CS578 - Speech Signal Processing

Lecture : Quasi-Harmonic Models of Speech

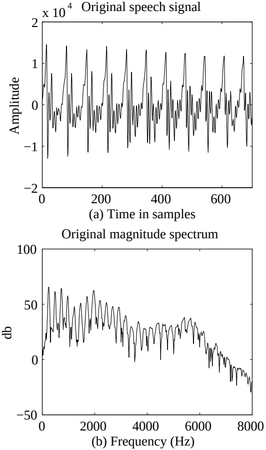
Yannis Pantazis
pantazis@iacm.forth.gr
(based on material from Prof. Stylianou)

Speech Signal Processing Lab
Computer Science Department - University of Crete

16 March 2022

Outline

Motivation for HNM



Brief overview of HNM

- HNM (Stylianou 1995 [1]) is a pitch-synchronous harmonic plus noise representation of the speech signal.
- Speech spectrum is divided into a low and a high band delimited by the so-called *maximum voiced frequency*
- The *lower* band of the spectrum (below the maximum voiced frequency) is represented solely by harmonically related sine waves.
- The *upper* band is modeled as a noise component modulated by a time-domain amplitude envelope.
- HNM allows high-quality copy synthesis and prosodic modifications.

Brief overview of HNM

- HNM (Stylianou 1995 [1]) is a pitch-synchronous harmonic plus noise representation of the speech signal.
- Speech spectrum is divided into a low and a high band delimited by the so-called *maximum voiced frequency*
- The *lower* band of the spectrum (below the maximum voiced frequency) is represented solely by harmonically related sine waves.
- The *upper* band is modeled as a noise component modulated by a time-domain amplitude envelope.
- HNM allows high-quality copy synthesis and prosodic modifications.

Brief overview of HNM

- HNM (Stylianou 1995 [1]) is a pitch-synchronous harmonic plus noise representation of the speech signal.
- Speech spectrum is divided into a low and a high band delimited by the so-called *maximum voiced frequency*
- The *lower* band of the spectrum (below the maximum voiced frequency) is represented solely by harmonically related sine waves.
- The *upper* band is modeled as a noise component modulated by a time-domain amplitude envelope.
- HNM allows high-quality copy synthesis and prosodic modifications.

Brief overview of HNM

- HNM (Stylianou 1995 [1]) is a pitch-synchronous harmonic plus noise representation of the speech signal.
- Speech spectrum is divided into a low and a high band delimited by the so-called *maximum voiced frequency*
- The *lower* band of the spectrum (below the maximum voiced frequency) is represented solely by harmonically related sine waves.
- The *upper* band is modeled as a noise component modulated by a time-domain amplitude envelope.
- HNM allows high-quality copy synthesis and prosodic modifications.

Brief overview of HNM

- HNM (Stylianou 1995 [1]) is a pitch-synchronous harmonic plus noise representation of the speech signal.
- Speech spectrum is divided into a low and a high band delimited by the so-called *maximum voiced frequency*
- The *lower* band of the spectrum (below the maximum voiced frequency) is represented solely by harmonically related sine waves.
- The *upper* band is modeled as a noise component modulated by a time-domain amplitude envelope.
- HNM allows high-quality copy synthesis and prosodic modifications.

HNM in equations

- Harmonic part:

$$h(t) = \sum_{k=-L(t)}^{L(t)} A_k(t) e^{j2\pi k f_0(t) t}$$

  where $A_k(t)$ and $f_0(t)$ are the instantaneous complex amplitude and real frequency, respectively

- Noise part:

$$n(t) = e(t) \left[v(\tau, t) \star g(t) \right]$$

  where $e(t), v(\tau, t), g(t)$ are a time envelope, an estimation of the PSD (filter), and white gaussian noise, respectively

- Speech:

$$s(t) = h(t) + n(t)$$

HNM IN EQUATIONS

- Harmonic part:

$$h(t) = \sum_{k=-L(t)}^{L(t)} A_k(t) e^{j2\pi k f_0(t) t}$$

  where $A_k(t)$ and $f_0(t)$ are the instantaneous complex amplitude and real frequency, respectively

- Noise part:

$$n(t) = e(t)\left[v(\tau, t) \star g(t)\right]$$

  where $e(t), v(\tau, t), g(t)$ are a time envelope, an estimation of the PSD (filter), and white gaussian noise, respectively

- Speech:

$$s(t) = h(t) + n(t)$$

HNM IN EQUATIONS

- Harmonic part:

$$h(t) = \sum_{k=-L(t)}^{L(t)} A_k(t) e^{j2\pi k f_0(t) t}$$

  where $A_k(t)$ and $f_0(t)$ are the instantaneous complex amplitude and real frequency, respectively

- Noise part:

$$n(t) = e(t) \left[v(\tau, t) \star g(t) \right]$$

  where $e(t), v(\tau, t), g(t)$ are a time envelope, an estimation of the PSD (filter), and white gaussian noise, respectively

- Speech:

$$s(t) = h(t) + n(t)$$

Outline

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, [2] [3]])
  - Least Squares (LS) method
- Frequency mismatch (eg, $\hat{f}_k := k\hat{f}_0$ in HNM):

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

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, [2] [3]])
  - Least Squares (LS) method
- Frequency mismatch (eg, $\hat{f}_k := k\hat{f}_0$ in HNM):

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

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, [2] [3]])
  - Least Squares (LS) method
- Frequency mismatch (eg, $\hat{f}_k := k\hat{f}_0$ in HNM):

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

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, [2] [3]])
  - Least Squares (LS) method
- Frequency mismatch (eg, $\hat{f}_k := k\hat{f}_0$ in HNM):

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

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, [2] [3]])
  - Least Squares (LS) method
- Frequency mismatch (eg, $\hat{f}_k := k\hat{f}_0$ in HNM):

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

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 [4] (1989), Stylianou 1993, Pantazis [5] (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:

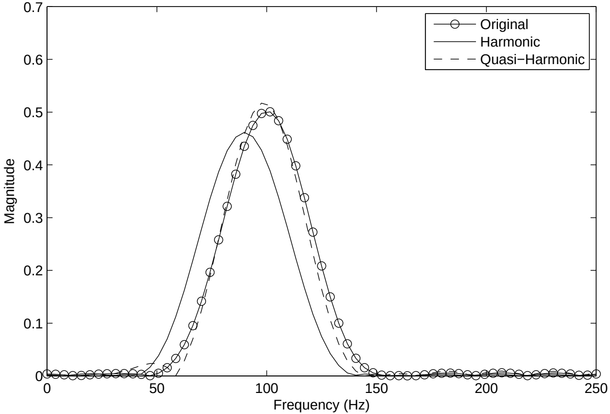
$$x(t) = \left(\sum_{k=-K}^{K} a_k e^{j2\pi \hat{f}_k t} \right) w(t)$$

- QHM (de Prony 1795, Laroche [4] (1989), Stylianou 1993, Pantazis [5] (2008, 2011)):

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

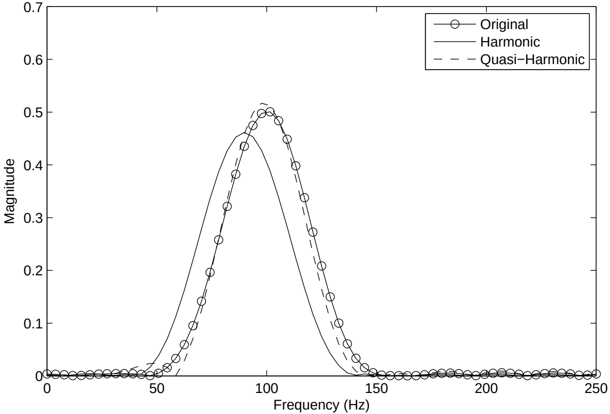
Quasi-Harmonic Model, QHM [5]

- HM versus QHM in frequency estimation - pure tone @ 100 Hz
- Given frequency for both models: 90 Hz



Quasi-Harmonic Model, QHM [5]

- HM versus QHM in frequency estimation - pure tone @ 100 Hz
- Given frequency for both models: 90 Hz



Quasi-Harmonic Model, QHM [5]

Time domain properties:

- Inst. amplitude:
  $M_k(t) = |a_k + tb_k| = \sqrt{(a_k^R + tb_k^R)^2 + (a_k^I + tb_k^I)^2}$
- Inst. phase: $\Phi_k(t) = 2\pi \hat{f}_k t + \tan^{-1} \frac{a_k^I + tb_k^I}{a_k^R + tb_k^R}$
- Inst. frequency: $F_k(t) = \frac{1}{2\pi}\Phi'(t) = \hat{f}_k + \frac{1}{2\pi} \frac{a_k^R b_k^I - a_k^I b_k^R}{M_k^2(t)}$
- where x^R, x^I denote the real and imaginary part of x

Quasi-Harmonic Model, QHM [5]

Time domain properties:

- Inst. amplitude:
  $M_k(t) = |a_k + tb_k| = \sqrt{(a_k^R + tb_k^R)^2 + (a_k^I + tb_k^I)^2}$
- Inst. phase: $\Phi_k(t) = 2\pi \hat{f}_k t + \tan^{-1} \frac{a_k^I + tb_k^I}{a_k^R + tb_k^R}$
- Inst. frequency: $F_k(t) = \frac{1}{2\pi}\Phi'(t) = \hat{f}_k + \frac{1}{2\pi}\frac{a_k^R b_k^I - a_k^I b_k^R}{M_k^2(t)}$
- where x^R, x^I denote the real and imaginary part of x

Quasi-Harmonic Model, QHM [5]

Time domain properties:

- Inst. amplitude:
  $M_k(t) = |a_k + tb_k| = \sqrt{(a_k^R + tb_k^R)^2 + (a_k^I + tb_k^I)^2}$
- Inst. phase: $\Phi_k(t) = 2\pi \hat{f}_k t + \tan^{-1} \dfrac{a_k^I + tb_k^I}{a_k^R + tb_k^R}$
- Inst. frequency: $F_k(t) = \dfrac{1}{2\pi}\Phi'(t) = \hat{f}_k + \dfrac{1}{2\pi}\dfrac{a_k^R b_k^I - a_k^I b_k^R}{M_k^2(t)}$
- where x^R, x^I denote the real and imaginary part of x

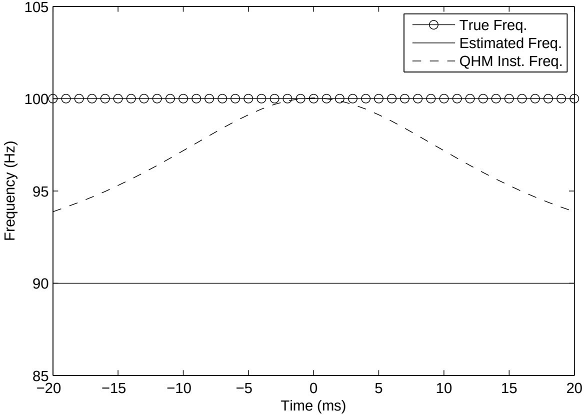
Quasi-Harmonic Model, QHM [5]

Time domain properties:

- Inst. amplitude:
  $M_k(t) = |a_k + tb_k| = \sqrt{(a_k^R + tb_k^R)^2 + (a_k^I + tb_k^I)^2}$
- Inst. phase: $\Phi_k(t) = 2\pi \hat{f}_k t + \tan^{-1} \dfrac{a_k^I + tb_k^I}{a_k^R + tb_k^R}$
- Inst. frequency: $F_k(t) = \dfrac{1}{2\pi}\Phi'(t) = \hat{f}_k + \dfrac{1}{2\pi}\dfrac{a_k^R b_k^I - a_k^I b_k^R}{M_k^2(t)}$
- where x^R, x^I denote the real and imaginary part of x

Quasi-Harmonic Model, QHM [5]

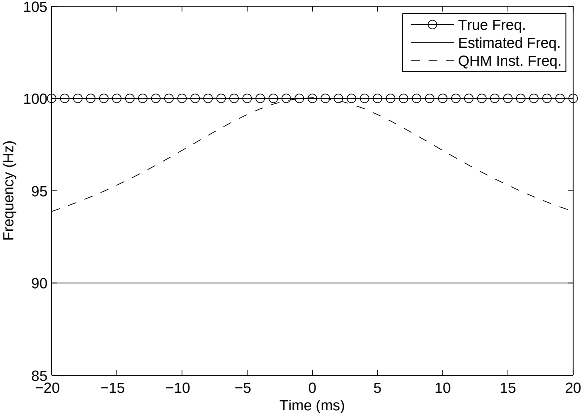
- HM vs QHM inside analysis window - pure tone @ 100 Hz:
- Given frequency for both models: 90 Hz



- Highlight: **frequency correction mechanism**
- Let's discuss a bit on that...

Quasi-Harmonic Model, QHM [5]

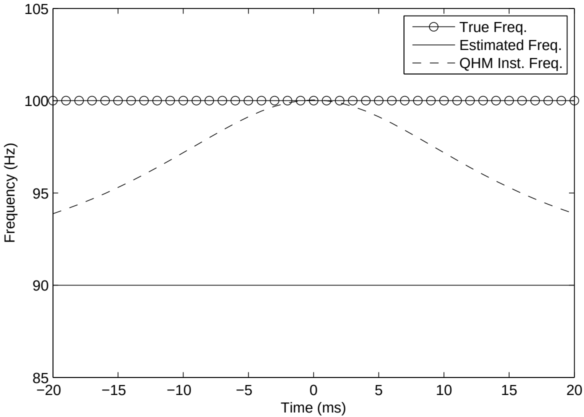
- HM vs QHM inside analysis window - pure tone @ 100 Hz:
- Given frequency for both models: 90 Hz



- Highlight: **frequency correction mechanism**
- Let's discuss a bit on that...

Quasi-Harmonic Model, QHM [5]

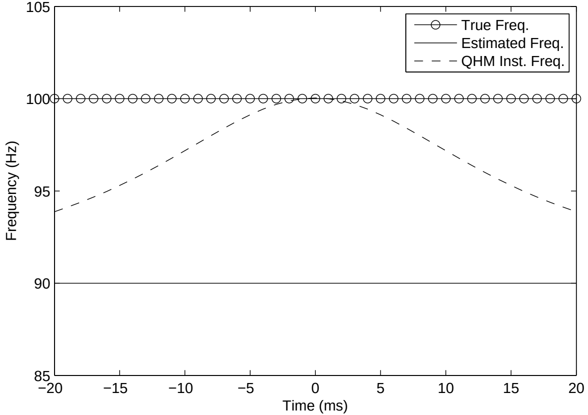
- HM vs QHM inside analysis window - pure tone @ 100 Hz:
- Given frequency for both models: 90 Hz



- Highlight: **frequency correction mechanism**
- Let's discuss a bit on that...

Quasi-Harmonic Model, QHM [5]

- HM vs QHM inside analysis window - pure tone @ 100 Hz:
- Given frequency for both models: 90 Hz



- Highlight: **frequency correction mechanism**
- Let's discuss a bit on that...

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:

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

- **Approximation of** the k-th 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:

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

- **Approximation of** the k-th 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:

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

- **Approximation of** the k-th 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:

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

- **Approximation of** the k-th 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:

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

- **Approximation of** the k-th 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**:

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

Key

- In other words, QHM suggests a frequency correction to the input frequencies $\hat{f}_k$ (or a frequency estimator).
- However, this suggestion is conditional on the magnitude of $\rho_{2,k}$ and the value of term $W''(f)$ at f_k.
- Also, the correction term depends on the window main lobe's width

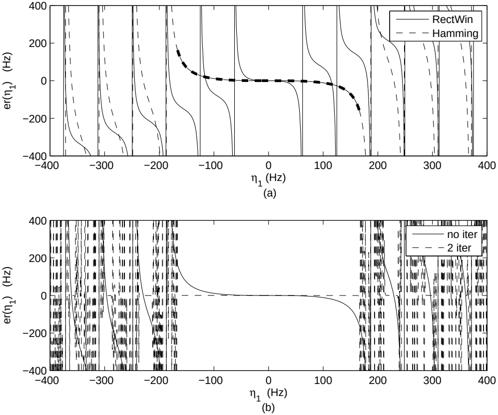
Key

- In other words, QHM suggests a frequency correction to the input frequencies $\hat{f}_k$ (or a frequency estimator).
- However, this suggestion is conditional on the magnitude of $\rho_{2,k}$ and the value of term $W''(f)$ at f_k.
- Also, the correction term depends on the window main lobe's width

Key

- In other words, QHM suggests a frequency correction to the input frequencies $\hat{f}_k$ (or a frequency estimator).
- However, this suggestion is conditional on the magnitude of $\rho_{2,k}$ and the value of term $W''(f)$ at f_k.
- Also, the correction term depends on the window main lobe's width

Single Sinusoid



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

Outline

ITERATIVE QHM, iQHM [6]

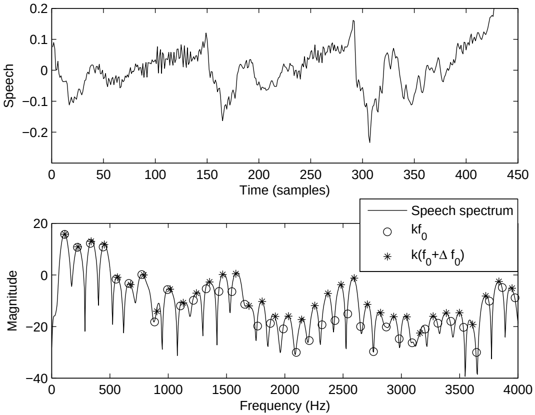
- This frequency updating mechanism provides frequencies which can be used in the model iteratively and result in better parameter estimation (a_k, b_k)
- This iterative parameter estimation is referred to as the *iterative QHM*

ITERATIVE QHM, IQHM [6]

- This frequency updating mechanism provides frequencies which can be used in the model iteratively and result in better parameter estimation (a_k, b_k)
- This iterative parameter estimation is referred to as the *iterative QHM*

ITERATIVE QHM, iQHM [6]

HM versus iQHM in frequency estimation - speech signal:



Robustness against additive noise

- Signal contaminated by noise:

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

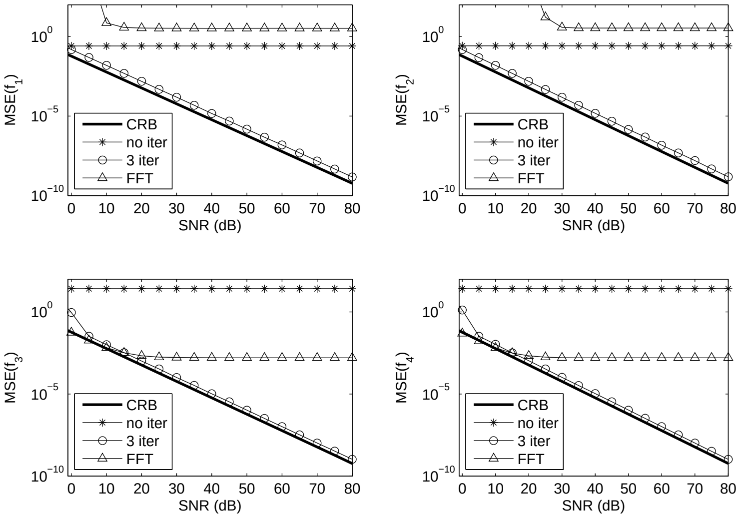
- Mean Squared Error (MSE):

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

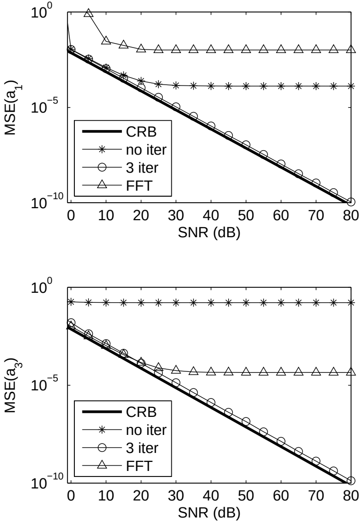
$$MSE\{\hat{a}_k\} = \frac{1}{M}\sum_{i=1}^{M} |\hat{a}_k(i) - a_k|^2$$

- 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

Notes on QHM

- QHM has been shown to be closely related to:
  - Gauss-Newton frequency estimation method
  - Reassigned Spectrogram
  - AM-FM decomposition

Outline

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 [7]

- 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 [7]

- 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):

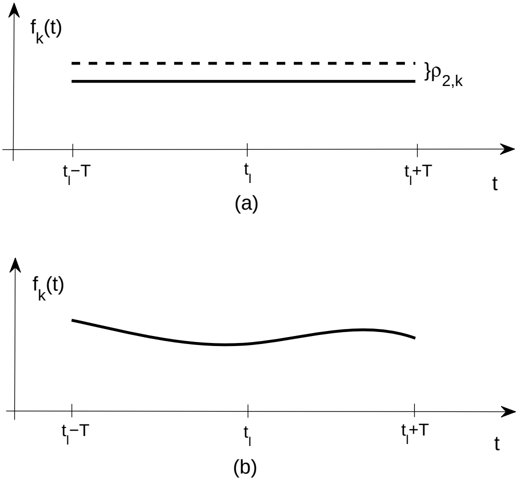
$$x(t) = \left(\sum_{k=-K}^{K} (a_k + tb_k) e^{j\tilde{\phi}_k(t)} \right) w(t)$$

where

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

is the estimated instantaneous phase.

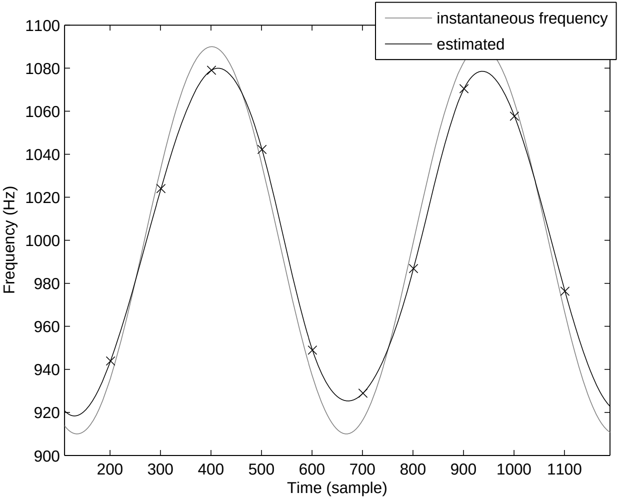
From QHM to aQHM; Graphically



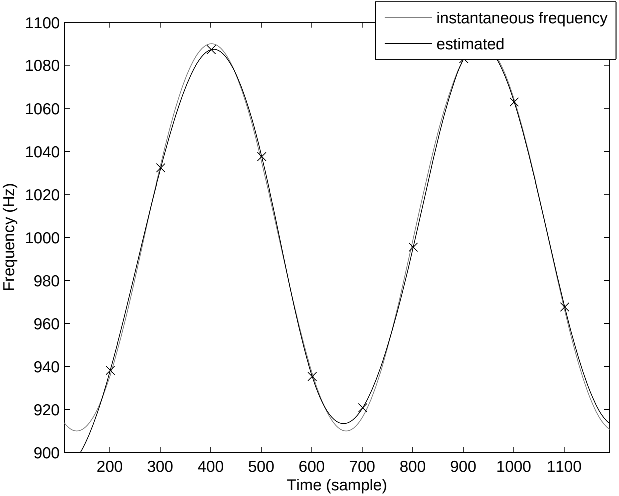
aQHM, in practice

- 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

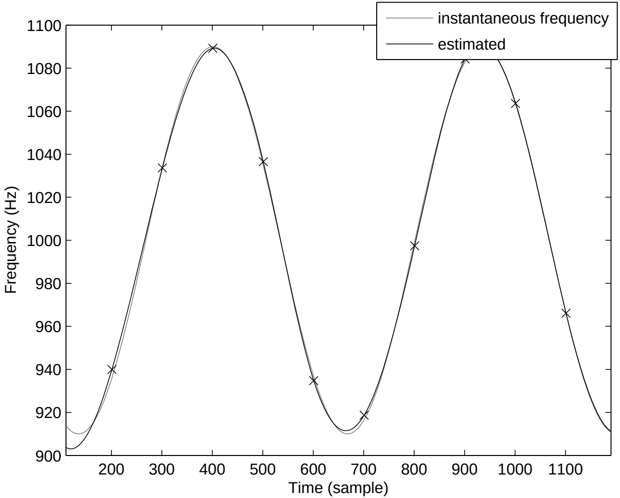
Example of estimation in aQHM: no iteration (QHM)



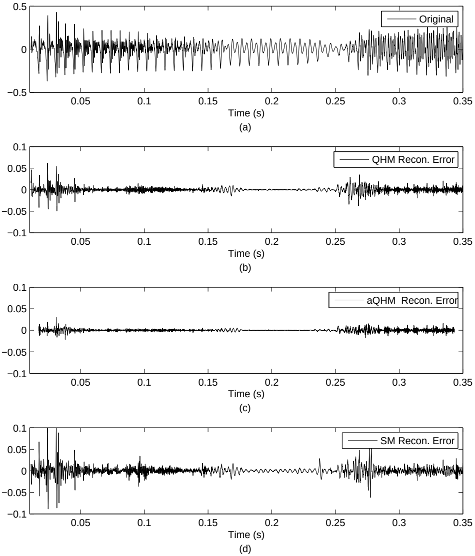
Example of estimation in aQHM: one iteration



Example of estimation in aQHM: two iterations



Reconstruction errors with QHM, aQHM, SM



Outline

Extended aQHM

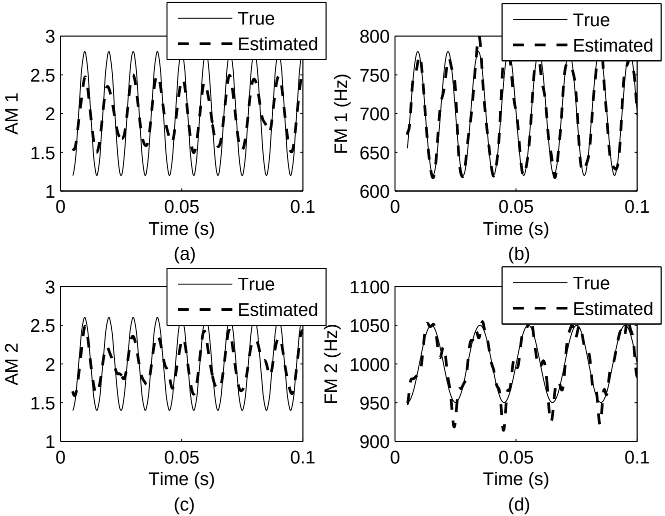
- Recall aQHM:

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

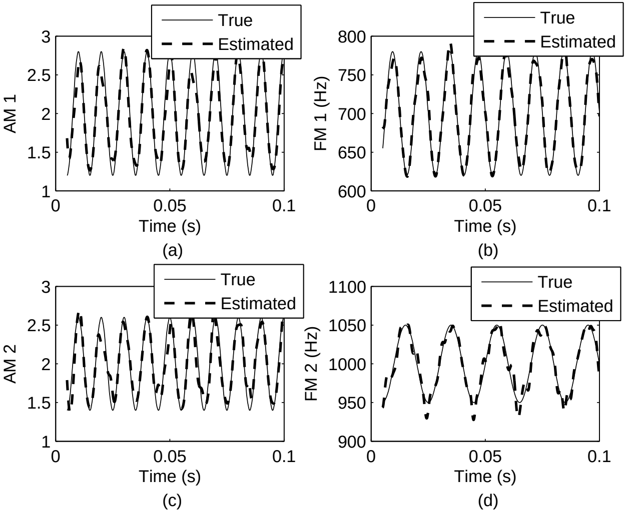
- Extended aQHM:

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

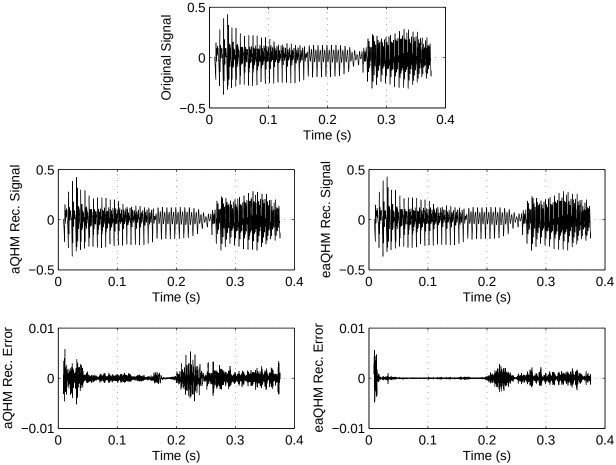
AM-FM modeling: aQHM



AM-FM modeling: extended aQHM



Comparing Adaptive Models



Outline

THANK YOU

for your attention

Outline

References I

Y. Stylianou, *Harmonic plus Noise Models for Speech, combined with Statistical Methods, for Speech and Speaker Modification.*
PhD thesis, Ecole Nationale Supèrieure des Télécommunications, Jan 1996.

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," in *Interspeech*, (Brisbane), Sep 2008.

Y. Pantazis, O. Rosec, and Y. Stylianou, "Iterative Estimation of Sinusoidal Signal Parameters," *IEEE Signal Processing Letters*, vol. 17, no. 5, pp. 461–464, 2010.

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