Statistical Speech Analysis and Nonlinear Modeling

Nassos Katsamanis & Petros Maragos

National Technical University of Athens

School of Electrical & Computer Engineering

CVSP Group

COST277 Meeting,

Limerick, 16th April 2003

- *The problem*:

  Speech Analysis based on the AM-FM model

- *The method*:

  Statistical Multiband Demodulation using Kalman Filters and Energy Separation

References:

- AFM Speech Model and ESA,
  - *Maragos, Kaiser, Quatieri*
    IEEE Trans. Audio Speech 1993
- Statistical Multiband Demodulation,
  - *Lu, Doerschuk*
    IEEE Trans. Signal Proc. 1996
  - *Pai, Doerschuk*
    IEEE Trans. Signal Proc. 2000

- Introduction
- **Contents**
- Algorithm Description
  - ✓ Kalman-MBDA
  - ✓ Kalman-ESA
  - ✓ IF & IA Estimation
  - ✓ Basic Variation (bKES)
  - ✓ Modified Meas. Equation (mKES)
  - ✓ Comparison
- Application to Speech
  - ✓ Speech Analysis
  - ✓ Formant Tracking
- Conclusions – Further Research

Contents

- Description of the Algorithm
  - ✓ Kalman-MBDA
  - ✓ The Kalman-ESA model
  - ✓ Instantaneous Frequency and Amplitude Estimation
  - ✓ Basic Variation
  - ✓ Modified Measurement Equation
  - ✓ Comparison
- Application to Speech
  - ✓ Speech Analysis
  - ✓ Formant Tracking
- Conclusions – Further Research

- Introduction
- Contents
- **Algorithm Description**
  - ✓ Kalman-MBDA
  - ✓ Kalman-ESA
  - ✓ IF & IA Estimation
  - ✓ Basic Variation (bKES)
  - ✓ Modified Meas. Equation (mKES)
  - ✓ Comparison
- Application to Speech
  - ✓ Speech Analysis
  - ✓ Formant Tracking
- Conclusions – Further Research

Based on the AM-FM model:

$$y_i(t) = a_i(t)\cos(\underbrace{2\pi f_i t + 2\pi f_{m,i}\int_0^t q_i(\tau)d\tau + \theta_i}_{\phi_i(t)})$$

For a longer time interval:

$$\phi_i(t) = \theta_i + 2\pi\int_0^t [\, f_i(\tau) + v_i(\tau)]d\tau$$

The speech signal (multicomponent):

$$s(t) = \sum_{i=1}^{K} a_i(t)\cos\phi_i(t)$$

Dynamic System, for each component:

$$f_i(n) = f_i(n-1) + q_{f_i} w_{f_i}(n-1)$$

$$a_i(n) = \alpha_{a_i} a_i(n-1) + q_{a_i} w_{a_i}(n-1)$$

$$v_i(n) = \alpha_{v_i} v_i(n-1) + q_{v_i} w_{v_i}(n-1)$$

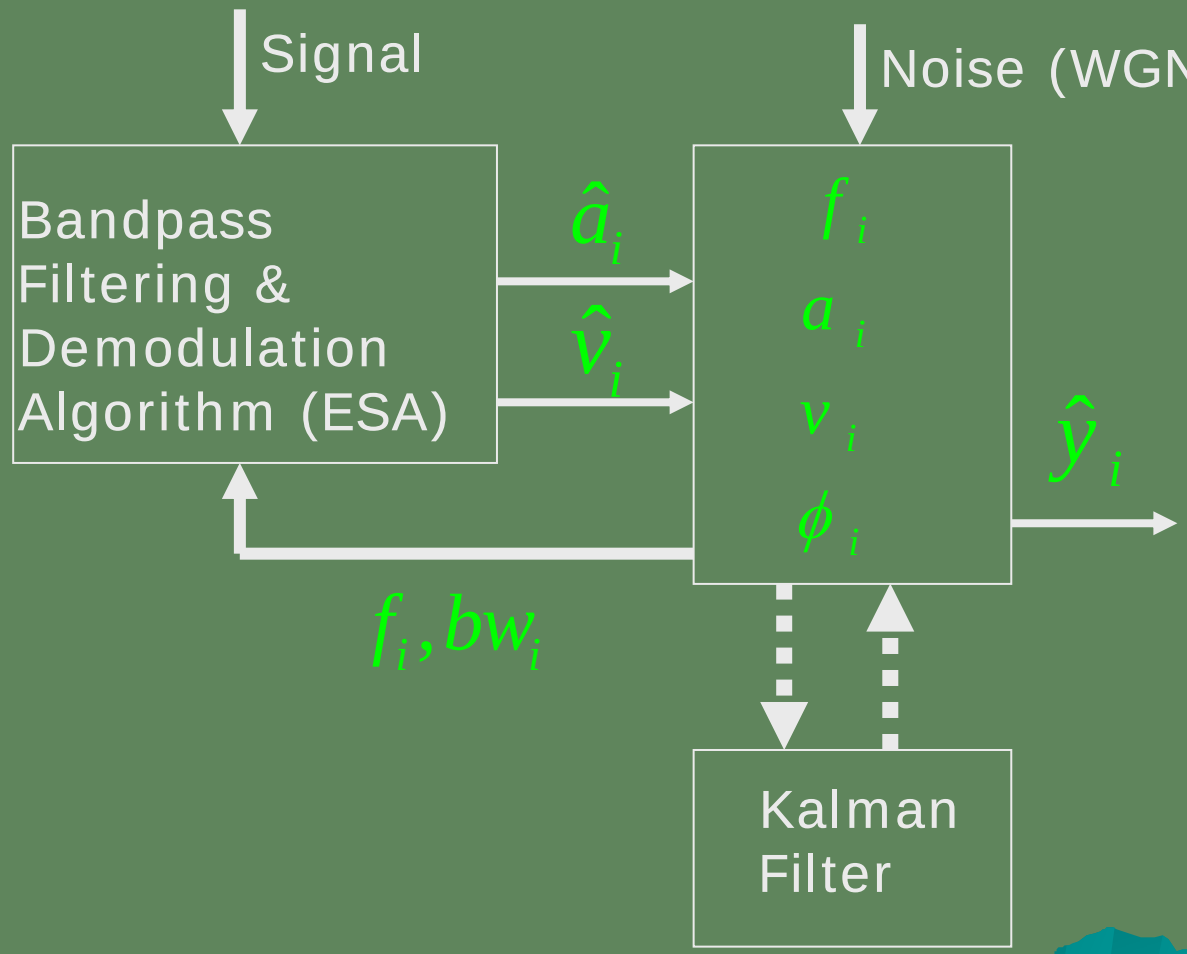
$$\phi_i(n) = \phi_i(n-1) + 2\pi T_s (f_i(n-1) + v_i(n-1))$$

The measurement equation:

$$y(n) = \sum_{i=1}^{K} a_i(n) \cos(\phi_i(n)) + ru(n)$$

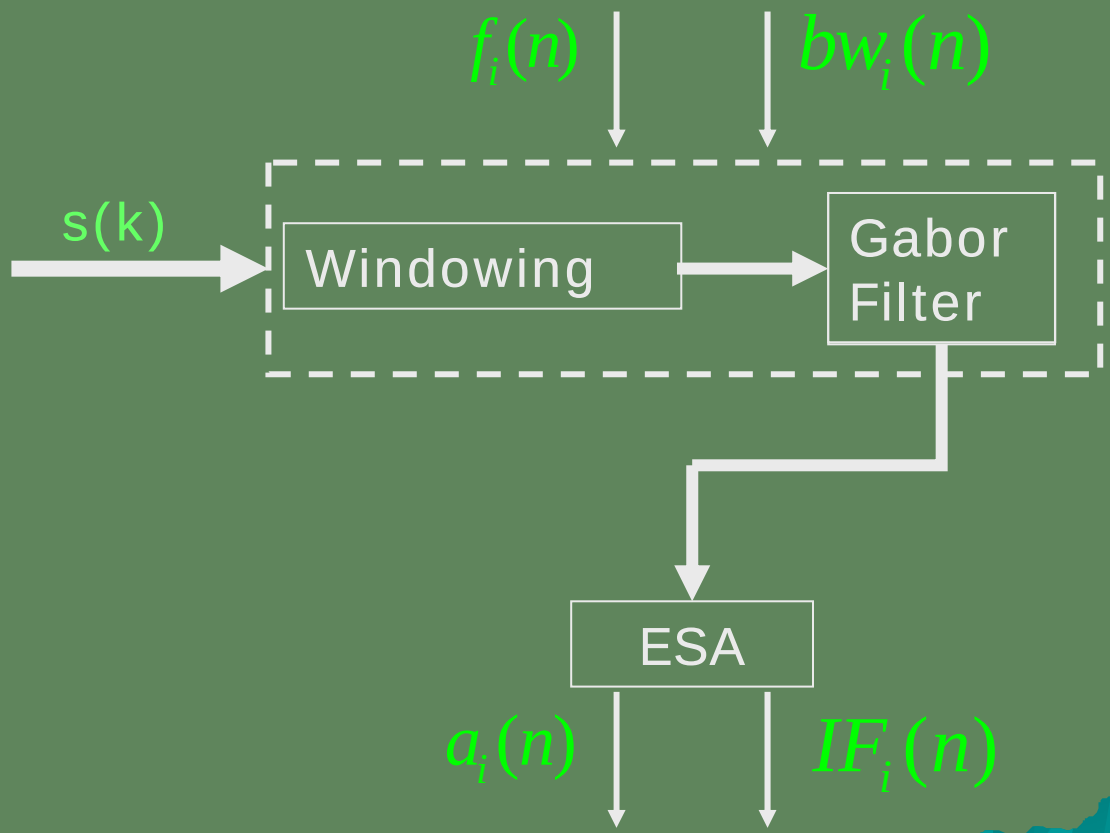
- Introduction
- Contents
- Algorithm Description
  - ✓ Kalman-MBDA
  - ✓ **Kalman-ESA**
  - ✓ IF & IA Estimation
  - ✓ Basic Variation (bKES)
  - ✓ Modified Meas. Equation (mKES)
  - ✓ Comparison
- Application to Speech
  - ✓ Speech Analysis
  - ✓ Formant Tracking
- Conclusions – Further Research

For each component:



Instantaneous Frequency & Amplitude Estimation – (IFAE)

s(k) : multicomponent AFM signal



- Introduction
- Contents
- Algorithm Description
  - ✓ Kalman-MBDA
  - ✓ Kalman-ESA
  - ✓ IF & IA Estimation
  - ✓ **Basic Variation (bKES)**
  - ✓ Modified Meas. Equation (mKES)
  - ✓ Comparison
- Application to Speech
  - ✓ Speech Analysis
  - ✓ Formant Tracking
- Conclusions – Further Research

State Equations:

$$f_i(n) = f_i(n-1) + q_{f_i} w_{f_i}(n-1)$$

$$\begin{pmatrix} a_i(n) \\ IF_i(n) \end{pmatrix} = IFAE(s, f_i(n-1), bw_i, n) + Q^{IFAE} w_{IFAE}(n)$$

$$\phi_i(n) = \phi_i(n-1) + 2\pi T_s IF_i(n-1)$$

Measurement Equation:

$$y(n) = \sum_{i=1}^{K} a_i(n) \cos(\phi_i(n)) + ru(n)$$

State Equations:

$$f_i(n) = f_i(n-1) + q_{f_i} w_{f_i}(n-1)$$

$$a_i(n) = \alpha_{a_i} a_i(n-1) + q_{a_i} w_{a_i}(n-1)$$

$$v_i(n) = \alpha_{v_i} v_i(n-1) + q_{v_i} w_{v_i}(n-1)$$

$$\phi_i(n) = \phi_i(n-1) + 2\pi T_s (f_i(n-1) + v_i(n-1))$$

Measurement Equations:

$$y(n) = \sum_{i=1}^{K} a_i(n)\cos(\phi_i(n)) + ru(n)$$

$$y_{i,1}(n) = f_i(n) + v_i(n) + r_{i,1}u_{i,1}(n)$$

$$y_{i,2}(n) = |a_i(n)| + r_{i,2}u_{i,2}(n)$$

- Introduction
- Contents
- Algorithm Description
  - ✓ Kalman-MBDA
  - ✓ Kalman-ESA
  - ✓ IF & IA Estimation
  - ✓ Basic Variation (bKES)
  - ✓ Modified Meas. Equation (mKES)
  - ✓ **Comparison**
- Application to Speech
  - ✓ Speech Analysis
  - ✓ Formant Tracking
- Conclusions – Further Research

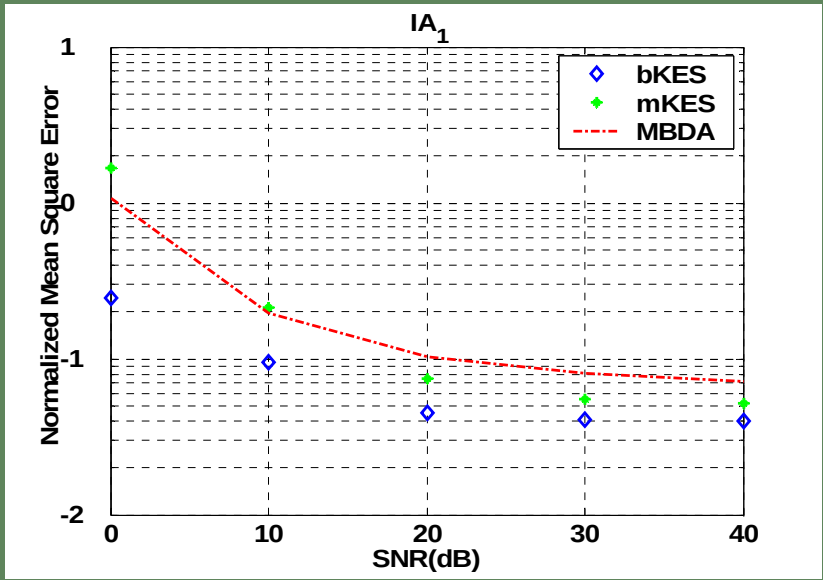
The signal:

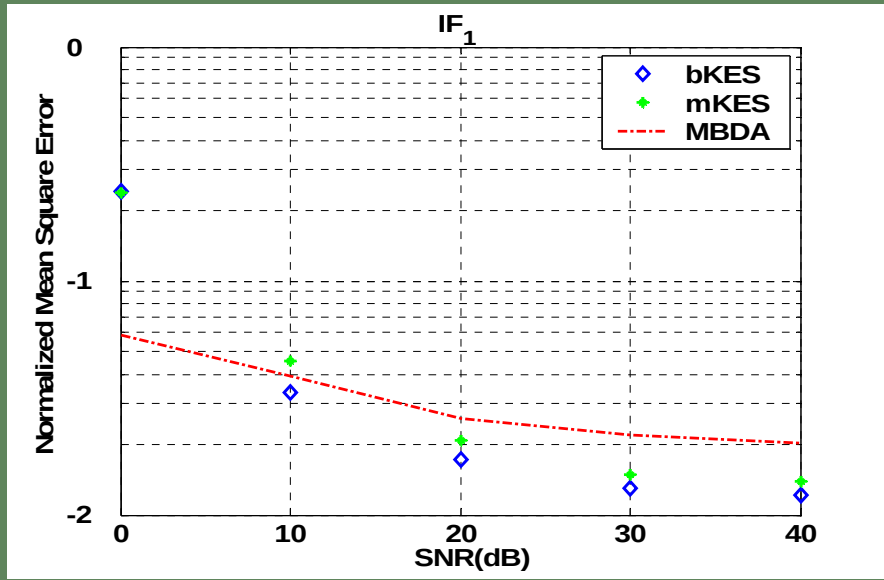
$$y(t) = \sum_{i=1}^{2} [1 - \kappa \cos(2\pi f_{AM} t)]$$

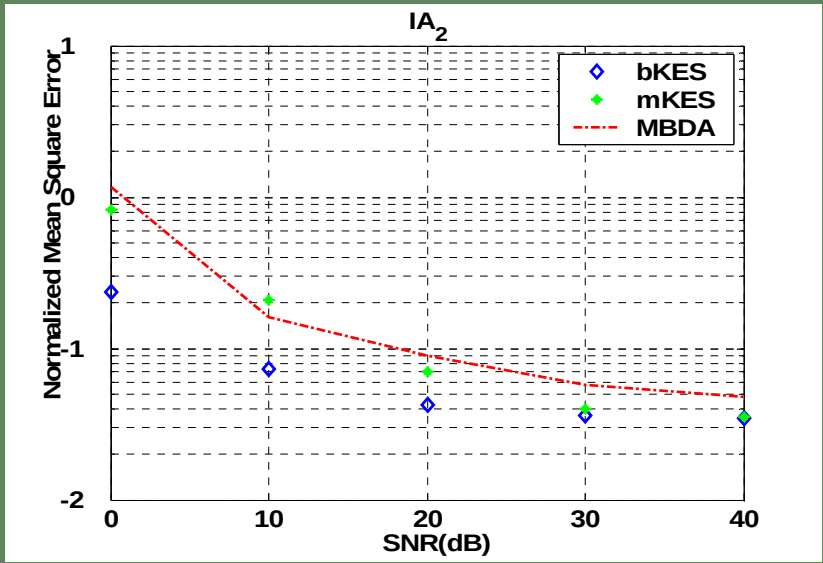
$$\cos\left(2\pi\left(f_i t + \beta \int_0^t \cos(2\pi f_{FM,i} \tau) d\tau\right)\right)$$

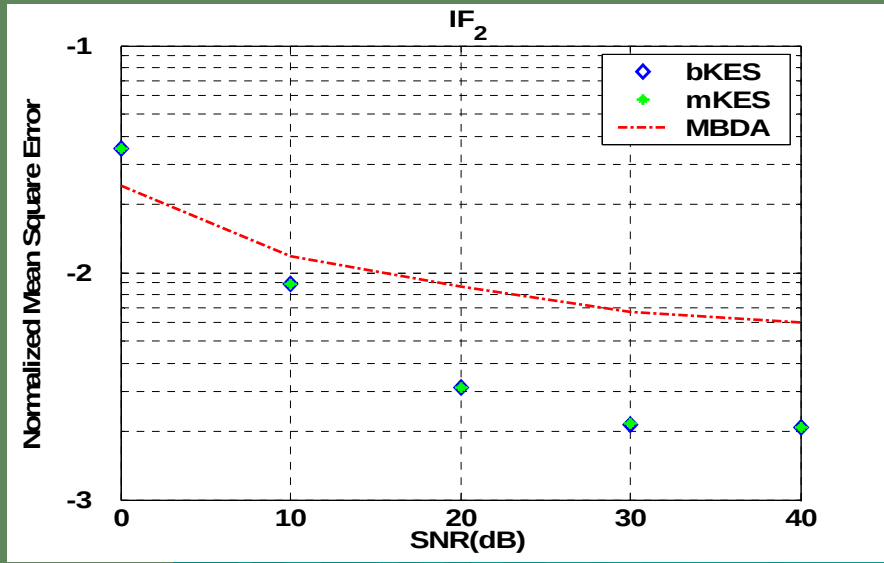
- Effect of Noise
- Effect of the AM index
- Effect of the FM index

Effect of Noise

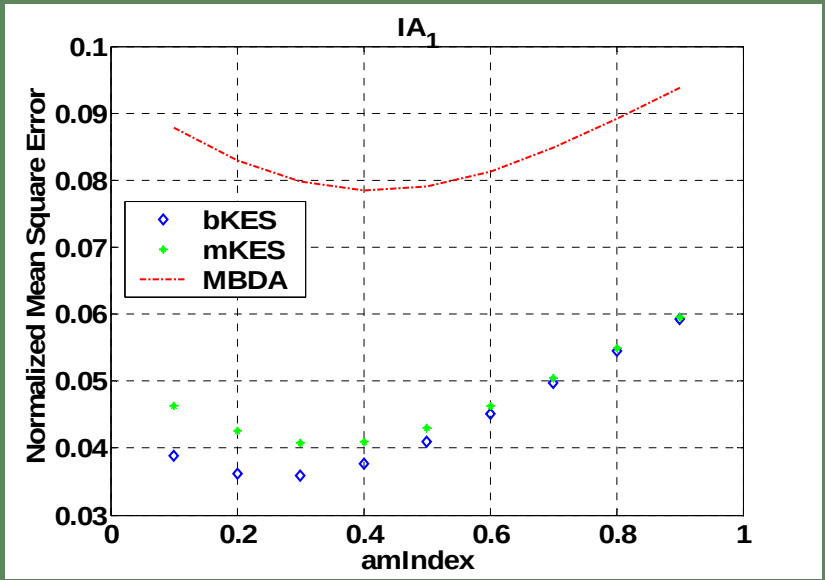


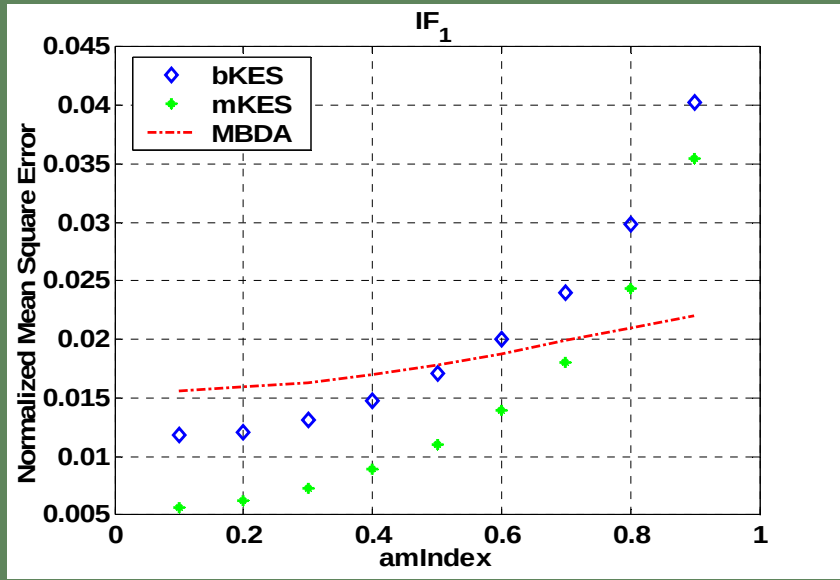


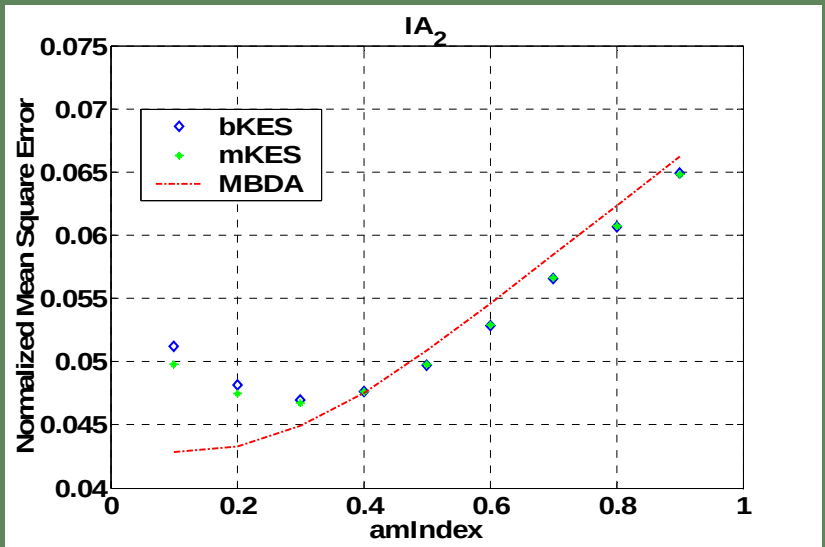


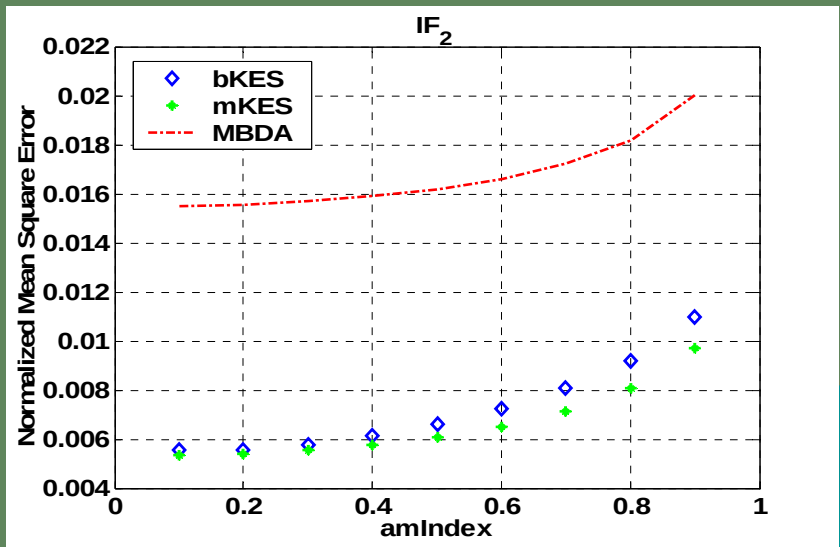


Effect of the AM index

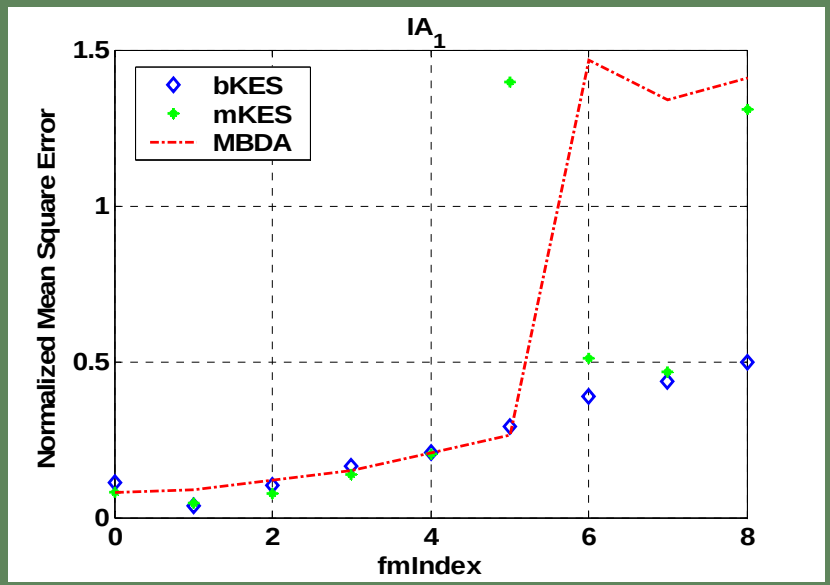


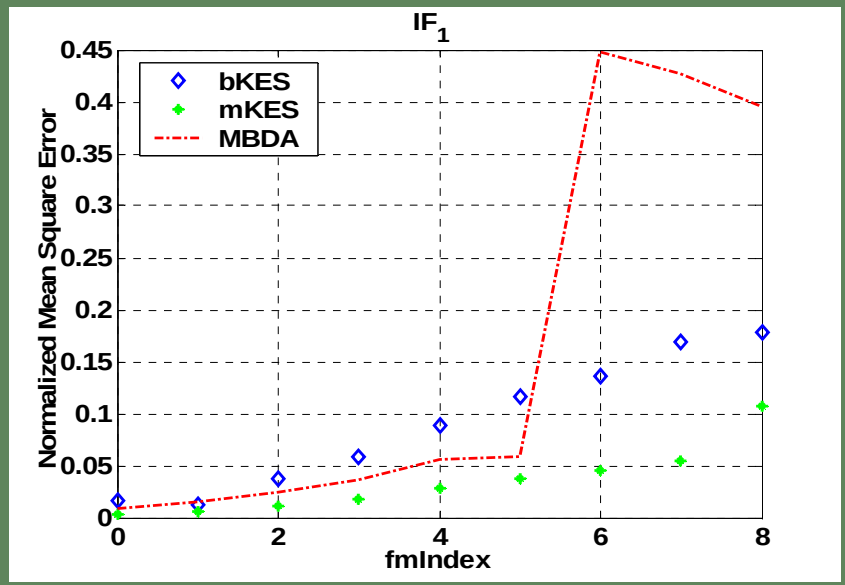


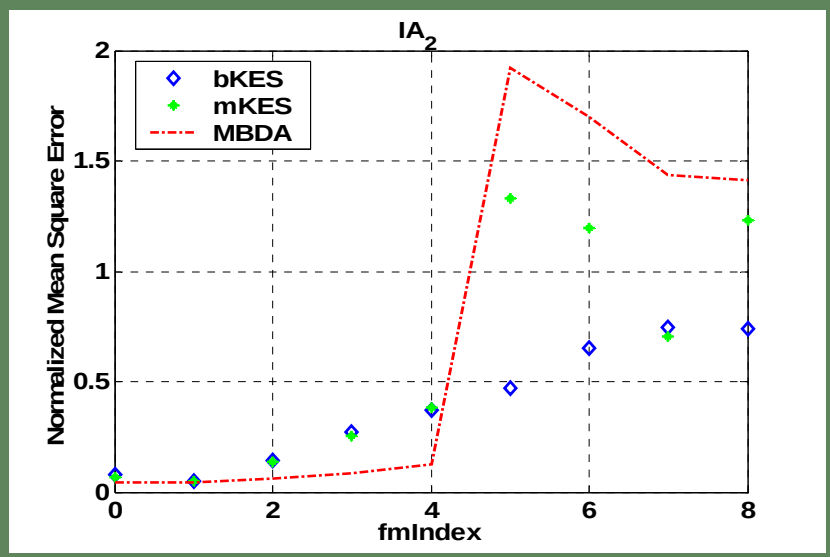


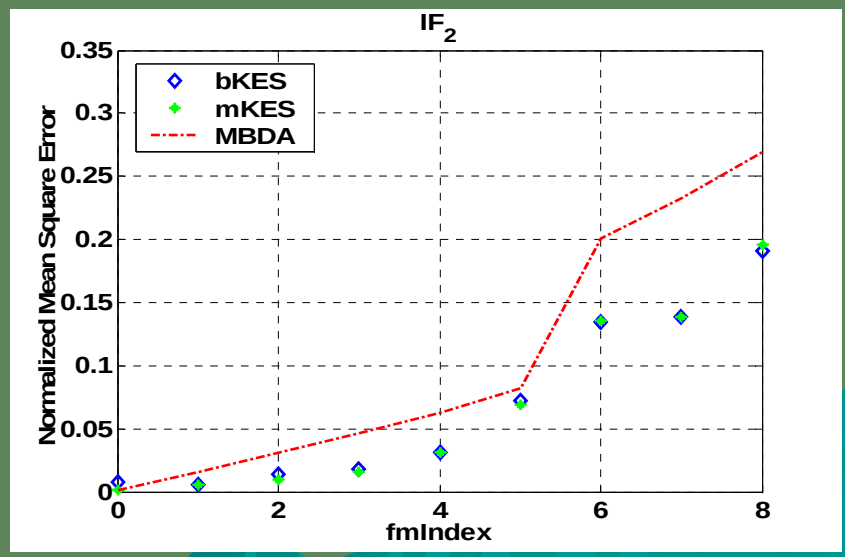


Effect of the FM index





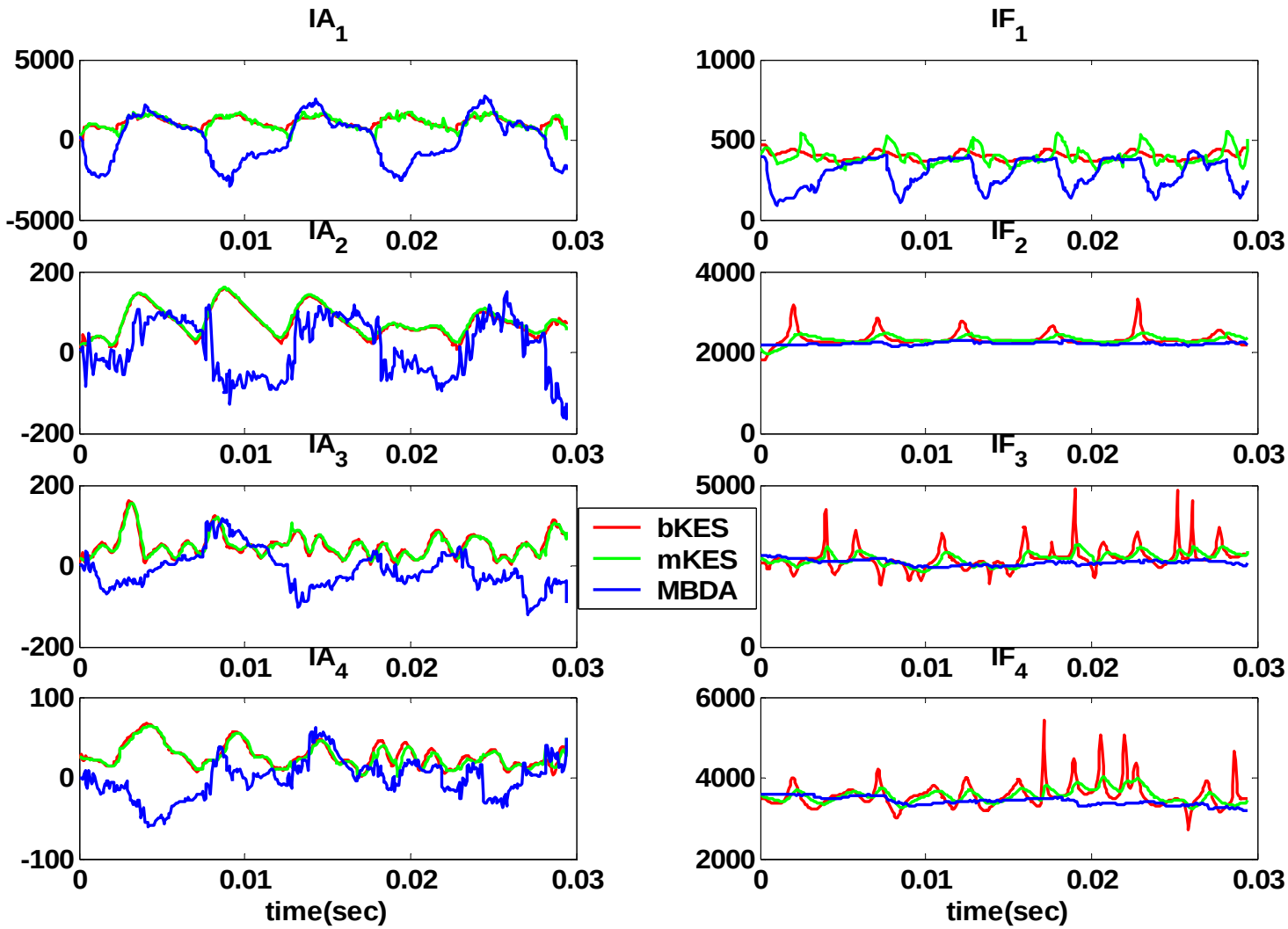




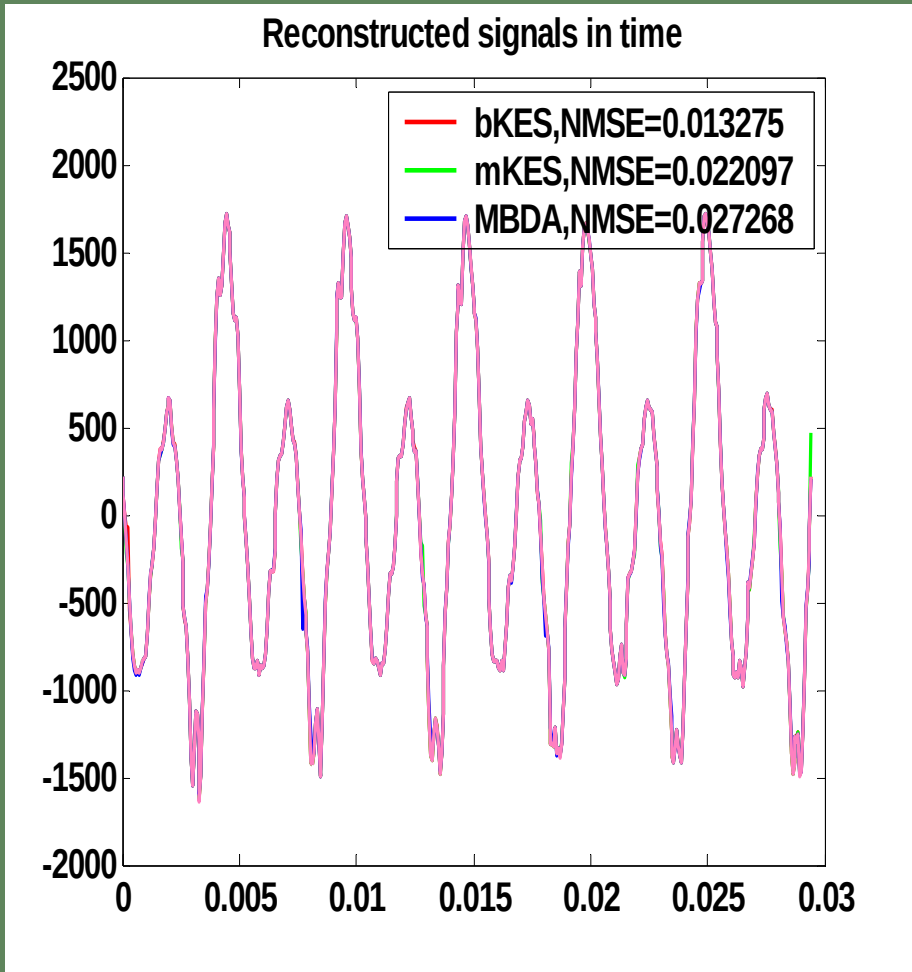
- Speech Analysis
  - Demodulation of the separate components
  - Phoneme /ee/ in the word m/ee/ting
  - Signal Reconstruction
- Formant Tracking
  - Voiced Sentence

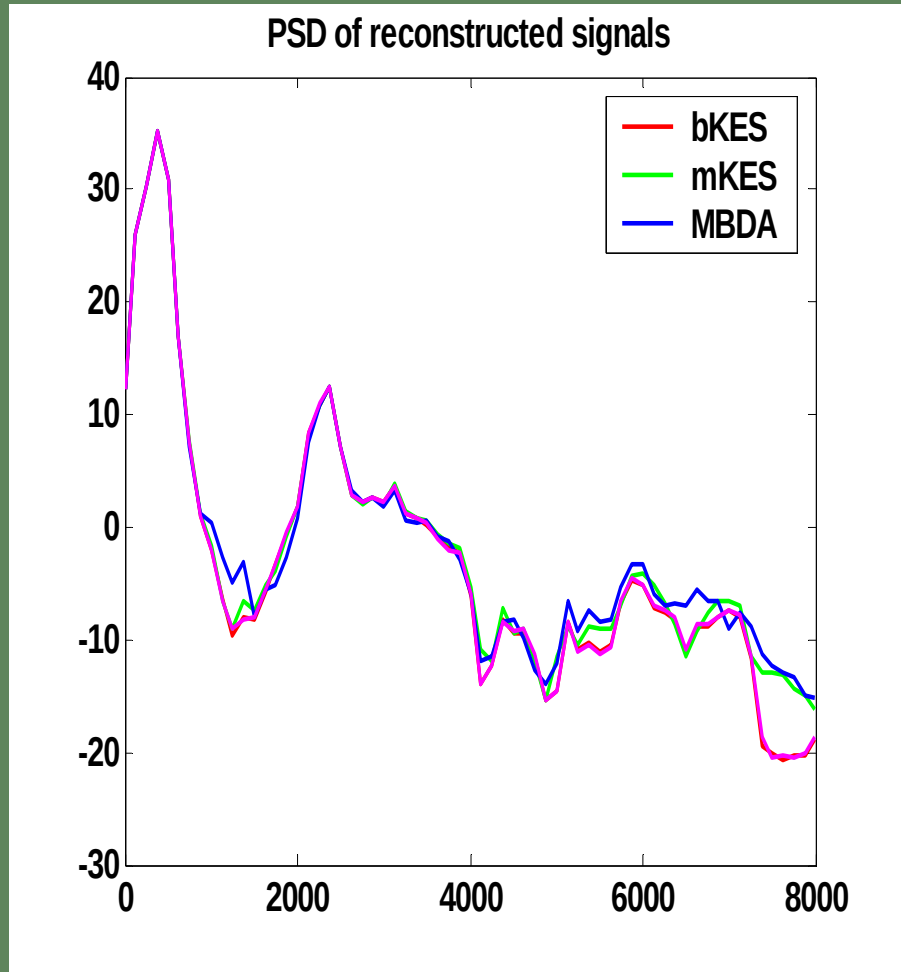
    “Where were you while we were away?”

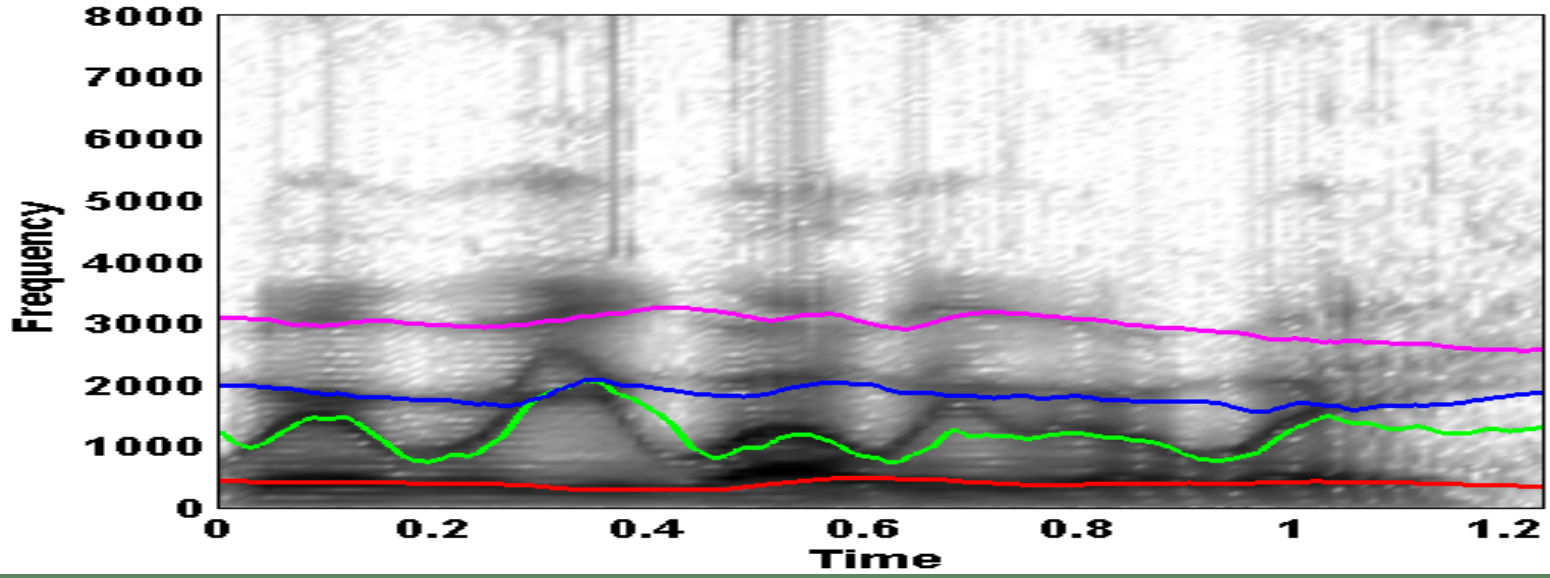
    -Signal Reconstruction

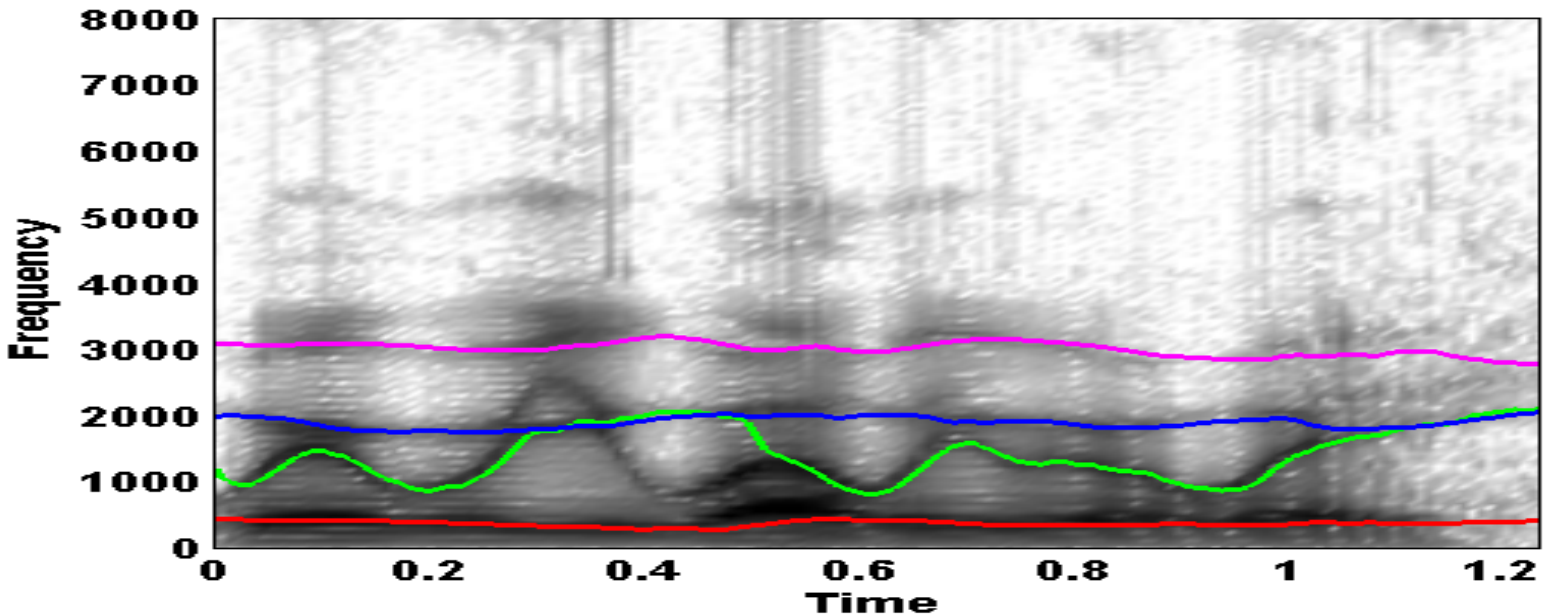

IA1
IF1
IA2
IF2
IA3
IF3
IA4
IF4
bKES
mKES
MBDA
time(sec)
time(sec)

Reconstructed Signal









Original Signal:

- Enhancement by means of energy estimation
  - Inclusion of noise
  - Multiband signal processing
  - Better Demodulation
  - Natural background
- Comparing with Kalman-MBDA

Further Research

- Application to Speech
  - Formant Evolution Modeling
  - Bandwidth Modeling
- Theoretical Research
  - System Identification
  - Better Estimation
- Complexity Reduction