CS578- Speech Signal Processing

Lecture 6: Adaptive Sinusoidal Models

Yannis Stylianou

University of Crete, Computer Science Dept., Multimedia Informatics Lab
yannis@csd.uoc.gr

Univ. of Crete

Outline

Sinusoidal Model

- Sinusoidal Model:

$$x(t) = \sum_{k=-K}^{K} \gamma_k e^{j2\pi f_k t}$$

- Special case, Harmonic Model:

$$x(t) = \sum_{k=-K}^{K} \gamma_k e^{j2\pi k f_0 t}$$

- Estimation of parameters (Linear approach):

$$\gamma = \mathcal{F}_{f_k} \mathbf{s}$$

- Model Evaluation, Mean-Squared Error (MSE):

$$\epsilon = \int_{w} |s(t) - x(t)|^2 \, dt$$

Sinusoidal Model

- Sinusoidal Model:

$$x(t) = \sum_{k=-K}^{K} \gamma_k e^{j2\pi f_k t}$$

- Special case, Harmonic Model:

$$x(t) = \sum_{k=-K}^{K} \gamma_k e^{j2\pi k f_0 t}$$

- Estimation of parameters (Linear approach):

$$\gamma = \mathcal{F}_{f_k} \mathbf{s}$$

- Model Evaluation, Mean-Squared Error (MSE):

$$\epsilon = \int_w |s(t) - x(t)|^2 \, dt$$

Sinusoidal Model

- Sinusoidal Model:

$$x(t) = \sum_{k=-K}^{K} \gamma_k e^{j2\pi f_k t}$$

- Special case, Harmonic Model:

$$x(t) = \sum_{k=-K}^{K} \gamma_k e^{j2\pi k f_0 t}$$

- Estimation of parameters (Linear approach):

$$\gamma = \mathcal{F}_{f_k} \mathbf{s}$$

- Model Evaluation, Mean-Squared Error (MSE):

$$\epsilon = \int_w |s(t) - x(t)|^2 \, dt$$

Sinusoidal Model

- Sinusoidal Model:

$$x(t) = \sum_{k=-K}^{K} \gamma_k e^{j2\pi f_k t}$$

- Special case, Harmonic Model:

$$x(t) = \sum_{k=-K}^{K} \gamma_k e^{j2\pi k f_0 t}$$

- Estimation of parameters (Linear approach):

$$\gamma = \mathcal{F}_{f_k} \mathbf{s}$$

- Model Evaluation, Mean-Squared Error (MSE):

$$\epsilon = \int_{w} |s(t) - x(t)|^2 \, dt$$

Outline

Towards Adaptive Sinusoidal Models

Estimating Sinusoidal parameters

- Sinusoidal representation for a speech/signal frame:

$$x(t) = \left(\sum_{k=-K}^{K} a_k e^{j2\pi f_k t} \right) w(t)$$

- Methods:
  - FFT-based methods (i.e., QIFFT [Abe et al., 2004-05, [1] [2]])
  - Subspace methods
  - Least Squares (LS) method
- Frequency mismatch:

$$\hat{f}_k = f_k + \eta_k$$

Estimating Sinusoidal parameters

- Sinusoidal representation for a speech/signal frame:

$$x(t) = \left(\sum_{k=-K}^{K} a_k e^{j2\pi f_k t} \right) w(t)$$

- Methods:
  - FFT-based methods (i.e., QIFFT [Abe et al., 2004-05, [1] [2]])
  - Subspace methods
  - Least Squares (LS) method
- Frequency mismatch:

$$\hat{f}_k = f_k + \eta_k$$

Estimating Sinusoidal parameters

- Sinusoidal representation for a speech/signal frame:

$$x(t) = \left(\sum_{k=-K}^{K} a_k e^{j2\pi f_k t} \right) w(t)$$

- Methods:
  - FFT-based methods (i.e., QIFFT [Abe et al., 2004-05, [1] [2]])
  - Subspace methods
  - Least Squares (LS) method
- Frequency mismatch:

$$\hat{f}_k = f_k + \eta_k$$

Estimating Sinusoidal parameters

- Sinusoidal representation for a speech/signal frame:

$$x(t) = \left(\sum_{k=-K}^{K} a_k e^{j2\pi f_k t} \right) w(t)$$

- Methods:
  - FFT-based methods (i.e., QIFFT [Abe et al., 2004-05, [1] [2]])
  - Subspace methods
  - Least Squares (LS) method
- Frequency mismatch:

$$\hat{f}_k = f_k + \eta_k$$

Estimating Sinusoidal parameters

- Sinusoidal representation for a speech/signal frame:

$$x(t) = \left(\sum_{k=-K}^{K} a_k e^{j2\pi f_k t} \right) w(t)$$

- Methods:
  - FFT-based methods (i.e., QIFFT [Abe et al., 2004-05, [1] [2]])
  - Subspace methods
  - Least Squares (LS) method
- Frequency mismatch:

$$\hat{f}_k = f_k + \eta_k$$

Estimating Sinusoidal parameters

- Sinusoidal representation for a speech/signal frame:

$$x(t) = \left(\sum_{k=-K}^{K} a_k e^{j2\pi f_k t} \right) w(t)$$

- Methods:
  - FFT-based methods (i.e., QIFFT [Abe et al., 2004-05, [1] [2]])
  - Subspace methods
  - Least Squares (LS) method
- Frequency mismatch:

$$\hat{f}_k = f_k + \eta_k$$

Quasi-Harmonic Model, QHM [5]

- Frequency mismatch:

$$x(t) = \left(\sum_{k=-K}^{K} a_k e^{j2\pi \hat{f}_k t} \right) w(t)$$

- QHM (de Prony 1795, Laroche [3] (1989), Stylianou 1993, Pantazis [4] (2008, 2011)):

$$x(t) = \left(\sum_{k=-K}^{K} (a_k + t b_k) e^{j2\pi \hat{f}_k t} \right) w(t)$$

Quasi-Harmonic Model, QHM [5]

- Frequency mismatch:

$$x(t) = \left(\sum_{k=-K}^{K} a_k e^{j2\pi \hat{f}_k t} \right) w(t)$$

- QHM (de Prony 1795, Laroche [3] (1989), Stylianou 1993, Pantazis [4] (2008, 2011)):

$$x(t) = \left(\sum_{k=-K}^{K} (a_k + t b_k) e^{j2\pi \hat{f}_k t} \right) w(t)$$

A few details of QHM

- QHM in the frequency domain:

$$X_k(f) = a_k W(f - \hat{f}_k) + j\frac{b_k}{2\pi} W'(f - \hat{f}_k)$$

- **Decomposition of** b_k: $b_k = \rho_{1,k} a_k + \rho_{2,k} j a_k$
- Then

$$X_k(f) = a_k \left[W(f - \hat{f}_k) - \frac{\rho_{2,k}}{2\pi} W'(f - \hat{f}_k) + j\frac{\rho_{1,k}}{2\pi} W'(f - \hat{f}_k) \right]$$

- and taking into account:

$$\begin{aligned} W(f - \hat{f}_k - \frac{\rho_{2,k}}{2\pi}) = \quad & W(f - \hat{f}_k) - \frac{\rho_{2,k}}{2\pi} W'(f - \hat{f}_k) + \\ & O(\rho_{2,k}^2 W''(f - \hat{f}_k)) \end{aligned}$$

- **Approximation of** the *kth* component of QHM

$$X_k(f) \approx a_k \left[W(f - \hat{f}_k - \frac{\rho_{2,k}}{2\pi}) + j\frac{\rho_{1,k}}{2\pi} W'(f - \hat{f}_k) \right]$$

A few details of QHM

- QHM in the frequency domain:

$$X_k(f) = a_k W(f - \hat{f}_k) + j\frac{b_k}{2\pi} W'(f - \hat{f}_k)$$

- **Decomposition of** b_k: $b_k = \rho_{1,k} a_k + \rho_{2,k} j a_k$
- Then

$$X_k(f) = a_k \left[W(f - \hat{f}_k) - \frac{\rho_{2,k}}{2\pi} W'(f - \hat{f}_k) + j\frac{\rho_{1,k}}{2\pi} W'(f - \hat{f}_k) \right]$$

- and taking into account:

$$\begin{aligned} W(f - \hat{f}_k - \frac{\rho_{2,k}}{2\pi}) = \quad & W(f - \hat{f}_k) - \frac{\rho_{2,k}}{2\pi} W'(f - \hat{f}_k) + \\ & O(\rho_{2,k}^2 W''(f - \hat{f}_k)) \end{aligned}$$

- **Approximation of** the *kth* component of QHM

$$X_k(f) \approx a_k \left[W(f - \hat{f}_k - \frac{\rho_{2,k}}{2\pi}) + j\frac{\rho_{1,k}}{2\pi} W'(f - \hat{f}_k) \right]$$

A few details of QHM

- QHM in the frequency domain:

$$X_k(f) = a_k W(f - \hat{f}_k) + j\frac{b_k}{2\pi} W'(f - \hat{f}_k)$$

- **Decomposition of** b_k: $b_k = \rho_{1,k} a_k + \rho_{2,k} j a_k$
- Then

$$X_k(f) = a_k \left[W(f - \hat{f}_k) - \frac{\rho_{2,k}}{2\pi} W'(f - \hat{f}_k) + j\frac{\rho_{1,k}}{2\pi} W'(f - \hat{f}_k) \right]$$

- and taking into account:

$$\begin{aligned} W(f - \hat{f}_k - \frac{\rho_{2,k}}{2\pi}) = \quad & W(f - \hat{f}_k) - \frac{\rho_{2,k}}{2\pi} W'(f - \hat{f}_k) + \\ & O(\rho_{2,k}^2 W''(f - \hat{f}_k)) \end{aligned}$$

- **Approximation of** the *kth* component of QHM

$$X_k(f) \approx a_k \left[W(f - \hat{f}_k - \frac{\rho_{2,k}}{2\pi}) + j\frac{\rho_{1,k}}{2\pi} W'(f - \hat{f}_k) \right]$$

A few details of QHM

- QHM in the frequency domain:

$$X_k(f) = a_k W(f - \hat{f}_k) + j\frac{b_k}{2\pi} W'(f - \hat{f}_k)$$

- **Decomposition of** b_k: $b_k = \rho_{1,k} a_k + \rho_{2,k} j a_k$
- Then

$$X_k(f) = a_k \left[W(f - \hat{f}_k) - \frac{\rho_{2,k}}{2\pi} W'(f - \hat{f}_k) + j\frac{\rho_{1,k}}{2\pi} W'(f - \hat{f}_k) \right]$$

- and taking into account:

$$\begin{aligned} W(f - \hat{f}_k - \frac{\rho_{2,k}}{2\pi}) = \; & W(f - \hat{f}_k) - \frac{\rho_{2,k}}{2\pi} W'(f - \hat{f}_k) + \\ & O(\rho_{2,k}^2 W''(f - \hat{f}_k)) \end{aligned}$$

- **Approximation of** the *kth* component of QHM

$$X_k(f) \approx a_k \left[W(f - \hat{f}_k - \frac{\rho_{2,k}}{2\pi}) + j\frac{\rho_{1,k}}{2\pi} W'(f - \hat{f}_k) \right]$$

A few details of QHM

- QHM in the frequency domain:

$$X_k(f) = a_k W(f - \hat{f}_k) + j\frac{b_k}{2\pi} W'(f - \hat{f}_k)$$

- **Decomposition of** b_k: $b_k = \rho_{1,k} a_k + \rho_{2,k} j a_k$
- Then

$$X_k(f) = a_k \left[W(f - \hat{f}_k) - \frac{\rho_{2,k}}{2\pi} W'(f - \hat{f}_k) + j\frac{\rho_{1,k}}{2\pi} W'(f - \hat{f}_k) \right]$$

- and taking into account:

$$\begin{aligned} W(f - \hat{f}_k - \frac{\rho_{2,k}}{2\pi}) = \quad & W(f - \hat{f}_k) - \frac{\rho_{2,k}}{2\pi} W'(f - \hat{f}_k) + \\ & O(\rho_{2,k}^2 W''(f - \hat{f}_k)) \end{aligned}$$

- **Approximation of** the *kth* component of QHM

$$X_k(f) \approx a_k \left[W(f - \hat{f}_k - \frac{\rho_{2,k}}{2\pi}) + j\frac{\rho_{1,k}}{2\pi} W'(f - \hat{f}_k) \right]$$

Approximations in QHM

- we found previously:

$$X_k(f) \approx a_k \left[W(f - \hat{f}_k - \frac{\rho_{2,k}}{2\pi}) + j\frac{\rho_{1,k}}{2\pi} W'(f - \hat{f}_k) \right]$$

- Back to the time-domain:

$$x_k(t) \approx a_k \left[e^{j(2\pi \hat{f}_k + \rho_{2,k})t} + \rho_{1,k} t e^{j2\pi \hat{f}_k t} \right] w(t)$$

- Initially, we assumed:

$$x_k(t) = a_k \left[e^{j(2\pi(\hat{f}_k + \eta_k))t} \right] w(t)$$

- in other words, **it is suggested**:

$$\hat{\eta}_k = \rho_{2,k}/2\pi$$

Approximations in QHM

- we found previously:

$$X_k(f) \approx a_k \left[W(f - \hat{f}_k - \frac{\rho_{2,k}}{2\pi}) + j \frac{\rho_{1,k}}{2\pi} W'(f - \hat{f}_k) \right]$$

- Back to the time-domain:

$$x_k(t) \approx a_k \left[e^{j(2\pi \hat{f}_k + \rho_{2,k})t} + \rho_{1,k} t e^{j2\pi \hat{f}_k t} \right] w(t)$$

- Initially, we assumed:

$$x_k(t) = a_k \left[e^{j(2\pi(\hat{f}_k + \eta_k))t} \right] w(t)$$

- in other words, **it is suggested**:

$$\hat{\eta}_k = \rho_{2,k}/2\pi$$

Approximations in QHM

- we found previously:

$$X_k(f) \approx a_k \left[W(f - \hat{f}_k - \frac{\rho_{2,k}}{2\pi}) + j \frac{\rho_{1,k}}{2\pi} W'(f - \hat{f}_k) \right]$$

- Back to the time-domain:

$$x_k(t) \approx a_k \left[e^{j(2\pi \hat{f}_k + \rho_{2,k})t} + \rho_{1,k} t e^{j2\pi \hat{f}_k t} \right] w(t)$$

- Initially, we assumed:

$$x_k(t) = a_k \left[e^{j(2\pi(\hat{f}_k + \eta_k))t} \right] w(t)$$

- in other words, **it is suggested**:

$$\hat{\eta}_k = \rho_{2,k} / 2\pi$$

Approximations in QHM

- we found previously:

$$X_k(f) \approx a_k \left[W(f - \hat{f}_k - \frac{\rho_{2,k}}{2\pi}) + j\frac{\rho_{1,k}}{2\pi} W'(f - \hat{f}_k) \right]$$

- Back to the time-domain:

$$x_k(t) \approx a_k \left[e^{j(2\pi \hat{f}_k + \rho_{2,k})t} + \rho_{1,k} t e^{j2\pi \hat{f}_k t} \right] w(t)$$

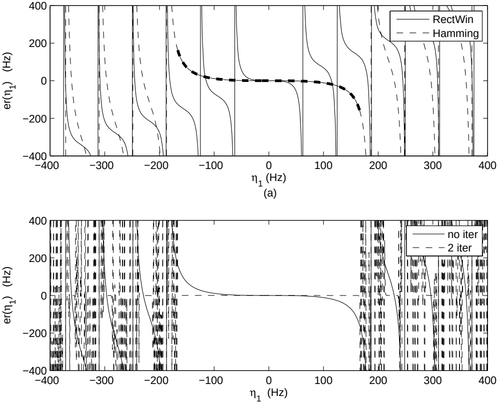
- Initially, we assumed:

$$x_k(t) = a_k \left[e^{j(2\pi(\hat{f}_k + \eta_k))t} \right] w(t)$$

- in other words, **it is suggested**:

$$\hat{\eta}_k = \rho_{2,k}/2\pi$$

Single Sinusoid



(a)

(b)

- Iteratively, the bias can be removed when $|\eta| < B/3$, where B is the bandwidth of the squared analysis window.

Robustness against additive noise

- Signal contaminated by noise:

$$y(t) = \sum_{k=1}^{4} a_k e^{j2\pi f_k} + v(t)$$

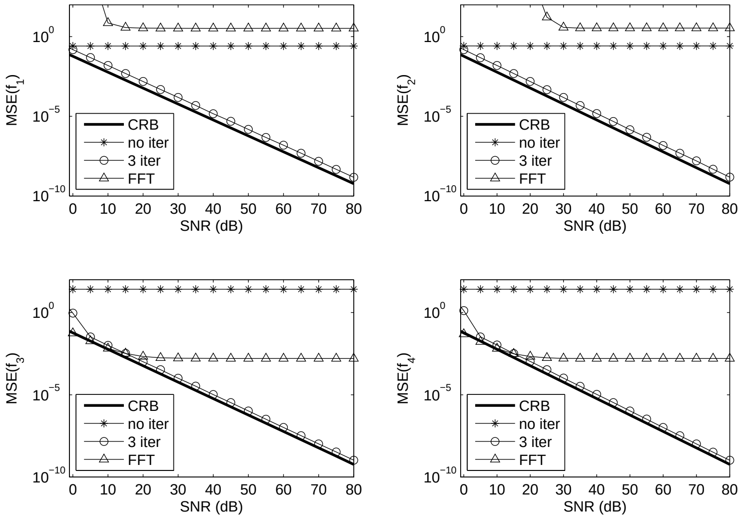
- Mean Squared Error (MSE):

$$MSE\{\hat{f}_k\} = \frac{1}{\mathrm{M}} \sum_{i=1}^{\mathrm{M}} |\hat{f}_k(i) - f_k|^2$$

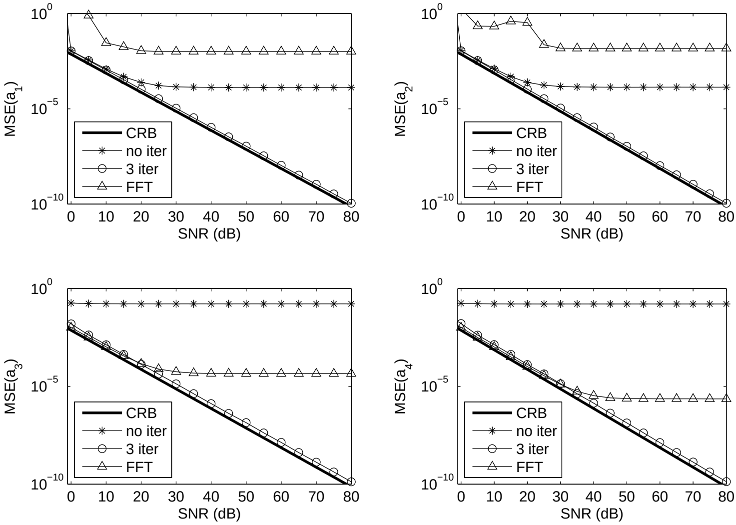
$$MSE\{\hat{a}_k\} = \frac{1}{\mathrm{M}} \sum_{i=1}^{\mathrm{M}} |\hat{a}_k(i) - a_k|^2$$

- Comparison with Cramer-Rao Bounds (CRB) and QIFFT (Abe et al. 2004)
- 10000 Monte Carlo simulations

MSE of frequencies as a function of SNR.



MSE of amplitudes as a function of SNR.



Notes on QHM

QHM has been shown to be closely related to:

- Gauss-Newton frequency estimation method
- Reassigned Spectrogram
- AM-FM decomposition

Notes on QHM

QHM has been shown to be closely related to:

- Gauss-Newton frequency estimation method
- Reassigned Spectrogram
- AM-FM decomposition

Notes on QHM

QHM has been shown to be closely related to:

- Gauss-Newton frequency estimation method
- Reassigned Spectrogram
- AM-FM decomposition

High resolution AM-FM decomposition

- AM-FM signal

$$y(t) = \sum_{k=1}^{K(t)} a_k(t) cos(\phi_k(t)),$$

- Taylor series expansion of the instantaneous phase of kth component:

$$\phi_k(t) = 2\pi\zeta_k t + \sum_{i=0}^{\infty} \phi_{k,i} \frac{t^i}{i!}$$

- Instantaneous frequency of the kth component at $t = 0$:

$$\nu_k(0) = \zeta_k + \frac{\phi_{k,1}}{2\pi}$$

- ... and previously we had:

$$F_k(0) = f_k + \frac{\rho_{2,k}}{2\pi}$$

High resolution AM-FM decomposition

- AM-FM signal

$$y(t) = \sum_{k=1}^{K(t)} a_k(t) cos(\phi_k(t)),$$

- Taylor series expansion of the instantaneous phase of kth component:

$$\phi_k(t) = 2\pi\zeta_k t + \sum_{i=0}^{\infty} \phi_{k,i} \frac{t^i}{i!}$$

- Instantaneous frequency of the kth component at $t = 0$:

$$\nu_k(0) = \zeta_k + \frac{\phi_{k,1}}{2\pi}$$

- ... and previously we had:

$$F_k(0) = f_k + \frac{\rho_{2,k}}{2\pi}$$

High resolution AM-FM decomposition

- AM-FM signal

$$y(t) = \sum_{k=1}^{K(t)} a_k(t) cos(\phi_k(t)),$$

- Taylor series expansion of the instantaneous phase of kth component:

$$\phi_k(t) = 2\pi\zeta_k t + \sum_{i=0}^{\infty} \phi_{k,i} \frac{t^i}{i!}$$

- Instantaneous frequency of the kth component at $t = 0$:

$$\nu_k(0) = \zeta_k + \frac{\phi_{k,1}}{2\pi}$$

- ... and previously we had:

$$F_k(0) = f_k + \frac{\rho_{2,k}}{2\pi}$$

High resolution AM-FM decomposition

- AM-FM signal

$$y(t) = \sum_{k=1}^{K(t)} a_k(t) cos(\phi_k(t)),$$

- Taylor series expansion of the instantaneous phase of kth component:

$$\phi_k(t) = 2\pi\zeta_k t + \sum_{i=0}^{\infty} \phi_{k,i} \frac{t^i}{i!}$$

- Instantaneous frequency of the kth component at $t = 0$:

$$\nu_k(0) = \zeta_k + \frac{\phi_{k,1}}{2\pi}$$

- ... and previously we had:

$$F_k(0) = f_k + \frac{\rho_{2,k}}{2\pi}$$

From QHM to Adaptive QHM, aQHM [6]

- QHM (**stationarity assumption**):

$$x(t) = \left(\sum_{k=-K}^{K} (a_k + tb_k) e^{2\pi j f_k t} \right) w(t)$$

- Adaptive QHM (aQHM):

$$x(t) = \left(\sum_{k=-K}^{K} (a_k + tb_k) e^{j\tilde{\phi}_k(t)} \right) w(t)$$

where

$$\tilde{\phi}_k(t) = 2\pi \int_0^t f_k(s) ds + \varphi, \ \ t \in [0, T]$$

is the estimated instantaneous phase.

From QHM to Adaptive QHM, aQHM [6]

- QHM (**stationarity assumption**):

$$x(t) = \left(\sum_{k=-K}^{K} (a_k + tb_k) e^{2\pi j f_k t} \right) w(t)$$

- Adaptive QHM (aQHM):

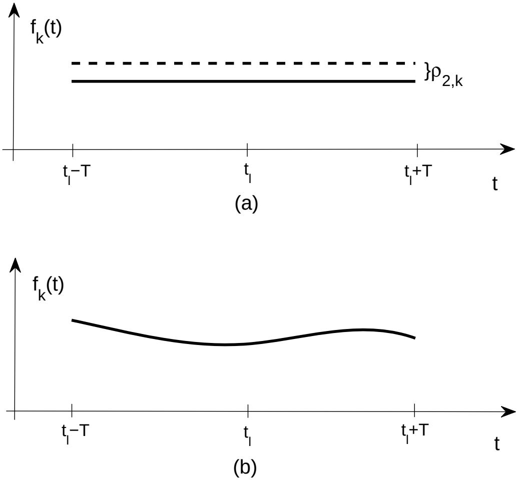
$$x(t) = \left(\sum_{k=-K}^{K} (a_k + tb_k) e^{j\tilde{\phi}_k(t)} \right) w(t)$$

where

$$\tilde{\phi}_k(t) = 2\pi \int_0^t f_k(s) ds + \varphi, \ \ t \in [0, T]$$

is the estimated instantaneous phase.

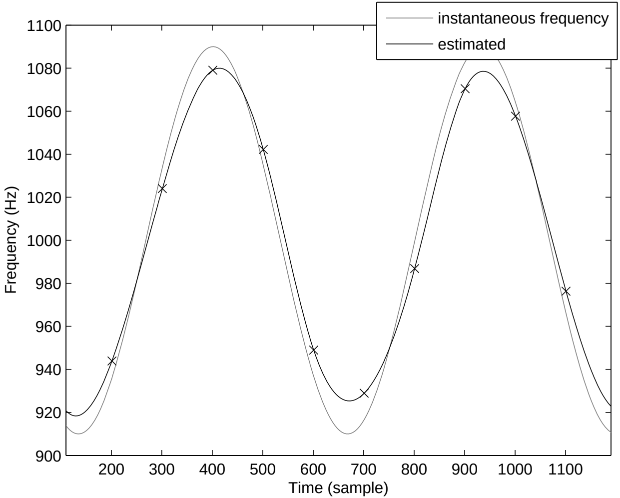
From QHM to aQHM; Graphically



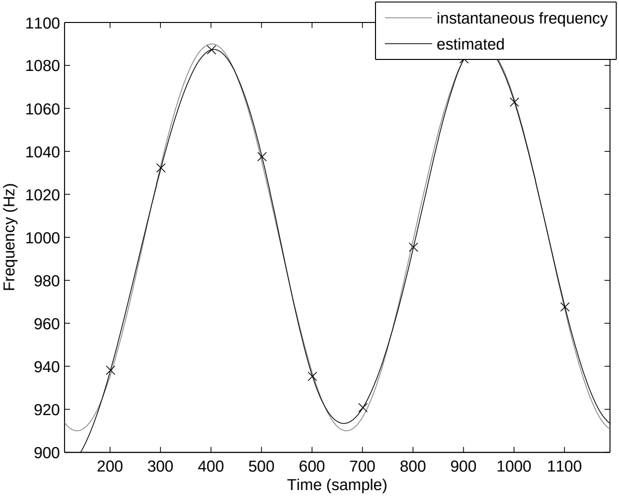
aQHM, in practice

- One sample: no interpolation between estimations
- Higher rates (i.e., 5ms, 10ms): Interpolation between estimates is required:
  - Amplitudes are linearly interpolated
  - Frequencies are interpolated with splines
  - Phases are interpolated by integration of instantaneous frequency

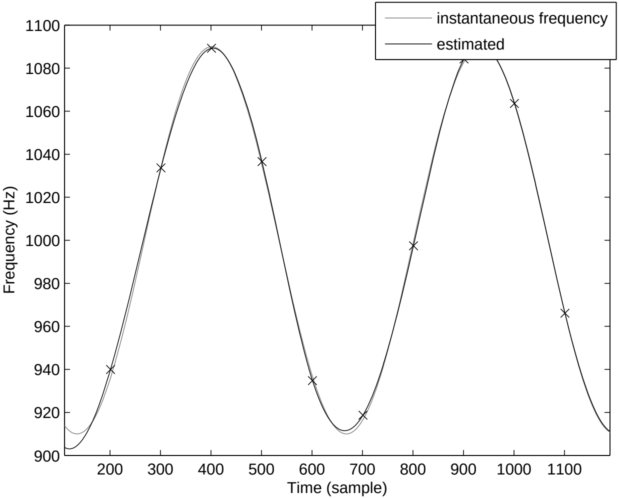
Example of estimation in aQHM: no iteration (QHM)



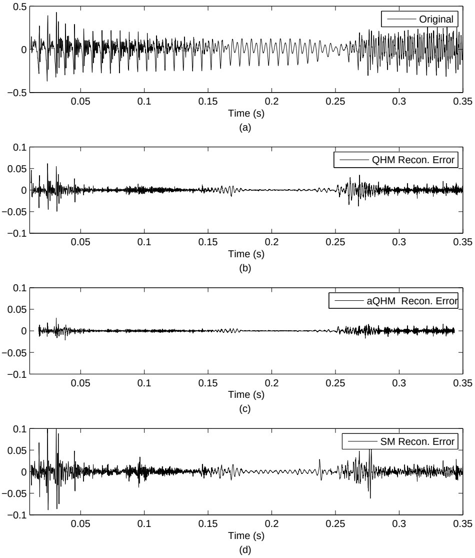
Example of estimation in aQHM: one iteration



Example of estimation in aQHM: two iterations



Reconstruction errors with QHM, aQHM, SM



Outline

Extended aQHM

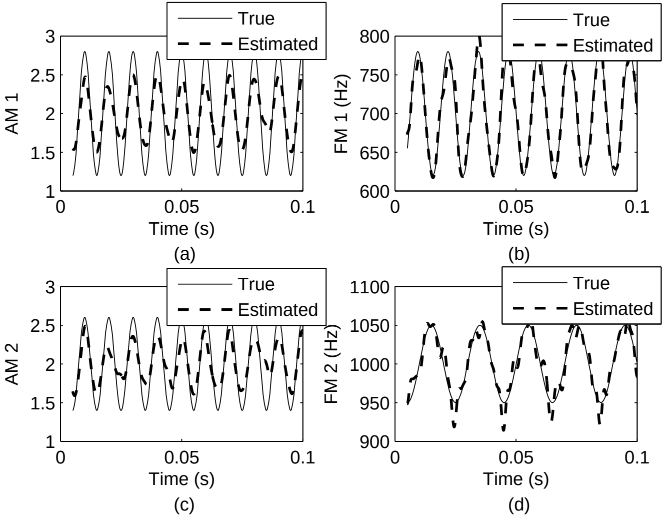
- Recall aQHM:

$$x(t) = \sum_{k=-K}^{K} (a_k + tb_k) e^{j\tilde{\phi}_k(t)}$$

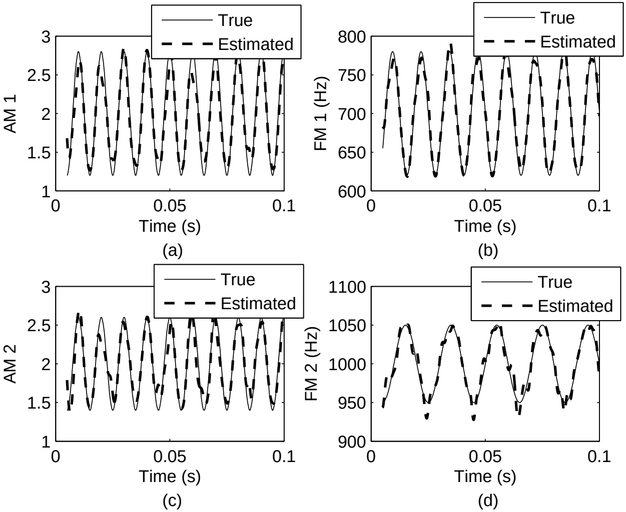
- Extended aQHM:

$$x(t) = \sum_{k=-K}^{K} (a_k + tb_k) \alpha(t) e^{j\tilde{\phi}_k(t)}$$

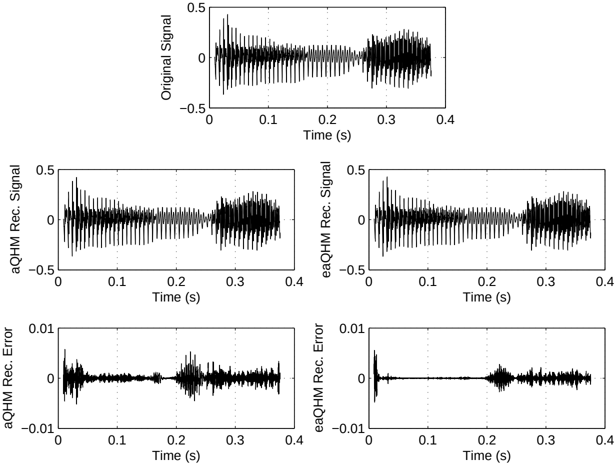
AM-FM modeling: aQHM



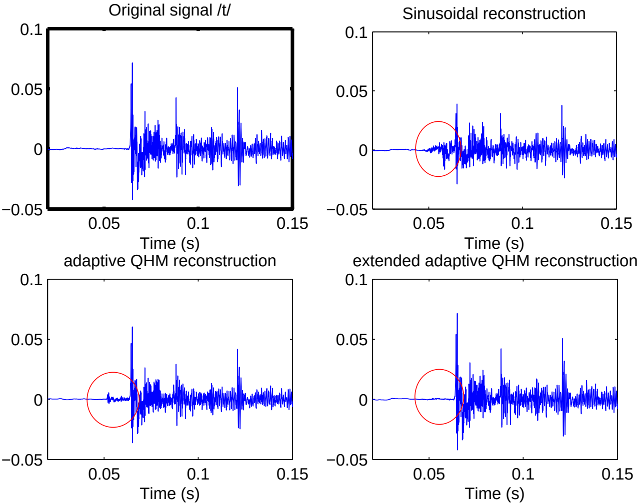
AM-FM modeling: extended aQHM



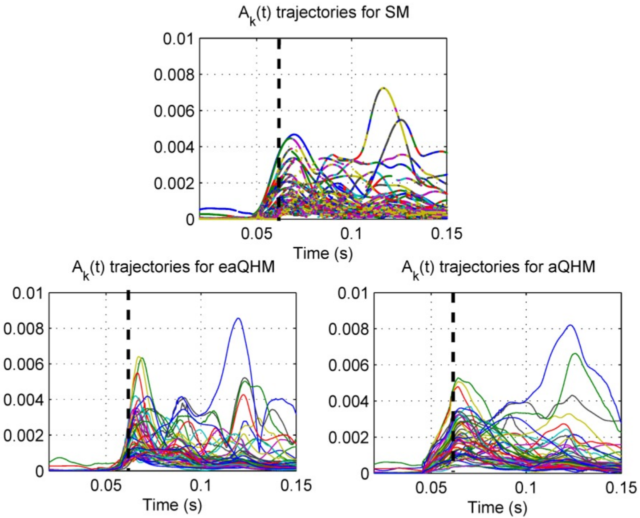
Comparing Adaptive Models



Stop modeling



Stop modeling

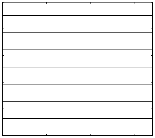


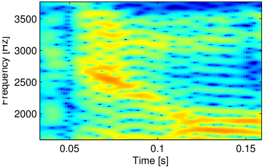
Time-Frequency Representations

DFT $s(t) = \sum_{k=0}^{K} a_k \cdot e^{j\phi_k(t)}$

$\phi_k(t) = k \cdot (2\pi/K) \cdot t$

Constant frequency basis

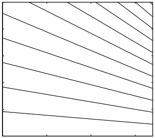


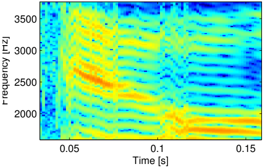


FChT[1] $s(t) = \sum_{k=0}^{K} a_k \cdot e^{j\phi_k(t)}$

$\phi_k(t) = k \cdot (2\pi/K + \alpha \cdot t) \cdot$

Linear frequency basis



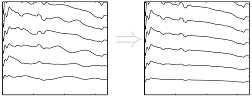


[1] M. Kepesi and L. Weruaga, *Adaptive Chirp-based time-frequency analysis of speech signals*, Speech communication, 2006.

Ideas

1) From FChT, harmonics exist in high frequencies
   - ⇒ Use a full-band representation

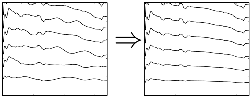
2) Quasi-harmonicity can be useful for analysis but maybe not necessary for encoding/decoding
   - ⇒ Use the strict harmonicity and keep the adaptivity



IDEAS

1) From FChT, harmonics exist in high frequencies
   - $\Rightarrow$ Use a full-band representation

2) Quasi-harmonicity can be useful for analysis but maybe not necessary for encoding/decoding
   - $\Rightarrow$ Use the strict harmonicity and keep the adaptivity

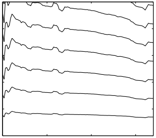


aQHNM$\Rightarrow$aHM

The Adaptive Harmonic Model (aHM)

aHM $s(t) = \sum_{k=-K}^{K} a_k(t) \cdot e^{j\phi_k(t)}$

$$\phi_k(t) = k \cdot \frac{2\pi}{f_s} \int_0^t f_0(\tau) d\tau$$



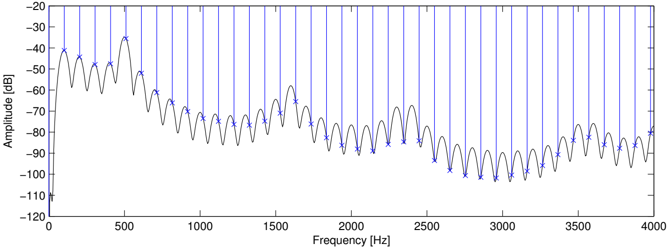
$a_k(t)$ Amplitude and phase (complex-valued function)
Interpolated from $a_k^{t_i}$ at time t_i

$f_0(t)$ Fundamental frequency
Interpolated from $f_0^{t_i}$ at time t_i

Parameters at a time t_i: $\{f_0^{t_i}, a_k^{t_i}\}$ $\quad k \in \{0, ..., K_i\}$

The problem of estimation for full-band representation

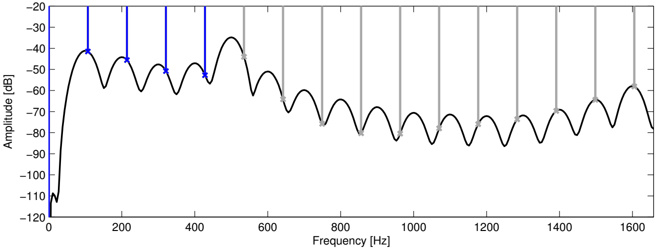
A small f_0 error propagates by multiplication: $f_k = k \cdot f_0$



Question How to estimate harmonics up to Nyquist ?

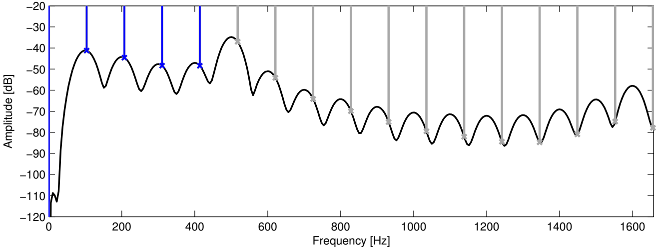
The Adaptive Iterative Refinement (AIR)

Assume first the f_0 error is small for low harmonics



The Adaptive Iterative Refinement (AIR)

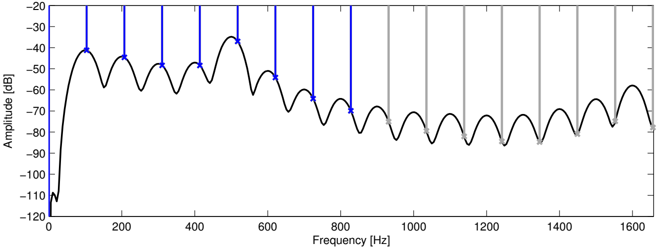
Then the frequency correction mechanism of QHM[1] can be used



[1] Y. Pantazis, O. Rosec and Y. Stylianou, *Iterative Estimation of Sinusoidal Signal Parameters*, IEEE Signal Processing Letters, 2010.

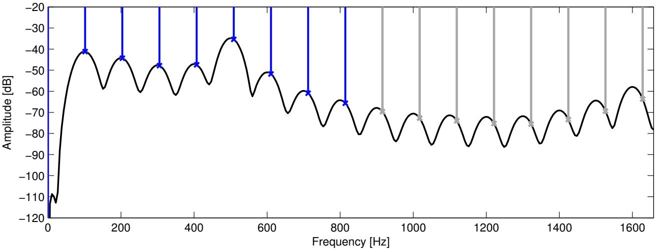
The Adaptive Iterative Refinement (AIR)

We can therefore increase the harmonic level



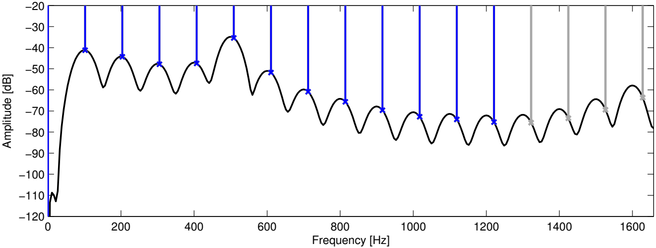
The Adaptive Iterative Refinement (AIR)

Correct the frequencies



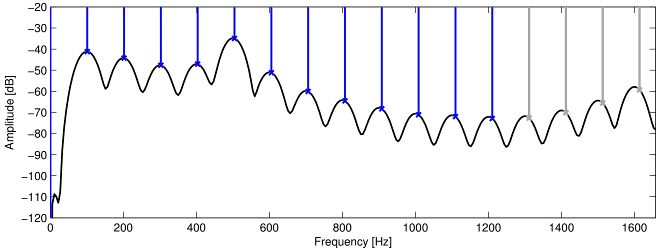
The Adaptive Iterative Refinement (AIR)

Increase the harmonic level



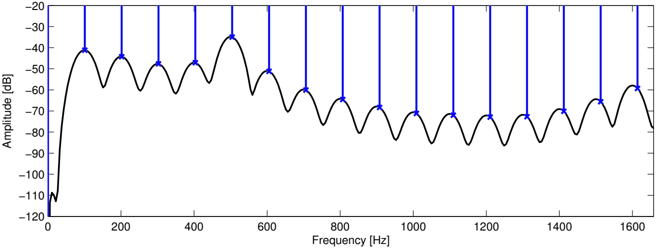
The Adaptive Iterative Refinement (AIR)

Correct the frequencies



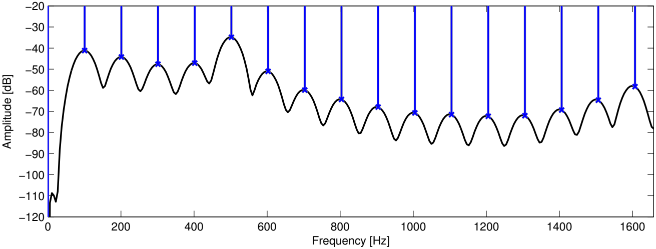
The Adaptive Iterative Refinement (AIR)

Increase the harmonic level



The Adaptive Iterative Refinement (AIR)

Correct the frequencies



Key papers to read

- Y. Pantazis, O. Rosec, and Y. Stylianou: *Adaptive AM-FM Signal Decomposition with Application to Speech Analysis* IEEE TASLP, Vol.19, No.2, Feb 2011, pp 290-300.
- Y. Pantazis, O. Rosec, and Y. Stylianou: *Iterative Estimation of Sinusoidal Signal Parameters*. IEEE SPL, Vol.17, No.5, May 2010, pp. 461-464
- P. Babu and P. Stoica: *Comments on Iterative Estimation of Sinusoidal Signal Parameters.* Comments on IEEE SPL, pp.1022-1023, Vol.17, No.12, Dec 2010
- Y. Pantazis, O. Rosec, and Y. Stylianou: *Reply to "Comments on 'Iterative Estimation of Sinusoidal Signal Parameters' "* IEEE SPL, pp. 1024-1026, Vol.17, No.12, Dec 2010
- G. Degottex and Y. Stylianou: *Analysis and Synthesis of Speech using an Adaptive Full-band Harmonic Model* IEEE TASLP, Vol.21, Issue 10, pp 2085-2095, 2013

Outline

Speech Modifications

- High quality pitch scale modification
- Within glottal cycle formant tracking and modifications towards voice conversion
- Free pre-echo effect in time scaling
- Emotion detection and control

Speech Modifications

- High quality pitch scale modification
- Within glottal cycle formant tracking and modifications towards voice conversion
- Free pre-echo effect in time scaling
- Emotion detection and control

Speech Modifications

- High quality pitch scale modification
- Within glottal cycle formant tracking and modifications towards voice conversion
- Free pre-echo effect in time scaling
- Emotion detection and control

Speech Modifications

- High quality pitch scale modification
- Within glottal cycle formant tracking and modifications towards voice conversion
- Free pre-echo effect in time scaling
- Emotion detection and control

- Unit selection based: fast synthesis, smooth trajectories, high quality prosodic modifications, voice conversion
- Statistical modeling: *naturally* smooth trajectories, articulators tracking, adaptation and controllability, modulations
- Sinusoidal models work nicely with DNNs

Speech Synthesis

- Unit selection based: fast synthesis, smooth trajectories, high quality prosodic modifications, voice conversion
- Statistical modeling: *naturally* smooth trajectories, articulators tracking, adaptation and controllability, modulations
- Sinusoidal models work nicely with DNNs

Speech Synthesis

- Unit selection based: fast synthesis, smooth trajectories, high quality prosodic modifications, voice conversion
- Statistical modeling: *naturally* smooth trajectories, articulators tracking, adaptation and controllability, modulations
- Sinusoidal models work nicely with DNNs

Speech Recognition

- Robust to noise filter bank estimation
- Data driven dynamic information
- High-resolution modulations
- Fits well to the DNN framework.

Speech Recognition

- Robust to noise filter bank estimation
- Data driven dynamic information
- High-resolution modulations
- Fits well to the DNN framework.

Speech Recognition

- Robust to noise filter bank estimation
- Data driven dynamic information
- High-resolution modulations
- Fits well to the DNN framework.

Speech Recognition

- Robust to noise filter bank estimation
- Data driven dynamic information
- High-resolution modulations
- Fits well to the DNN framework.

- Jitter and shimmer estimators
- Vocal fatigue
- Robust estimation in spasmodic dysphonia
- Vocal tremor.

Speech Pathology

- Jitter and shimmer estimators
- Vocal fatigue
- Robust estimation in spasmodic dysphonia
- Vocal tremor.

Speech Pathology

- Jitter and shimmer estimators
- Vocal fatigue
- Robust estimation in spasmodic dysphonia
- Vocal tremor.

Speech Pathology

- Jitter and shimmer estimators
- Vocal fatigue
- Robust estimation in spasmodic dysphonia
- Vocal tremor.

Outline

Acknowledgments

- My colleagues: Yannis Pantazis, Olivier Rosec (Voxygen), Vassilis Tsiaras, Gilles Degottex, Giorgos Kafentzis

THANK YOU

for your attention

Outline

M. Abe and J. S. III, “CQIFFT: Correcting Bias in a Sinusoidal Parameter Estimator based on Quadratic Interpolation of FFT Magnitude Peaks,” Tech. Rep. STAN-M-117, Stanford University, California, Oct 2004.

M. Abe and J. S. III, “AM/FM Estimation for Time-varying Sinusoidal Modeling,” in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing*, (Philadelphia), pp. III 201–204, 2005.

J. Laroche, “A new analysis/synthesis system of musical signals using Prony’s method. Application to heavily damped percussive sounds.,” in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing*, (Glasgow, UK), pp. 2053–2056, May 1989.

Y. Pantazis, O. Rosec, and Y. Stylianou, “On the Properties of a Time-Varying Quasi-Harmonic Model of Speech,” (Brisbane), Sep 2008.

Y. Pantazis, O. Rosec, and Y. Stylianou, “Iterative Estimation of Sinusoidal Signal Parameters,” vol. 17, no. 5, pp. 461–464, 2010.

Y. Pantazis, O. Rosec, and Y. Stylianou, “Adaptive AMFM signal decomposition with application to speech analysis,” *IEEE Trans. on Audio, Speech, and Language Processing*, vol. 19, pp. 290–300, 2011.