


# Estimation of Instantaneous Pitch Frequency in Speech Signals



Team Curiosity

Chepuri Shivani 2018122004

D Sai Siddarth 2018122006

# Contents and Workflow

- **Objectives**
- **About Pitch Estimation and its importance**
- **Abstract of the Main paper**
- **Algorithm outline**
- **Methodology**
- **More technical details**
- **Results and Plots**
- **Comparisons**

Video presentation Link:

[https://drive.google.com/file/d/10FXc8ZIRr\\_qIFjlleSVzH3vUzpqI7oY0/view](https://drive.google.com/file/d/10FXc8ZIRr_qIFjlleSVzH3vUzpqI7oY0/view)

# Overall Objectives

1. To Study and Explore the Algorithm proposed in the Main paper and cited papers for instantaneous pitch estimation.
2. To explore and implement VMD and VMD based optimisation algorithm in matlab.
3. To achieve proper input parameters for VMD.
4. To plot and compare results with inbuilt matlab functions and other methods for pitch estimation (for final ppt)
5. Strike some comparisons (VMD vs EMD, performance measures, etc) (for final ppt)
6. Video Presentation



## Datasets

- CMU Arctic DB - Scottish Male, US Female - 16KHz
- Cmu\_us\_awb\_arctic-0.90-release,  
cmu\_us\_slt\_arctic-0.95-release
- NOISEX-92

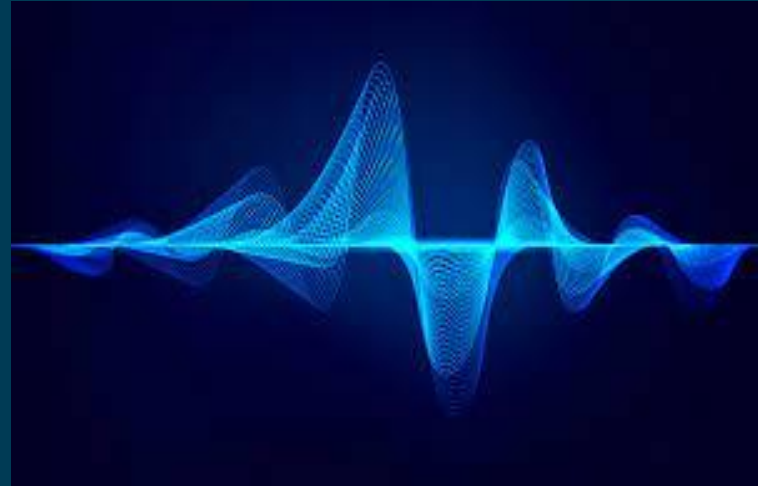


## User Inputs and Outputs

- A 1D (mono channel) speech signal
- VMD parameters - K, alpha, tau, K, DC, init, tol
  
- IMFs or VMFs, residuals (of all iterations)
- Component Centre Frequencies,  $F_0$  component,  $F_0$  envelope
- V/UV separated components
- Instantaneous Pitch Frequency

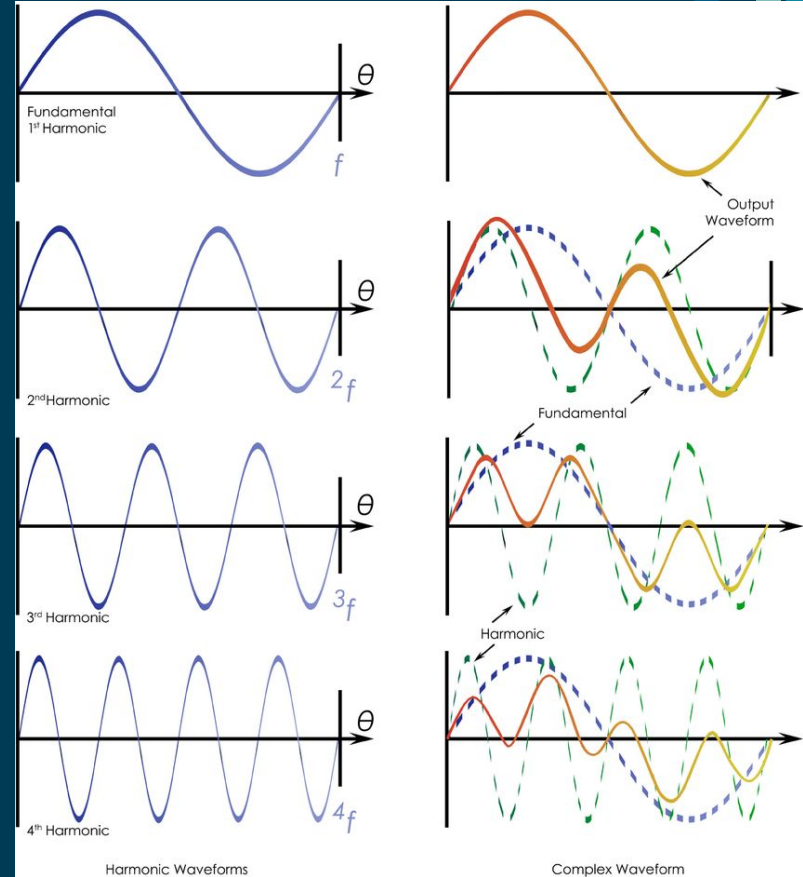
# Why is pitch very important?

- Frequency determines it
- Perceptual Property - periodicity
- Important Speaker feature
- Pitch Shifting, Time scaling
- Speech processing
- **Def:** It is the fundamental frequency of the vocal cord vibrations.
- Perceived by Human Brain



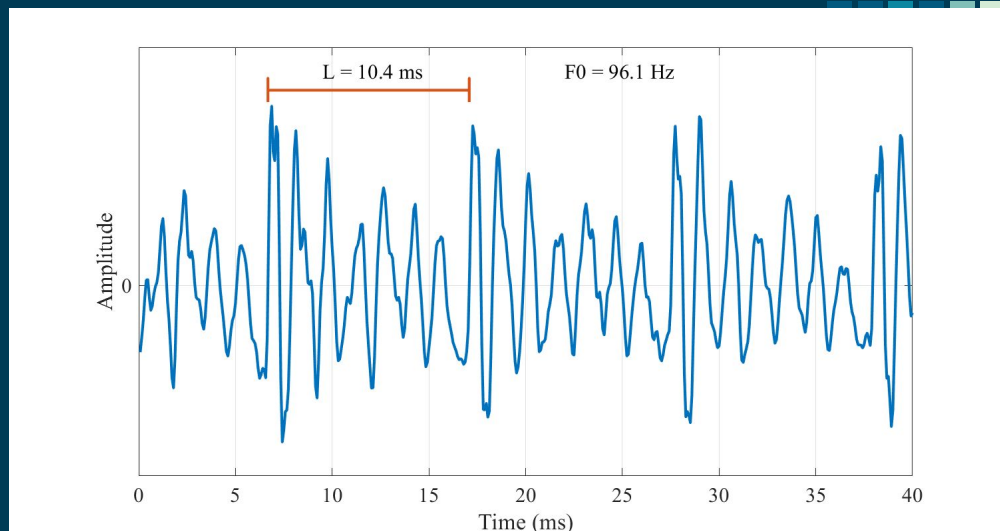
# Fundamental Frequency

- The lowest frequency of a periodic waveform.
- The average number of oscillations per second (Hz).
- $F_0$  of speech can vary from 40 Hz for low-pitched voices to 600 Hz for high-pitched voices.



# Fundamental Frequency

- Segment of a speech signal
- Fundamental period length  $T$
- $F_0 = 1/T$ .
- $F_s$  samples in one  $T$
- $L = F_s * T = F_s / F_0$ .



# Instantaneous Pitch Frequency

- The temporal derivative of the oscillation phase  $\phi$
- Useful for describing polychromatic (multiple frequencies) signals.
- The instantaneous frequency of a sinusoidal signal is constant and equals the oscillation frequency

$$z(t) = d(t) + jd_H(t) = A(t)e^{j\phi(t)} \quad (4)$$

In (4),  $d_H(t)$  is the Hilbert transform of signal  $d(t)$ . The amplitude envelope and instantaneous phase of analytic signal  $z(t)$  denoted as  $A(t)$  and  $\phi(t)$ , respectively and can be computed as follows [31]:

$$A(t) = \sqrt{d^2(t) + d_H^2(t)} \quad (5)$$

$$\phi(t) = \arctan \left[ \frac{d_H(t)}{d(t)} \right] \quad (6)$$

The instantaneous pitch frequency of the analytic signal  $z(t)$  is determined as follows:

$$\omega(t) = \frac{d\phi(t)}{dt} \quad (7)$$



# Pitch Estimation History/Techniques – PDAs

## Time domain detection

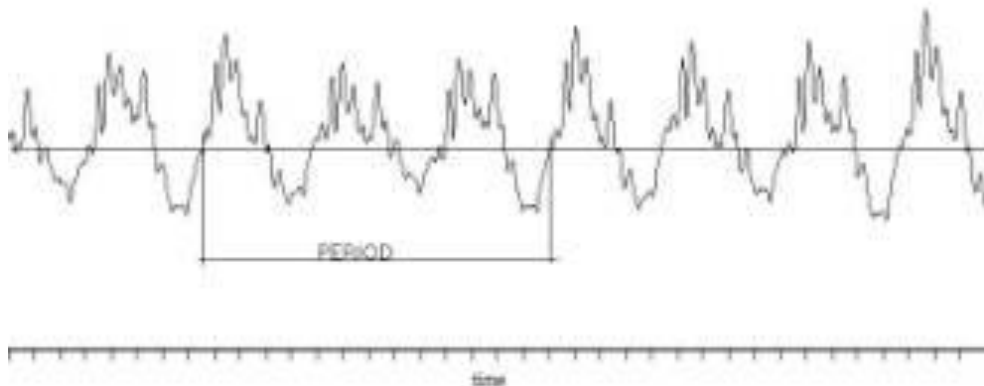
- ZCR
- Autocorrelation
- Maximum Likelihood
- Adaptive Filters Based
- Super Resolution Pitch Determination
- more

## Frequency Domain Detection

- Harmonic Product Spectrum
- Cepstrum
- Maximum Likelihood
- more

# Pitch Estimation History - PDAs

**Intuitive Approach:** Get the *zero crossing rate*. Does not work with overtones or with noise or quasi stationary signals.



**Autocorrelation Algorithms:** Highly accurate but prone to false detections. Bad with polyphonic and noisy signals

\*Will be explored more in final presentation for result comparison purposes

# Abstract of the Main Paper

- An algorithm based on VMD and Hilbert Transform.
- VMD is applied iteratively until centre frequencies converge.
- Specific input parameters are needed.
- Fundamental frequency component and its envelope are extracted.
- Voiced and Unvoiced (V/UV) regions from a given speech signal are detected for  $F_0$ .
- Instantaneous pitch Freq is obtained by Hilbert Transform of Voiced speech component of  $F_0$ .

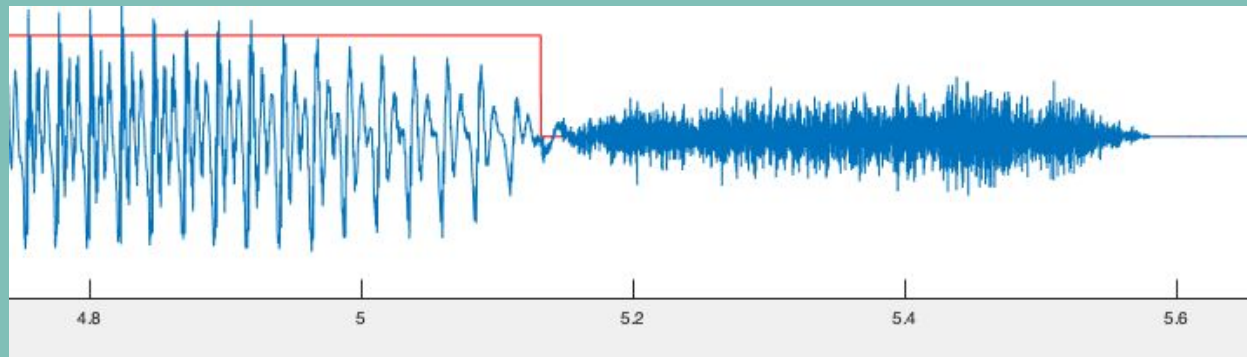


# V/UV speech components

Speech can be decomposed into numerous V/UV segments.

Voiced Speech - Lower ZCR, higher energy & amplitude, almost constant frequency tones, periodic (therefore can be identified & extracted)

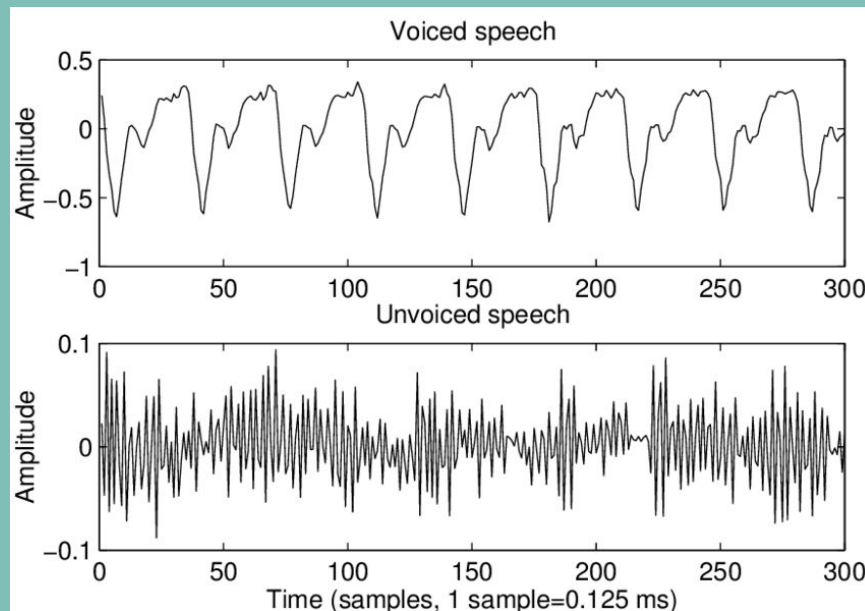
Unvoiced Speech - High ZCR, not periodic, random-like, energy & amplitude much lower



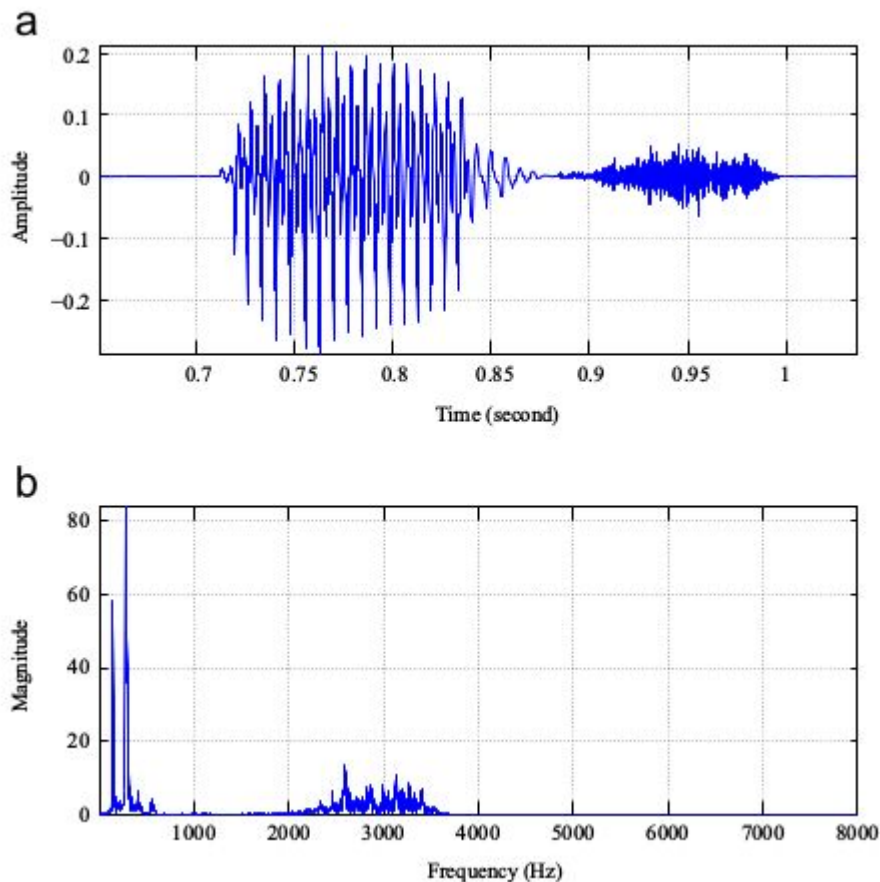
# V/UV speech components

**Voiced speech** - produced when periodic pulses of air generated by the vibrating glottis resonate through the vocal tract, at frequencies dependent on the vocal tract shape.

**Unvoiced speech** - caused by air passing through a narrow constriction of the vocal tract as when consonants are spoken.



V/UV segments  
amplitude[speech  
signal]/magnitud  
e spectrum

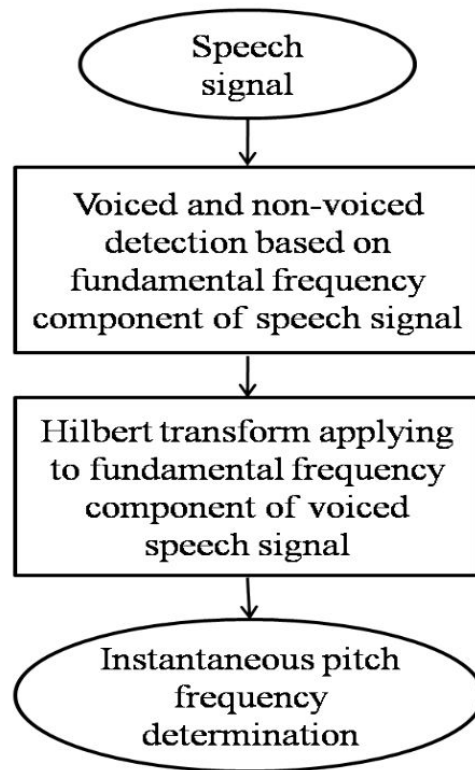


# Algorithm Flow chart & Methodology

**First Step:** Extraction of Voiced Segment of  $F_0$  of the given speech segment (most rigorous, involves more sub-steps) using VMD

**Second Step:** Hilbert Transform to get Analytical Signal

**Third Step:** Derivative of phase to get Instantaneous Pitch



# Variational Mode Decomposition

- VMD is used to decompose a real valued signal to sub-signals or modes ( $y_k$ ). It replaces traditional EMD.
- Unlike the IMFs which are AM-FM signals with non decreasing phase, these modes are almost considered as compact around the corresponding center frequencies.
- Performance depends on the number of modes given and other input parameters
- It is an adaptive signal processing method (used in time varying systems)



# Variational Mode Decomposition

1. The real signal is converted to the analytic signal using the Hilbert transform such that one-sided frequency spectrum of the signal is obtained.
2. The frequency spectrum of the component is shifted to baseband regions, using modulation property to the respective estimated center frequencies.
3. The bandwidth of a component is estimated through the H1 Gaussian smoothness of the demodulated signal, i.e. the squared L2-norm of the gradient.

# Variational Mode Decomposition

The resulting constrained variational problem is,

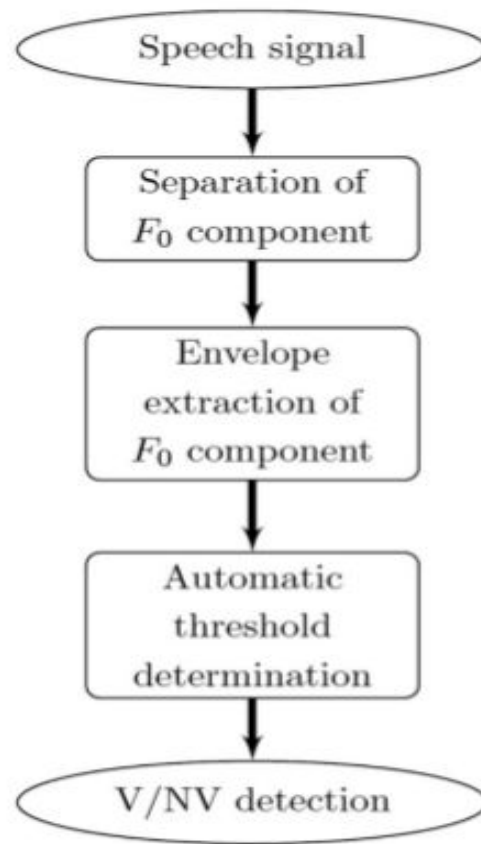
$$\min_{\{y_k\}, \{\omega_k\}} \left\{ \sum_{k=1}^K \left\| \partial_t \left[ \left( \delta(t) + \frac{j}{\pi t} \right) * y_k(t) \right] e^{-j\omega_k t} \right\|_2^2 \right\}$$

such that  $\sum_{k=1}^K y_k(t) = y(t)$

Here  $y_k$  denotes  $k$ th component of the set of modes  $\{y_k\}$ .  $\omega_k$  represents center frequency of  $K$ th mode of the signal and  $\{\omega_k\}$  represents the set of center frequencies

# Step 1 a- Determination of Fundamental Frequency Component

- Apply VMD on the signal with  $K = 2$  (we will have 2 components as the output) and choose the component with min central frequency for next iteration.
- We go on with the iterations until the center frequencies converge.
- Now, how to achieve the center frequencies is the question. For that we have ADMM optimisation algorithm (alternating direction method of multipliers)

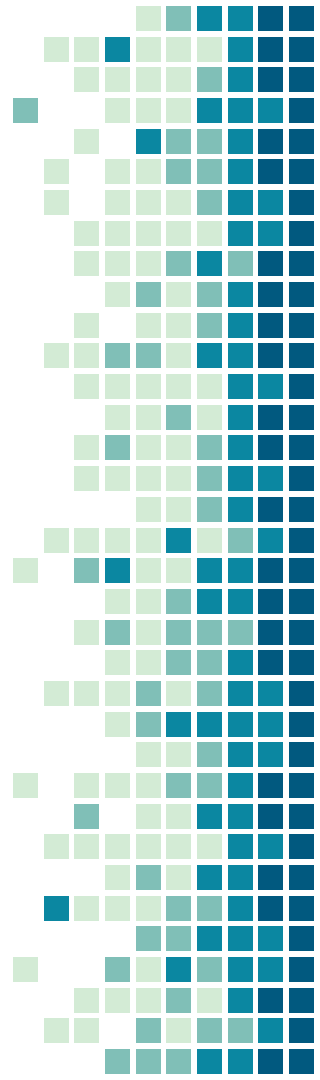


# Parameters for the proposed Algorithm

## The input parameters (#6):

1. The balancing parameter of the data-fidelity constraint ( $\alpha$ )
2. The time step of the dual ascent ( $\tau$ )
3. The number of components ( $K$ ) to be extracted
4. The tolerance of convergence criterion ( $\text{tol}$ )
5. Number of DC components
6. The initialization of center frequencies  $\omega$  (init)

The values of input parameters namely  $\text{tol}$ ,  $\tau$ , init,  $\alpha$ , and  $K$  are fixed to  $10^{20}$ , 0, 0, 80 and 2, respectively.



# Algorithm

## Constrained variational problem with parameters and lagrangians

$$\begin{aligned}\mathcal{L}(\{y_k\}, \{\omega_k\}, \lambda) := & \alpha \sum_k \left\| \partial_t \left[ \left( \delta(t) + \frac{j}{\pi t} \right) * y_k(t) \right] e^{-j\omega_k t} \right\|_2^2 \\ & + \left\| y(t) - \sum_k y_k(t) \right\|_2^2 + \left\langle \lambda(t), y(t) - \sum_k y_k(t) \right\rangle\end{aligned}$$

From ADMM, we get the following updates (alpha is low)

$$\omega_k^{n+1} = \frac{\int_0^\infty \omega |\hat{Y}_k(\omega)|^2 d\omega}{\int_0^\infty |\hat{Y}_k(\omega)|^2 d\omega} \quad \hat{Y}_k^{n+1}(\omega) = \frac{\hat{Y}(\omega) - \sum_{i \neq k} \hat{Y}_i(\omega) + \frac{\hat{\lambda}(\omega)}{2}}{1 + 2\alpha(\omega - \omega_k)^2}$$

# Step 1 a- Optimisation Algorithm

Minimisation is done wrt  $u_k$  and  $w_k$

Note  $u_k \sim Y_k$

## Convergence Criteria:

- The final selected component should posses the energy (or center frequency) nearly equal to the component in the previous iteration.
- Parseval relation to convert the signals and computations from time to frequency domains
- In this case, center frequency less than 50 Hz components are discarded

---

### Algorithm 1: ADMM optimization concept for VMD

---

Initialize  $\{u_k^1\}, \{\omega_k^1\}, \lambda^1, n \leftarrow 0$

**repeat**

$n \leftarrow n + 1$

**for**  $k = 1 : K$  **do**

Update  $u_k$ :

$$u_k^{n+1} \leftarrow \arg \min_{u_k} \mathcal{L}(\{u_{i < k}^{n+1}\}, \{u_{i \geq k}^n\}, \{\omega_i^n\}, \lambda^n) \quad (16)$$

**end for**

**for**  $k = 1 : K$  **do**

Update  $\omega_k$ :

$$\omega_k^{n+1} \leftarrow \arg \min_{\omega_k} \mathcal{L}(\{u_i^{n+1}\}, \{\omega_{i < k}^{n+1}\}, \{\omega_{i \geq k}^n\}, \lambda^n) \quad (17)$$

**end for**

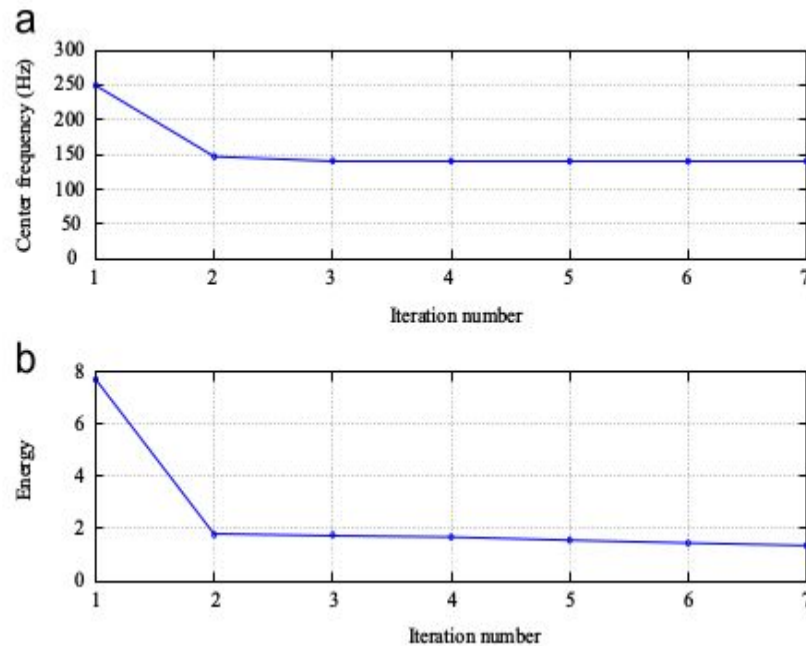
Dual ascent:

$$\lambda^{n+1} \leftarrow \lambda^n + \tau \left( f - \sum_k u_k^{n+1} \right) \quad (18)$$

**until** convergence:  $\sum_k \|u_k^{n+1} - u_k^n\|_2^2 / \|u_k^n\|_2^2 < \epsilon$ .

---

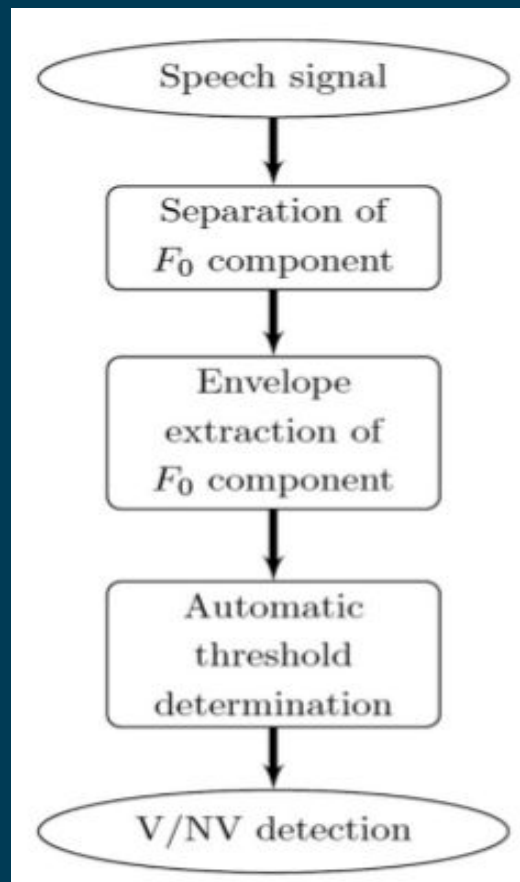
- Center frequency and energy of selected components of iter3 and iter4 are nearly equal.
- So, iter4 selected component is chosen as  $F_0$



# Steps yet to be implemented in main scope

1. Step 1b - Envelope Extraction
2. Step 1c- Automatic Thresholding & V/UV detection
3. Steps 2 & 3 - Apply hilbert transform and get instant pitch
4. Discuss Advantages and Disadvantages of VMD and the proposed method

(details and results from these steps will be done in final ppt)



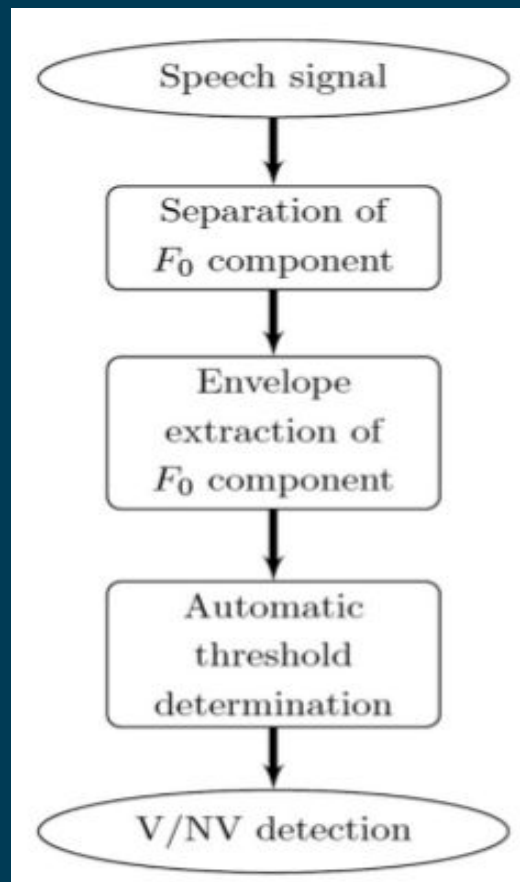


# Envelope Extraction & Thresholding

**Step 1b** - The main paper uses a technique called SDOF(based on cross correlation) and we are looking into its implementation and trying to look for a crisp method for this purpose.

**Step 1c** - The automatic threshold for V/NV detection is to be computed based on the paper ref [8]. We are also looking at a method which used ACFs, and an easier method

(In the final ppt, We will provide comparisons, if any)



# Results so far – $F_0$ and center freq convergence is achieved

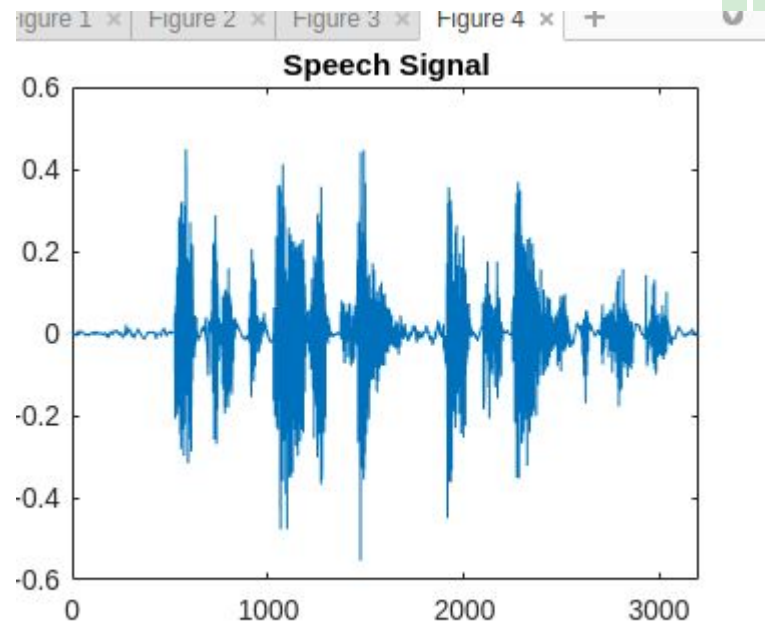
We got the  $F_0$  component from VMD. We had written an elaborate code for estimation of  $F_0$  from VMD.

We also computed the envelope of  $F_0$

We need to find automatic threshold next for V/UV and main paper scope will be done.

We need to verify energy convergence (it will as center freq are converging already)

Need to plot Hilbert Transform, Analytical signal etc



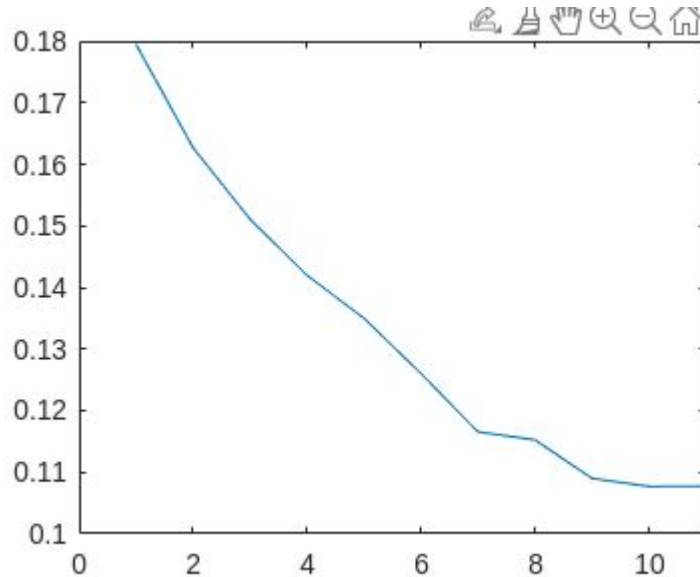
# Plots – Center Frequency Convergence

The convergence parameter  $\text{diff} = 0.001$  for these plots.

We experimented with 0.1, 0.01 also but it was not quite converging.

Value of Convergence is around 0.11 rad/s

We are plotting only time domain plots. Magnitude Spectrum, Threshold function, etc will plot in later ppt



# IMFS of all Iterations - Plot of $F_0$

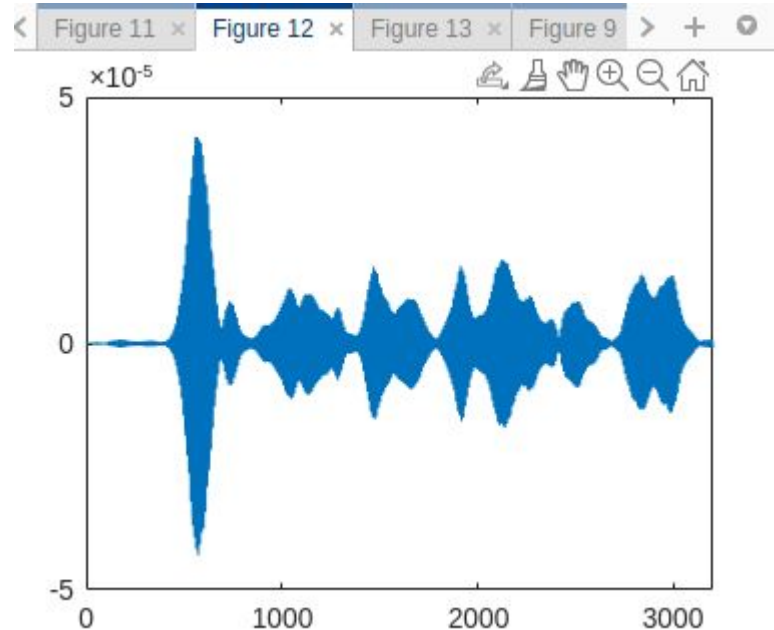
Number of IMF or modes achieved for each iteration of called vmd is 2 ( $k=2$ )

Number of iterations depends upon the convergence criteria.

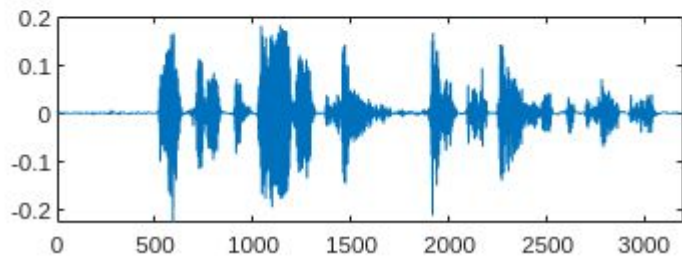
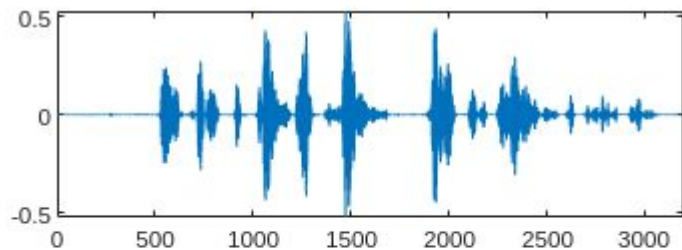
The component with least center frequency of all these achieved components is the  $F_0$ .

Note: This  $F_0$  may not be fully accurate and it depends on 'diff' parameter. In this case, it is pretty good

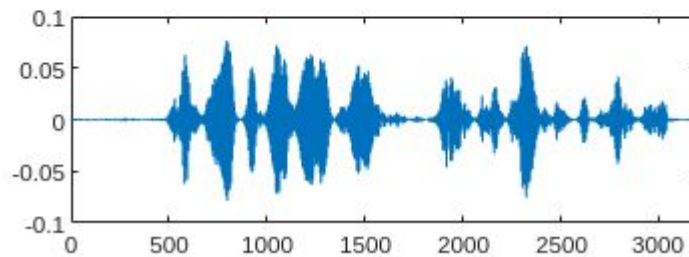
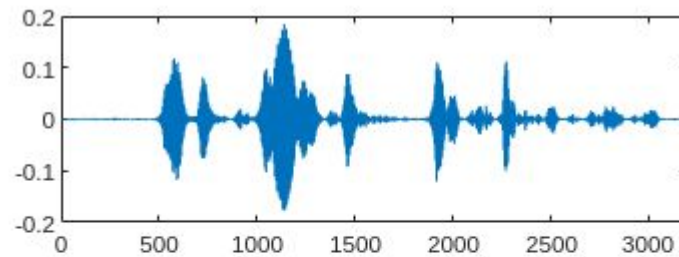
We will do some performance analysis in the next ppt to support this



# Intermediate IMF plots

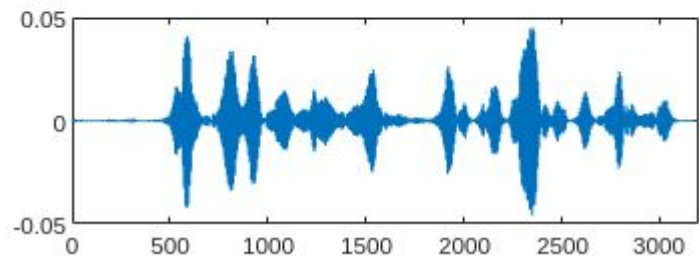
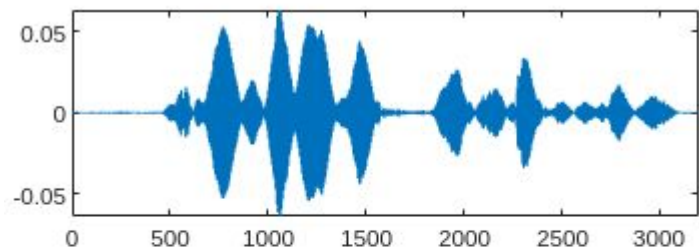


Iter 1

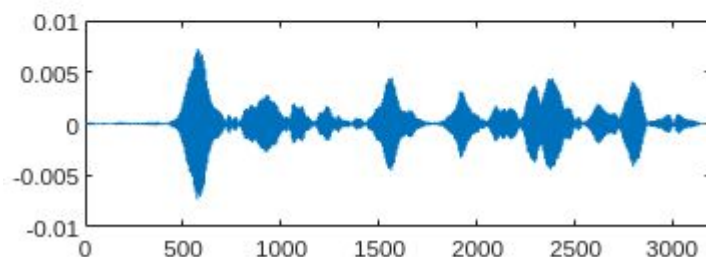
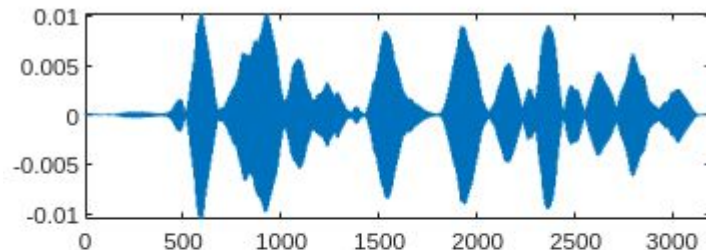


Iter 2

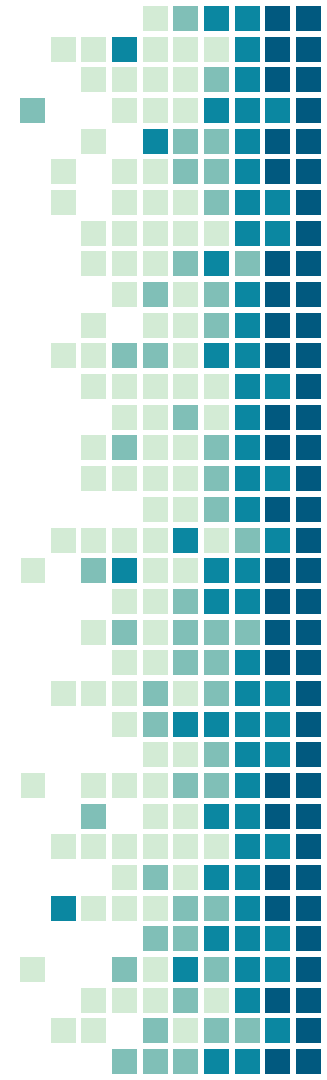
# Intermediate IMF plots



Iter 3

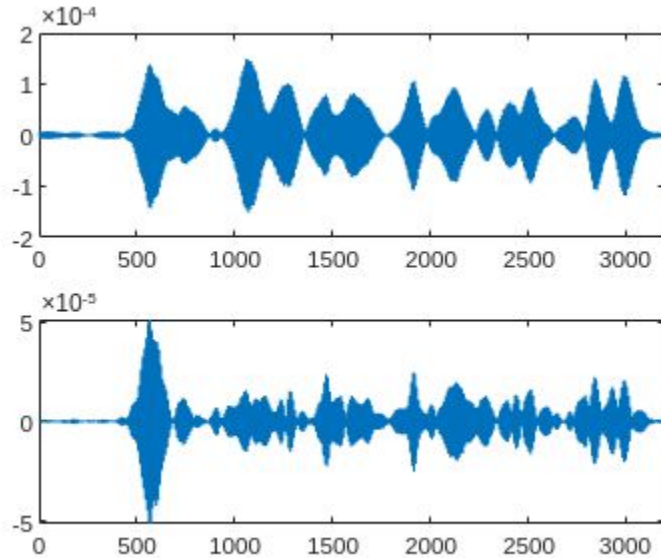


Iter 5

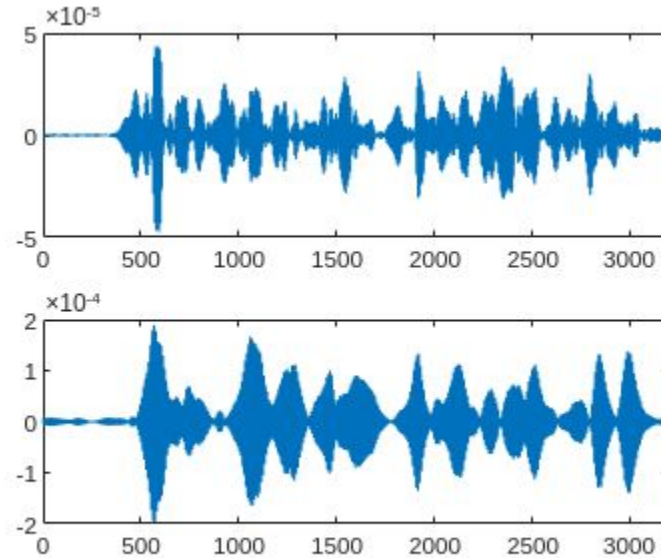


# Intermediate IMF plots

Iter 9



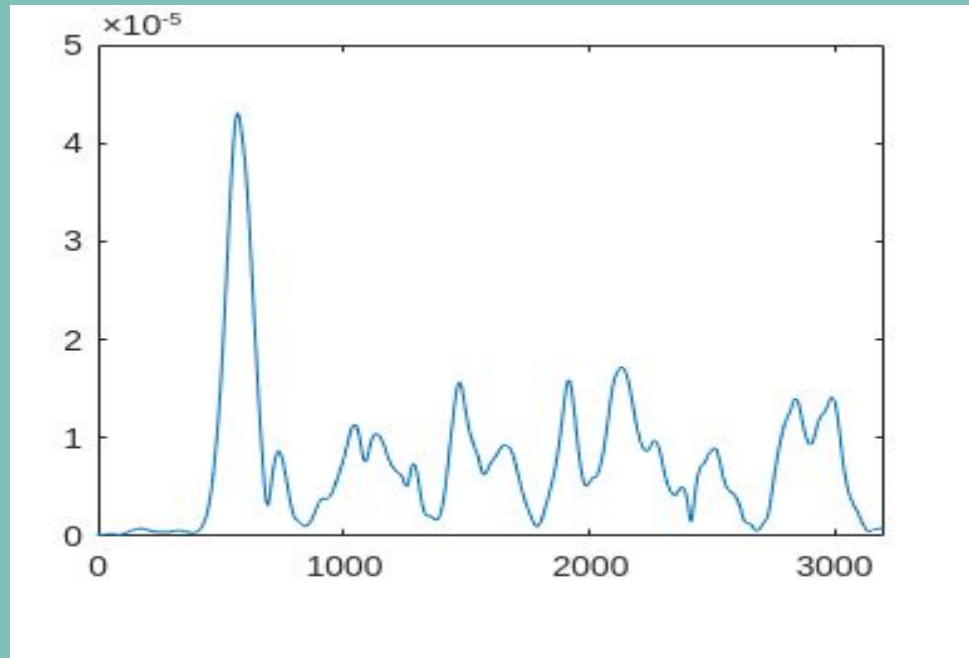
Iter 8



## Observations

We can clearly notice that with increased number of iterations, the frequency of the components is reducing (FO is basically the smallest frequency in the signal)

# F0 envelope





# Code Snippet

```
if v3.CentralFrequencies(1) <= v3.CentralFrequencies(2)
    comp = 1;
else
    comp = 2;
end

if comp == 1
    y_iter = v1(:,1);
else
    y_iter = v1(:,2);
end

if i == 1
    CF_arr = v3.CentralFrequencies(comp);
else
    CF_arr = [CF_arr; v3.CentralFrequencies(comp)];
end

if (i >= 2)
    if abs(CF_arr(i) - CF_arr(i-1)) <= diff
        y_F0 = y_iter;
        y_CF = CF_arr(i);
        break;
    end

    if(i>100)
        y_F0 = y_iter;
        y_CF = CF_arr(i);
        break;
    end
end

end
```

## Bringing more innovation and expanding scope - Plan

- V/UV classification using EMD to compare with the existing method
- Try More pitch estimation methods (eg. CWT, Autocorrelation, EMD etc)
- Come up with some good Performance Measures and compare these methods and their results
- Try VMD with parameter adaptation (yet to be explored)
- Try reducing the computation time for the overall algorithm (downsampling, etc)
- Introduce Noise into the entire process and check results
- VMD vs EMD
- Explore applications for this project
- Produce results and points mentioned in slide 24 and other missing objectives.

Video presentation Link:

[https://drive.google.com/file/d/10FXc8ZIRr\\_qIFjIleSVzH3vUzpqI7oY0/view](https://drive.google.com/file/d/10FXc8ZIRr_qIFjIleSVzH3vUzpqI7oY0/view)

## Contributions:

- Matlab Code for VMD part and Optimisation part have been done by both the members equally.
- Both of us experimented with input parameters, results and debugging.
- Both of us helped each other in understanding various concepts involved in the algorithm
- Presentation slides mostly done by Shivani, some are done by Siddharth.

We both studied all the concepts and steps but for the sake of presenting in the video, we have shared the topics randomly.

# References

1. [http://www.festvox.org/cmu\\_arctic/](http://www.festvox.org/cmu_arctic/)
2. <https://ieeexplore.ieee.org/document/7369574> (main)
3. Y. Li, B. Xue, H. Hong, and X. Zhu, "Instantaneous pitch estimation based on empirical wavelet transform," in 19th International Conference on Digital Signal Processing. IEEE, 2014, pp. 250–253.
4. A. Upadhyay and R.B. Pachori, "Instantaneous voiced/non-voiced detection in speech signals based on variational mode decomposition," Journal of the Franklin Institute, vol. 352, no. 7, pp.2679–2707, 2015.
5. K. Dragomiretskiy and D. Zosso, "Variational mode decomposition," IEEE Transactions on Signal Processing, vol. 62, no. 3, pp. 531–544, Feb. 2014.
6. [https://en.wikipedia.org/wiki/Pitch\\_detection\\_algorithm](https://en.wikipedia.org/wiki/Pitch_detection_algorithm)
7. <https://ccrma.stanford.edu/~pdelac/154/m154paper.htm>
8. <https://sci-hub.do/10.1016/j.jfranklin.2013.01.002>

# THANK YOU!

Please find 1 matlab code file, named  
'curiosity.m' (zipped while submission)

Video presentation Link:

[https://drive.google.com/file/d/10FXc8ZIRr\\_gIFjIleSVzH3vUzpqI7oY0/view](https://drive.google.com/file/d/10FXc8ZIRr_gIFjIleSVzH3vUzpqI7oY0/view)