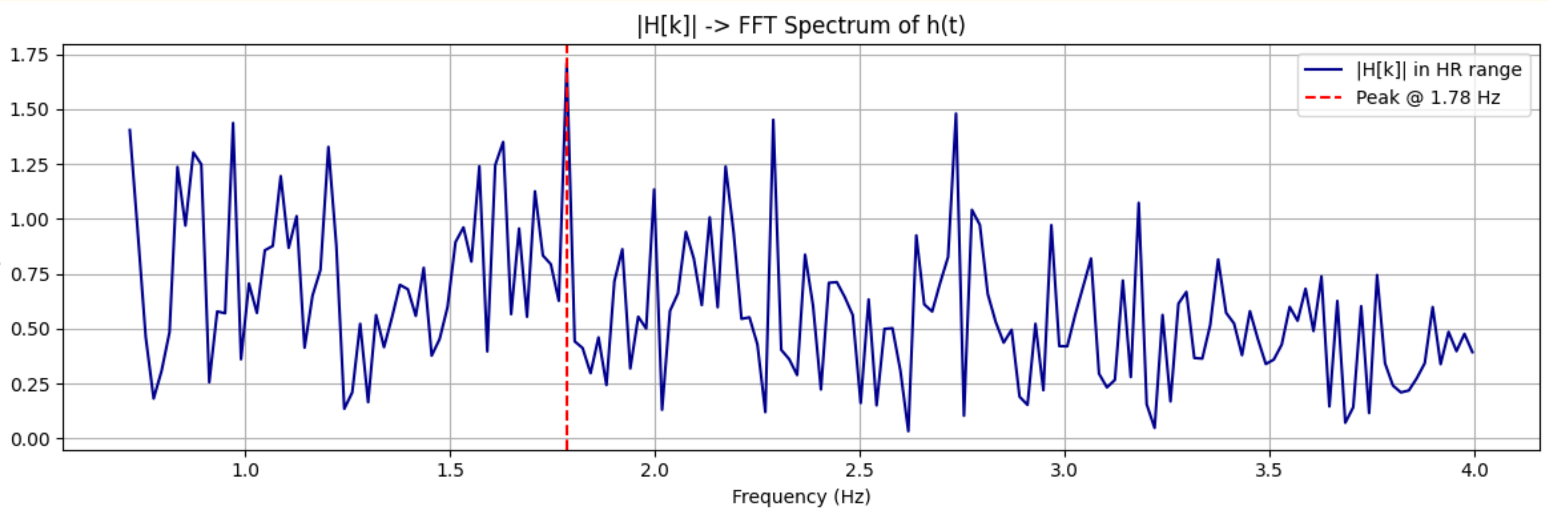
| | k | f_k (Hz) | H[k] | (complex) | \|H[k]\| |
|---|---|---|---|---|---|
| 0 | 0 | 0.000000 | 5.468348e-12-0.000000e+ | 00j | 5.468348e-12 |
| 1 | 1 | 0.019392 | -8.501821e+00-1.386808e+ | 00j | 8.614186e+00 |
| 2 | 2 | 0.038785 | 2.406633e+00+1.131701e+ | 00j | 2.659441e+00 |
| 3 | 3 | 0.058177 | 5.949433e-01-5.461370e- | 01j | 8.076033e-01 |
| 4 | 4 | 0.077569 | -1.224793e-01+2.319030e- | 01j | 2.622598e-01 |
| 5 | 5 | 0.096962 | -2.133134e-01+8.615722e- | 01j | 8.875862e-01 |
| 6 | 6 | 0.116354 | -2.240612e-01-2.661160e+ | 00j | 2.670576e+00 |
| 7 | 7 | 0.135747 | -3.501056e-01-1.716806e+ | 00j | 1.752141e+00 |
| 8 | 8 | 0.155139 | 1.018283e+00+5.700549e- | 01j | 1.166989e+00 |
| 9 | 9 | 0.174531 | 1.160186e+00-1.053608e+ | 00j | 1.567202e+00 |




Note: Only the frequency withing the valid human heart rate range (0.7 Hz – 4.0 Hz) are consider for HR estimate.



POS Output h(t) Over Time
Amplitude
Time (s)

|H[k]| -> FFT Spectrum of h(t)
|H[k]| in HR range
Peak @ 1.78 Hz
Frequency (Hz)