Advances in Statistical Estimation and Tracking of AM-FM Speech Components

Athanassios Katsamanis & Petros Maragos

School of Electrical and Computer Engineering,

National Technical University of Athens

[nkatsam, maragos]@cs.ntua.gr

Overview

- **Motivation**
  - Speech AM-FM Model

    $y(k) = \sum_i \alpha_i(k) cos[\phi_i(k)]$
  - Speech Nonstationarity
  - Varying and Noisy Environments
- **Our approach**
  - State-Space Model for Instantaneous Amplitude and Instantaneous Frequency and Extended Kalman Filtering (EKF)
  - Local Regularization of EKF estimates using a standard Demodulation Algorithm
  - Allow time varying power and bandwidth of speech resonances in a Particle Filtering framework

References

- *AM-FM Speech Model and Energy Based Demodulation*, Maragos, Kaiser, Quatieri (1991-1993)
- *Statistical Model Based Demodulation Algorithm*, Lu, Doerschuk (1996)
- *Particle Methods for Speech*, Vermaak, Andrieu, Doucet, Godsill (2002)

Outline

- Background
  - The Energy Separation Algorithm and Gabor filtering
  - The Statistical Model Based Demodulation Algorithm (MBDA)
- Regularization of the estimates of the Multiband Demodulation Algorithm (MBDA) by applying the ESA
  - Modified state-space model
  - Experimental Results
- Particle Filtering for time-varying parameters
  - Particle Filtering
  - Parameter Estimation
  - Experimental Results
- Conclusions

Demodulation via Energy Separation

Monocomponent AM-FM signal

$$x(t) = \alpha(t)\cos(\varphi(t))$$

$$\omega(t) \triangleq \frac{d\varphi(t)}{dt}$$

Teager-Kaiser Energy Operator

$$\Psi[x(t)] \triangleq [\dot{x}(t)]^2 - x(t)\ddot{x}(t)$$

Energy Separation Algorithm (ESA)

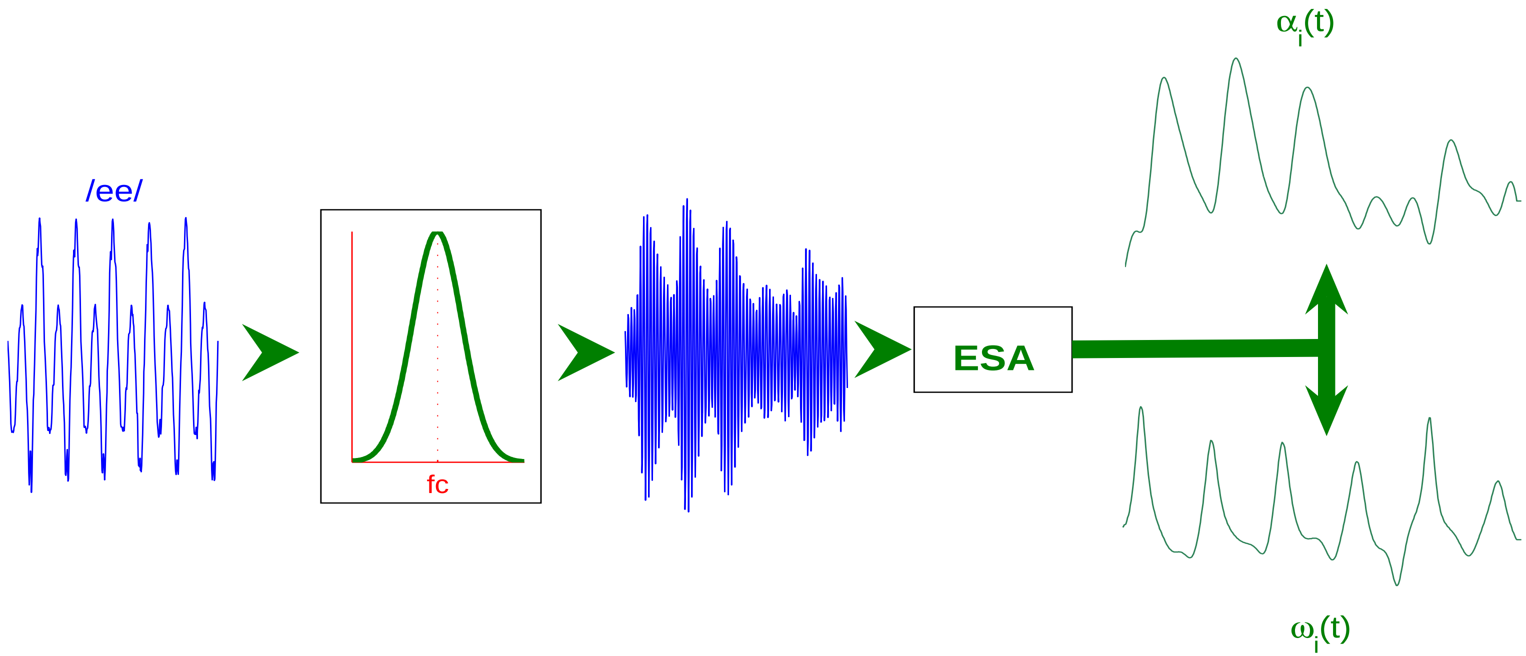
$$\omega(t) \approx \sqrt{\frac{\Psi[\dot{x}(t)]}{\Psi[x(t)]}}, \ |\alpha(t)| \approx \frac{\Psi[x(t)]}{\sqrt{\Psi[\dot{x}(t)]}}$$

Demodulation via Energy Separation

For multicomponent signals

$$x(t) = \sum \alpha_i(t) \cos(\varphi_i(t))$$

$$\omega_i(t) \triangleq \frac{d\varphi_i(t)}{dt}$$



Statistical Model Based Demodulation Algorithm (MBDA)

The state-space equations for each component:

$$\alpha_i(k+1) = \beta_{\alpha_i}\alpha_i(k) + q_{\alpha_i} w_{\alpha_i}(k)$$
$$\nu_i(k+1) = \beta_{\nu_i}\nu_i(k) + q_{\nu_i} w_{\nu_i}(k)$$
$$f_i(k+1) = f_i(k) + q_{f_i} w_{f_i}(k)$$

Instantaneous Phase

$$\varphi_i(k) = \varphi_i(0) + 2\pi T_s \sum_{m=0}^{k-1} [f_i(m) + \nu_i(m)]$$

Observation

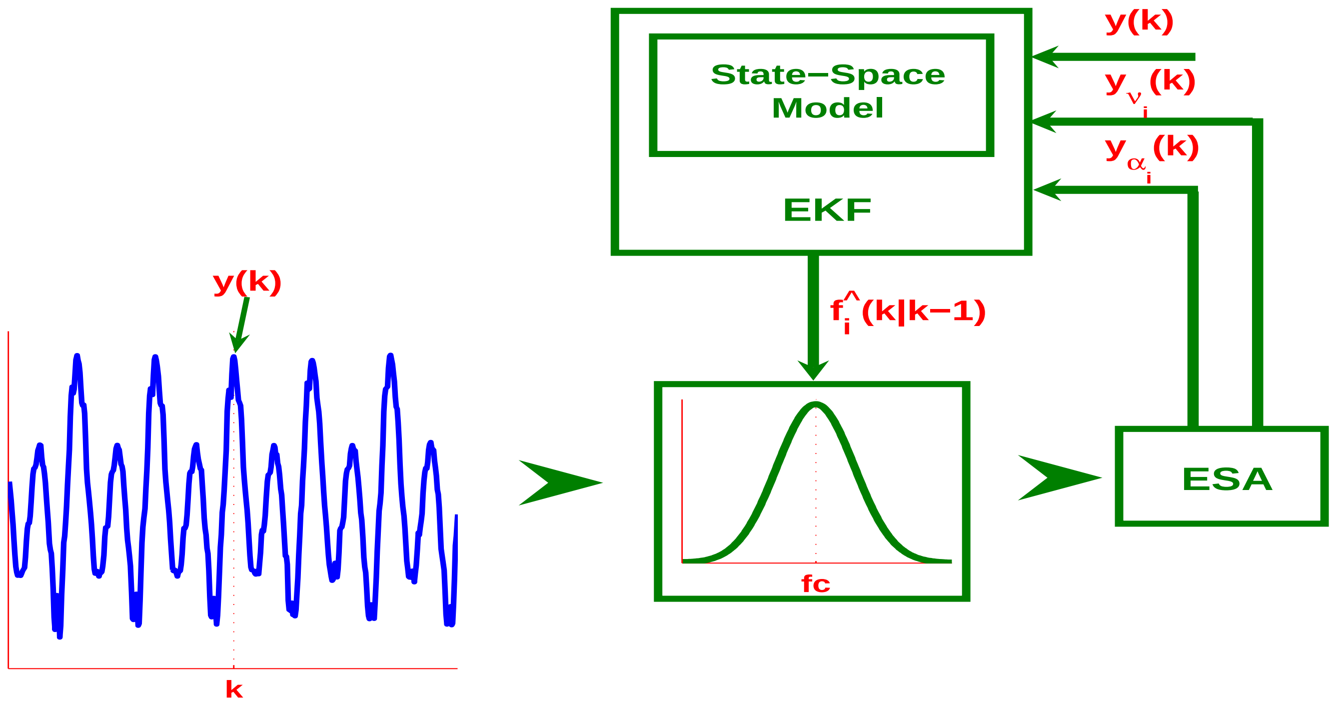
$$y(k) = \sum_{i=1}^{K} \alpha_i(k)\cos(\varphi_i(k)) + r u(k)$$

Local Regularization of EKF estimates

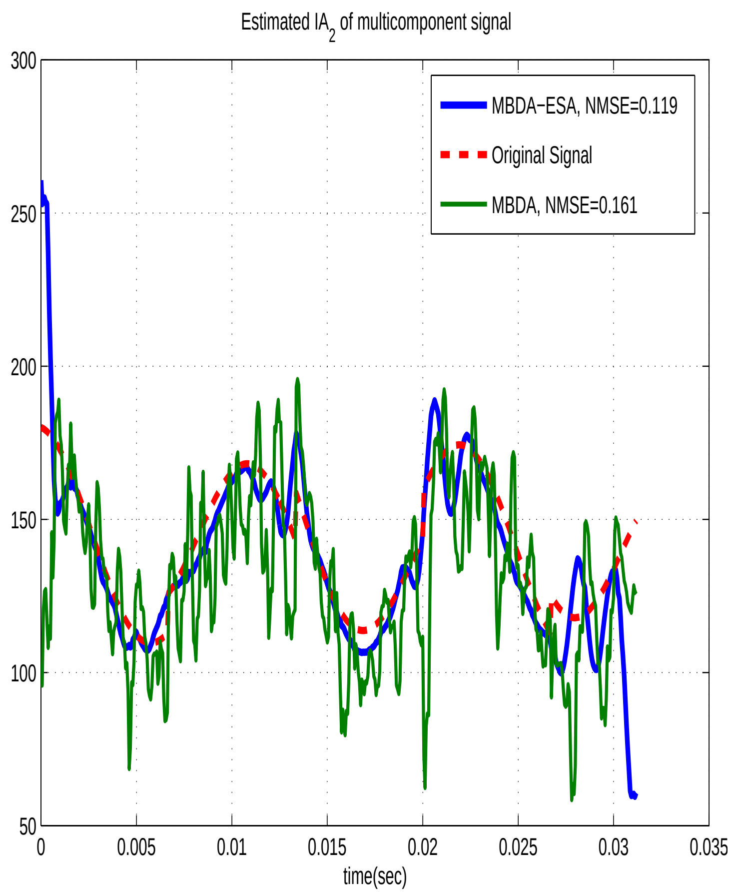
- MBDA is sensitive to initialization, spectral variations, outliers

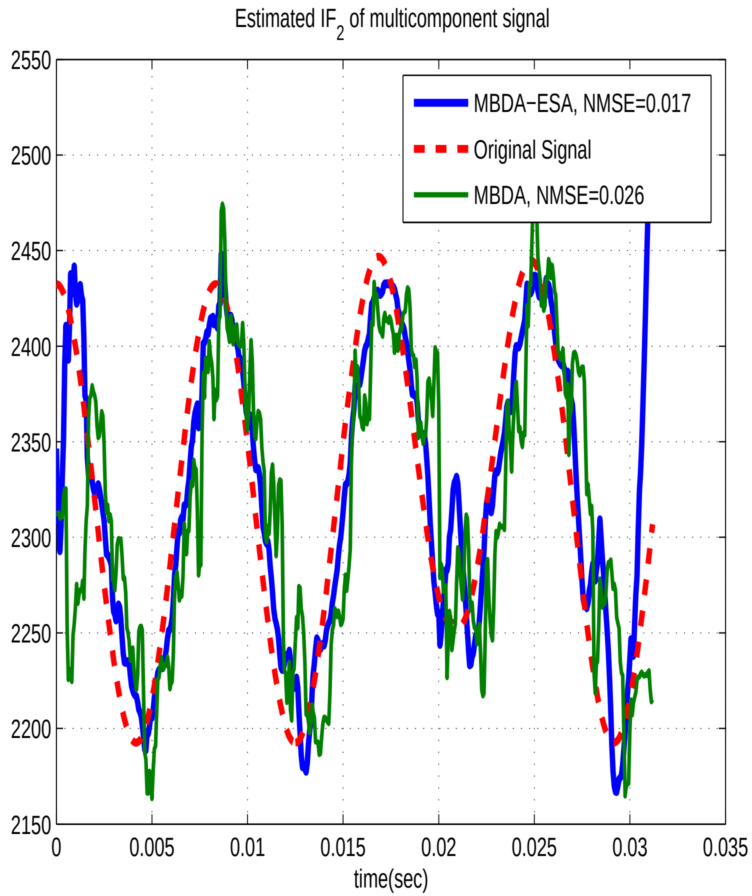
Additional observation equations:

$$y_{\alpha_i}(k) = |\alpha_i(k)| + r_{\alpha_i} u_{\alpha_i}(k)$$
$$y_{\nu_i}(k) = \nu_i(k) + f_i(k) + r_{\nu_i} u_{\nu_i}(k)$$



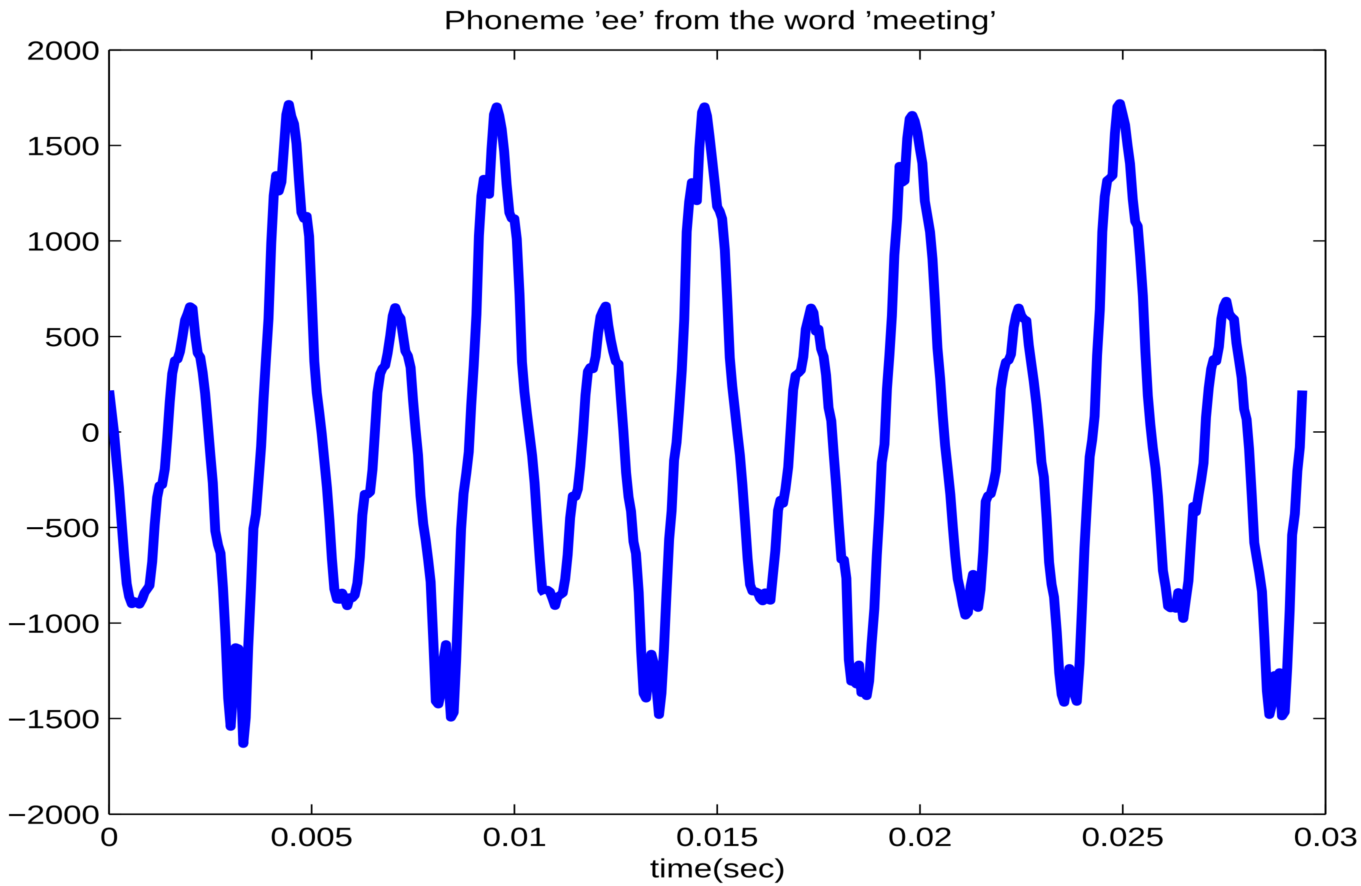
Experimental Results, Synthetic





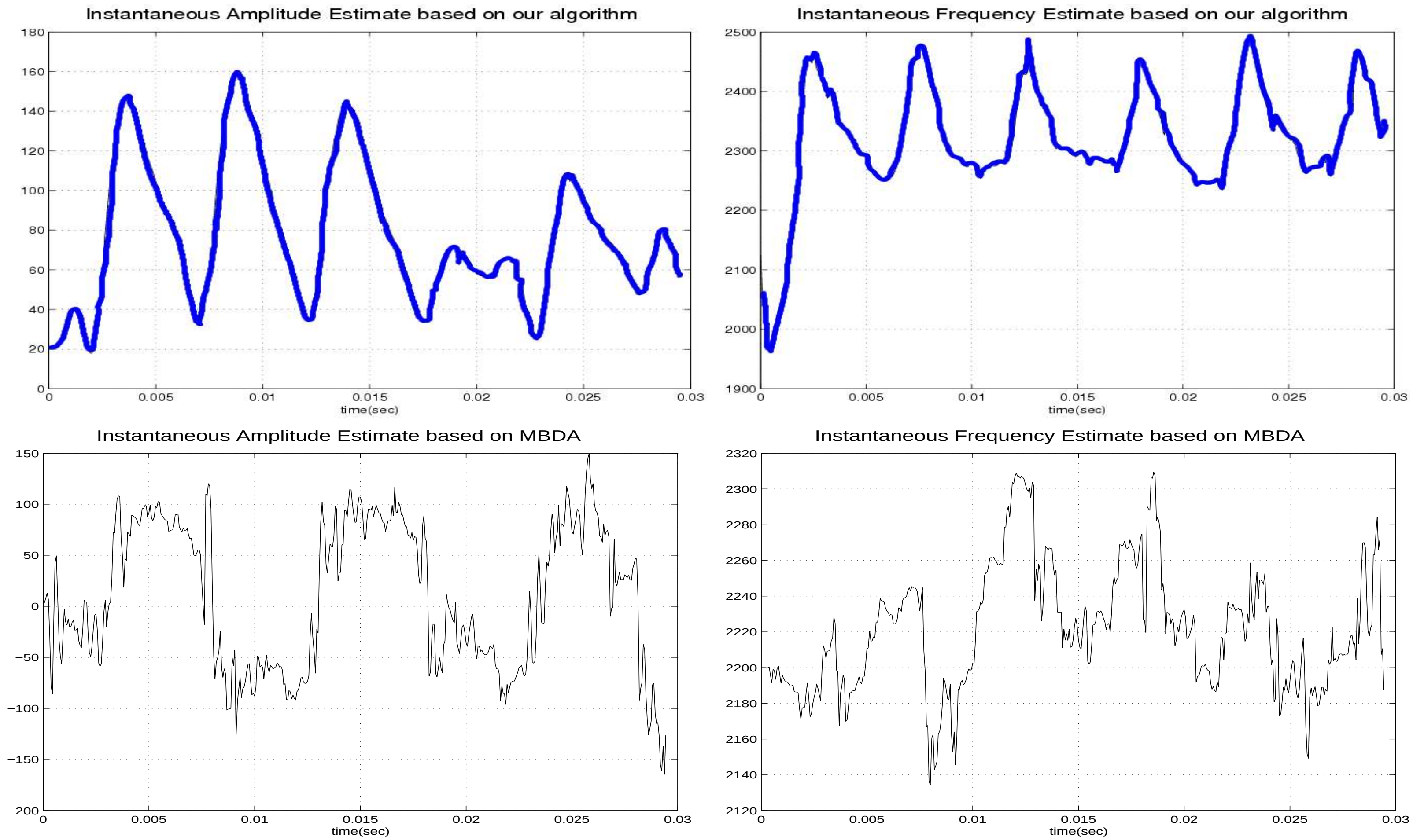
- 3-component synthetic signal, sum of AM-FM cosines

Experimental Results, Speech



- 4 speech resonances assumed in the model

Experimental Results, Speech



- IF_2: Instantaneous Frequency of the second speech resonance

Time Varying Model Parameters

- Spectral analysis of the MBDA state-space model
  - $q_{\alpha_i} \leftrightarrows$ Power of the Speech Resonance
  - $q_{\nu_i} \leftrightarrows$ Bandwidth of the Speech Resonance
- Nonstationarity of speech $\Rightarrow q_{\alpha_i}$ and q_{ν_i} are expected to vary
- Time varying $\lambda_{\alpha_i} = \log q^2_{\alpha_i}$ and $\lambda_{\nu_i} = \log q^2_{\nu_i}$

$$p(\lambda_{\alpha_i}(0)) = N(2\log q^0_{\alpha_i}, \delta^2_{\lambda^0_{\alpha_i}})$$

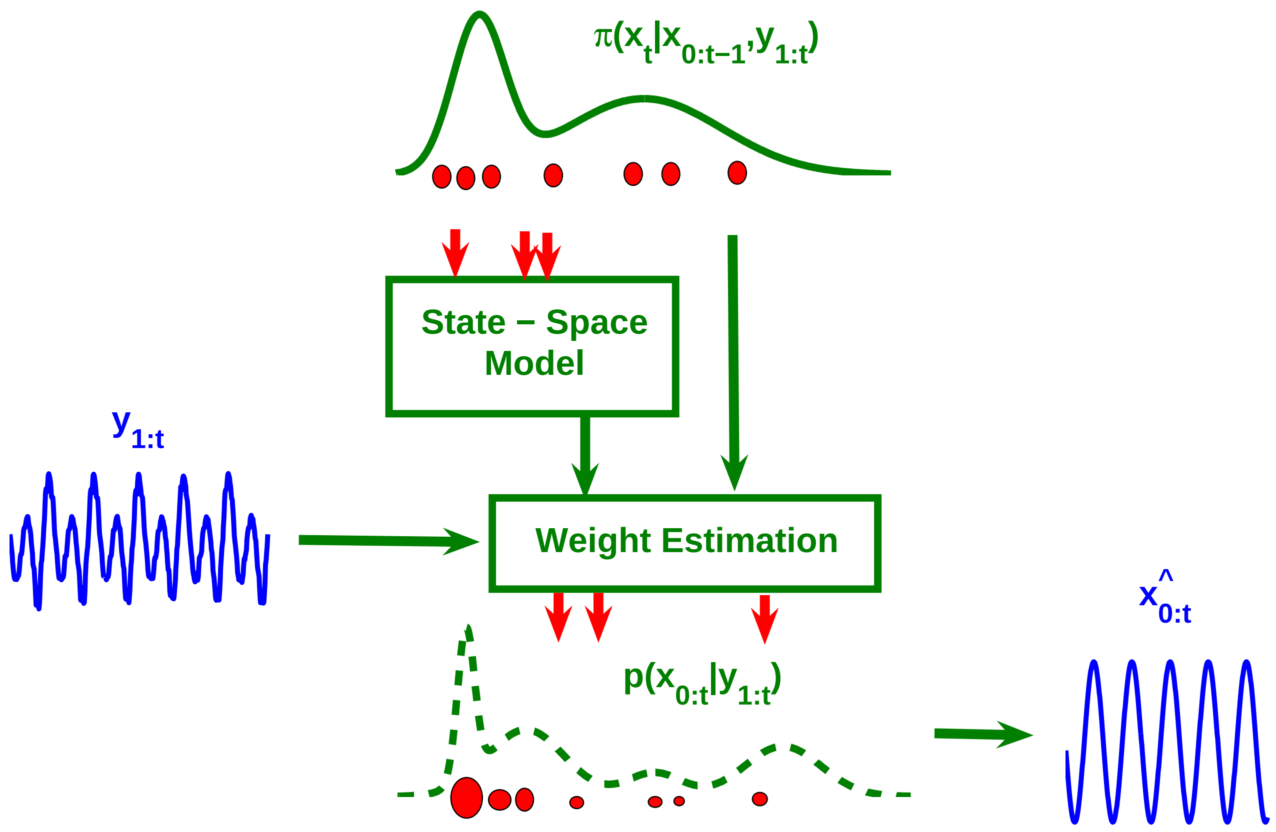
$$p(\lambda_{\alpha_i}(k)|\lambda_{\alpha_i}(k-1)) = N(\lambda_{\alpha_i}(k-1), \delta^2_{\lambda_{\alpha_i}})$$

$$p(\lambda_{\nu_i}(0)) = N(2\log q^0_{\nu_i}, \delta^2_{\lambda^0_{\nu_i}})$$

$$p(\lambda_{\nu_i}(k)|\lambda_{\nu_i}(k-1)) = N(\lambda_{\nu_i}(k-1), \delta^2_{\lambda_{\nu_i}})s$$

Particle Filtering

- Estimation and tracking become possible even when the assumed model is highly nonlinear or non-Gaussian



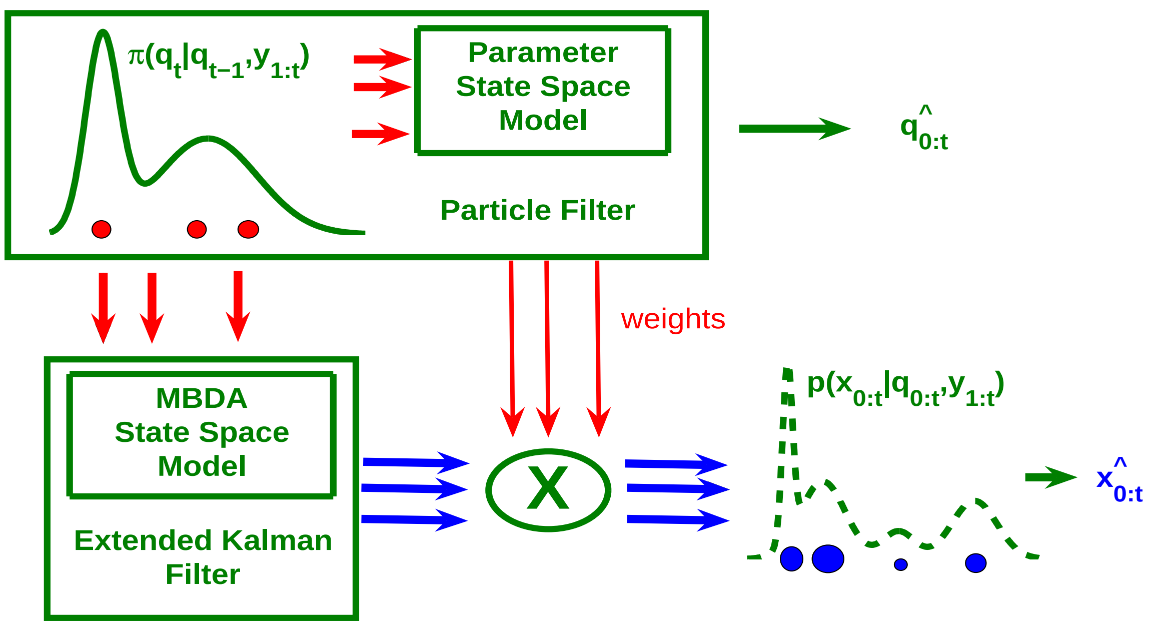
Extended Kalman and Particle Filtering

Conditional on the parameter vector $\mathbf{q}$ the state space equations are as in MBDA

$$\mathbf{x} = (\alpha_i, \nu_i, f_i | i = 1 \ldots K)$$

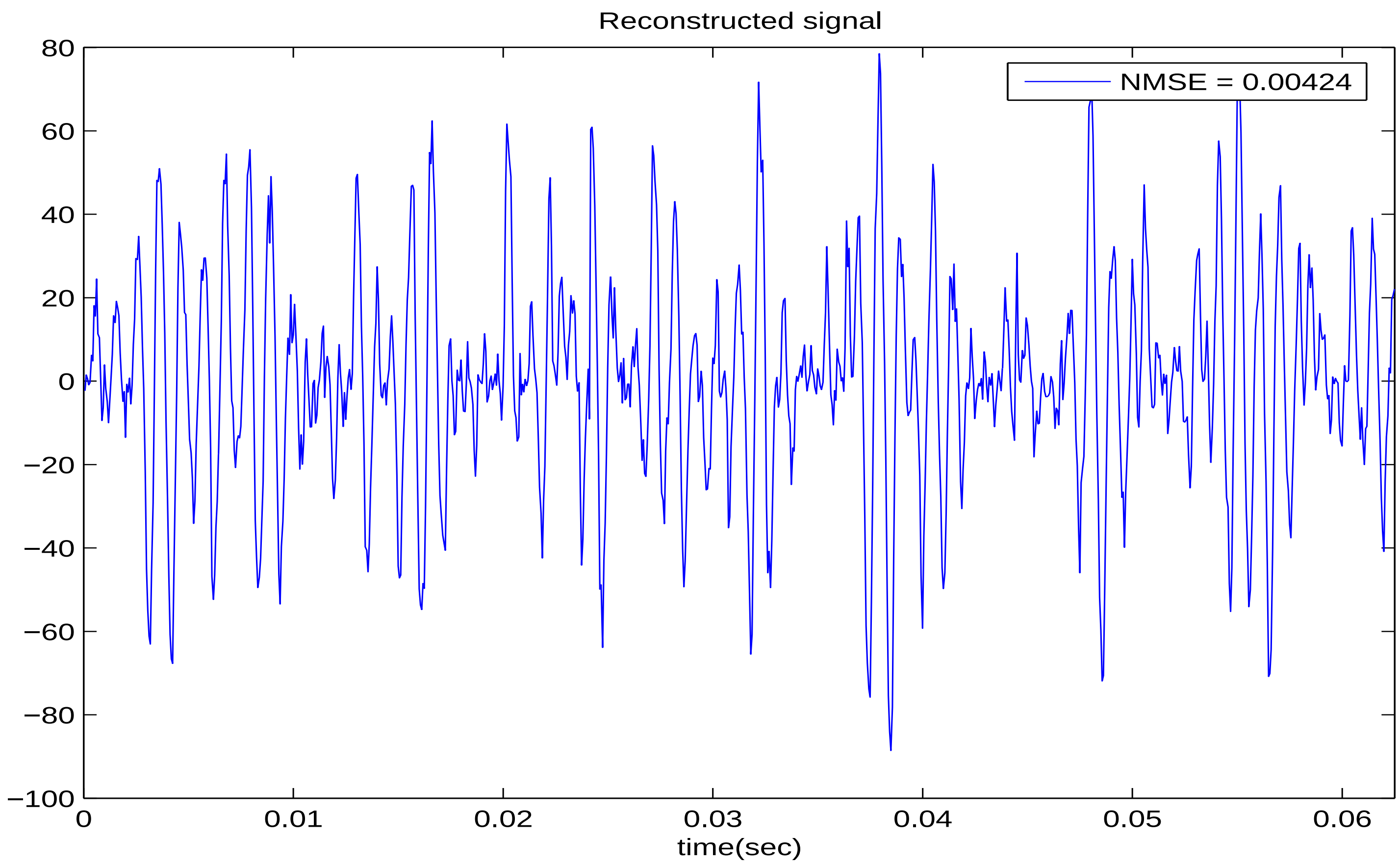
$$\mathbf{q} = (q_{\alpha_i}, q_{\nu_i} | i = 1 \ldots K)$$

$$p(\mathbf{x}_k, \mathbf{q}_{0:k} | y_{1:k}) = p(\mathbf{x}_k | \mathbf{q}_{0:k}, y_{1:k}) p(\mathbf{q}_{0:k} | y_{1:k})$$

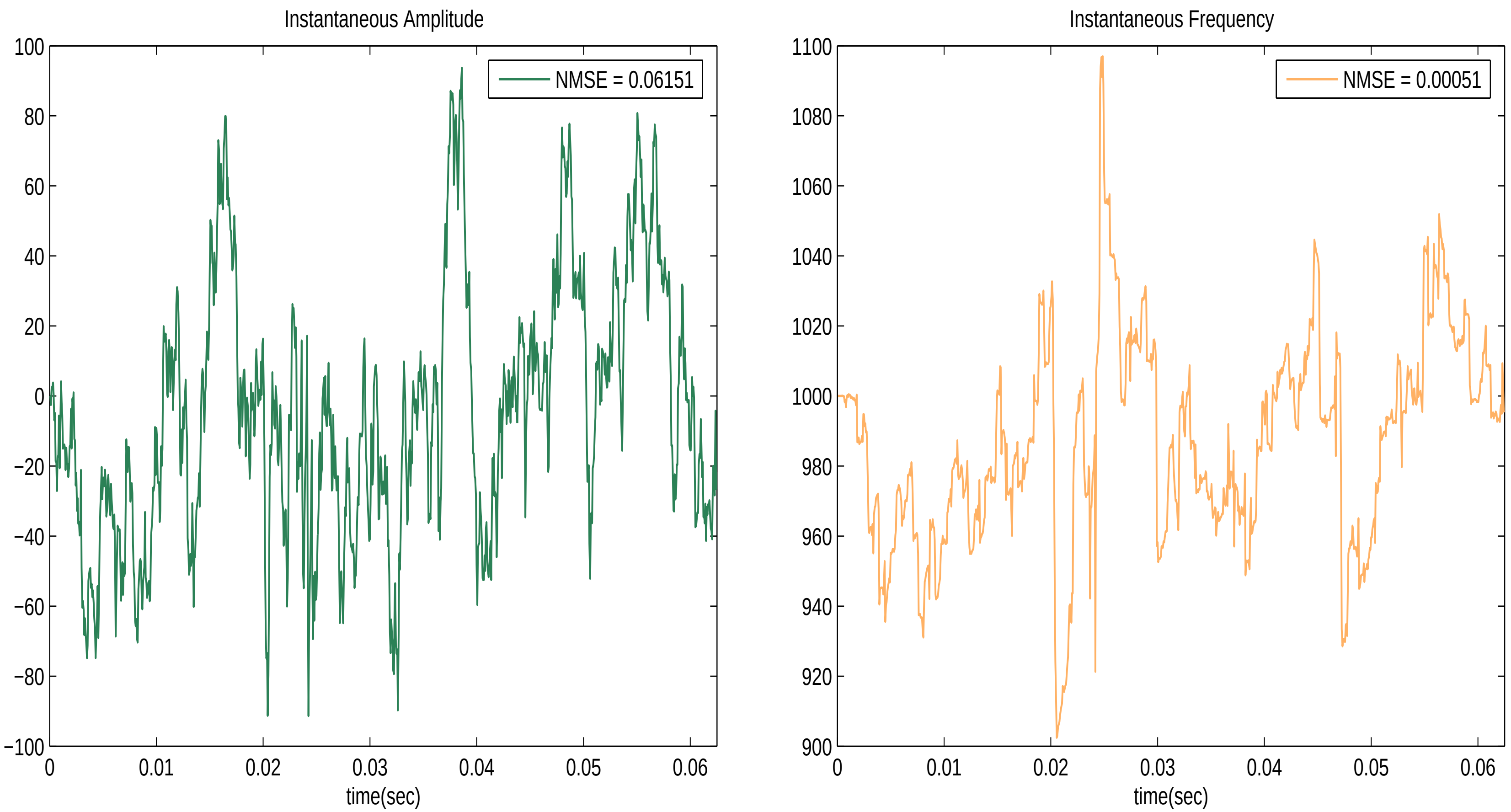


Experimental Results, Synthetic

- Synthetic signal randomly generated using the modified model

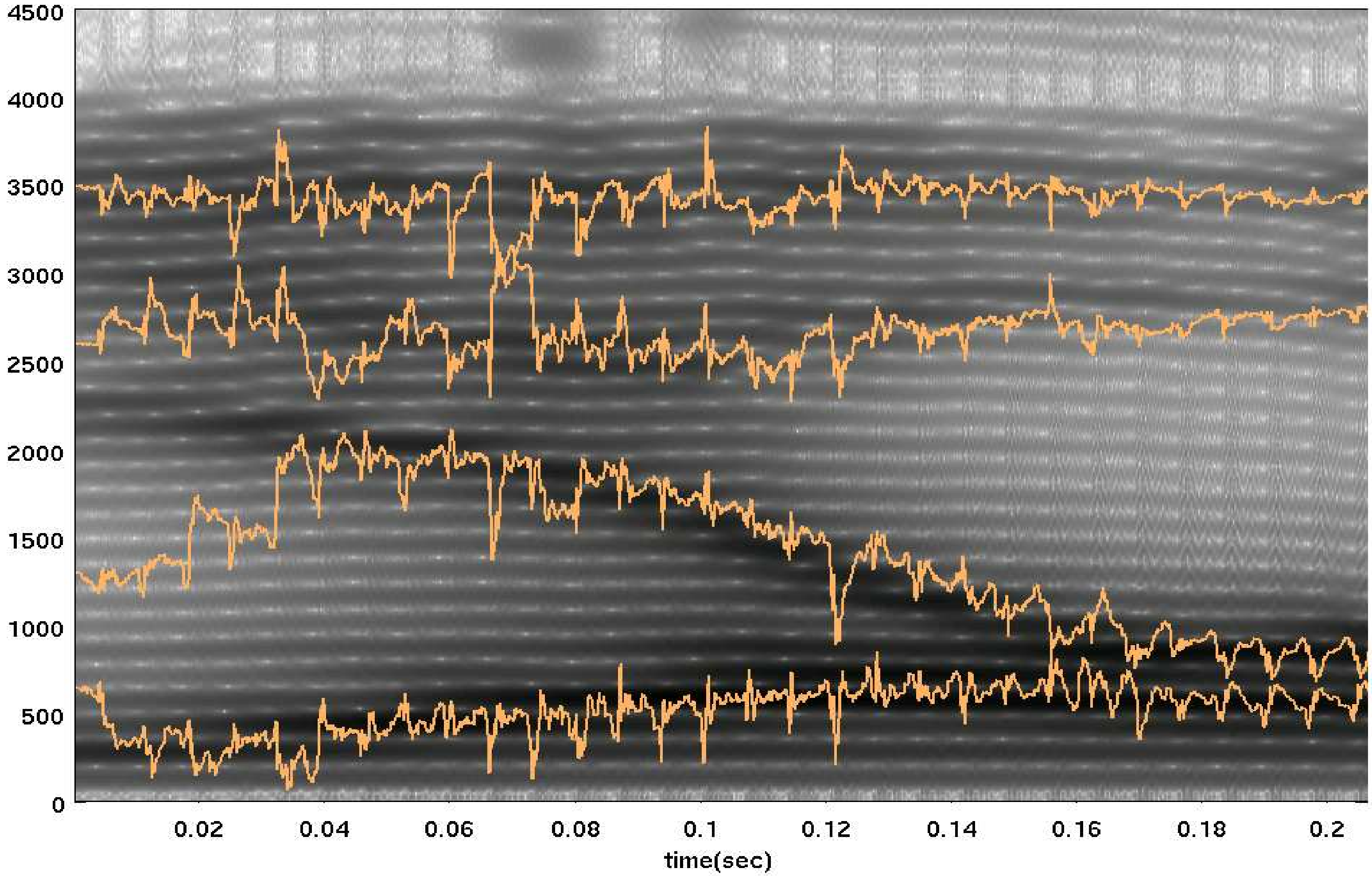


Experimental Results, Synthetic



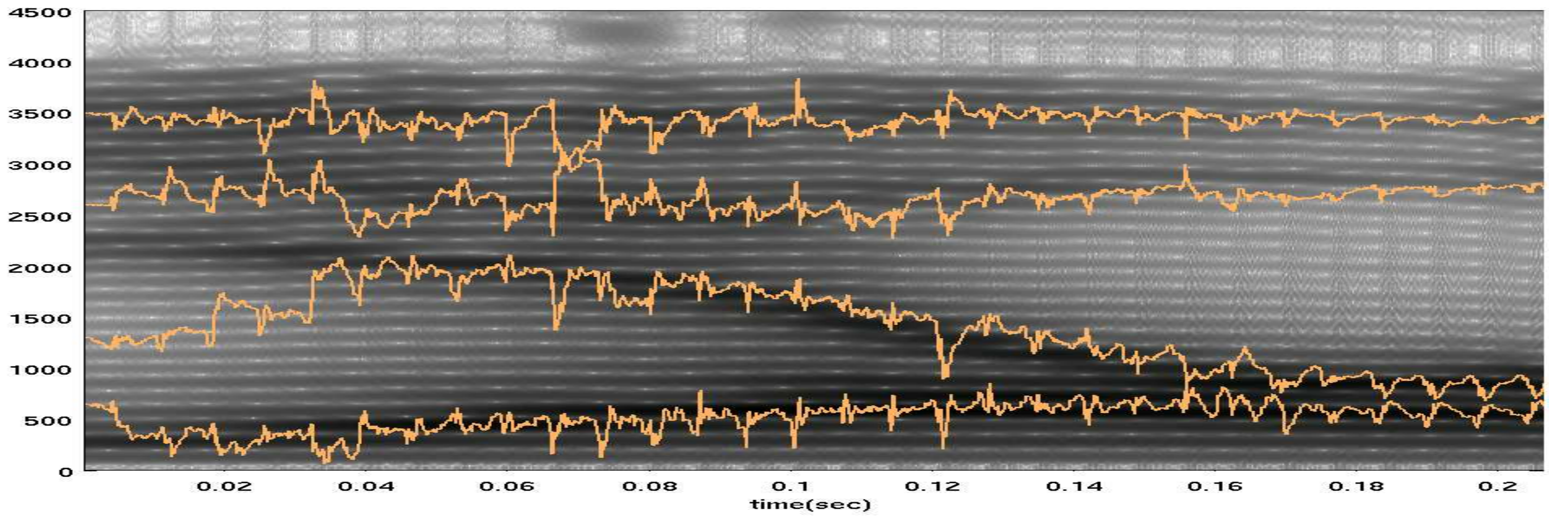
Experimental Results, Speech

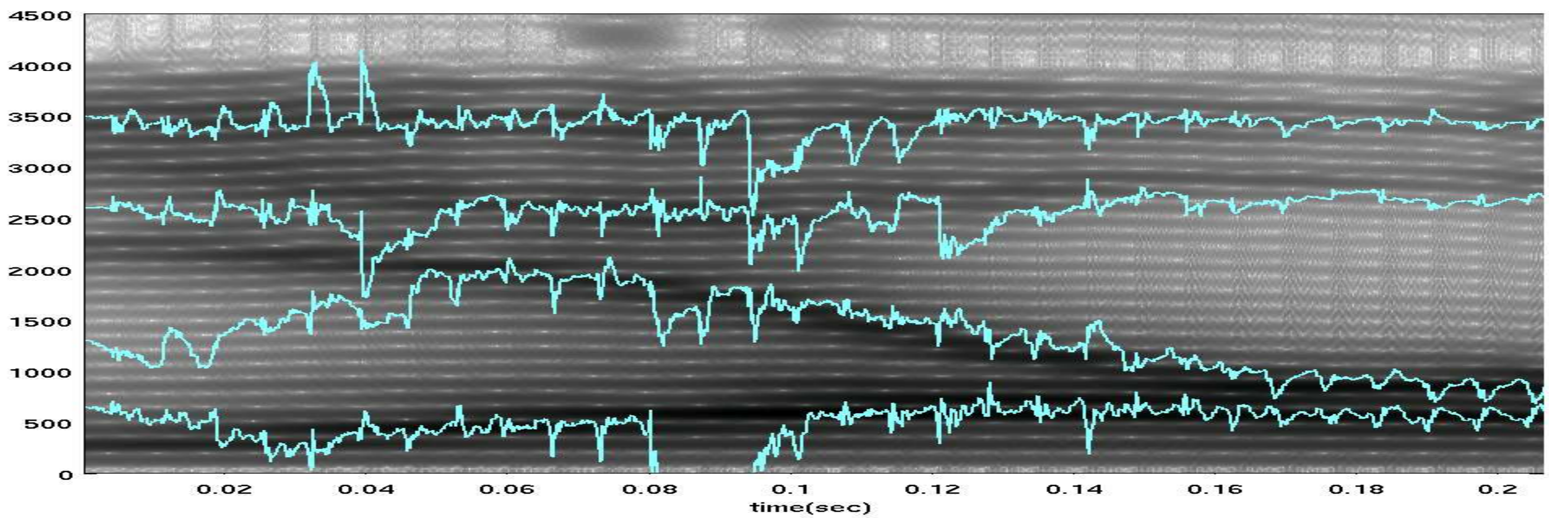
- Instantaneous Frequency, proposed approach (word “yell")



Experimental Results, Speech

- Instantaneous Frequency, proposed approach vs MBDA





Contributions and Conclusions

- Tracking and local regularization of Instantaneous Amplitude and Frequency of Speech Components
  - Improved robustness against possibly wrong initial configuration or rapid variations of the estimated magnitudes
- Tracking Spectrally Varying Speech Components using particle filtering techniques
  - Allow time varying model parameters
  - Robust against wide spectral variations
- Future Work
  - Possible applications in Speech Recognition, Speech Enhancement, Speech Synthesis

For further information on our work:

http://cvsp.cs.ntua.gr

powered by LaTeX