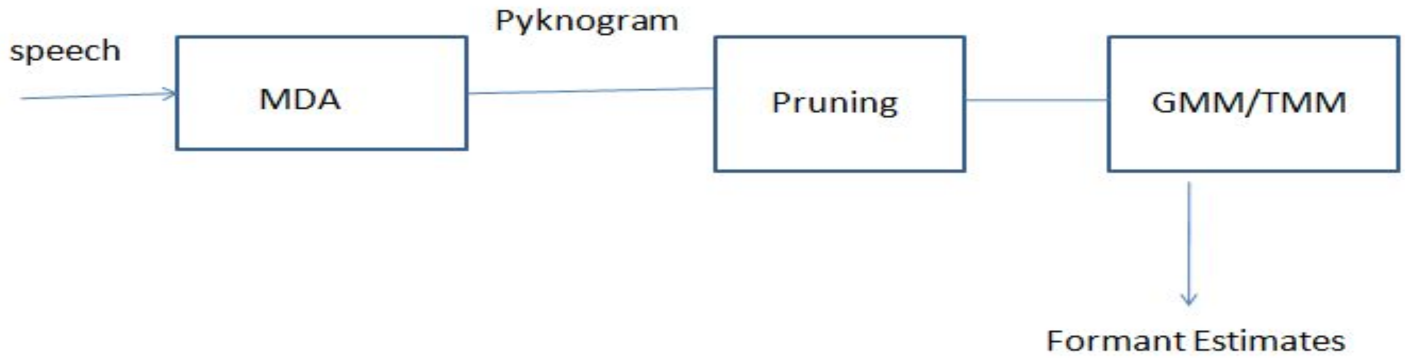
Mixture Model approach for Formant Tracking

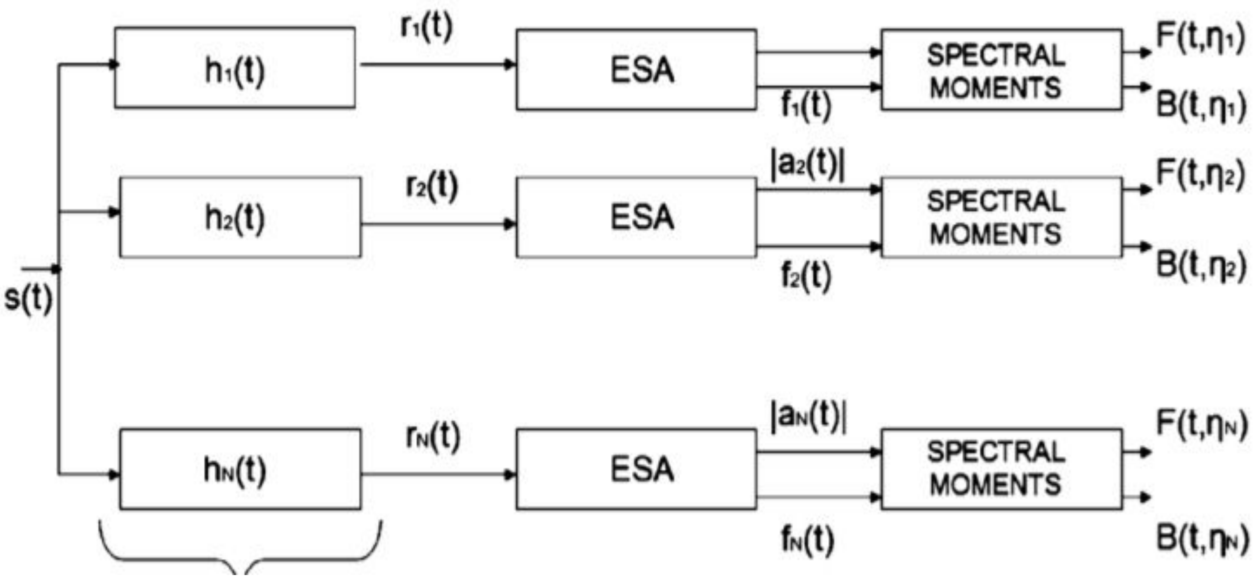
K.Parimala
K.Annapurna

Block Diagram



This approach consists of two main steps:
1.Computation of Pyknogram using multiband AM/FM demodulation
2.Statistical Modeling of the pyknogram

Pyknogram



Gabor filter for sub band decomposition
Specifications:- bandwidth of each filter is 400 Hz ,the spacing between successive filter’s center frequency is 50 Hz

Discrete Energy separation Algorithm

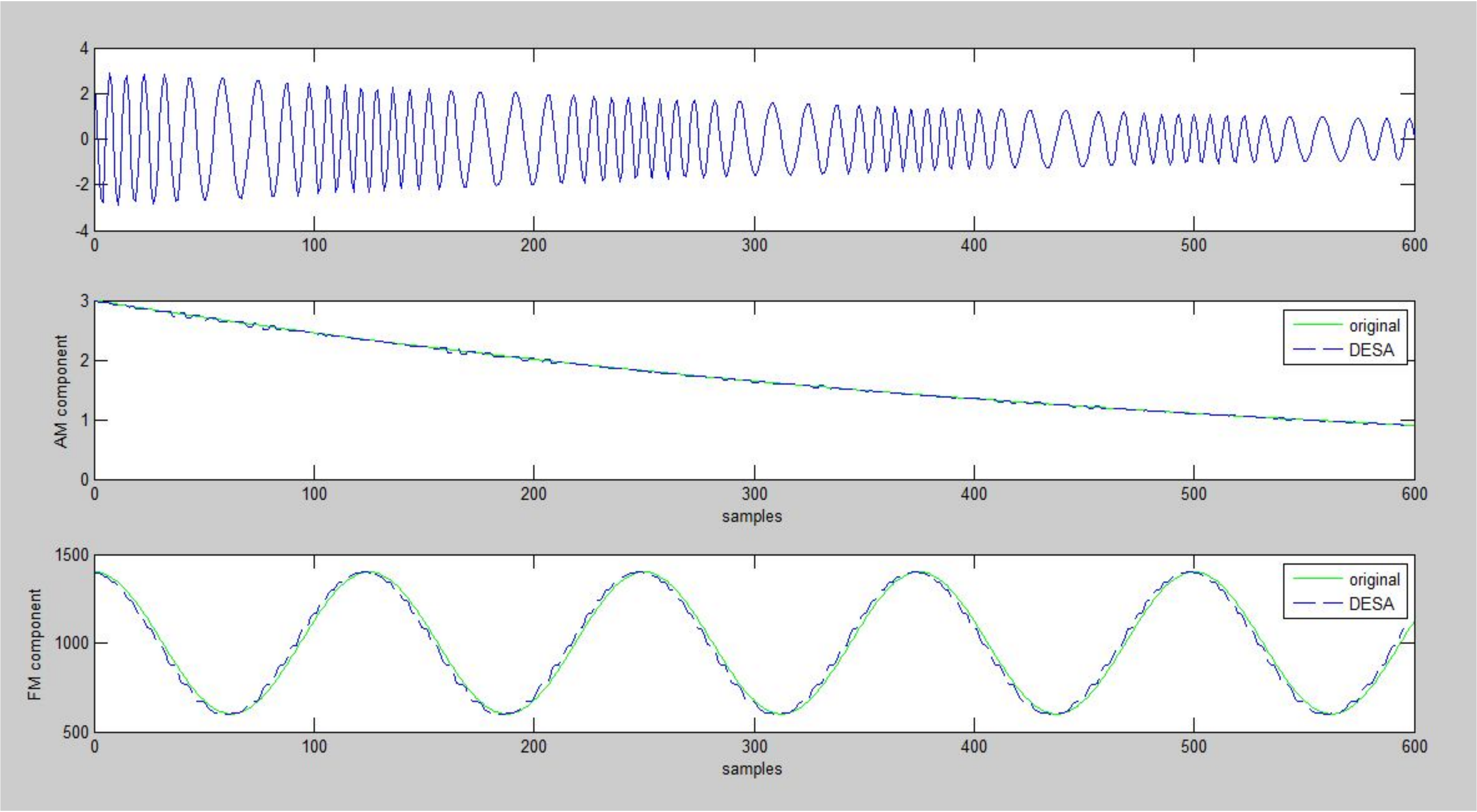
- This separates the signal into AM and FM components
- For a discrete signal x(n)

$$x(n) = a(n)\cos\left(\Omega_c n + \sum_{i=0}^{n} q(i)\frac{1}{T} + \theta\right),$$

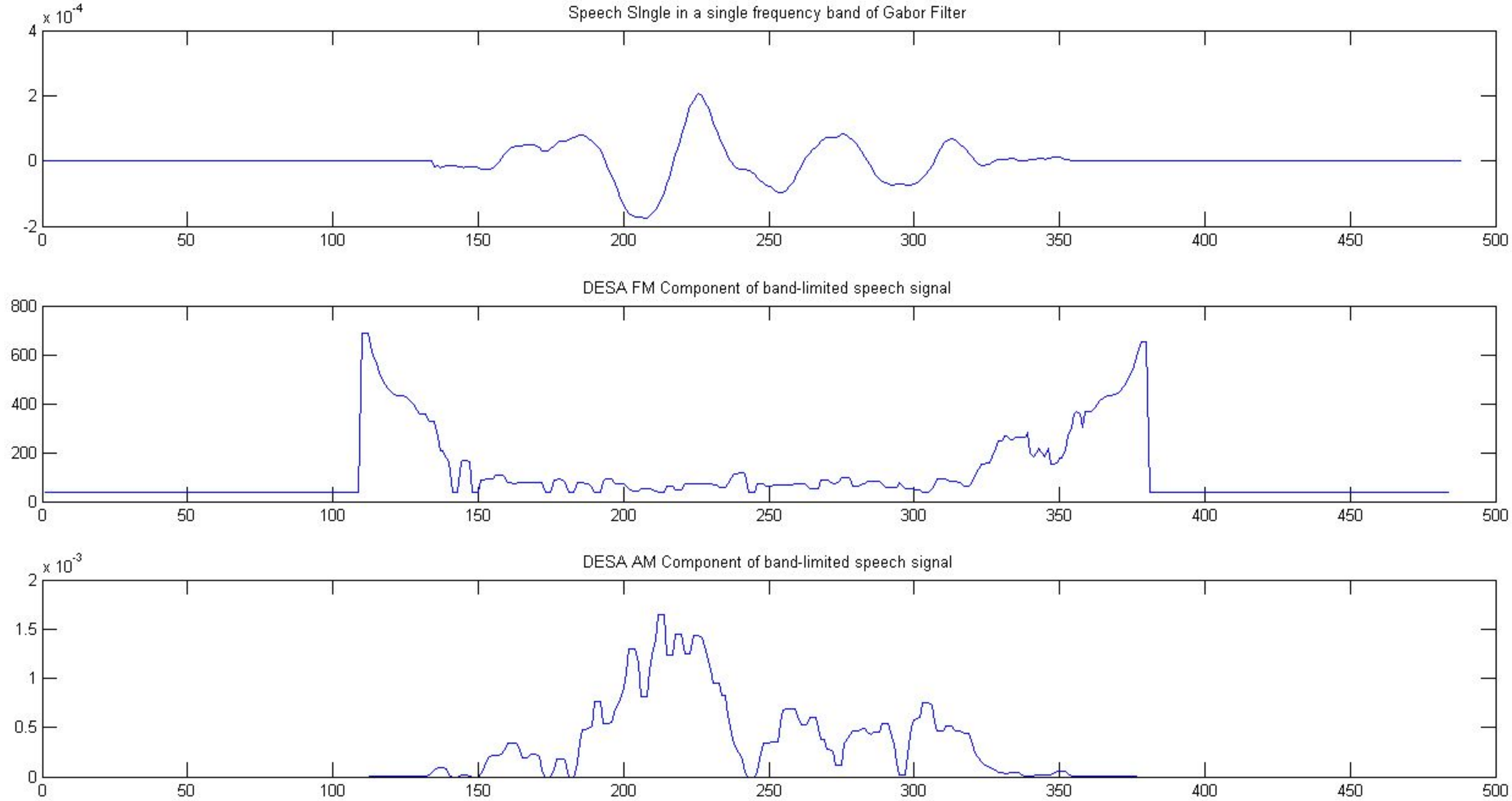
$$\Psi[x(n)] = x^2(n) - x(n-1)x(n+1).$$

$$\begin{aligned}\hat{\Omega}_i(n) &\approx \frac{1}{2}\arccos\left(1 - \frac{\Psi[x(n+1) - x(n-1)]}{2\Psi[x(n)]}\right) \\ &\approx \Omega_c + q(n) \\ |\hat{a}(n)| &\approx \frac{2\Psi[x(n)]}{\sqrt{\Psi[x(n+1) - x(n-1)]}}.\end{aligned}$$

Validation of DESA with Chirp Signal



DESA output for band passed speech signal centred at 300 HZ



Spectral Moments

$$F(t_0, \eta_k) = \frac{\int_{t_0}^{t_0+T} f_k(t)\, |a_k(t)|^2 \, \mathrm{d}t}{\int_{t_0}^{t_0+T} |a_k(t)|^2 \, \mathrm{d}t},$$

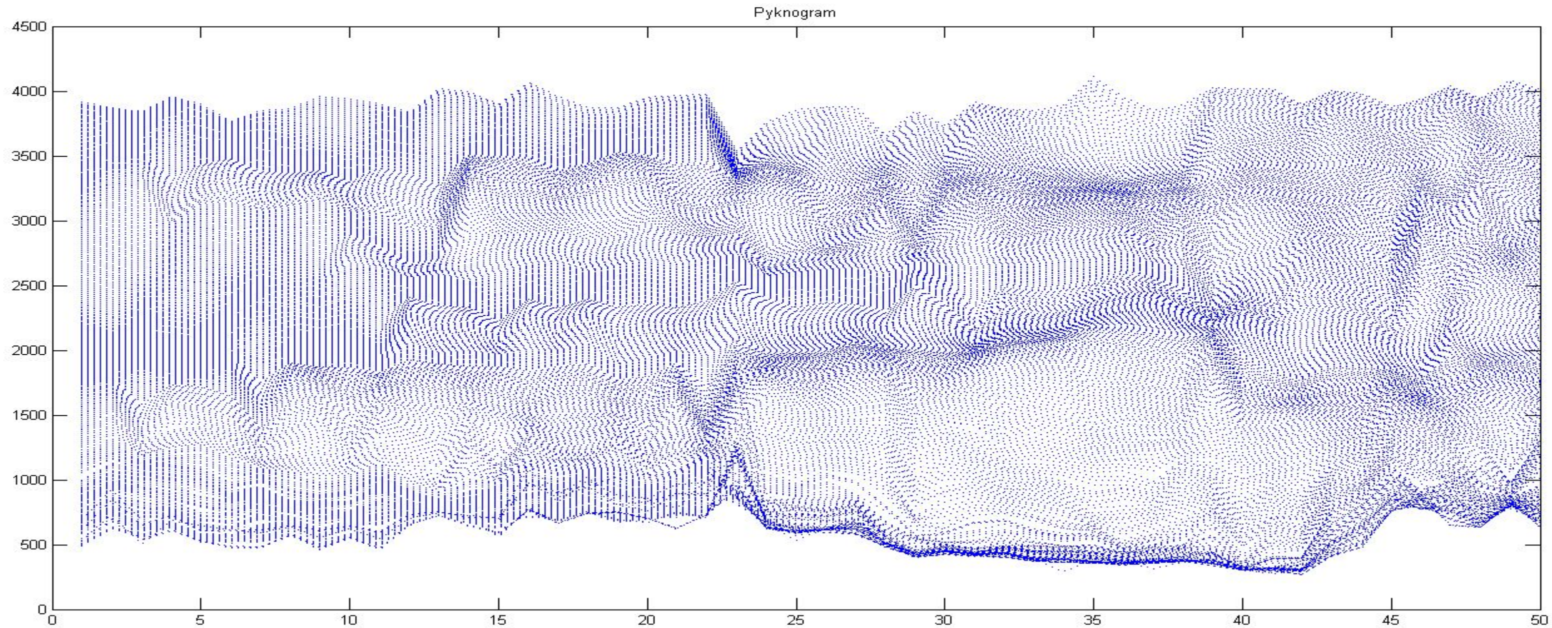
This denotes the average IF of the k th Band pass signal
The scatter plot of the F matrix is called Pyknogram

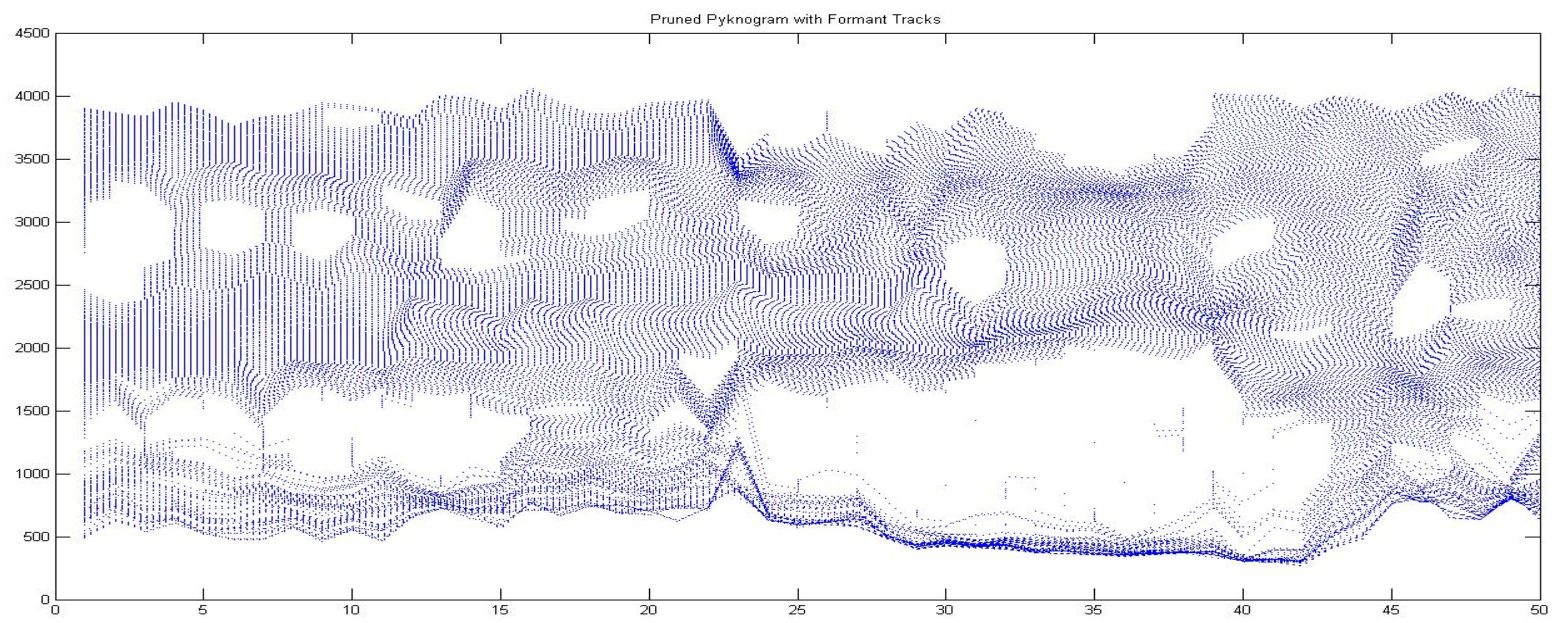
Pruning of Pyknogram

$F(t,n+1)-F(t, n) <$ threshold

Threshold should be selected to capture the dense regions in Pyknogram

Pyknogram before and after pruning





Modeling the pyknogram data

- Corresponding to each frame, the pyknogram data is viewed as the sampled distribution of a random variable.
- The L –component tMM at every time frame is then given as

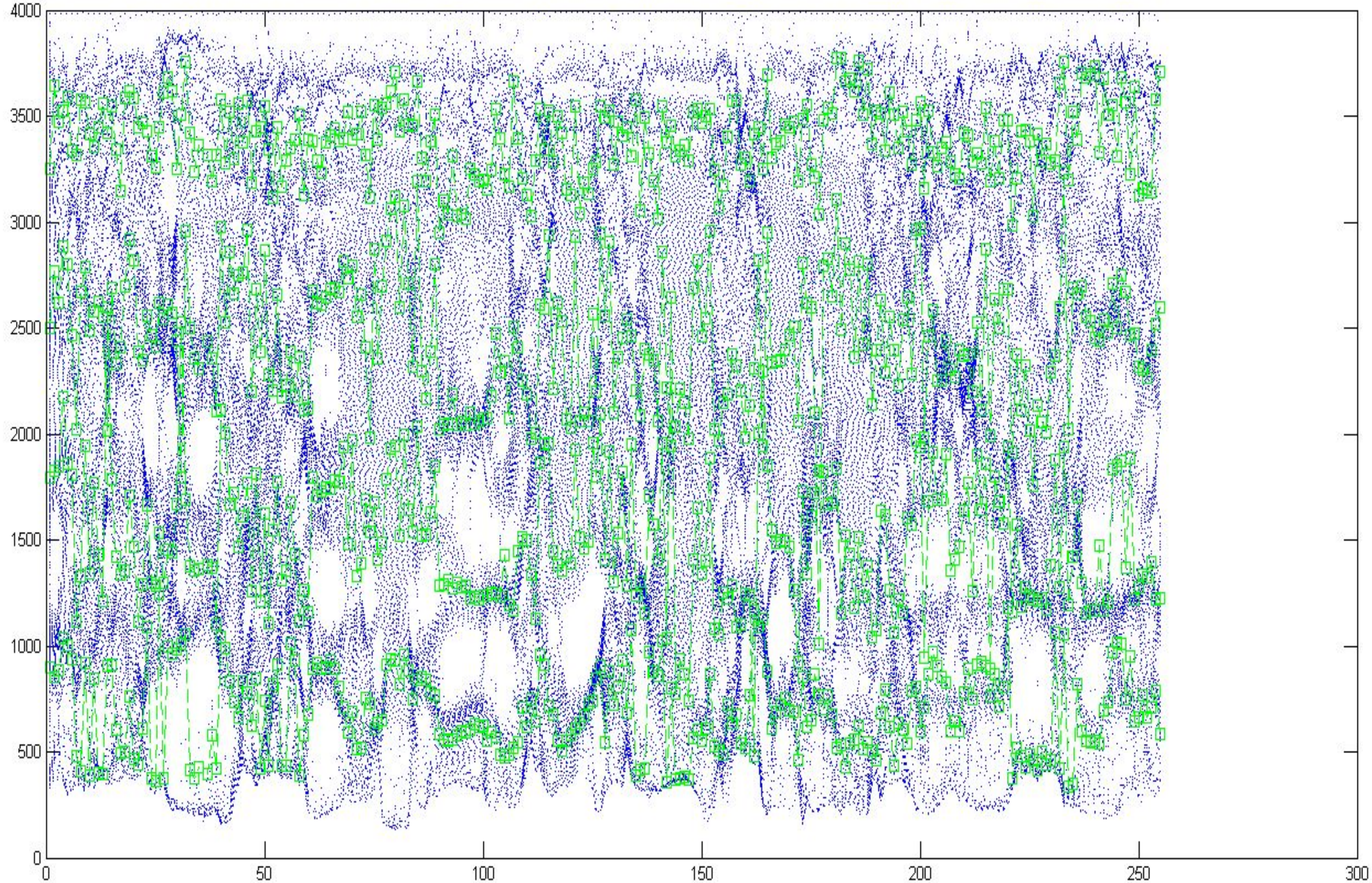
$$p\left(x(t);\Psi^t\right) = \sum_{i=1}^{L} \pi_i^t p_{\mathbf{X}}\left(x(t);\mu_i^t,\mathbf{\Sigma_i^t},\nu_i^t\right),$$

(L=4)-component tmm density function must be used for four formants

$$p_{\mathbf{X}}(x) = p(x;\mu,\mathbf{\Sigma},\nu),$$
$$= \frac{\Gamma\left(\frac{\nu+1}{2}\right)\|\mathbf{\Sigma}\|^{-\frac{1}{2}}}{(\pi\nu)^{\frac{1}{2}}\Gamma\left(\frac{\nu}{2}\right)\left(1+\frac{\delta(x;\mu,\mathbf{\Sigma})}{\nu}\right)^{\frac{\nu+1}{2}}},$$

$$\delta(x;\mu,\mathbf{\Sigma}) = (x-\mu)^\top \Sigma^{-1}(x-\mu),$$

Using GMM for modeling



Expectation Maximization

- Here the goal is to estimate the parameters of tmm given the raw formants ‘X’ for each frame t.
- Posterior probalities and updating equations

$$\tau_{ij}^{(t,k+1)} \triangleq \frac{\pi_i^{(t,k)} p\left(x_j(t); \boldsymbol{\mu}_i^{(t,k)}, \boldsymbol{\Sigma}_{\mathbf{i}}^{(t,k)}, \nu_i^{(t,k)}\right)}{\sum_{i=1}^{L} \pi_i^{(t,k)} p\left(x_j(t); \boldsymbol{\mu}_i^{(t,k)}, \boldsymbol{\Sigma}_{\mathbf{i}}^{(t,k)}, \boldsymbol{\nu}_i^{(t,k)}\right)},$$

$$u_{ij}^{(t,k+1)} \triangleq \frac{\nu_i^{(t,k)} + 1}{\nu_i^{(t,k)} + \boldsymbol{\delta}\left(x_j(t); \boldsymbol{\mu}_i^{(t,k)}, \boldsymbol{\Sigma}_{\mathbf{i}}^{(t,k)}\right)}.$$

- <u>Updating Equations for all the parameters</u>

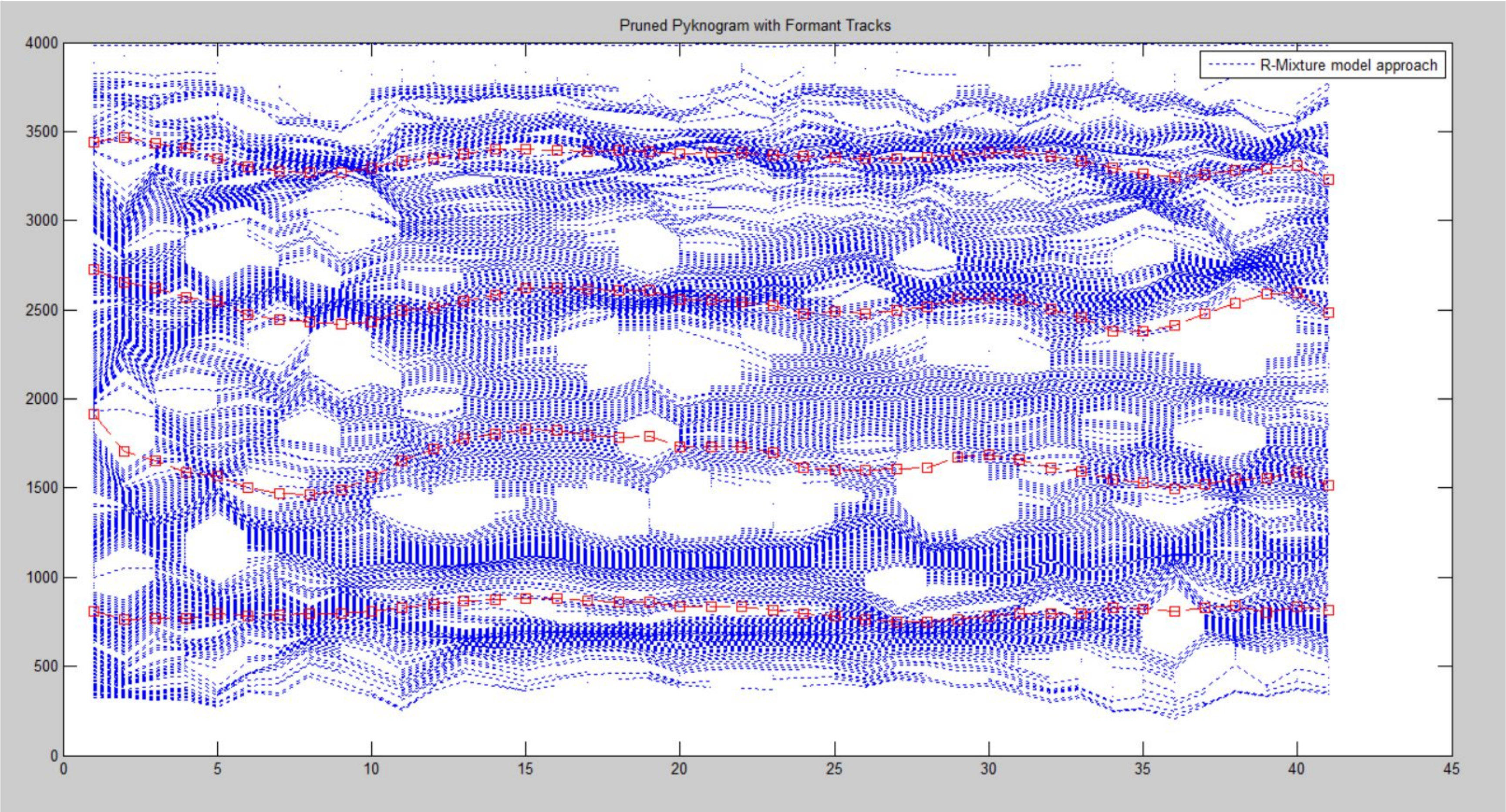
$$\pi_i^{(t,k+1)} = \sum_{j=1}^{n(t)} \frac{\tau_{ij}^{(t,k+1)}}{n(t)}, \quad i = 1, \cdots, L,$$

$$\boldsymbol{\mu}_{\mathrm{i}}^{(t,k+1)} = \frac{\sum_{j=1}^{n(t)} \tau_{ij}^{(t,k+1)} u_{ij}^{(t,k+1)} x_j(t)}{\sum_{j=1}^{n(t)} \tau_{ij}^{(t,k+1)}},$$

$$\boldsymbol{\Sigma}_i^{(t,k+1)} = \frac{\sum_{j=1}^{n(t)} \tau_{ij}^{(t,k+1)} u_{ij}^{(t,k+1)} \beta_{ij}^{(t,k+1)}}{\sum_{j=1}^{n(t)} \tau_{ij}^{(t,k+1)}},$$

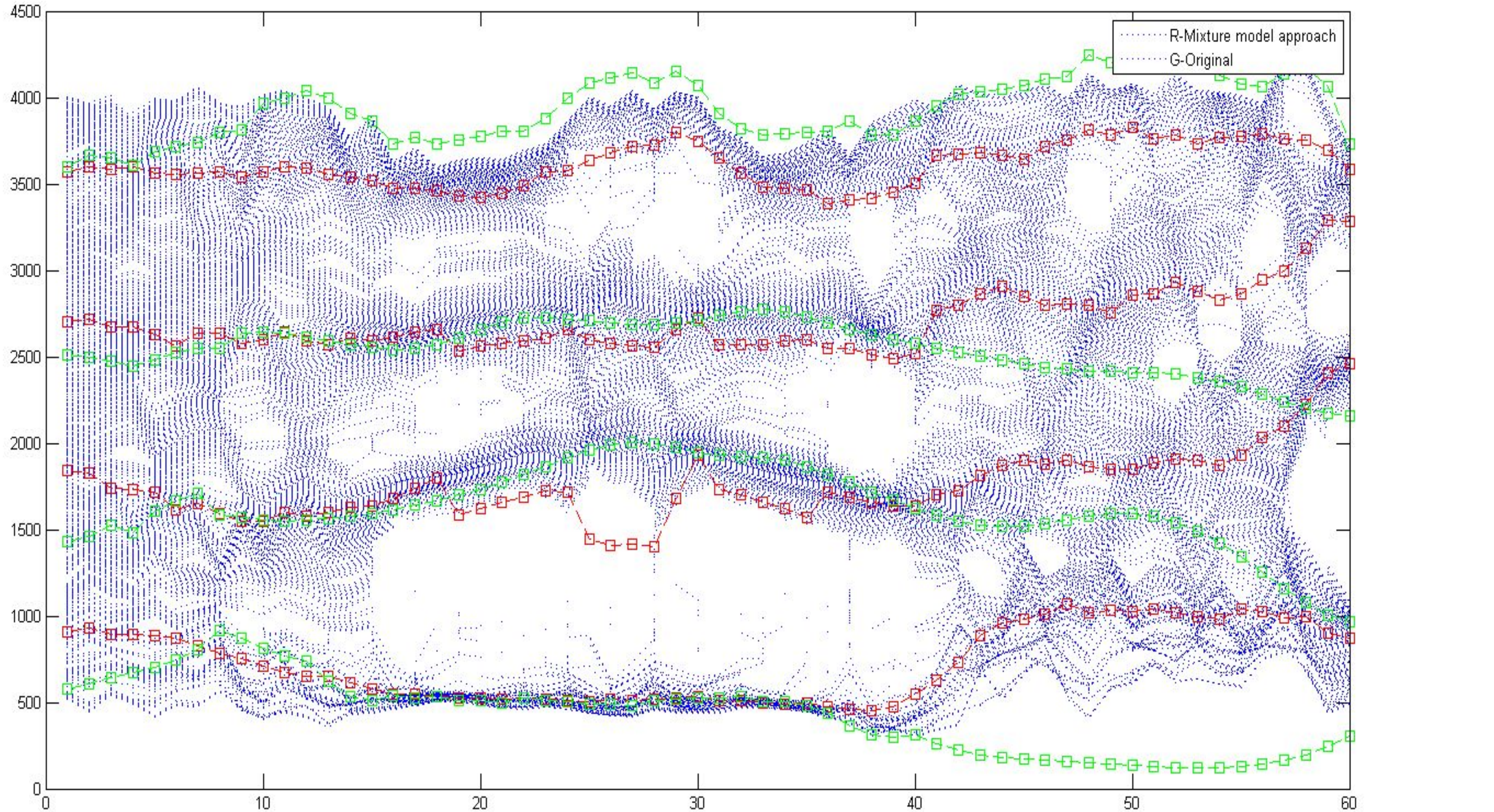
After convergence, the formant tracks {F1,F2,F3,F4} are means of multimodal density.

Results : Formants for phoneme \pa\

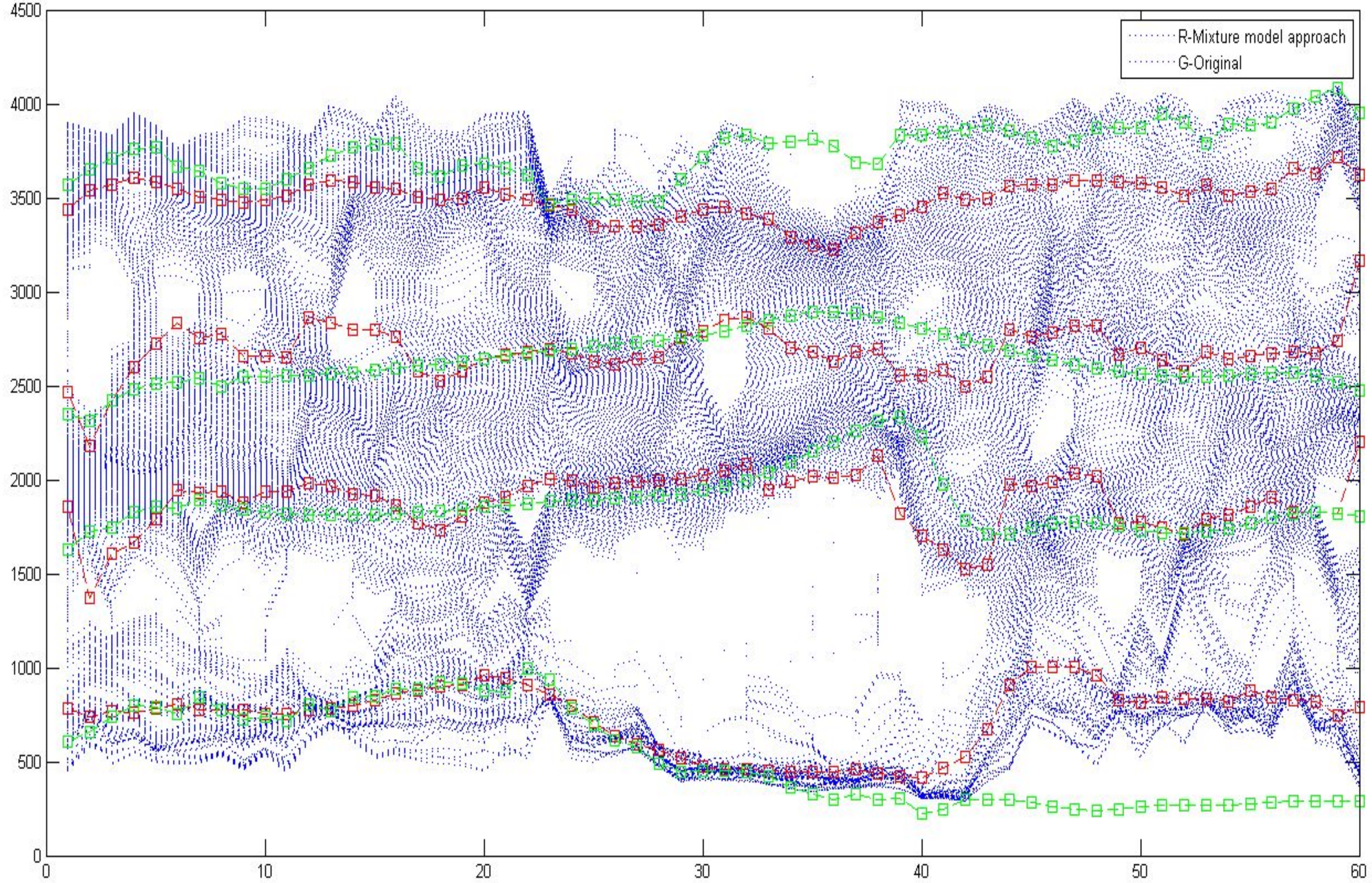


Results

Comparisons with VTR database



Result2:



R-Mixture model approach
G-Original
4500
4000
3500
3000
2500
2000
1500
1000
500
0
0
10
20
30
40
50
60

Results:

Percentage Deviations for each formant for various TIMIT speech signals :
The deviation in first formant is lower due to lower base values but the deviation values of first formant are lower.

F1	F2	F3	F4
10.2018	6.6943	5.8441	6.4053
16.7272	6.7807	3.1973	6.8096
17.5933	6.2059	4.7712	7.0685

F1	F2	F3	F4
54.31	105.90	177.71	246.48
64.95	167.59	83.106	211.46
89.15	115.67	97.25	365.069