Speech Modifications

Yannis Stylianou

Outline of the talk

QHM

Adaptive QHM

Noise part

A/S with aQHNM

Audio Examples

Discussion

Speech Modifications

Lecture 4: Adaptive Quasi-Harmonic Modeling, aQHM

Yannis Stylianou

Telecommunications and Network Lab, ICS- FORTH
styliano@ics.forth.gr

Vitoria, 2010 Sept 2nd

1. QHM
   - Introduction
   - Robustness of QHM
   - Notes
2. Adaptive QHM
   - Validation
   - AM-FM
   - aQHM and Speech
3. Noise part
4. A/S with aQHNM
   - Synthesis
5. Audio Examples
   - Notes
6. Discussion

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])
  - 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])
  - 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])
  - 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])
  - 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:

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

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

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

Introduce the Quasi-Harmonic Model, QHM

Speech Modifications

Yannis Stylianou

Outline of the talk

QHM
- Introduction
- Robustness of QHM
- Notes

Adaptive QHM

Noise part

A/S with aQHNM

Audio Examples

Discussion

- 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 1989, Pantazis et al. 2008]:

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

Introduce the Quasi-Harmonic Model, QHM

- 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 1989, Pantazis et al. 2008]:

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

A few details of QHM

- QHM in the frequency domain:

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

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

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

- 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 *kth* component of QHM

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

A few details of QHM

- QHM in the frequency domain:

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

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

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

- 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 *kth* component of QHM

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

A few details of QHM

- QHM in the frequency domain:

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

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

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

- 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 *kth* component of QHM

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

A few details of QHM

- QHM in the frequency domain:

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

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

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

- 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 *kth* component of QHM

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

A few details of QHM

- QHM in the frequency domain:

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

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

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

- 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 *kth* component of QHM

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

Approximations in QHM

- we found previously:

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

- Back to the time-domain:

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

- in other words, **we suggest**:

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

Approximations in QHM

- we found previously:

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

- Back to the time-domain:

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

- in other words, **we suggest**:

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

Approximations in QHM

Speech Modifications

Yannis Stylianou

Outline of the talk

QHM
Introduction
Robustness of QHM
Notes

Adaptive QHM

Noise part

A/S with aQHNM

Audio Examples

Discussion

- we found previously:

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

- Back to the time-domain:

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

- in other words, **we suggest**:

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

Influence of the Window

- We would like small value for $W''(f)$ at f_k
- $W''(f)$ is influenced by the length, T, of the analysis window, i.e., for rectangular window: $W''(f) \propto T^3$
- So ... we would like a **short analysis window**
- But ... **long analysis window** provides robustness
- Length of the window $\rightleftarrows$ bandwidth

Influence of the Window

- We would like small value for $W''(f)$ at f_k
- $W''(f)$ is influenced by the length, T, of the analysis window, i.e., for rectangular window: $W''(f) \propto T^3$
- So ... we would like a **short analysis window**
- But ... **long analysis window** provides robustness
- Length of the window $\rightleftarrows$ bandwidth

Influence of the Window

- We would like small value for $W''(f)$ at f_k
- $W''(f)$ is influenced by the length, T, of the analysis window, i.e., for rectangular window: $W''(f) \propto T^3$
- So ... we would like a **short analysis window**
- But ... **long analysis window** provides robustness
- Length of the window $\rightleftarrows$ bandwidth

Influence of the Window

- We would like small value for $W''(f)$ at f_k
- $W''(f)$ is influenced by the length, T, of the analysis window, i.e., for rectangular window: $W''(f) \propto T^3$
- So ... we would like a **short analysis window**
- But ... **long analysis window** provides robustness
- Length of the window $\rightleftarrows$ bandwidth

Influence of the Window

- We would like small value for $W''(f)$ at f_k
- $W''(f)$ is influenced by the length, T, of the analysis window, i.e., for rectangular window: $W''(f) \propto T^3$
- So ... we would like a **short analysis window**
- But ... **long analysis window** provides robustness
- Length of the window $\rightleftarrows$ bandwidth

Influence of frequency mismatch

- Assume

$$x(t) = \alpha e^{j(2\pi \hat{f} t + \eta t)}$$

- modeled by QHM as

$$s(t) = (a + tb)e^{j(2\pi \hat{f} t)} \quad -T \leq t \leq T$$

- LS solution provides:

$$a = \alpha \frac{sin(\eta T)}{\eta T}$$

$$b = \alpha\ 3j\left(\frac{sin(\eta T)}{\eta^2 T^3} - \frac{cos(\eta T)}{\eta T^2}\right)$$

- Frequency mismatch estimate:

$$\hat{\eta} = 3\left(\frac{1}{\eta T^2} - \frac{cot(\eta T)}{T}\right)$$

Influence of frequency mismatch

- Assume

$$x(t) = \alpha e^{j(2\pi \hat{f} t + \eta t)}$$

- modeled by QHM as

$$s(t) = (a + tb)e^{j(2\pi \hat{f} t)} \quad -T \leq t \leq T$$

- LS solution provides:

$$a = \alpha \frac{sin(\eta T)}{\eta T}$$

$$b = \alpha \, 3j(\frac{sin(\eta T)}{\eta^2 T^3} - \frac{cos(\eta T)}{\eta T^2})$$

- Frequency mismatch estimate:

$$\hat{\eta} = 3\left(\frac{1}{\eta T^2} - \frac{cot(\eta T)}{T}\right)$$

Influence of frequency mismatch

- Assume

$$x(t) = \alpha e^{j(2\pi \hat{f} t + \eta t)}$$

- modeled by QHM as

$$s(t) = (a + tb)e^{j(2\pi \hat{f} t)} \quad -T \leq t \leq T$$

- LS solution provides:

$$a = \alpha \frac{sin(\eta T)}{\eta T}$$

$$b = \alpha \; 3j\left(\frac{sin(\eta T)}{\eta^2 T^3} - \frac{cos(\eta T)}{\eta T^2}\right)$$

- Frequency mismatch estimate:

$$\hat{\eta} = 3\left(\frac{1}{\eta T^2} - \frac{cot(\eta T)}{T}\right)$$

Influence of frequency mismatch

- Assume

$$x(t) = \alpha e^{j(2\pi \hat{f} t + \eta t)}$$

- modeled by QHM as

$$s(t) = (a + tb) e^{j(2\pi \hat{f} t)} \quad -T \leq t \leq T$$

- LS solution provides:

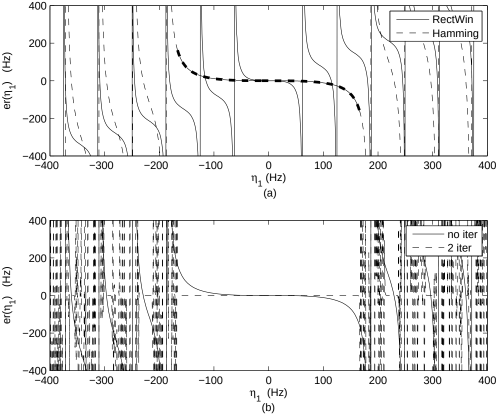
$$a = \alpha \frac{sin(\eta T)}{\eta T}$$

$$b = \alpha \ 3j\left(\frac{sin(\eta T)}{\eta^2 T^3} - \frac{cos(\eta T)}{\eta T^2}\right)$$

- Frequency mismatch estimate:

$$\hat{\eta} = 3\left(\frac{1}{\eta T^2} - \frac{cot(\eta T)}{T}\right)$$

Combining Influences



- 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:

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

where $v(t)$: complex additive white gaussian noise

Sinusoid	1st	2nd	3rd	4th
Frequency (Hz)	100	200	1000	2000
Amplitude	$e^{j\pi/10}$	$e^{j\pi/4}$	$e^{j\pi/3}$	$e^{j\pi/5}$
Freq. Mismatch (Hz)	± 10	± 10	± 100	± 100

Robustness against additive noise

- Signal contaminated by noise:

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

where $v(t)$: complex additive white gaussian noise

-

Sinusoid	1st	2nd	3rd	4th
Frequency (Hz)	100	200	1000	2000
Amplitude	$e^{j\pi/10}$	$e^{j\pi/4}$	$e^{j\pi/3}$	$e^{j\pi/5}$
Freq. Mismatch (Hz)	± 10	± 10	± 100	± 100

Evaluation

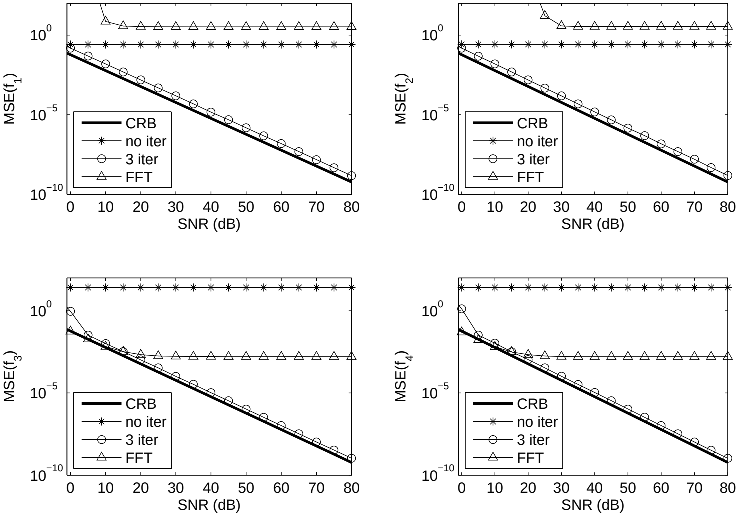
- 10000 Monte Carlo simulations
- Mean Squared Error (MSE):

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

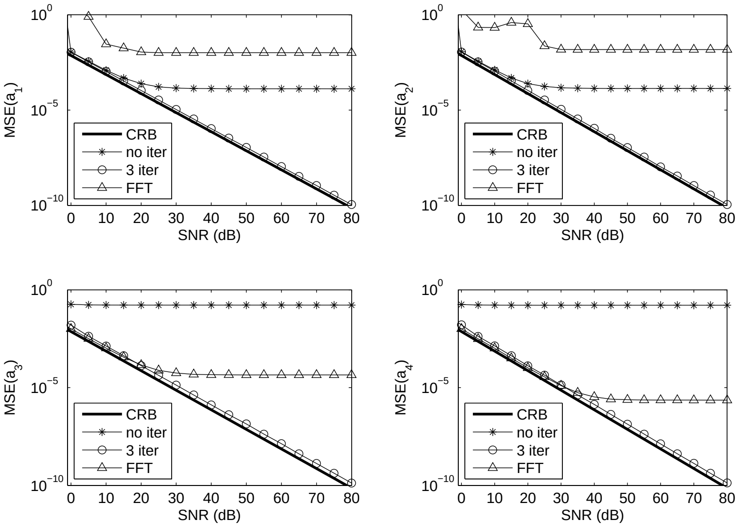
$$MSE\{\hat{a}_k\} = \frac{1}{M}\sum_{i=1}^{M} |\hat{a}_k(i) - a_k|^2$$

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

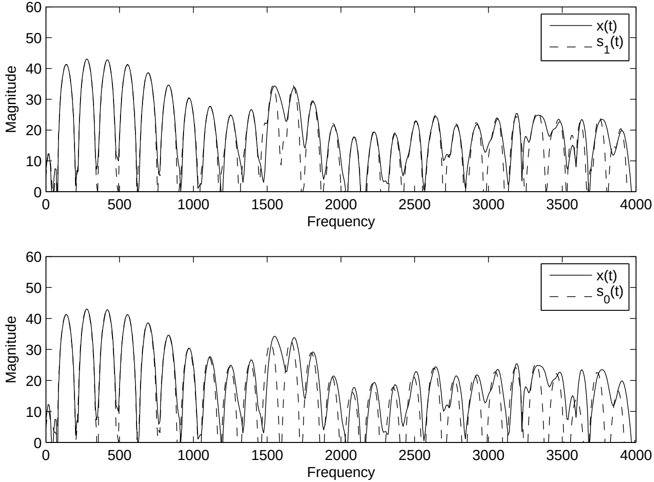
MSE of frequencies as a function of SNR.



MSE of amplitudes as a function of SNR.



Harmonic/Quasi-Harmonic representation for Speech



Notes on QHM

- The approximation made in QHM is valid provided that the frequency mismatch lies in a specific interval.
- This interval is a function of the bandwidth of the analysis window.
- Robustness of QHM against noise was tested and verified.
- It can be shown that iterative QHM is equivalent to (an approximate) Gauss-Newton method.

Notes on QHM

- The approximation made in QHM is valid provided that the frequency mismatch lies in a specific interval.
- This interval is a function of the bandwidth of the analysis window.
- Robustness of QHM against noise was tested and verified.
- It can be shown that iterative QHM is equivalent to (an approximate) Gauss-Newton method.

Notes on QHM

- The approximation made in QHM is valid provided that the frequency mismatch lies in a specific interval.
- This interval is a function of the bandwidth of the analysis window.
- Robustness of QHM against noise was tested and verified.
- It can be shown that iterative QHM is equivalent to (an approximate) Gauss-Newton method.

Notes on QHM

Speech Modifications

Yannis Stylianou

Outline of the talk

QHM
Introduction
Robustness of QHM
Notes

Adaptive QHM

Noise part

A/S with aQHNM

Audio Examples

Discussion

- The approximation made in QHM is valid provided that the frequency mismatch lies in a specific interval.
- This interval is a function of the bandwidth of the analysis window.
- Robustness of QHM against noise was tested and verified.
- It can be shown that iterative QHM is equivalent to (an approximate) Gauss-Newton method.

From QHM to Adaptive QHM, aQHM

- QHM (**stationarity assumption**):

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

- Adaptive QHM (aQHM):

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

where

$$\tilde{\phi}_k(t) = 2\pi \int_0^t f_k(u) du, \;\; t \in [0, T]$$

is the estimated instantaneous phase.

From QHM to Adaptive QHM, aQHM

- QHM (**stationarity assumption**):

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

- Adaptive QHM (aQHM):

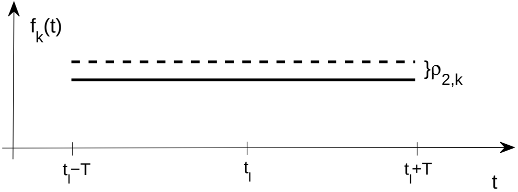
$$h(t) = \left(\sum_{k=-K}^{K} (a_k + t b_k) e^{j \tilde{\phi}_k(t)} \right) w(t)$$

where

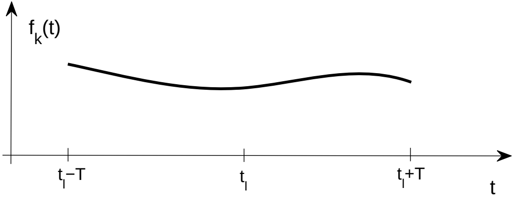
$$\tilde{\phi}_k(t) = 2\pi \int_0^t f_k(u) du, \;\; t \in [0, T]$$

is the estimated instantaneous phase.

From QHM to aQHM; Graphically



(a)



(b)

Frame Rate in aQHM

- 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

Frame Rate in aQHM

- 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

Frame Rate in aQHM

- 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

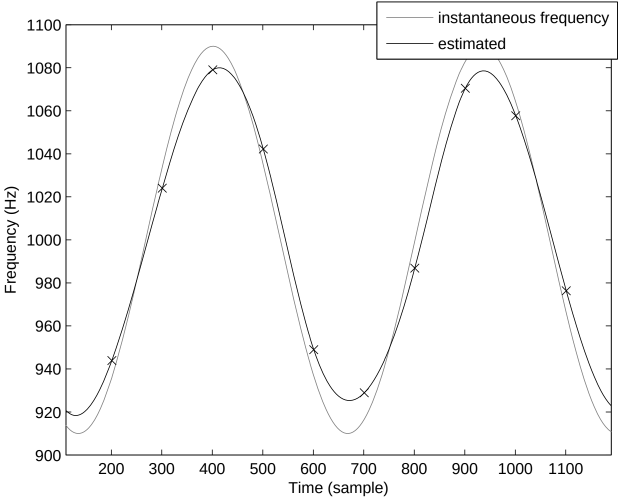
Frame Rate in aQHM

- 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

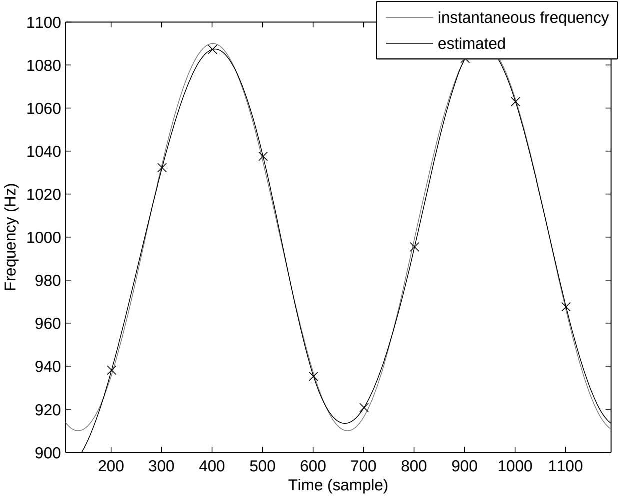
Frame Rate in aQHM

- 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

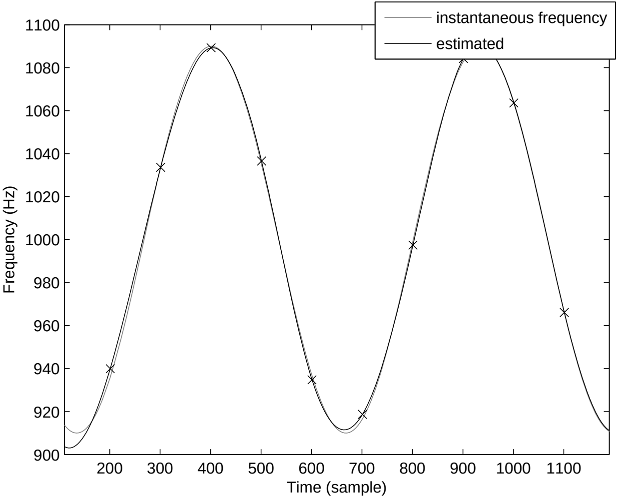
Example of estimation in aQHM: no iteration (QHM)



Example of estimation in aQHM: one iteration



Example of estimation in aQHM: two iterations



Validation: Example 1

- Let us consider as example:

$$x(t) = (11 - 340t + 4000t^2)e^{j2\pi(280t + 19500t^2)}$$

- and frequency mismatch: $|f_k - \zeta_k| = 35Hz$, using Hamming window
- Consider simplified approaches: Sinusoidal-based analysis (SM)
- Mean Absolute Error (MAE) [frame rate: one sample]

	AM	FM (Hz)
QHM, 10ms	0.46	2.15
QHM, 5ms	0.04	2.39
aQHM (1)	0.008	0.006
SM, 10ms	0.44	3.08

Validation: Example 1

- Let us consider as example:

$$x(t) = (11 - 340t + 4000t^2)e^{j2\pi(280t+19500t^2)}$$

- and frequency mismatch: $|f_k - \zeta_k| = 35Hz$, using Hamming window
- Consider simplified approaches: Sinusoidal-based analysis (SM)
- Mean Absolute Error (MAE) [frame rate: one sample]

	AM	FM (Hz)
QHM, 10ms	0.46	2.15
QHM, 5ms	0.04	2.39
aQHM (1)	0.008	0.006
SM, 10ms	0.44	3.08

Validation: Example 1

- Let us consider as example:

$$x(t) = (11 - 340t + 4000t^2)e^{j2\pi(280t+19500t^2)}$$

- and frequency mismatch: $|f_k - \zeta_k| = 35Hz$, using Hamming window
- Consider simplified approaches: Sinusoidal-based analysis (SM)
- Mean Absolute Error (MAE) [frame rate: one sample]

	AM	FM (Hz)
QHM, 10ms	0.46	2.15
QHM, 5ms	0.04	2.39
aQHM (1)	0.008	0.006
SM, 10ms	0.44	3.08

Validation: Example 1

- Let us consider as example:

$$x(t) = (11 - 340t + 4000t^2)e^{j2\pi(280t + 19500t^2)}$$

- and frequency mismatch: $|f_k - \zeta_k| = 35Hz$, using Hamming window
- Consider simplified approaches: Sinusoidal-based analysis (SM)
- Mean Absolute Error (MAE) [frame rate: one sample]

	AM	FM (Hz)
QHM, 10ms	0.46	2.15
QHM, 5ms	0.04	2.39
aQHM (1)	0.008	0.006
SM, 10ms	0.44	3.08

Validation: Example 2

- Let us consider as example:

$$s(t) = 2(1 + 0.4cos(2\pi 30t))e^{j(2\pi 700t + cos(2\pi 130t))} + 2(1 + 0.3cos(2\pi 50t))e^{j(2\pi 1000t + cos(2\pi 130t))}$$

- and frequency mismatch: $|f_k - \zeta_k| = 35Hz$, using Hamming window (16ms)
- Mean Absolute Error (MAE) [frame rate: one sample]

	AM1	AM2	FM1 (Hz)	FM2 (Hz)
QHM	0.36	0.37	69.99	69.74
aQHM (15)	0.05	0.11	21.43	20.14
SM	0.30	0.35	83.50	75.43

Validation: Example 2

- Let us consider as example:

$$s(t) = 2(1 + 0.4cos(2\pi 30t))e^{j(2\pi 700t + cos(2\pi 130t))} + 2(1 + 0.3cos(2\pi 50t))e^{j(2\pi 1000t + cos(2\pi 130t))}$$

- and frequency mismatch: $|f_k - \zeta_k| = 35Hz$, using Hamming window (16ms)
- Mean Absolute Error (MAE) [frame rate: one sample]

	AM1	AM2	FM1 (Hz)	FM2 (Hz)
QHM	0.36	0.37	69.99	69.74
aQHM (15)	0.05	0.11	21.43	20.14
SM	0.30	0.35	83.50	75.43

Validation: Example 2

- Let us consider as example:

$$s(t) = 2(1 + 0.4cos(2\pi 30t))e^{j(2\pi 700t + cos(2\pi 130t))} + 2(1 + 0.3cos(2\pi 50t))e^{j(2\pi 1000t + cos(2\pi 130t))}$$

- and frequency mismatch: $|f_k - \zeta_k| = 35Hz$, using Hamming window (16ms)
- Mean Absolute Error (MAE) [frame rate: one sample]

	AM1	AM2	FM1 (Hz)	FM2 (Hz)
QHM	0.36	0.37	69.99	69.74
aQHM (15)	0.05	0.11	21.43	20.14
SM	0.30	0.35	83.50	75.43

QHM reminder

- QHM:

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

- Instantaneous parameters:

$$\begin{aligned} A_k(t) &= \sqrt{(a_k^R + tb_k^R)^2 + (a_k^I + tb_k^I)^2} \\ \Phi_k(t) &= 2\pi f_k t + atan \frac{a_k^I + tb_k^I}{a_k^R + tb_k^R} \\ F_k(t) &= f_k + \frac{1}{2\pi} \frac{a_k^R b_k^I - a_k^I b_k^R}{M_k^2(t)} \end{aligned}$$

QHM REMINDER

Speech Modifications

Yannis Stylianou

Outline of the talk

QHM

Adaptive QHM

Validation

AM-FM

aQHM and Speech

Noise part

A/S with aQHNM

Audio Examples

Discussion

- QHM:

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

- Instantaneous parameters:

$$\begin{aligned} A_k(t) &= \sqrt{(a_k^R + t b_k^R)^2 + (a_k^I + t b_k^I)^2} \\ \Phi_k(t) &= 2\pi f_k t + atan \frac{a_k^I + t b_k^I}{a_k^R + t b_k^R} \\ F_k(t) &= f_k + \frac{1}{2\pi} \frac{a_k^R b_k^I - a_k^I b_k^R}{M_k^2(t)} \end{aligned}$$

QHM and AM-FM decomposition

- AM-FM signal

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

- 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:

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

QHM and AM-FM decomposition

- AM-FM signal

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

- 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:

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

QHM and AM-FM decomposition

- AM-FM signal

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

- 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:

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

QHM and AM-FM decomposition

- AM-FM signal

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

- 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:

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

Hypotheses and Suggestions

- Hypotheses:
  1. $\rho_{2,k}$ is small: frequency mismatch ($|f_k - \zeta_k|$) is small
  2. Window has large bandwidth (i.e., short-time window)
- Suggestions:

$$\hat{\phi}_{k,1} = \rho_{2,k}$$

Hypotheses and Suggestions

- Hypotheses:
  1. $\rho_{2,k}$ is small: frequency mismatch ($|f_k - \zeta_k|$) is small
  2. Window has large bandwidth (i.e., short-time window)
- Suggestions:

$$\hat{\phi}_{k,1} = \rho_{2,k}$$

Hypotheses and Suggestions

- Hypotheses:
  1. $\rho_{2,k}$ is small: frequency mismatch ($|f_k - \zeta_k|$) is small
  2. Window has large bandwidth (i.e., short-time window)
- Suggestions:

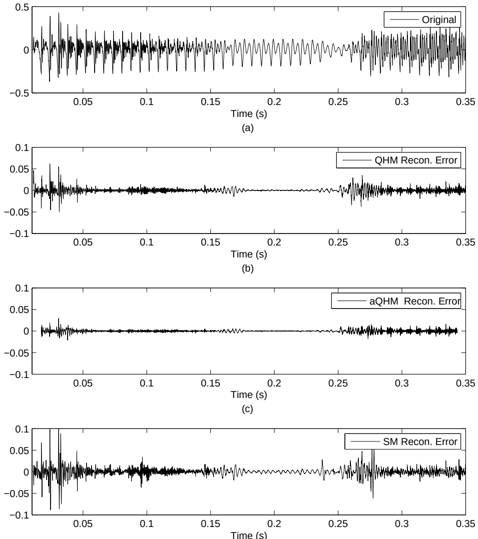
$$\hat{\phi}_{k,1} = \rho_{2,k}$$

Hypotheses and Suggestions

- Hypotheses:
  1. $\rho_{2,k}$ is small: frequency mismatch ($|f_k - \zeta_k|$) is small
  2. Window has large bandwidth (i.e., short-time window)
- Suggestions:

$$\hat{\phi}_{k,1} = \rho_{2,k}$$

Example of reconstruction errors with QHM, aQHM, SM



Speech Modifications
Yannis Stylianou
Outline of the talk
QHM
Adaptive QHM
Validation
AM-FM
aQHM and Speech
Noise part
A/S with aQHNM
Audio Examples
Discussion

- Signal Reconstruction

$$\hat{x}(t) = \sum_{k=1}^{K} \hat{A}_k(t) \cos\left(\hat{\Phi}_k(t)\right)$$

- Test on 5 minutes of 20 female and male voiced speech (TIMIT)
- Average Signal-to-Error Reconstruction Ratio (in dB)

	Male	Female
QHM	23.9	29.1
aQHM (3)	29.1	34.1
SM	17.2	21.1

aQHM: AM-FM decomposition of Voiced Speech

- Signal Reconstruction

$$\hat{x}(t) = \sum_{k=1}^{K} \hat{A}_k(t) \cos\left(\hat{\Phi}_k(t)\right)$$

- Test on 5 minutes of 20 female and male voiced speech (TIMIT)
- Average Signal-to-Error Reconstruction Ratio (in dB)

	Male	Female
QHM	23.9	29.1
aQHM (3)	29.1	34.1
SM	17.2	21.1

aQHM: AM-FM decomposition of Voiced Speech

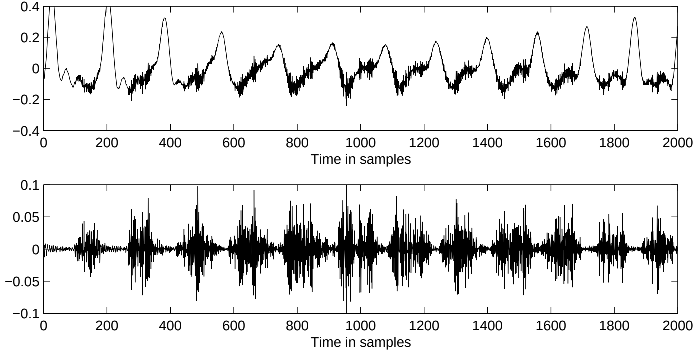
- Signal Reconstruction

$$\hat{x}(t) = \sum_{k=1}^{K} \hat{A}_k(t) \cos\left(\hat{\Phi}_k(t)\right)$$

- Test on 5 minutes of 20 female and male voiced speech (TIMIT)
- Average Signal-to-Error Reconstruction Ratio (in dB)

	Male	Female
QHM	23.9	29.1
aQHM (3)	29.1	34.1
SM	17.2	21.1

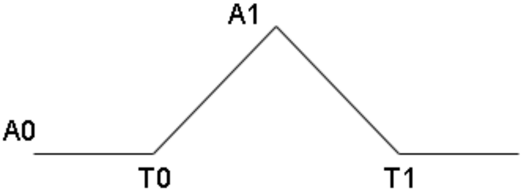
Noise part



0.4
0.2
0
−0.2
−0.4
0
200
400
600
800
1000
1200
1400
1600
1800
2000
Time in samples
0.1
0.05
0
−0.05
−0.1
0
200
400
600
800
1000
1200
1400
1600
1800
2000
Time in samples

A simple approach ...

So far in HNM, we mainly use the Triangular Envelope:



Signal envelope

There are many ways to obtain the “envelope” of a signal, as:

- Hilbert Transform (analytic signal)
- Low-pass local energy (energy envelope):

$$e[n] = \frac{1}{2N+1} \sum_{k=-N}^{N} |r[n-k]|$$

where $r[n]$ denotes the residual signal.

Signal envelope

There are many ways to obtain the “envelope” of a signal, as:

- Hilbert Transform (analytic signal)
- Low-pass local energy (energy envelope):

$$e[n] = \frac{1}{2N+1} \sum_{k=-N}^{N} |r[n-k]|$$

where $r[n]$ denotes the residual signal.

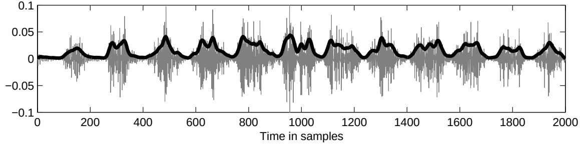
Hilbert envelope

We may also use the Hilbert envelope, computed as:

$$\tilde{e}_H[n] = \sum_{k=L-M+1}^{L} a_k e^{2\pi k (f_0/f_s) n}$$

Example of energy envelope

Example of Energy Envelope, with $N = 7$



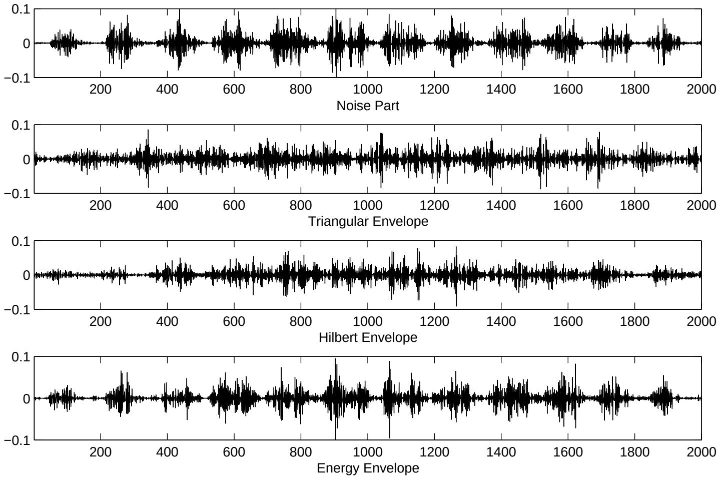
Energy envelope

The energy envelope can be efficiently parameterized with a few Fourier coefficients:

$$\hat{e}[n] = \sum_{k=-L_e}^{L_e} A_k e^{j2\pi k(f_0/f_s)n}$$

where L_e is set to be 3 to 4

Comparing Envelopes



Synthesis

- Deterministic Part: Interpolation of amplitudes, frequencies and phases much similar to aQHM analysis.
- Stochastic Part: Overlap Add (OLA) synthesis.

Audio Examples

	Original	SM	aQHNM
Male 1			
Male 2			
Female 1			
Female 2			

Notes

- An Analysis/Synthesis of speech, namely aQHNM, was presented.
- For the deterministic part, the local stationarity assumption is discarded.
- For the stochastic part, the time-domain envelope model captures the characteristics of noise.
- The reconstructed speech signal is perceptually indistinguishable from the original one.

Notes

- An Analysis/Synthesis of speech, namely aQHNM, was presented.
- For the deterministic part, the local stationarity assumption is discarded.
- For the stochastic part, the time-domain envelope model captures the characteristics of noise.
- The reconstructed speech signal is perceptually indistinguishable from the original one.

Notes

- An Analysis/Synthesis of speech, namely aQHNM, was presented.
- For the deterministic part, the local stationarity assumption is discarded.
- For the stochastic part, the time-domain envelope model captures the characteristics of noise.
- The reconstructed speech signal is perceptually indistinguishable from the original one.

Notes

- An Analysis/Synthesis of speech, namely aQHNM, was presented.
- For the deterministic part, the local stationarity assumption is discarded.
- For the stochastic part, the time-domain envelope model captures the characteristics of noise.
- The reconstructed speech signal is perceptually indistinguishable from the original one.

Discussion

- aQHM is a speech/signal analysis tool
- Transformations in the aQHM context is still an open question (e.g., Pitch modification)
- TTS and speech coding in the aQHM context
- Multiresolution Analysis for features estimation
- Multiresolution Analysis for speech modifications

Discussion

- aQHM is a speech/signal analysis tool
- Transformations in the aQHM context is still an open question (e.g., Pitch modification)
- TTS and speech coding in the aQHM context
- Multiresolution Analysis for features estimation
- Multiresolution Analysis for speech modifications

Discussion

- aQHM is a speech/signal analysis tool
- Transformations in the aQHM context is still an open question (e.g., Pitch modification)
- TTS and speech coding in the aQHM context
- Multiresolution Analysis for features estimation
- Multiresolution Analysis for speech modifications

Discussion

- aQHM is a speech/signal analysis tool
- Transformations in the aQHM context is still an open question (e.g., Pitch modification)
- TTS and speech coding in the aQHM context
- Multiresolution Analysis for features estimation
- Multiresolution Analysis for speech modifications

Discussion

- aQHM is a speech/signal analysis tool
- Transformations in the aQHM context is still an open question (e.g., Pitch modification)
- TTS and speech coding in the aQHM context
- Multiresolution Analysis for features estimation
- Multiresolution Analysis for speech modifications