SPE-L1.3

TANDEM-STRAIGHT:

A TEMPORALLY STABLE POWER SPECTRAL REPRESENTATION FOR PERIODIC SIGNALS AND APPLICATIONS TO INTERFERENCE-FREE SPECTRUM, F0, AND APERIODICITY ESTIMATION

Hideki Kawahara, Masanori Morise, Toru Takahashi
Ryuchi Nisimura, Toshio Irino and Hideki Banno[†]

Wakayama University, Wakayama Japan
[†]Meijyo University, Nagoya Japan

ICASSP 2008 Las Vegas, 30 March-4 April, 2008

Outline

- Reformulation of STRAIGHT
  (F0 adaptive time-frequency representation)
  - Power spectrum without temporal variations
    TANDEM
  - Consistent sampling for spectral envelope recovery
    STRAIGHT
- Unified approach based on TANDEM and STRAIGHT
  (F0 and aperiodicity detection)

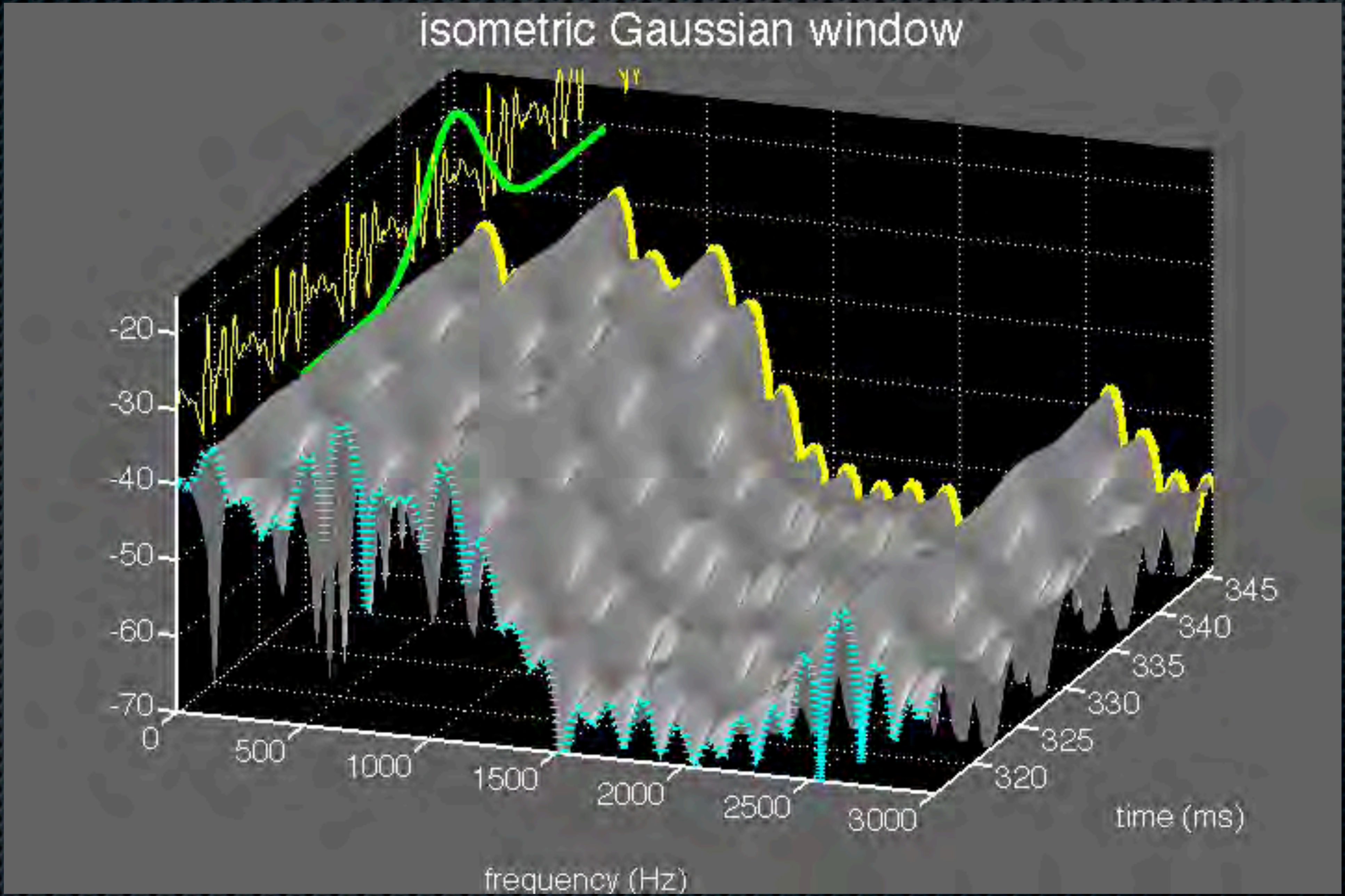
Outline

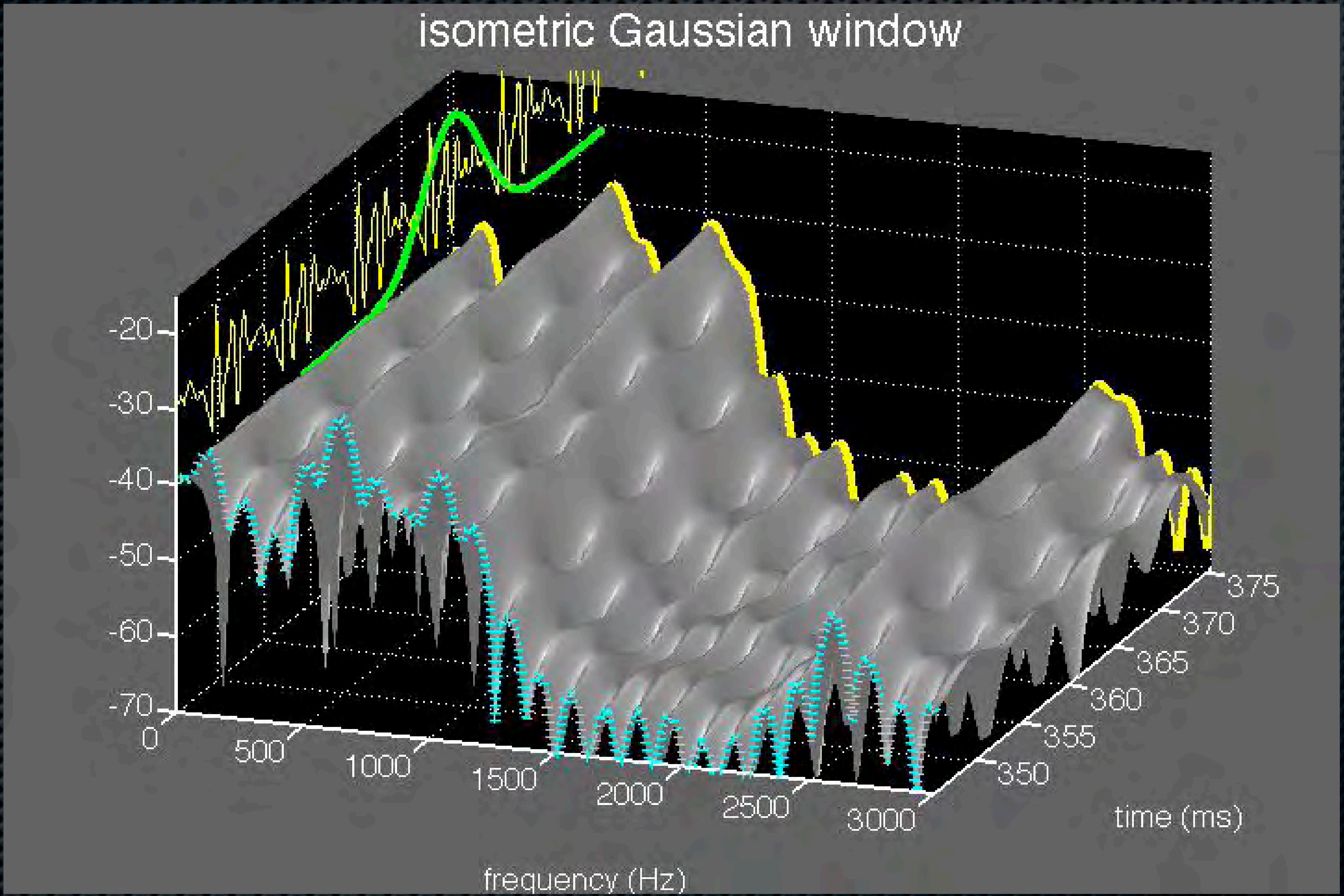
- Reformulation of STRAIGHT
  (F0 adaptive time-frequency representation)
  - Power spectrum without temporal variations
    TANDEM
  - Consistent sampling for spectral envelope recovery
    STRAIGHT
- Unified approach based on TANDEM and STRAIGHT
  (F0 and aperiodicity detection)

Outline

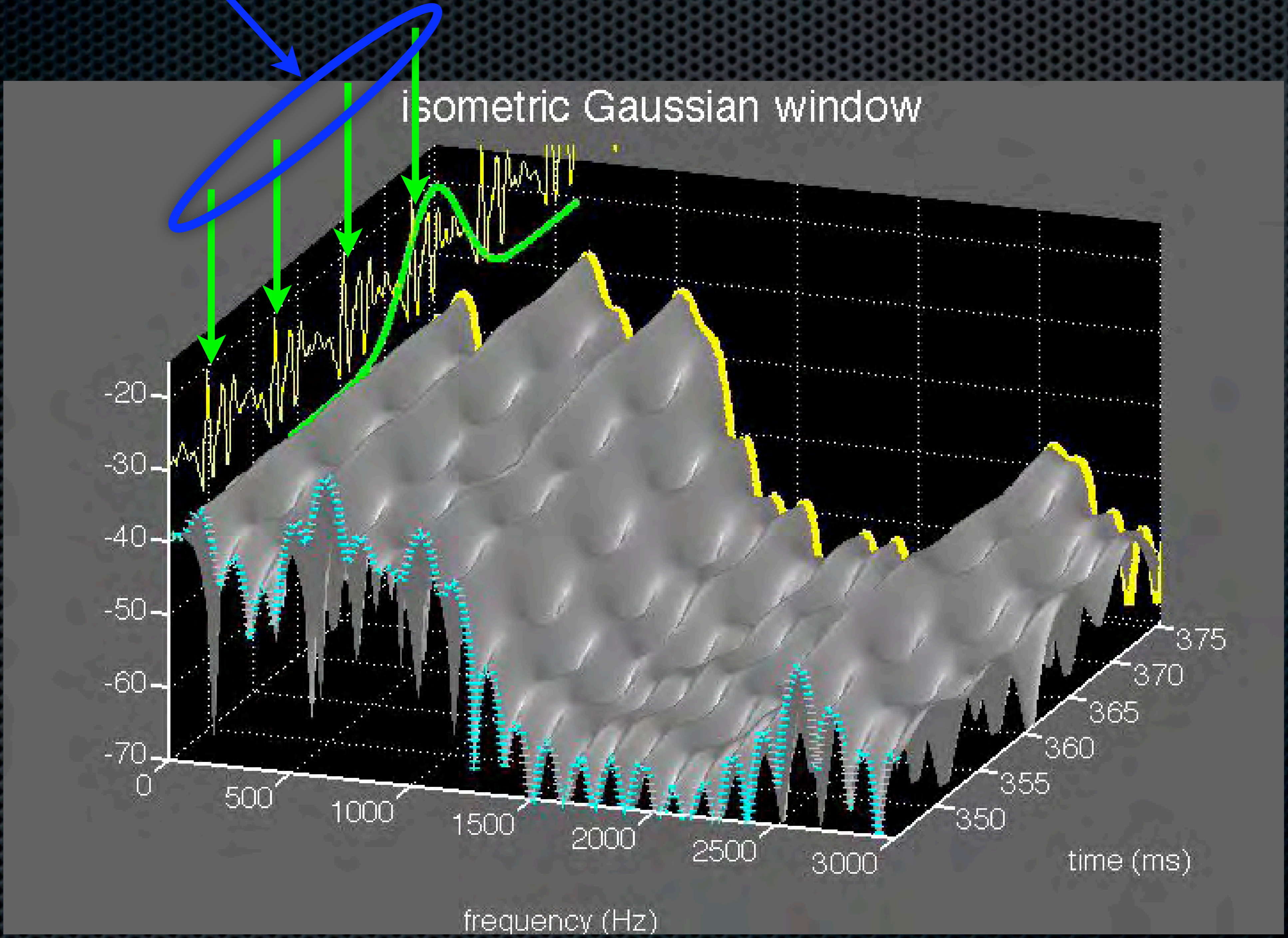
- Reformulation of STRAIGHT
  (F0 adaptive time-frequency representation)
  - Power spectrum without temporal variations
    TANDEM
  - Consistent sampling for spectral envelope recovery
    STRAIGHT
- Unified approach based on TANDEM and STRAIGHT
  (F0 and aperiodicity detection)

SFT-based spectrogram



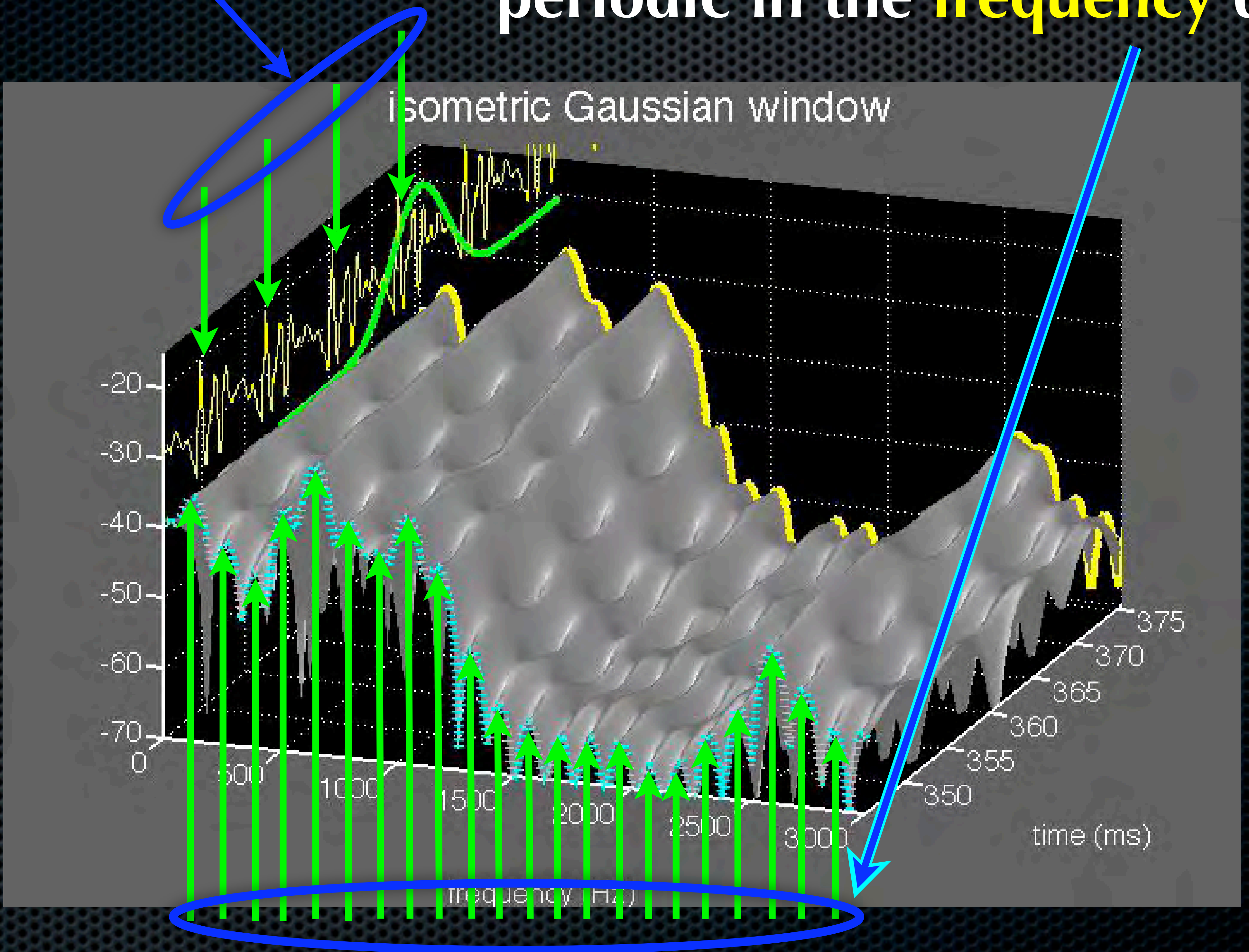

isometric Gaussian window
-20
-30
-40
-50
-60
-70
0
500
1000
1500
2000
2500
3000
frequency (Hz)
350
355
360
365
370
375
time (ms)

periodic in the time domain

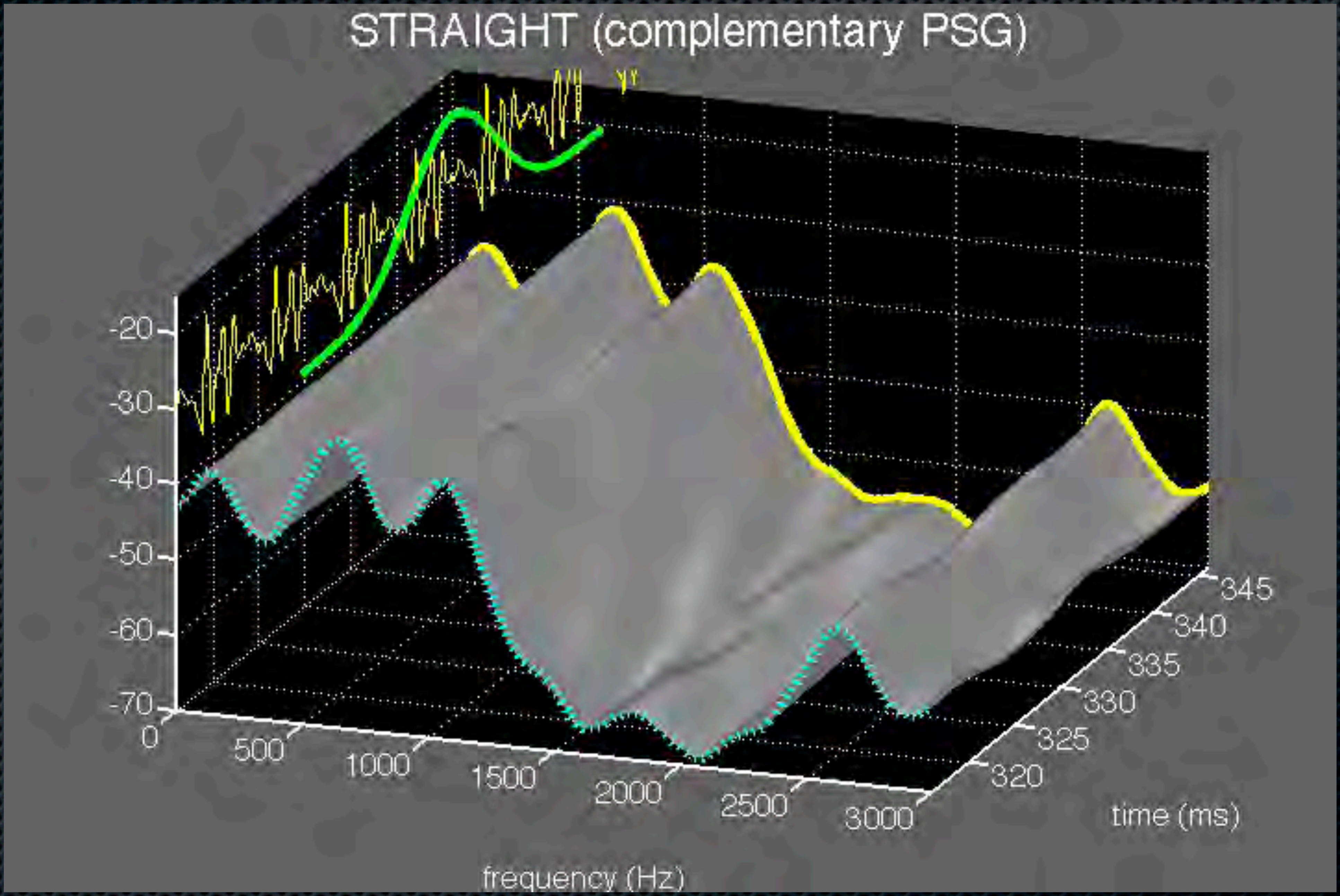


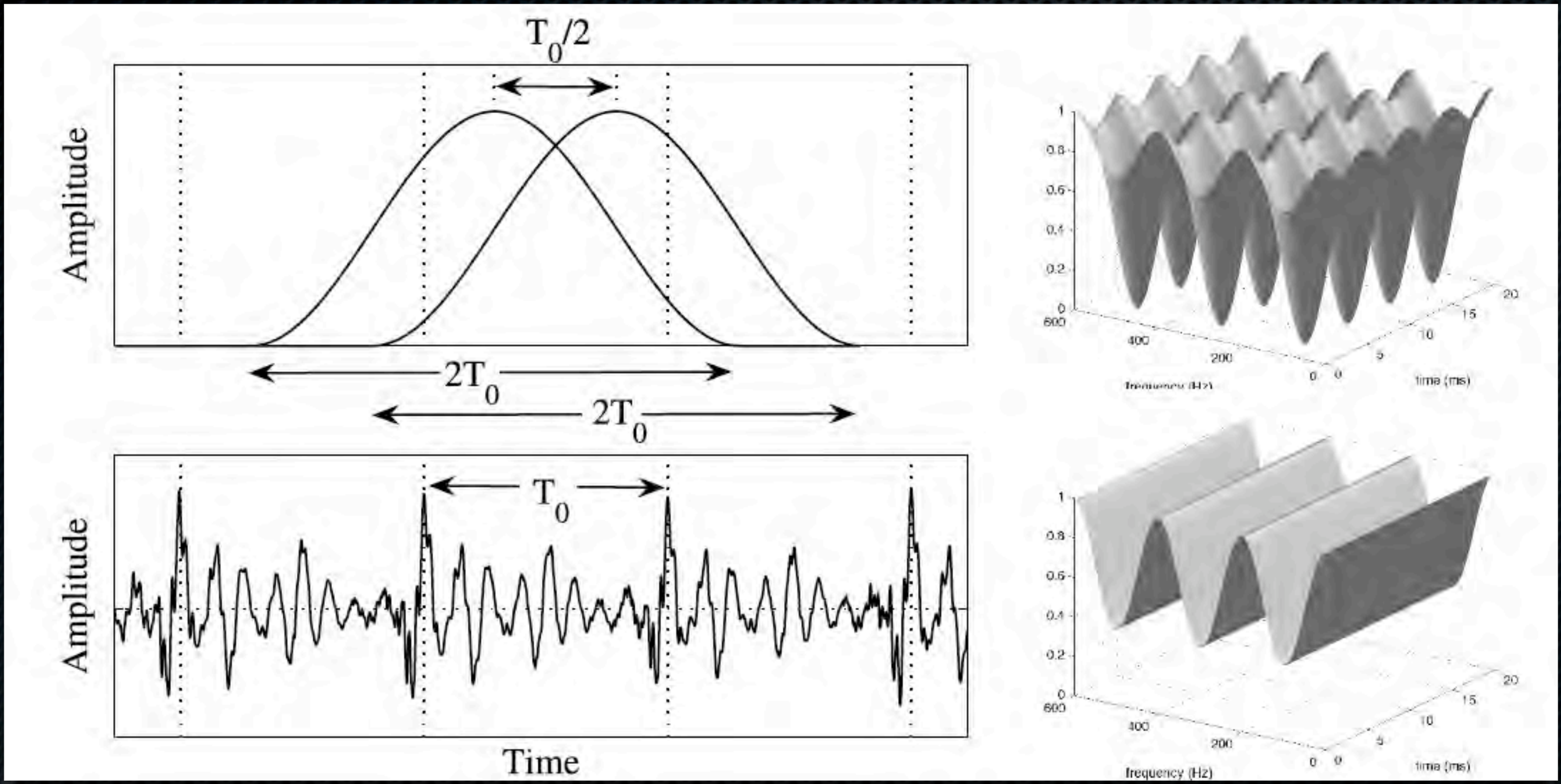
periodic in the **time** domain

periodic in the **frequency** domain



Spline-based optimum smoothing reconstructs the underlying smooth time-frequency representation





TANDEM spectrum

Power spectrum estimation without periodic variations

TANDEM: assumption

TANDEM: assumption

- Time window has low level side lobes.

TANDEM: assumption

- Time window has low level side lobes.
- Only two harmonic components exist in each equivalent band-pass filter.

TANDEM: assumption

- Time window has low level side lobes.
- Only two harmonic components exist in each equivalent band-pass filter.

$$\delta(\omega) + \alpha e^{j\beta}\delta(\omega - \omega_0)$$

$$\begin{aligned}|S(\omega,\tau)|^2 &= H^2(\omega) + \alpha^2 H^2(\omega - \omega_0) \\ &\quad + 2\alpha H(\omega) H(\omega - \omega_0)\cos(\omega_0\tau + \beta)\end{aligned}$$

$$|S(\omega,\tau)|^2 + |S(\omega,\tau + T_0/2)|^2$$

TANDEM: assumption

- Time window has low level side lobes.
- Only two harmonic components exist in each equivalent band-pass filter.

$$\delta(\omega) + \alpha e^{j\beta}\delta(\omega - \omega_0)$$

signal model

$$\begin{aligned} |S(\omega,\tau)|^2 = {} & H^2(\omega) + \alpha^2 H^2(\omega - \omega_0) \\ & + 2\alpha H(\omega) H(\omega - \omega_0) \cos(\omega_0 \tau + \beta) \end{aligned}$$

$$|S(\omega,\tau)|^2 + |S(\omega,\tau + T_0/2)|^2$$

TANDEM: assumption

- Time window has low level side lobes.
- Only two harmonic components exist in each equivalent band-pass filter.

signal model

$$\delta(\omega) + \alpha e^{j\beta}\delta(\omega - \omega_0)$$

power spectrum

$$|S(\omega,\tau)|^2 = H^2(\omega) + \alpha^2 H^2(\omega - \omega_0) + 2\alpha H(\omega) H(\omega - \omega_0)\cos(\omega_0\tau + \beta)$$

$$|S(\omega,\tau)|^2 + |S(\omega,\tau + T_0/2)|^2$$

TANDEM: assumption

- Time window has low level side lobes.
- Only two harmonic components exist in each equivalent band-pass filter.

signal model

$$\delta(\omega) + \alpha e^{j\beta}\delta(\omega - \omega_0)$$

power spectrum

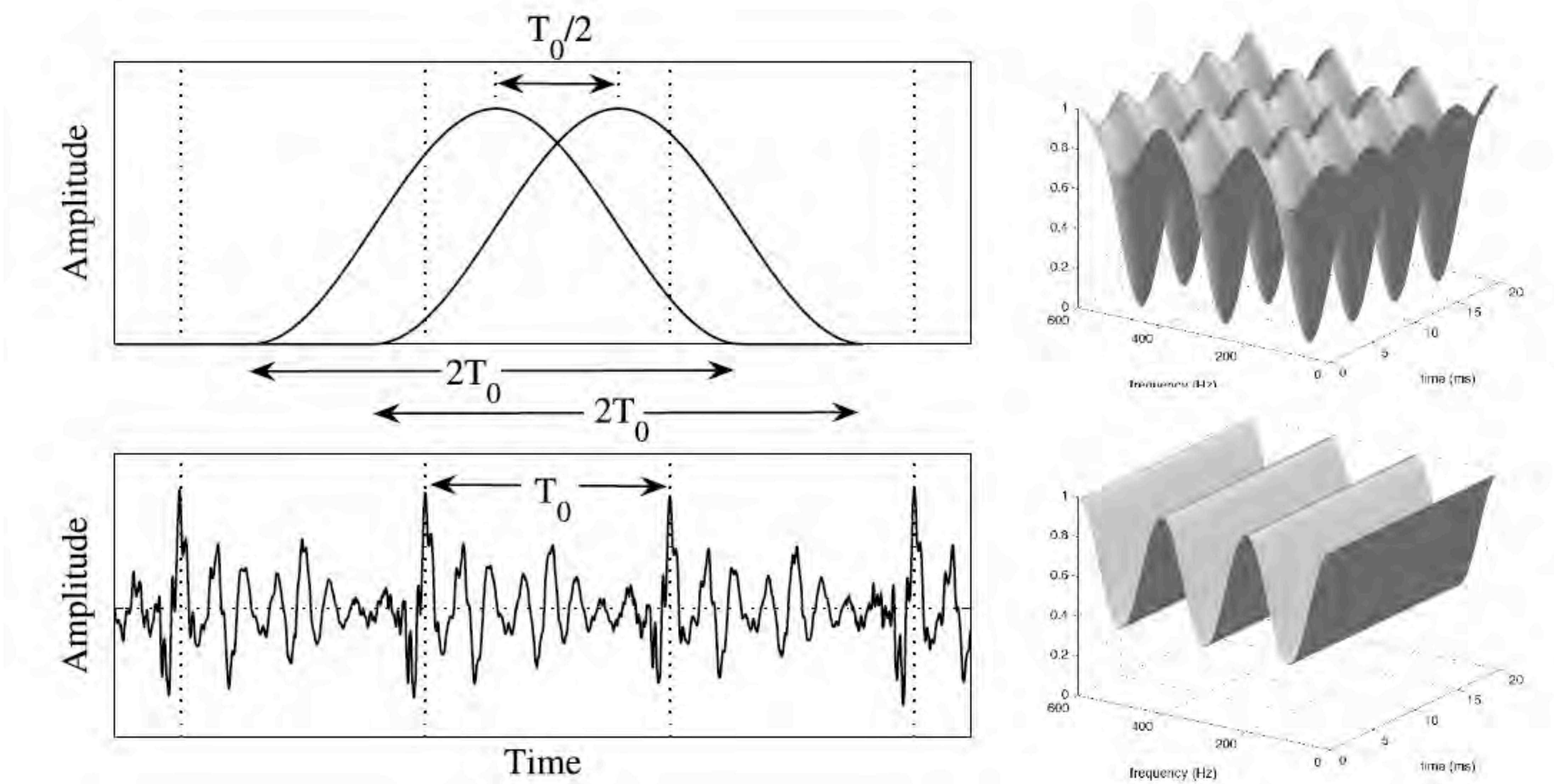
$$|S(\omega,\tau)|^2 = H^2(\omega) + \alpha^2 H^2(\omega - \omega_0) + \boxed{2\alpha H(\omega)H(\omega - \omega_0)\cos(\omega_0\tau + \beta)}$$

periodic variation

$$|S(\omega,\tau)|^2 + |S(\omega,\tau + T_0/2)|^2$$

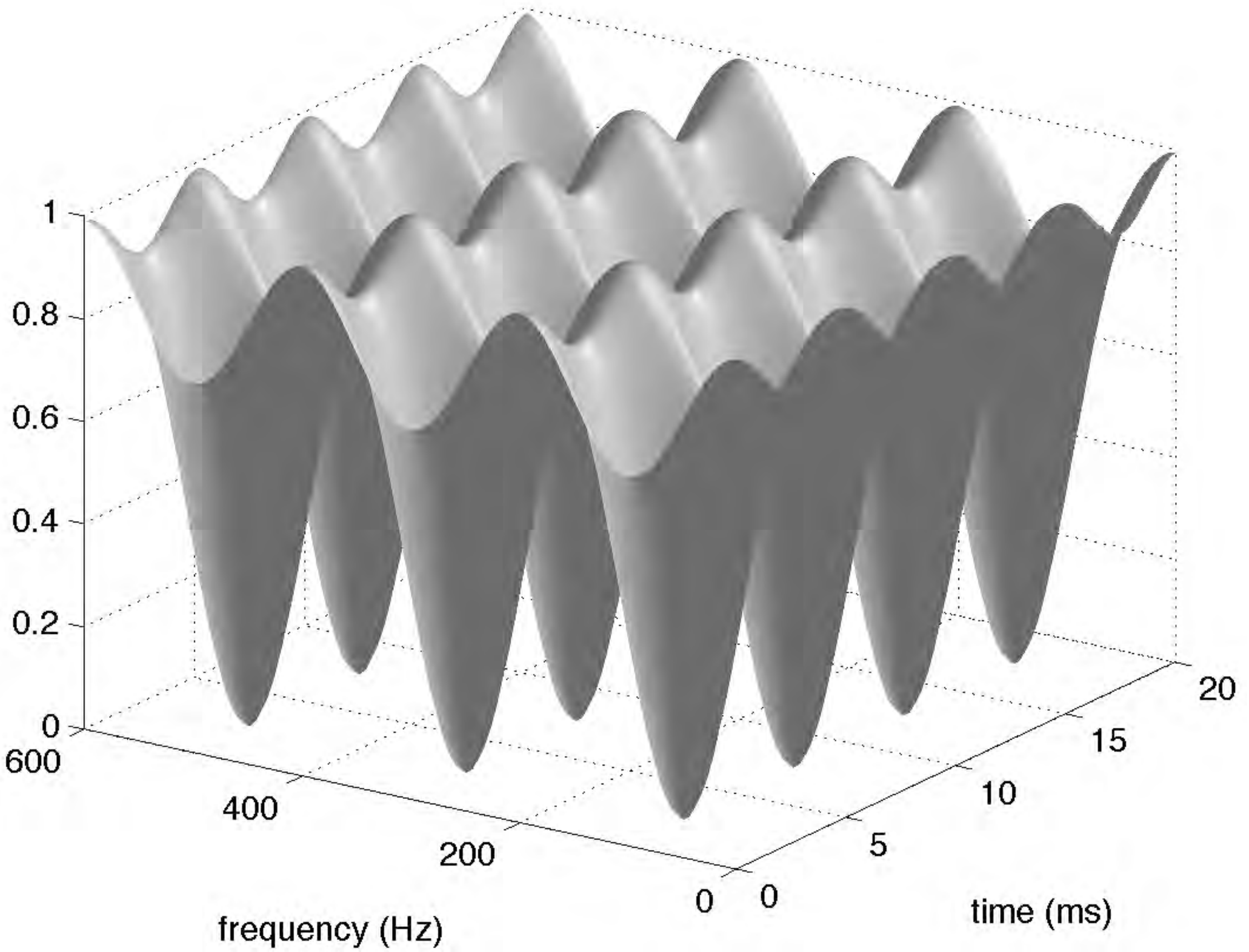
TANDEM spectrum

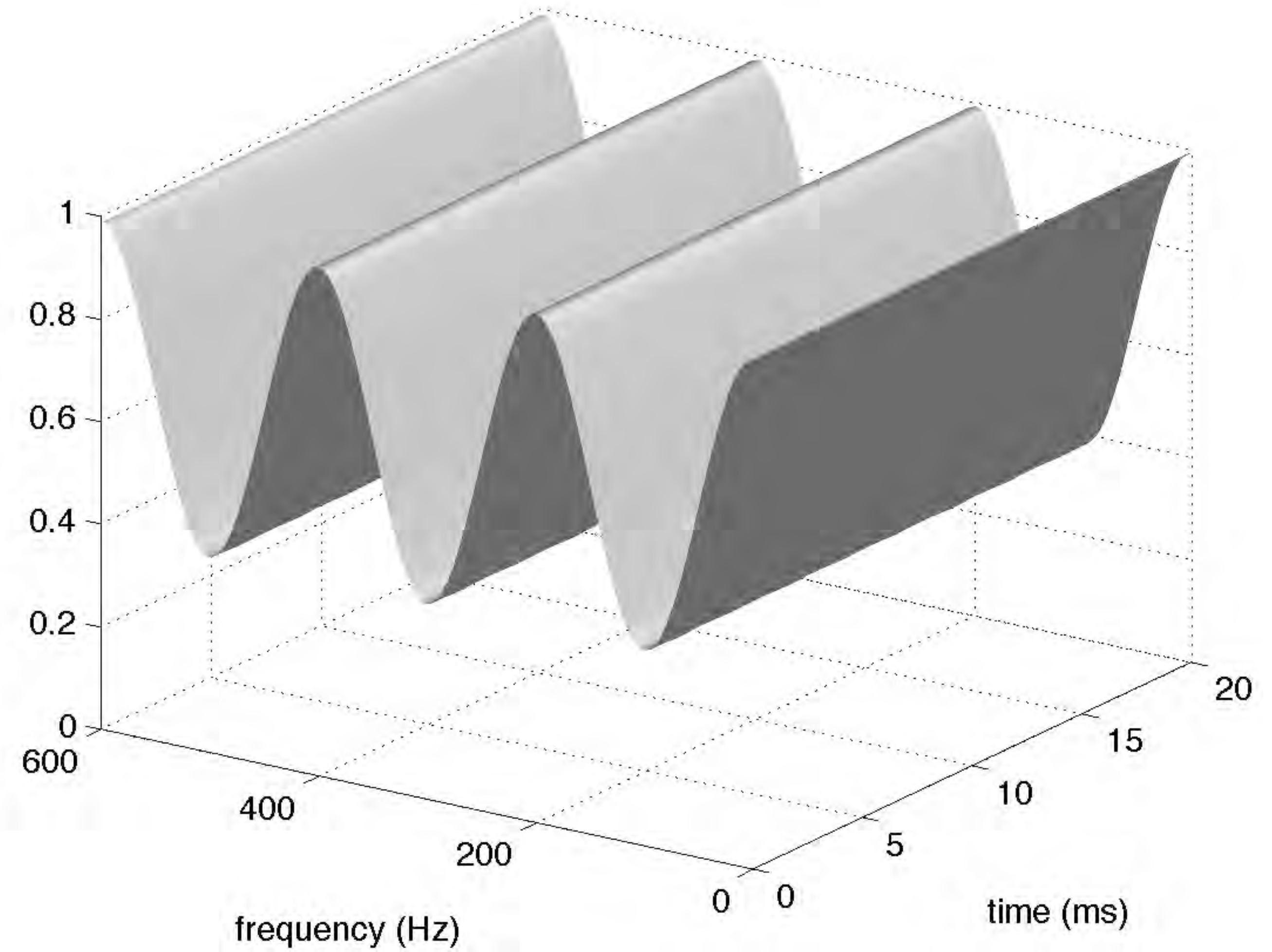
- Calculate power spectra of T0/2 apart and then average them.

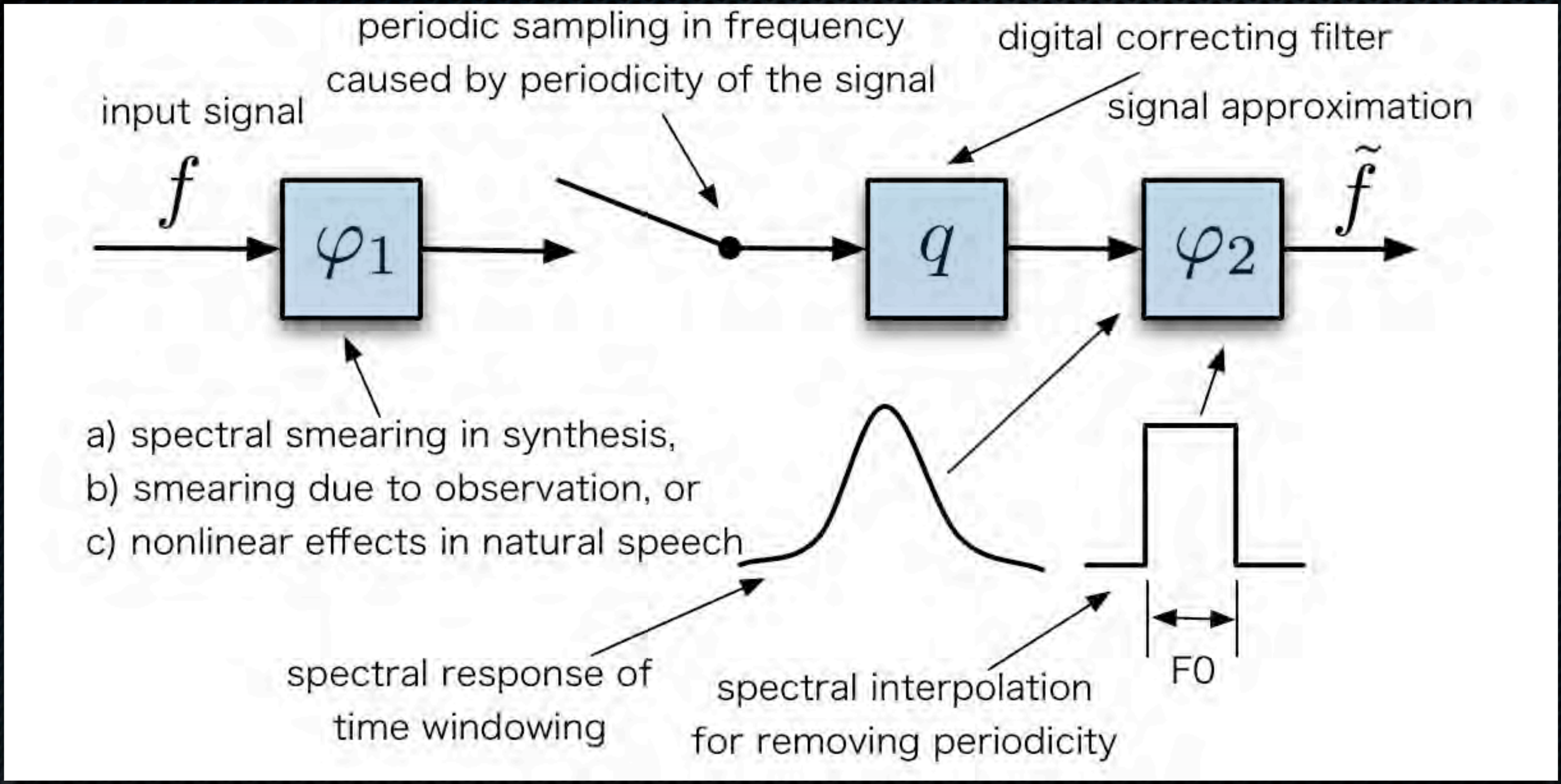


TANDEM spectrum

- Calculate power spectra of T0/2 apart and then average them.





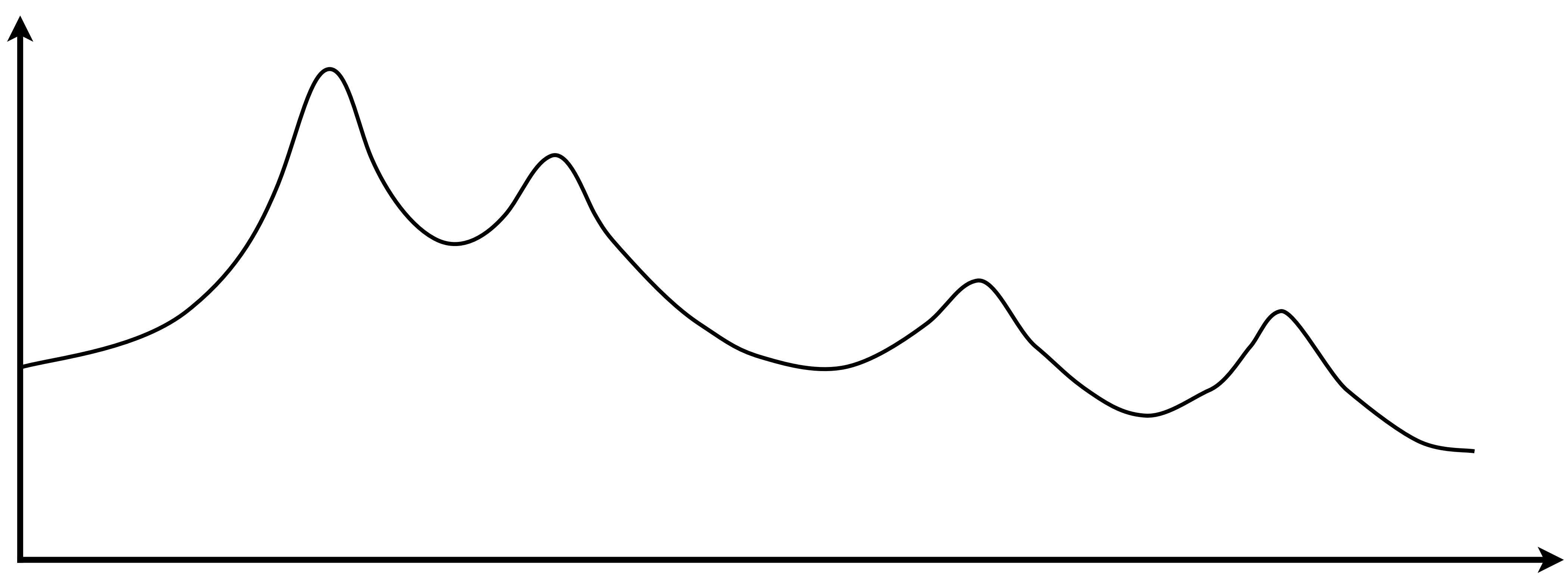


STRAIGHT spectrum

Reformulation based on consistent sampling

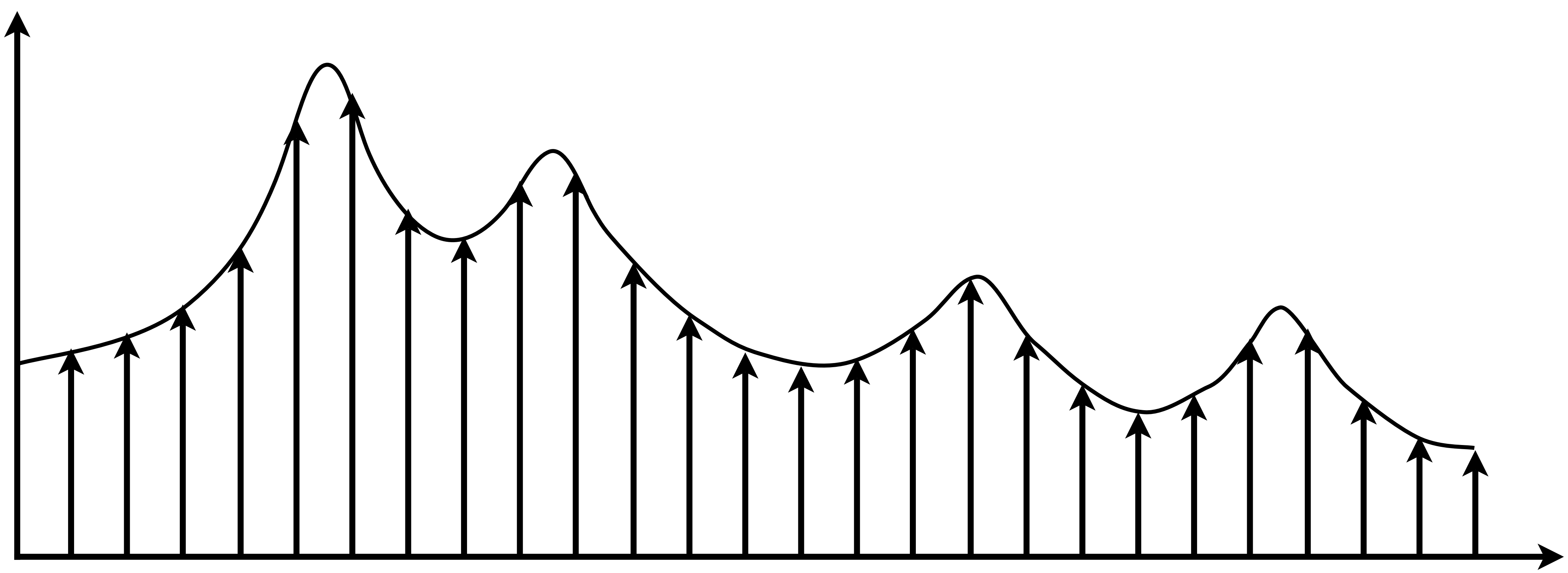
Vowel spectral model

- Periodic sampling also in the frequency domain
  - Continuous spectrum has to be recovered from sampled values (D/A conversion)



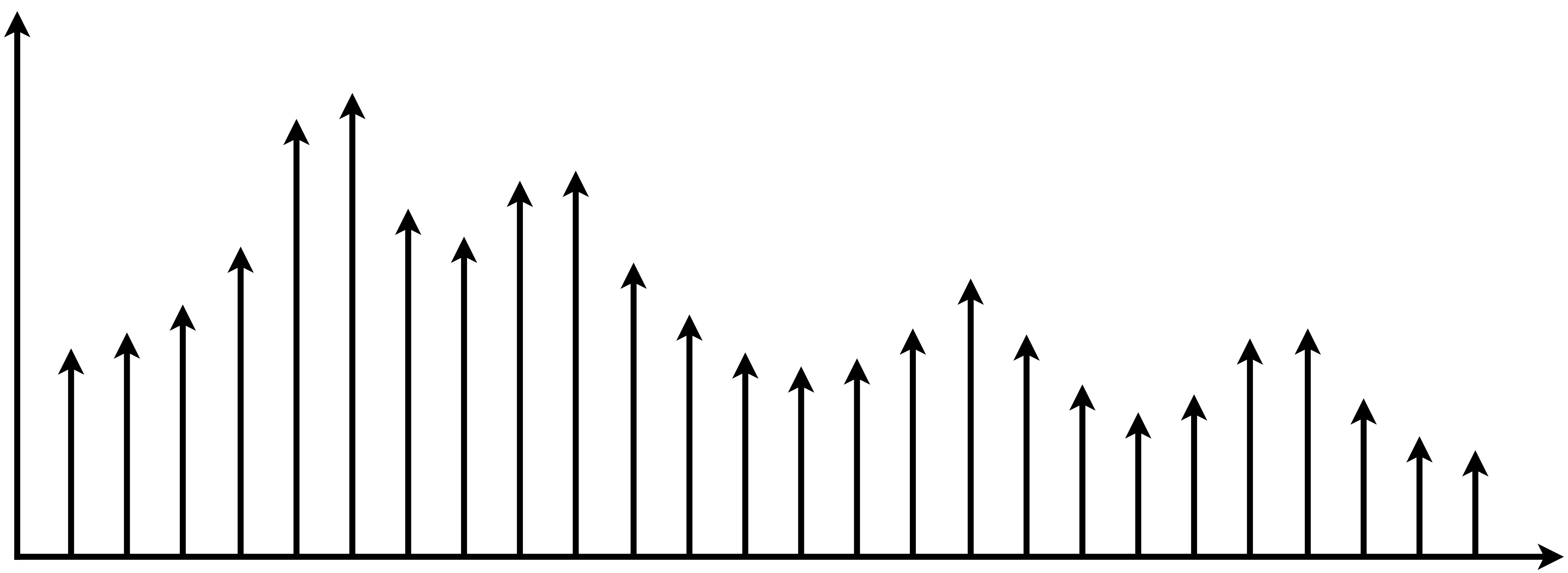
Vowel spectral model

- Periodic sampling also in the frequency domain
  - Continuous spectrum has to be recovered from sampled values (D/A conversion)



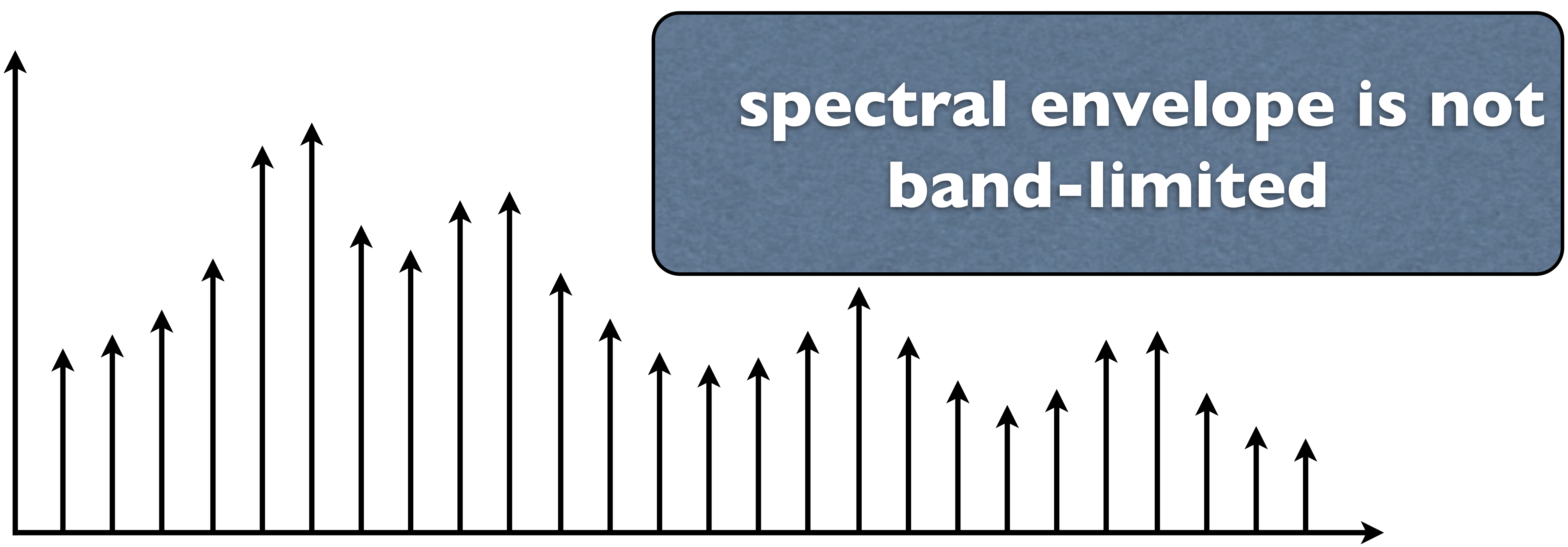
Vowel spectral model

- Periodic sampling also in the frequency domain
  - Continuous spectrum has to be recovered from sampled values (D/A conversion)



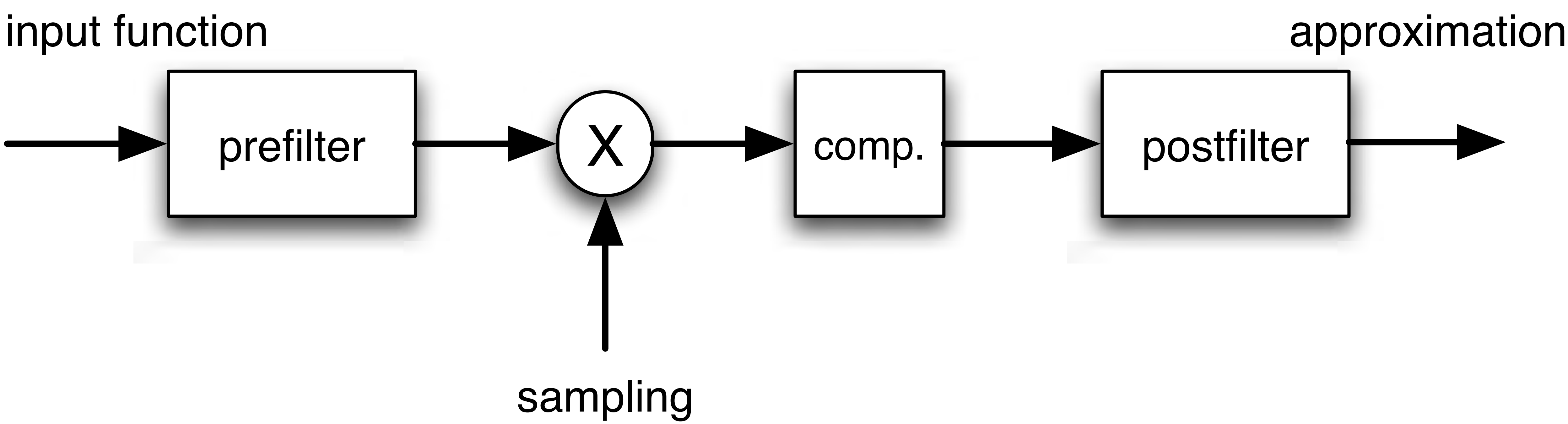
Vowel spectral model

- Periodic sampling also in the frequency domain
  - Continuous spectrum has to be recovered from sampled values (D/A conversion)



Consistent sampling

- Does not require perfect reconstruction
  - Resampled value required to be consistent with the initially sampled value



Sampling—50 Years After Shannon

MICHAEL UNSER, FELLOW, IEEE PROCEEDINGS OF THE IEEE, VOL. 88, NO. 4, APRIL 2000

$$a_{12}(k) = \langle \varphi_1(x-k), \varphi_2(x) \rangle \tag{22}$$

where φ_1 is the analysis function and where φ_2 is the generating (or synthesis) function on the reconstruction side.

Theorem 2 [127]: Let $f \in H$ be an unknown input function. Provided there exists $m > 0$ such that $|A_{12}(e^{j\omega})| \geq m$, a.e., then there is a unique signal approximation $\tilde{f}$ in $V(\varphi_2)$ that is consistent with f in the sense that

$$\forall f \in H,\ c_1(k) = \langle f, \varphi_1(x-k) \rangle = \langle \tilde{f}, \varphi_1(x-k) \rangle. \tag{23}$$

This signal approximation is given by

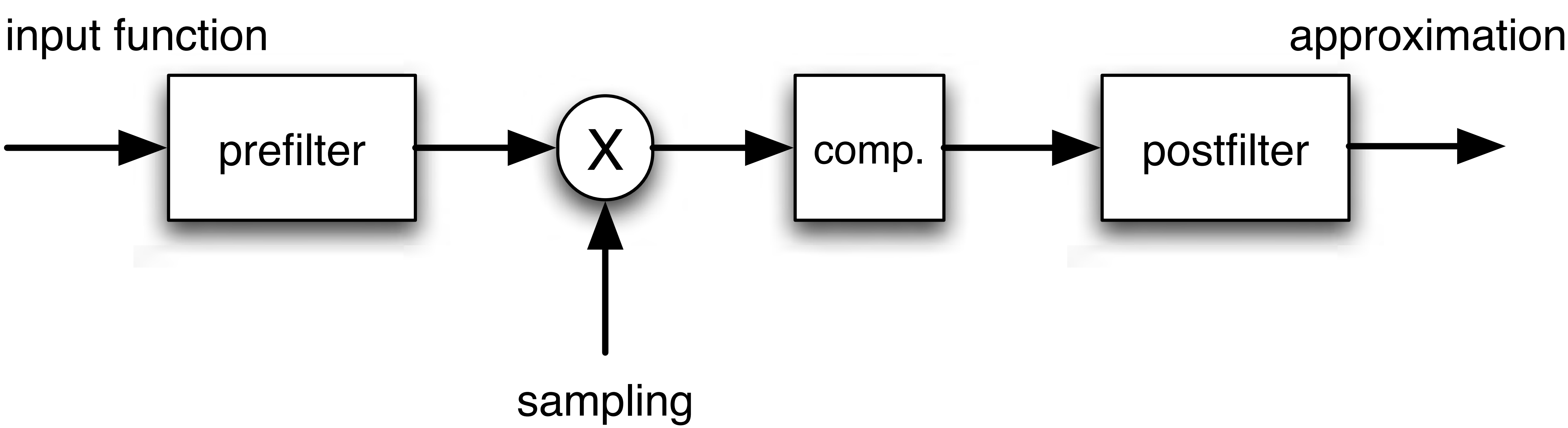
$$\tilde{f} = \tilde{P} f(x) = \sum_{k \in Z} (c_1 * q)(k) \varphi_2(x-k) \tag{24}$$

where

$$Q(z) = \frac{1}{\sum_{k \in Z} a_{12}(k) z^{-k}} \tag{25}$$

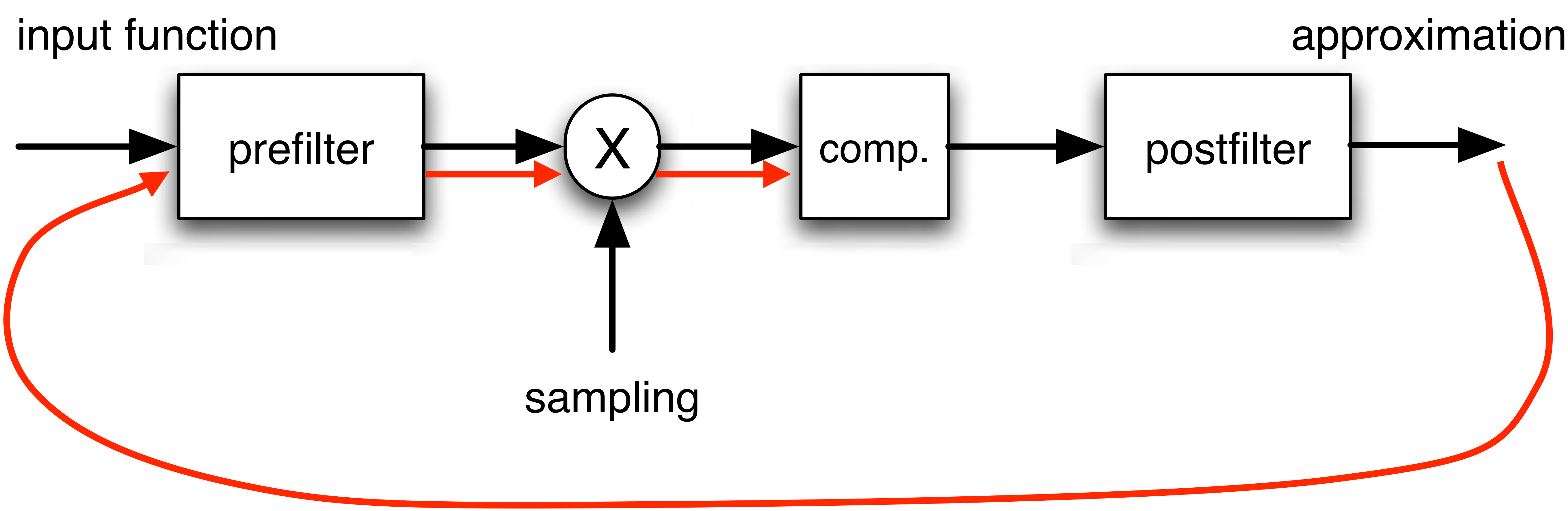
Consistent sampling

- Does not require perfect reconstruction
  - Resampled values are required to be consistent with the initially sampled value



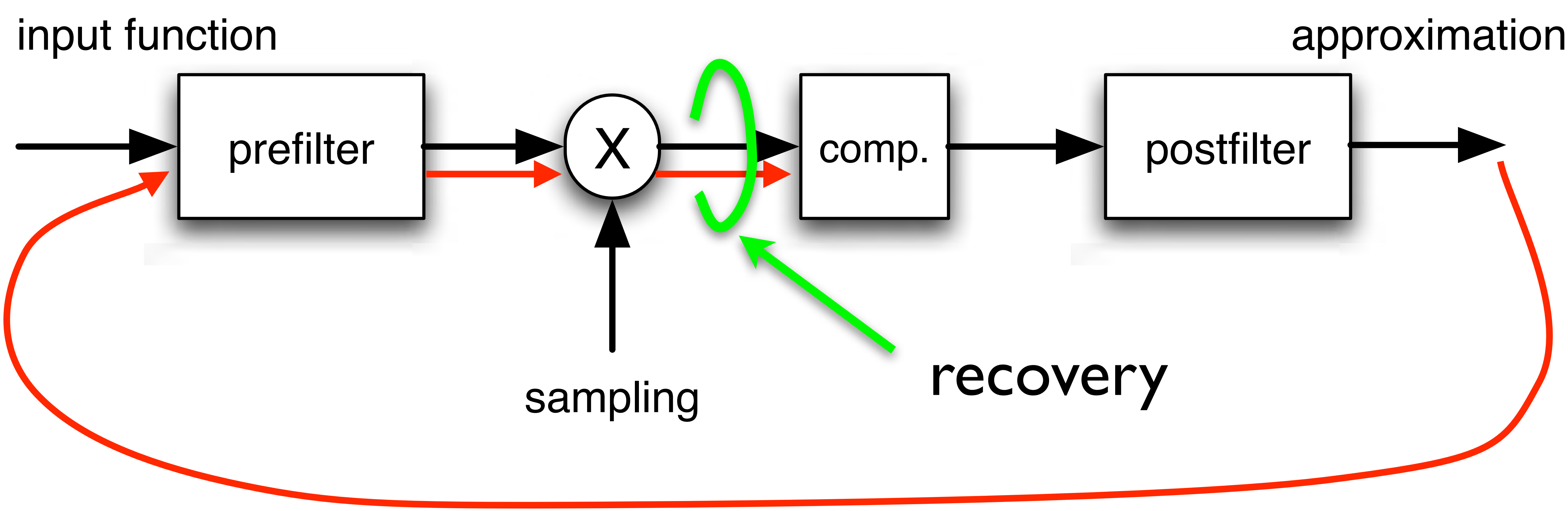
Consistent sampling

- Does not require perfect reconstruction
  - Resampled values are required to be consistent with the initially sampled value



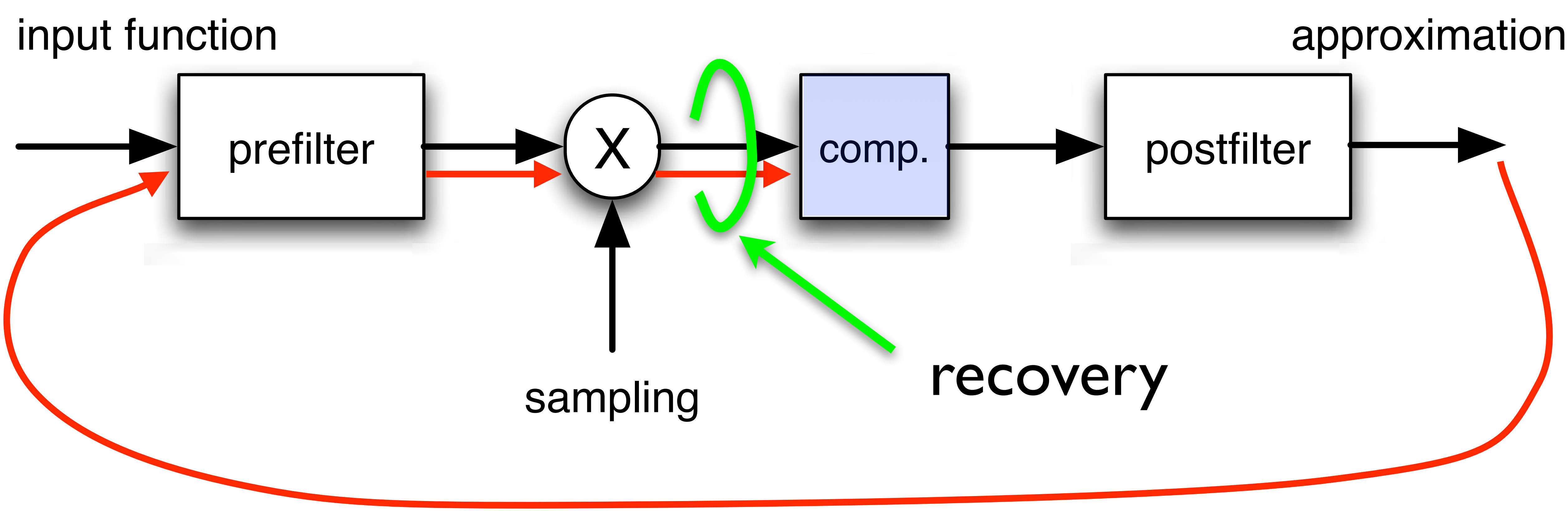
Consistent sampling

- Does not require perfect reconstruction
  - Resampled values are required to be consistent with the initially sampled value



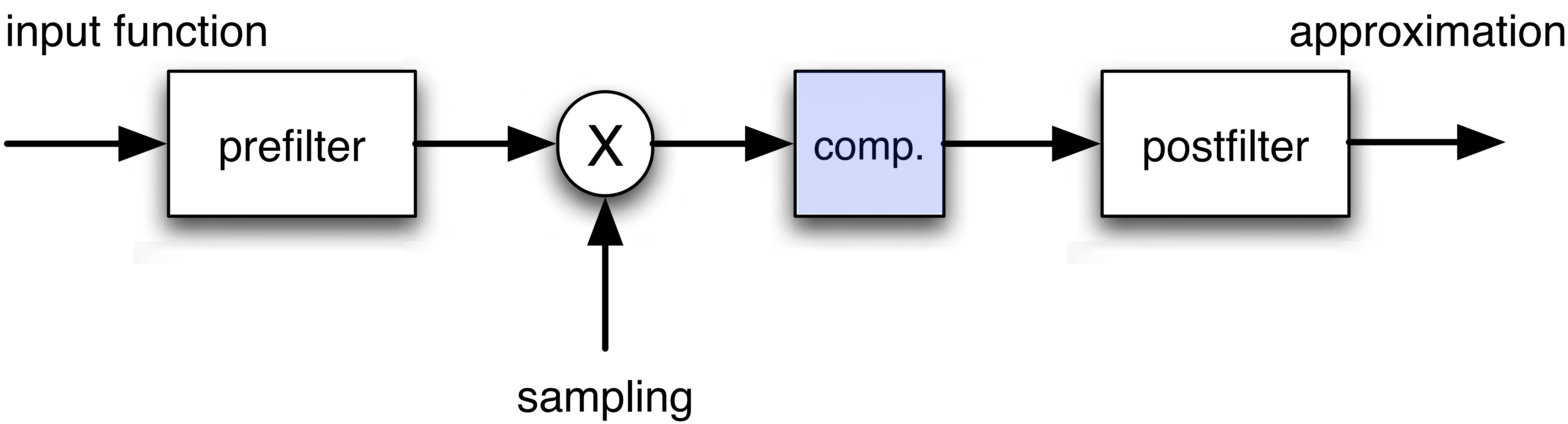
Consistent sampling

- Does not require perfect reconstruction
  - Resampled values are required to be consistent with the initially sampled value



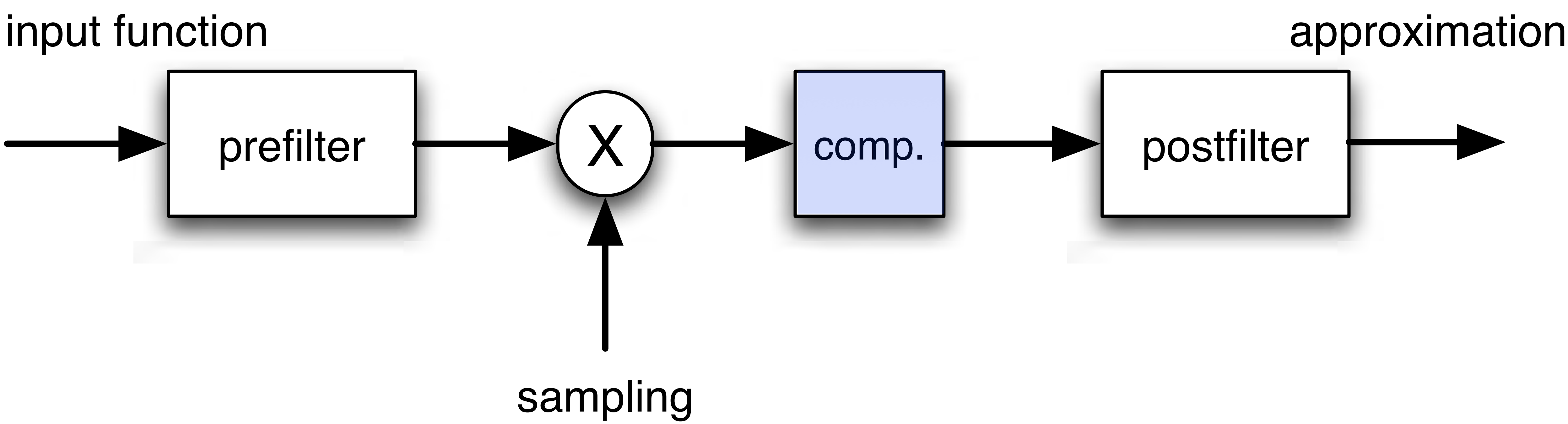
Consistent sampling

- Does not require perfect reconstruction
  - Resampled values are required to be consistent with the initially sampled value



Consistent sampling

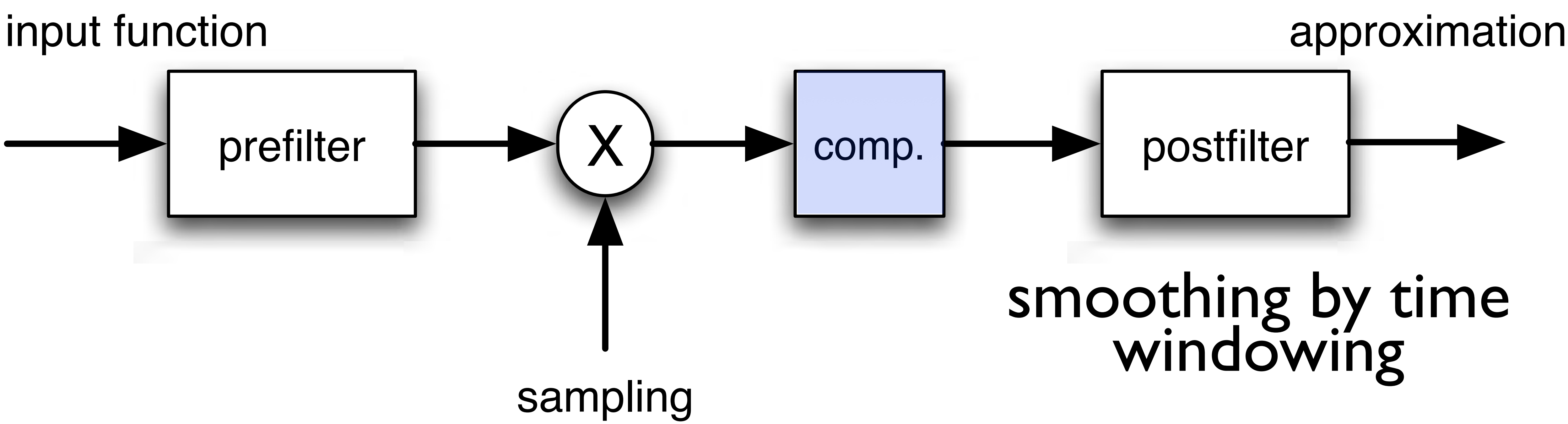
- Does not require perfect reconstruction
  - Resampled values are required to be consistent with the initially sampled value



sampling at harmonic frequencies

Consistent sampling

- Does not require perfect reconstruction
  - Resampled values are required to be consistent with the initially sampled value

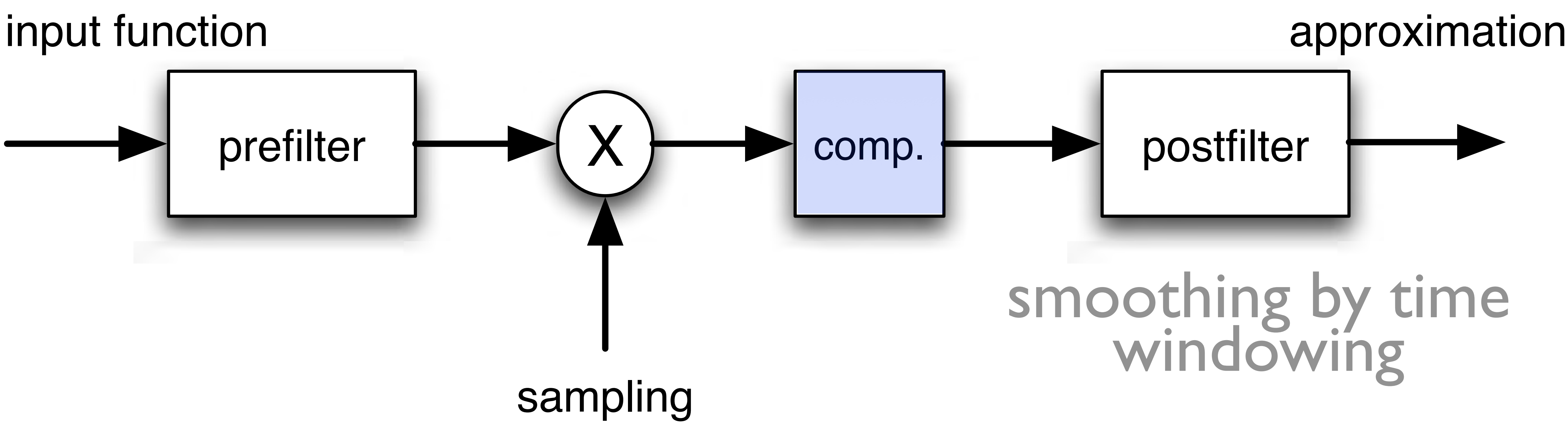


smoothing by time windowing

sampling at harmonic frequencies

Consistent sampling

- Does not require perfect reconstruction
  - Resampled values are required to be consistent with the initially sampled value



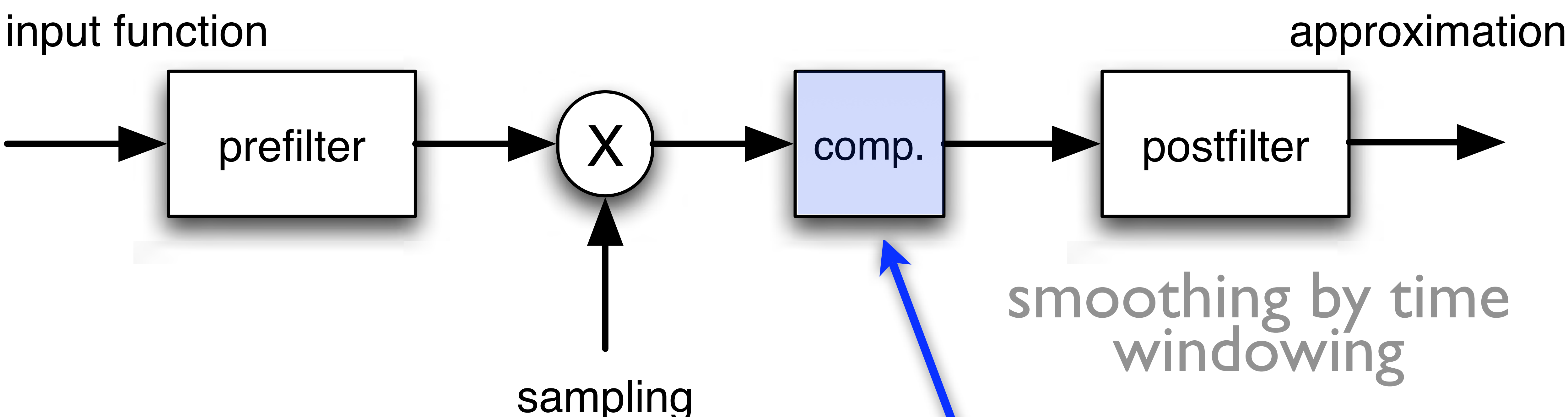
smoothing by time windowing

sampling at harmonic frequencies

sampling pulse removal by holding

Consistent sampling

- Does not require perfect reconstruction
  - Resampled values are required to be consistent with the initially sampled value



smoothing by time windowing

sampling at harmonic frequencies

sampling pulse removal by holding

sampled value recovery by a compensation filter

A simple implementation

- Holding using cumulative power spectrum and interpolation of differentiation

TANDEM spectrum

$$C(\omega) = \int_{\omega_L}^{\omega} P_T(\lambda) d\lambda$$

$$L_S(\omega) = \ln \left[C(\omega + \omega_0/2) - C(\omega - \omega_0/2) \right] - \ln \omega_0$$

$$P_{TST}(\omega) = e^{[\bar{q}_1 (L_S(\omega - \omega_0) + L_S(\omega + \omega_0)) + \bar{q}_0 L_S(\omega)]}.$$

A simple implementation

- Holding using cumulative power spectrum and interpolation of differentiation

TANDEM spectrum

$$C(\omega) = \int_{\omega_L}^{\omega} P_T(\lambda) d\lambda$$

$$L_S(\omega) = \ln\left[C(\omega + \omega_0/2) - C(\omega - \omega_0/2)\right] - \ln \omega_0$$

$$P_{TST}(\omega) = e^{[\bar{q}_1 (L_S(\omega - \omega_0) + L_S(\omega + \omega_0)) + \bar{q}_0 L_S(\omega)]}.$$

A simple implementation

- Holding using cumulative power spectrum and interpolation of differentiation

TANDEM spectrum

holding and interpolation

$$C(\omega) = \int_{\omega_L}^{\omega} P_T(\lambda) d\lambda$$

$$L_S(\omega) = \ln \boxed{[C(\omega + \omega_0/2) - C(\omega - \omega_0/2)]} - \ln \omega_0$$

$$P_{TST}(\omega) = e^{[\bar{q}_1 (L_S(\omega - \omega_0) + L_S(\omega + \omega_0)) + \bar{q}_0 L_S(\omega)]}.$$

A simple implementation

- Holding using cumulative power spectrum and interpolation of differentiation

TANDEM spectrum

holding and interpolation

$$C(\omega) = \int_{\omega_L}^{\omega} P_T(\lambda) d\lambda$$

$$L_S(\omega) = \ln \boxed{[C(\omega + \omega_0/2) - C(\omega - \omega_0/2)]} - \ln \omega_0$$

$$P_{TST}(\omega) = e^{[\bar{q}_1 (L_S(\omega - \omega_0) + L_S(\omega + \omega_0)) + \bar{q}_0 L_S(\omega)]}.$$

A simple implementation

- Holding using cumulative power spectrum and interpolation of differentiation
- Compensation on log-spectrum to assure results to be positive

TANDEM spectrum

$$C(\omega) = \int_{\omega_L}^{\omega} P_T(\lambda) d\lambda$$

holding and interpolation

$$L_S(\omega) = \ln \left[C(\omega + \omega_0/2) - C(\omega - \omega_0/2) \right] - \ln \omega_0$$

$$P_{TST}(\omega) = e^{[\bar{q}_1 (L_S(\omega - \omega_0) + L_S(\omega + \omega_0)) + \bar{q}_0 L_S(\omega)]}.$$

compensation on log-spectrum

A simple implementation

- Holding using cumulative power spectrum and interpolation of differentiation
- Compensation on log-spectrum to assure results to be positive

TANDEM spectrum

$$C(\omega) = \int_{\omega_L}^{\omega} P_T(\lambda) d\lambda$$

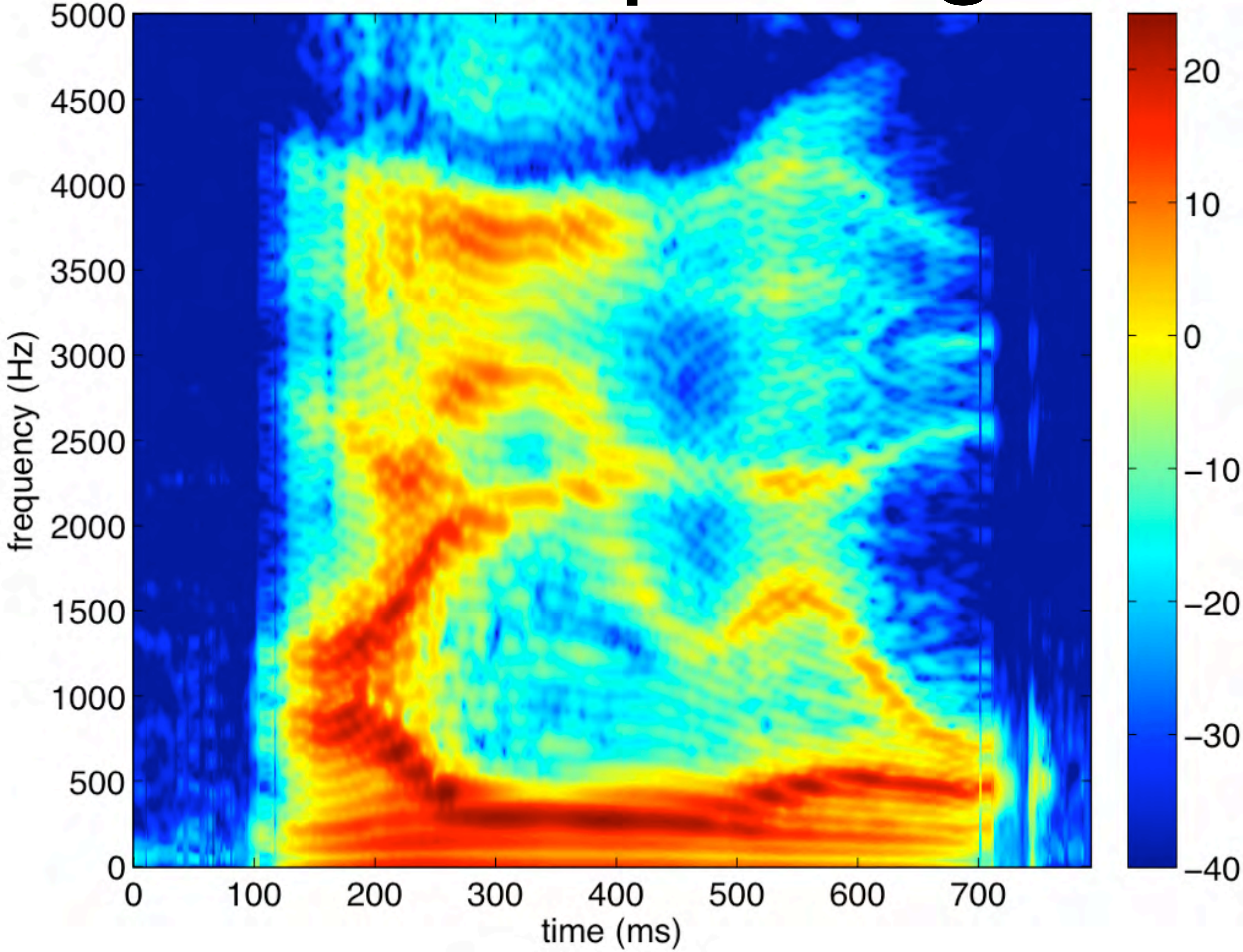
holding and interpolation

$$L_S(\omega) = \ln \left[C(\omega + \omega_0/2) - C(\omega - \omega_0/2)\right] - \ln \omega_0$$

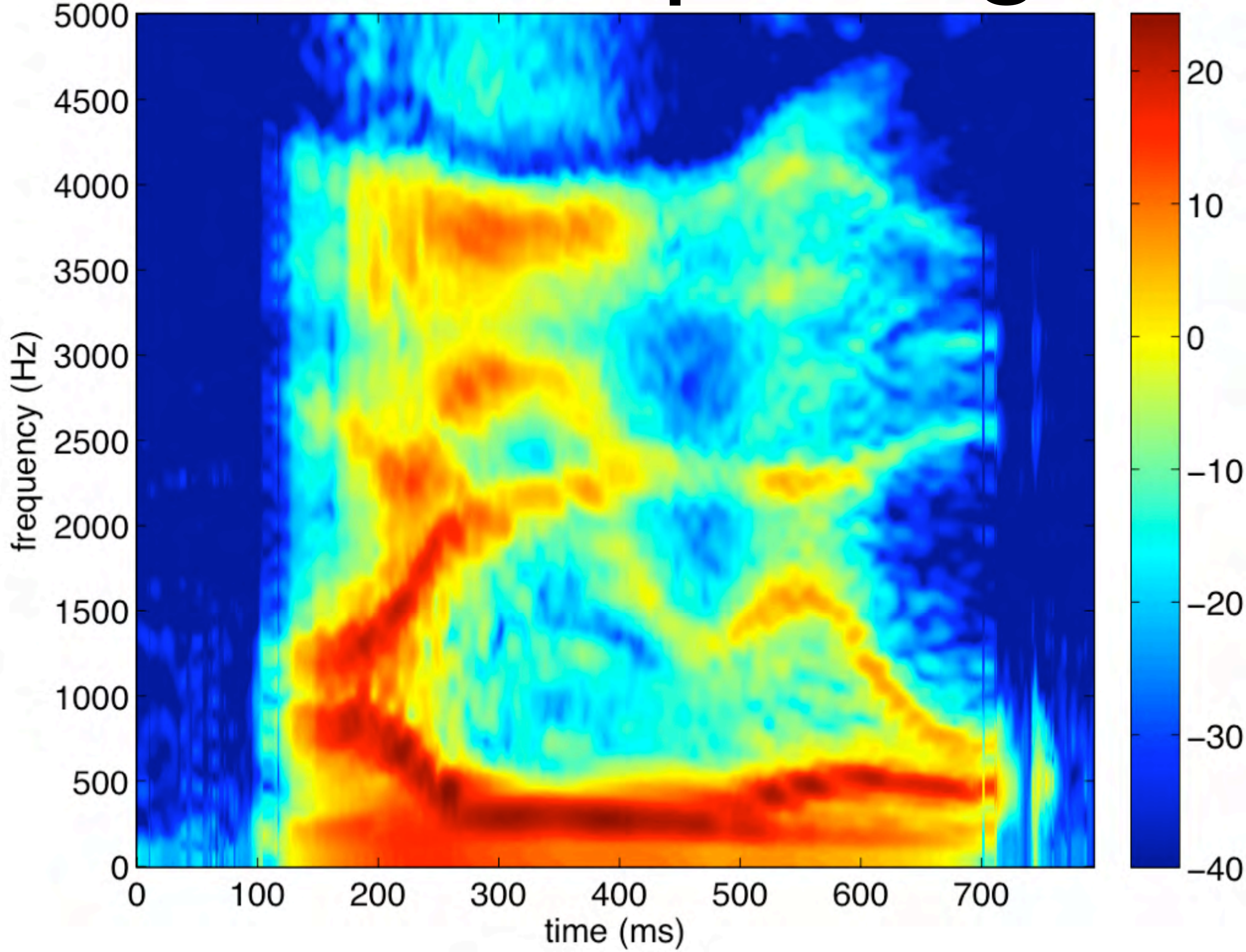
$$P_{TST}(\omega) = e^{[\bar{q}_1(L_S(\omega-\omega_0)+L_S(\omega+\omega_0))+\bar{q}_0 L_S(\omega)]}.$$

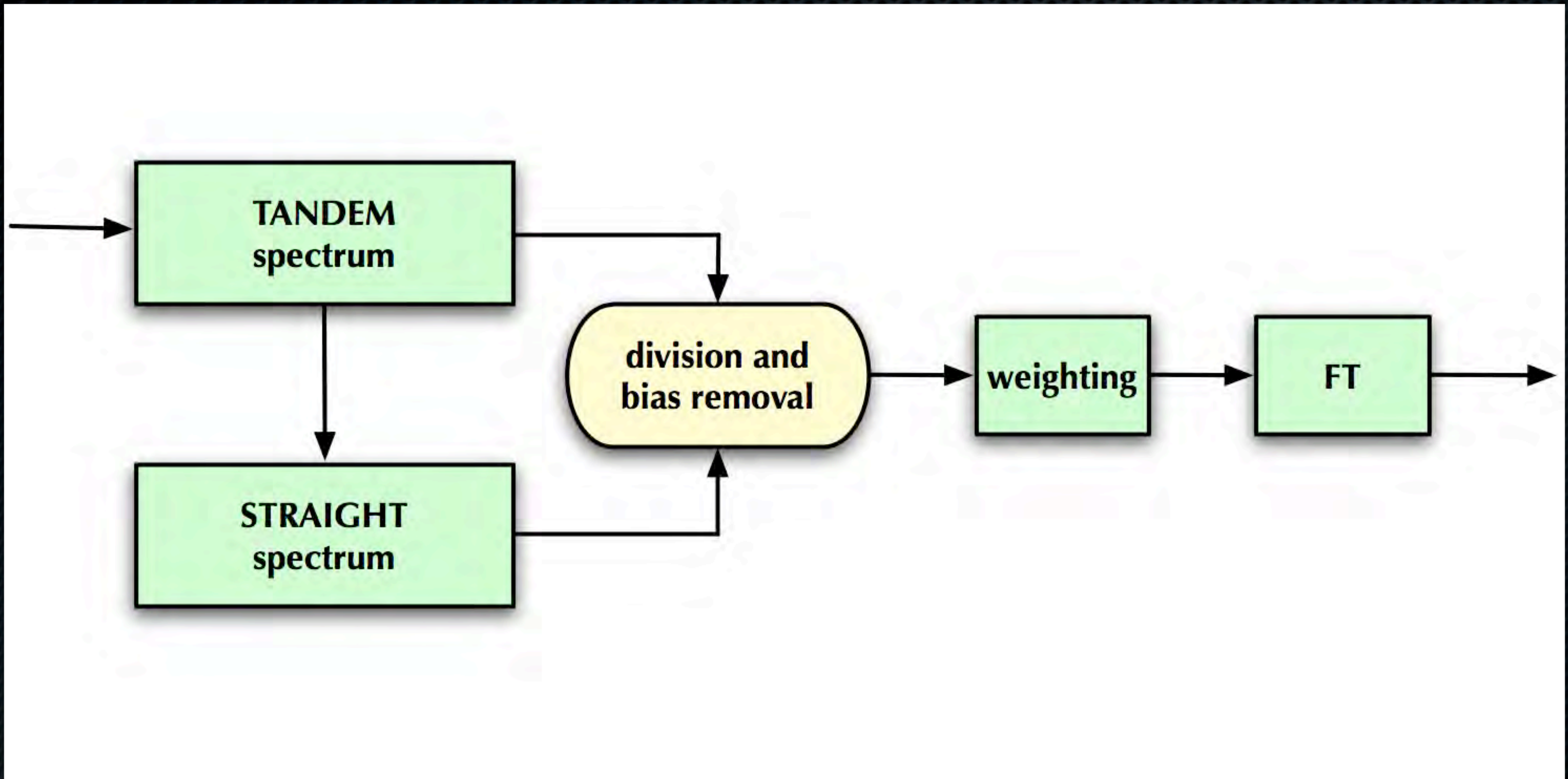
compensation on log-spectrum

TANDEM spectrogram



STRAIGHT spectrogram





Periodicity detection

Unified framework based on TANDEM and STRAIGHT spectra

Application of TANDEM and STRAIGHT to F0 extraction

Application of TANDEM and STRAIGHT to F0 extraction

- STRAIGHT spectrum --> envelope

Application of TANDEM and STRAIGHT to F0 extraction

- STRAIGHT spectrum --> envelope
- TANDEM spectrum --> envelope modulated by periodic excitation

Application of TANDEM and STRAIGHT to F0 extraction

- STRAIGHT spectrum --> envelope
- TANDEM spectrum --> envelope modulated by periodic excitation
- TANDEM spectrum divided by STRAIGHT spectrum yields bias plus modulation

Application of TANDEM and STRAIGHT to F0 extraction

- STRAIGHT spectrum --> envelope
- TANDEM spectrum --> envelope modulated by periodic excitation
- TANDEM spectrum divided by STRAIGHT spectrum yields bias plus modulation

variation spectrum

TANDEM spectrum

$$P_C(\omega) = \frac{P_T(\omega)}{P_{TST}(\omega)} - 1$$

STRAIGHT spectrum

How does it behave?

How does it behave?

- Band-pass filtering in the spatial frequency domain

How does it behave?

- Band-pass filtering in the spatial frequency domain
  - Time windowing --> low pass

How does it behave?

- Band-pass filtering in the spatial frequency domain
  - Time windowing --> low pass
  - Variation spectrum --> high pass

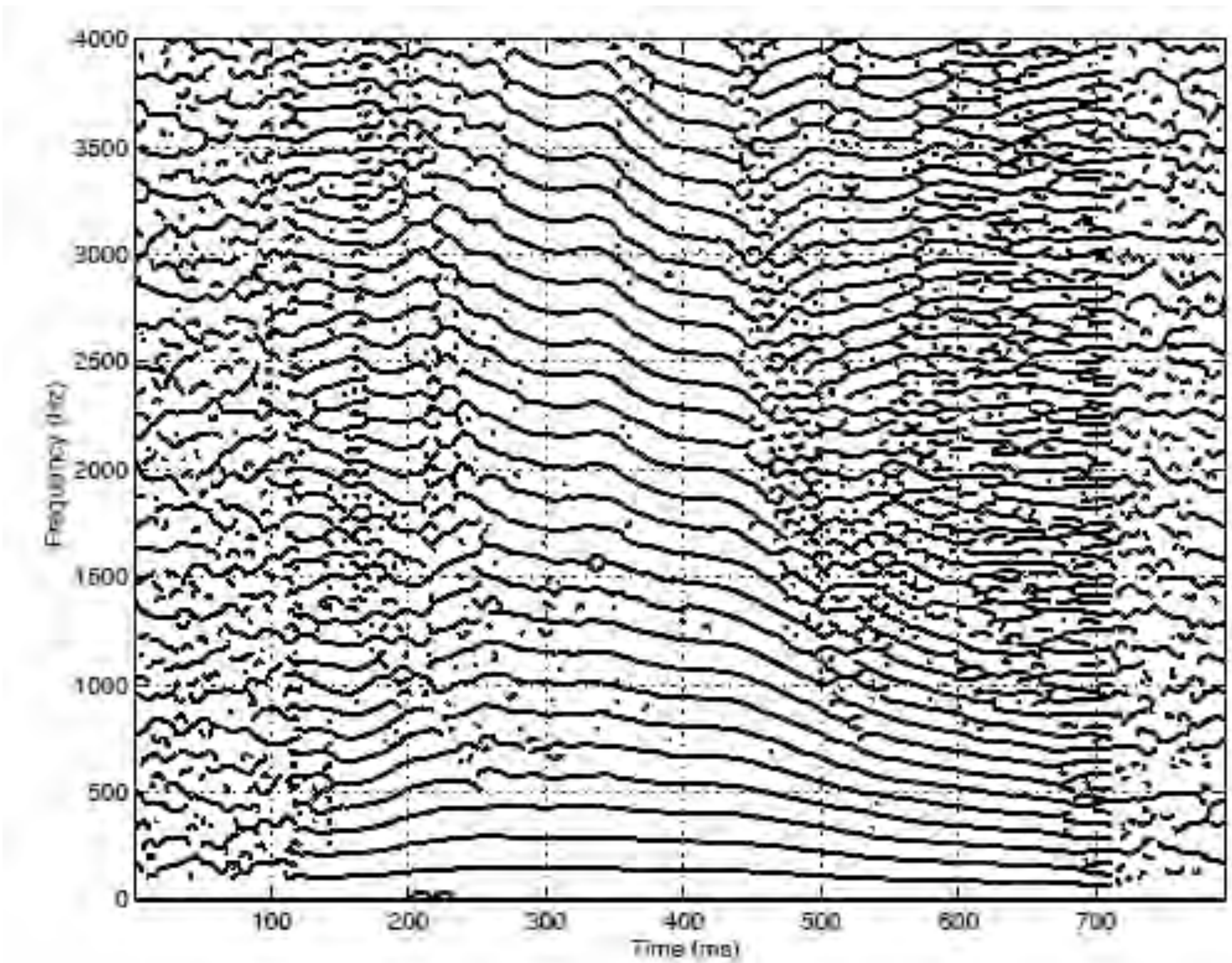
How does it behave?

- Band-pass filtering in the spatial frequency domain
  - Time windowing --> low pass
  - Variation spectrum --> high pass
  - Band-pass response --> tunable to F0

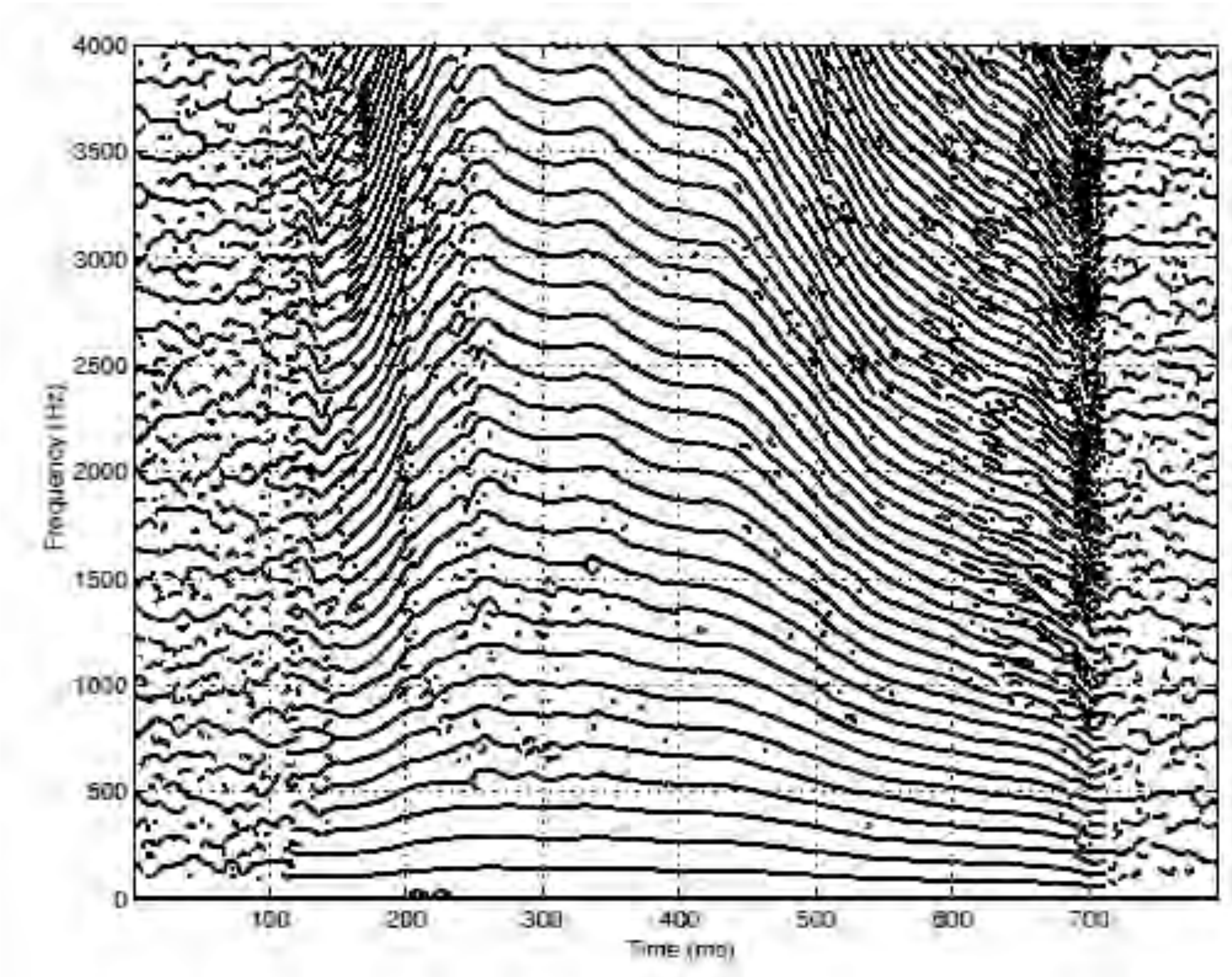
How does it behave?

- Band-pass filtering in the spatial frequency domain
  - Time windowing --> low pass
  - Variation spectrum --> high pass
  - Band-pass response --> tunable to F0
- In the higher (original) frequency region, harmonic structure is corrupted by moire effects

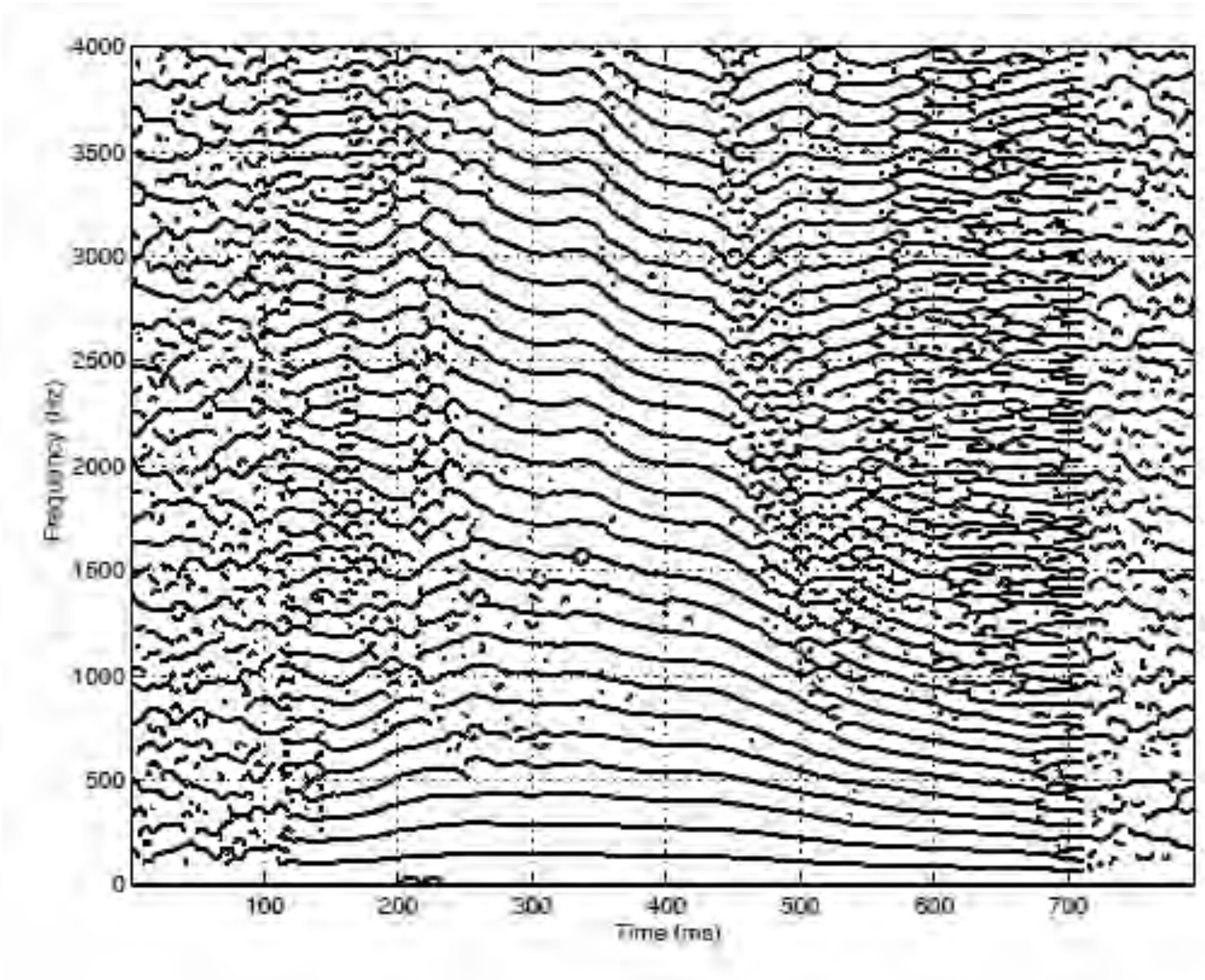
Moire effects



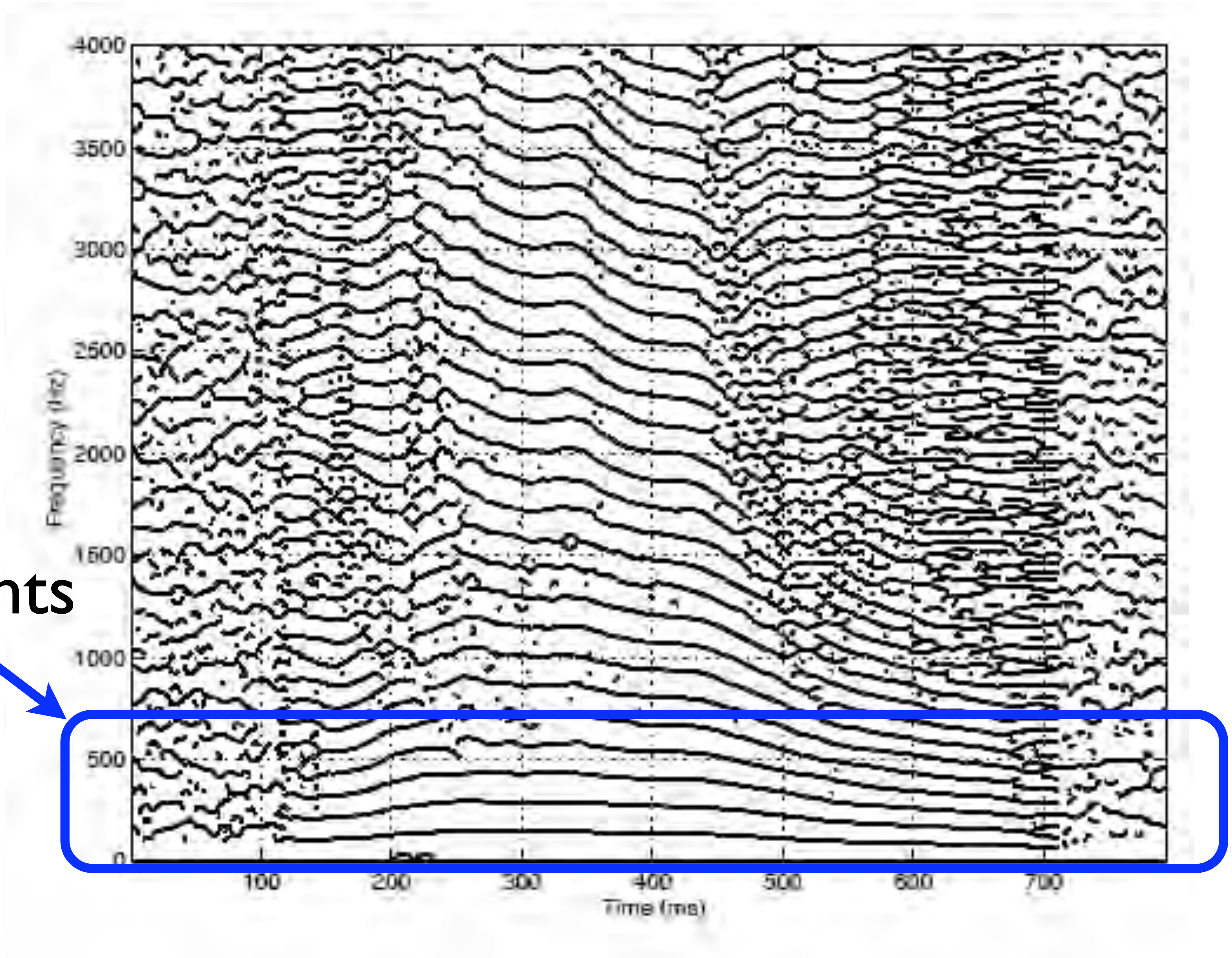
Moire effects



Moire effects

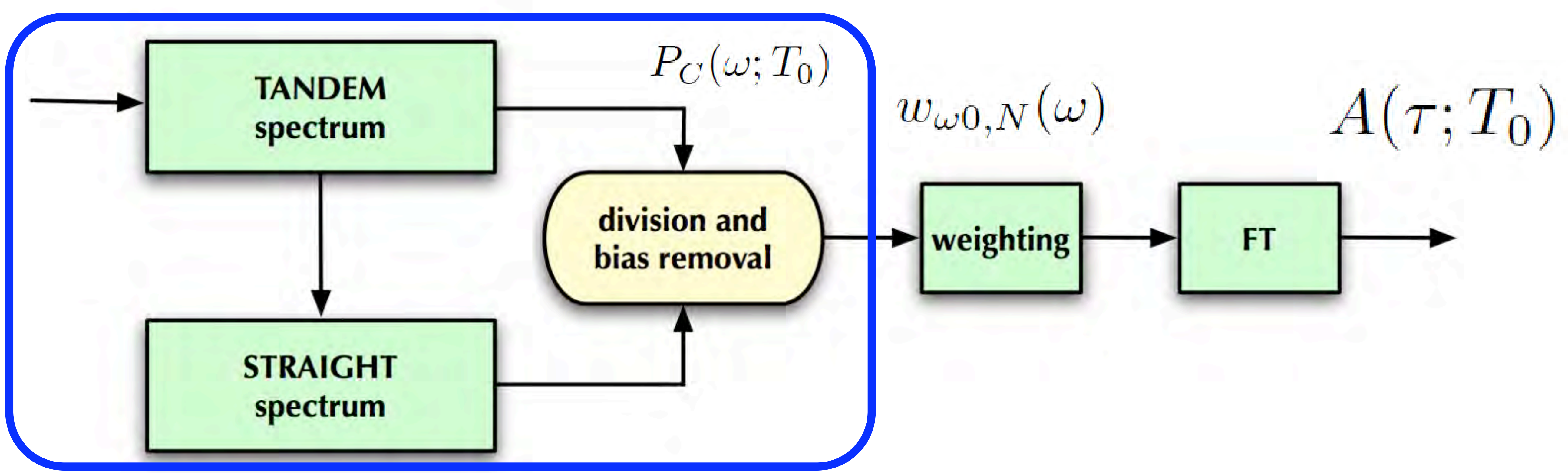


Moire effects



reliable components

F0 tuned F0 detector



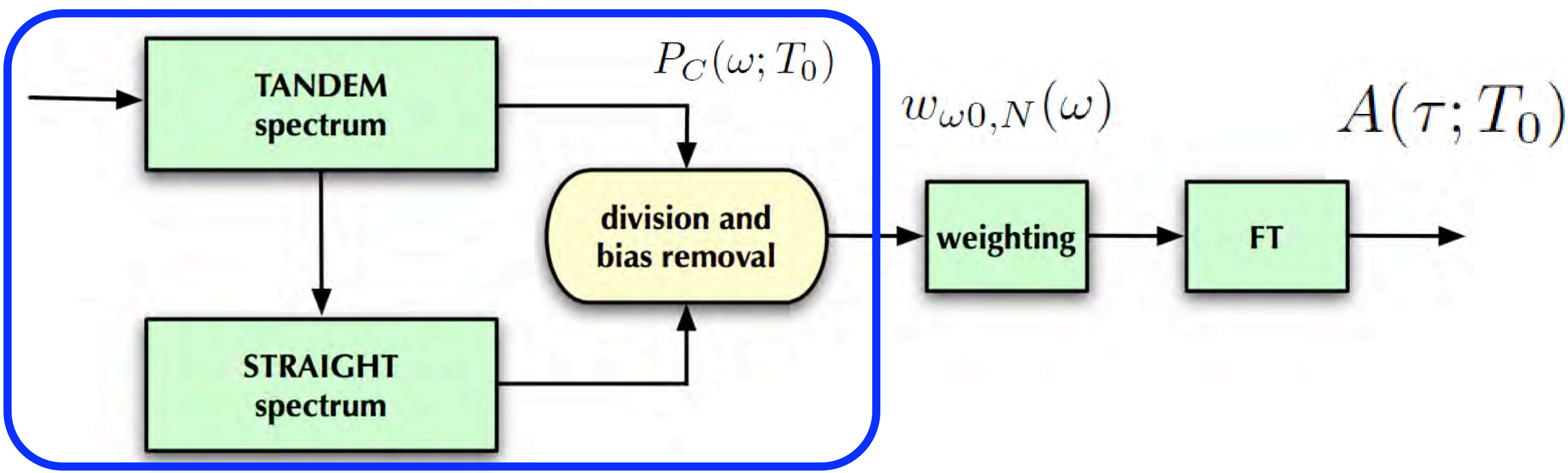
F0 tuned F0 detector

spatial frequency component

$$A(\tau; T_0) = \int_{-\infty}^{\infty} w_{\omega 0,N}(\omega) P_C(\omega; T_0) e^{-j\omega\tau} d\omega$$

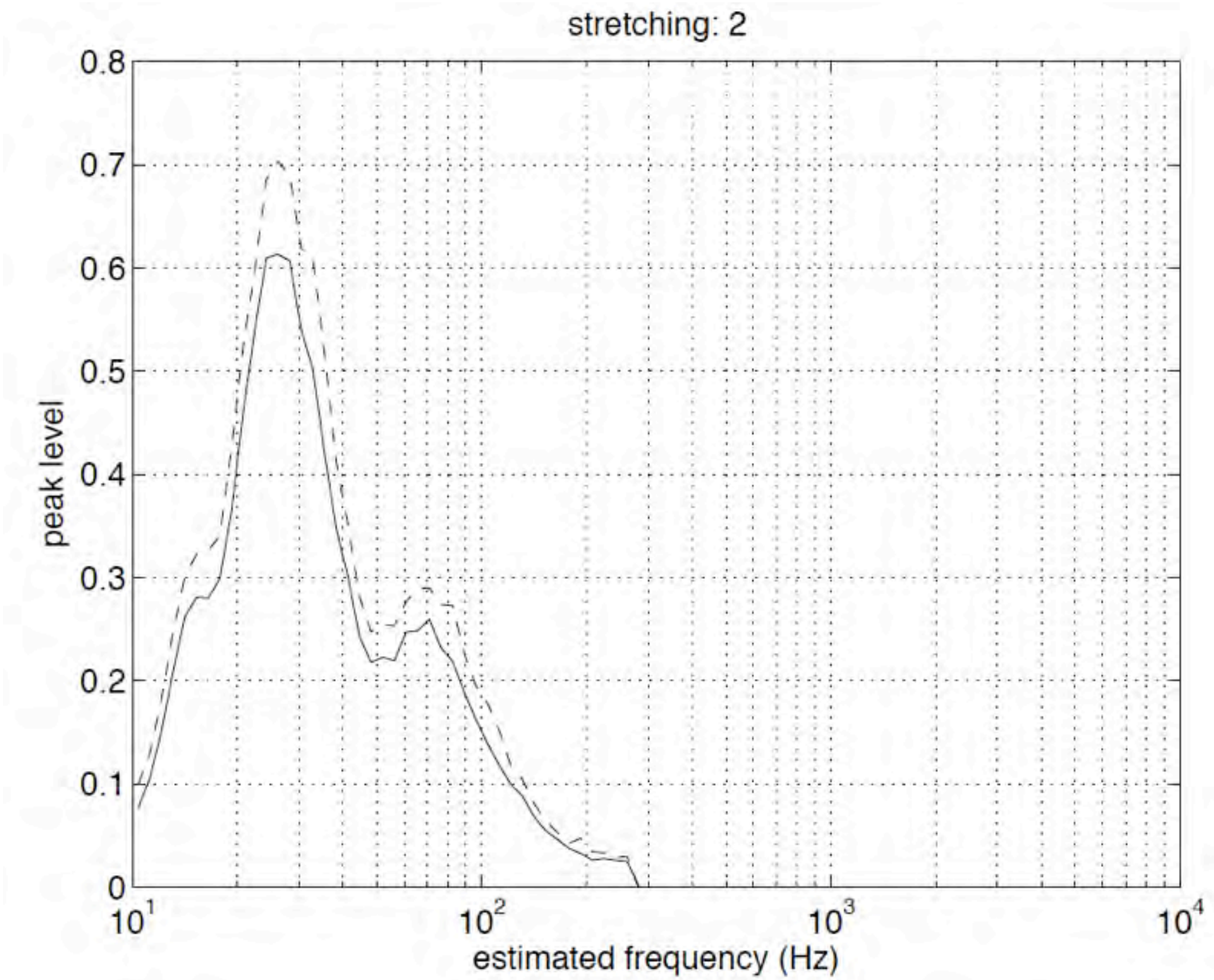
selector of lower frequency region

$$w_{\omega 0,N}(\omega) = c_0 \left(1 + \cos\left(\pi\omega / N\omega_0\right)\right)$$



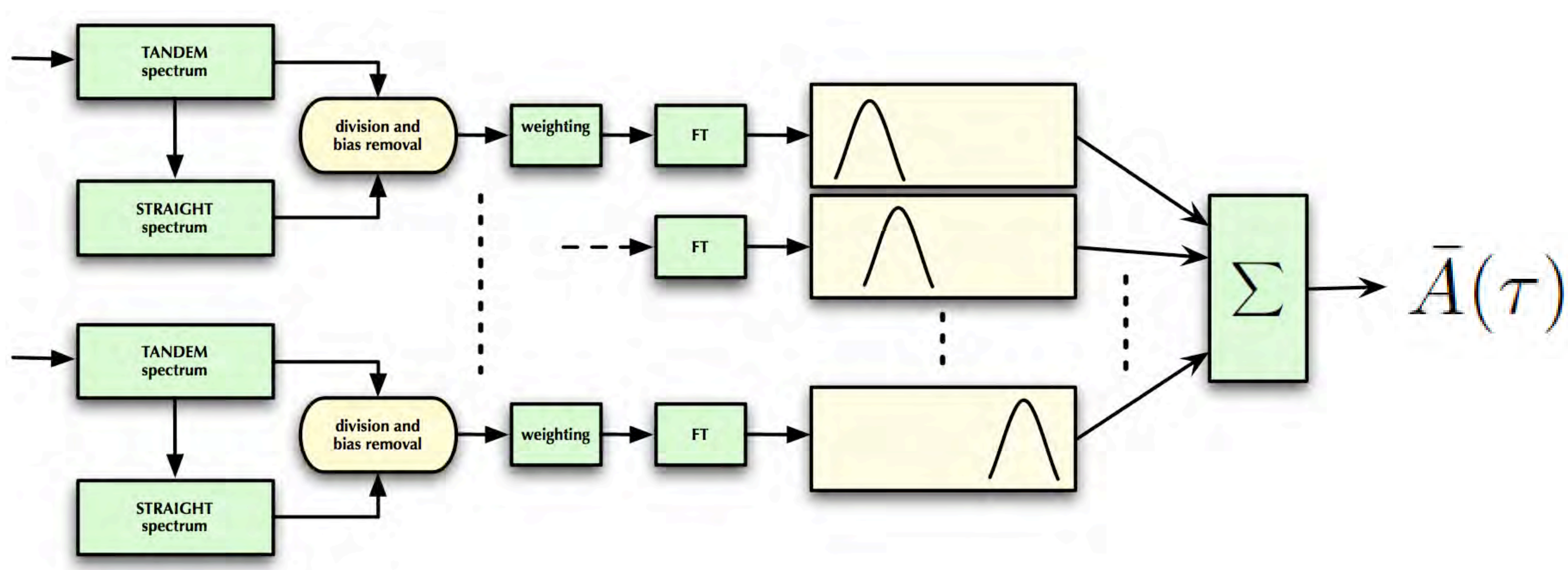
F0 tuned F0 detector

response to random inputs



Group of detectors

combination of detectors to cover wider region



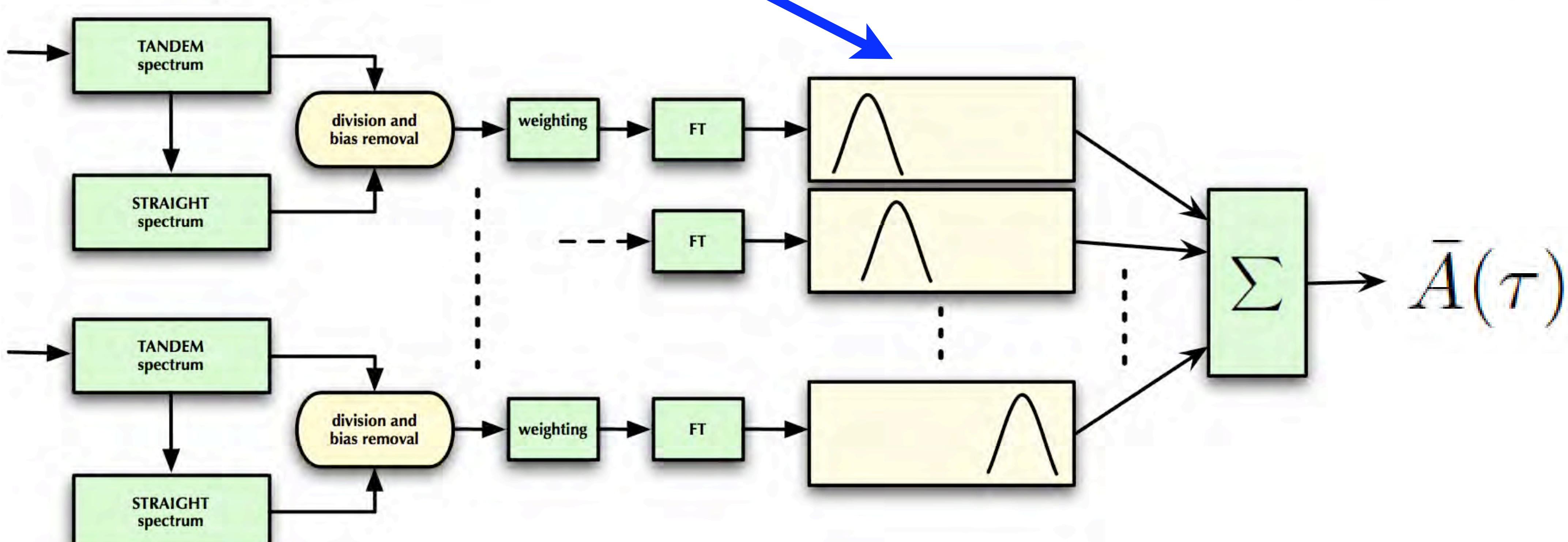
Group of detectors

combination of detectors to cover wider region

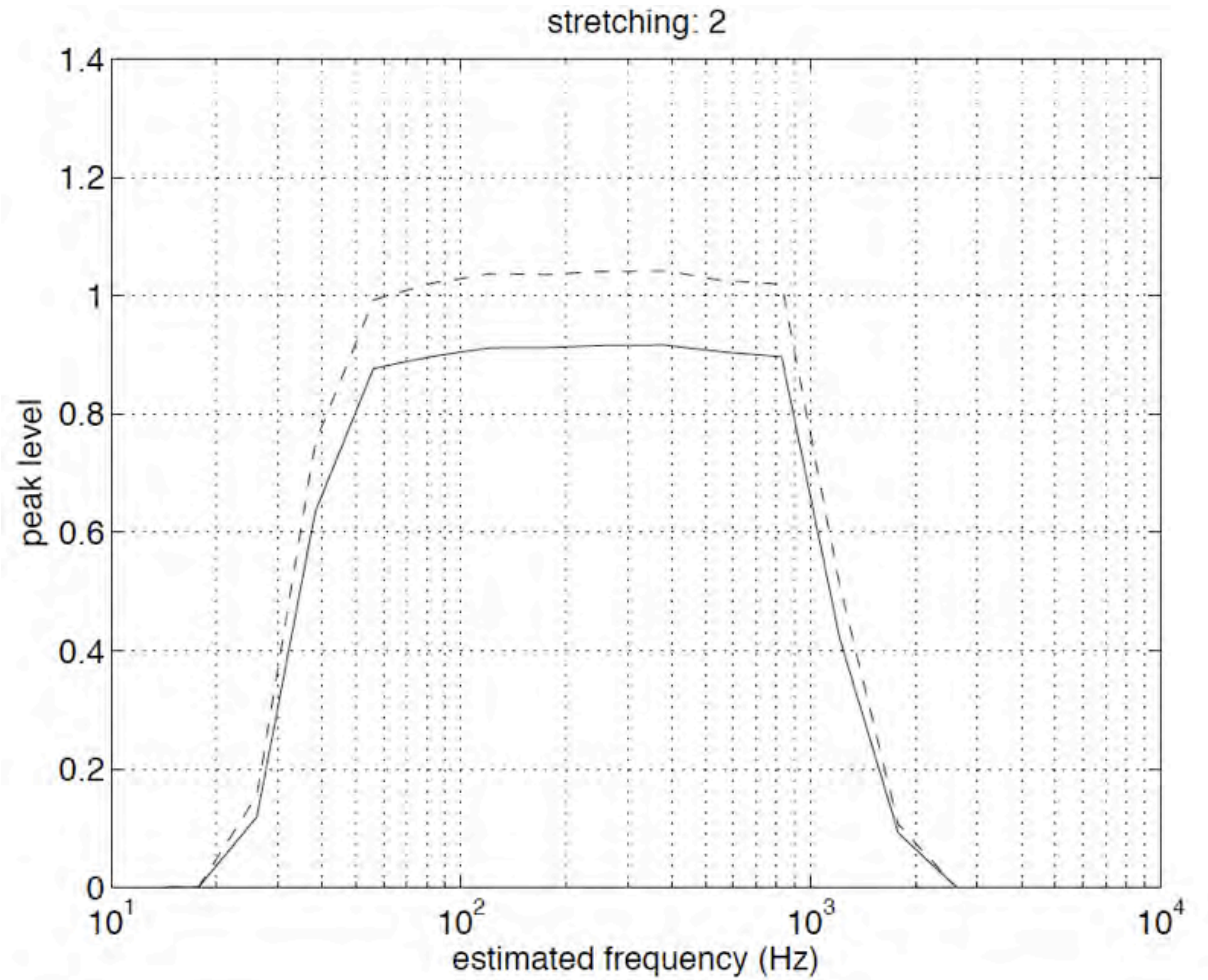
selector of a tuned lag region

$$w_{LAG}(\tau; T_0) = 0.5 + 0.5\cos\left(\pi \log_2\left(\frac{\tau}{T_0}\right)\right)$$

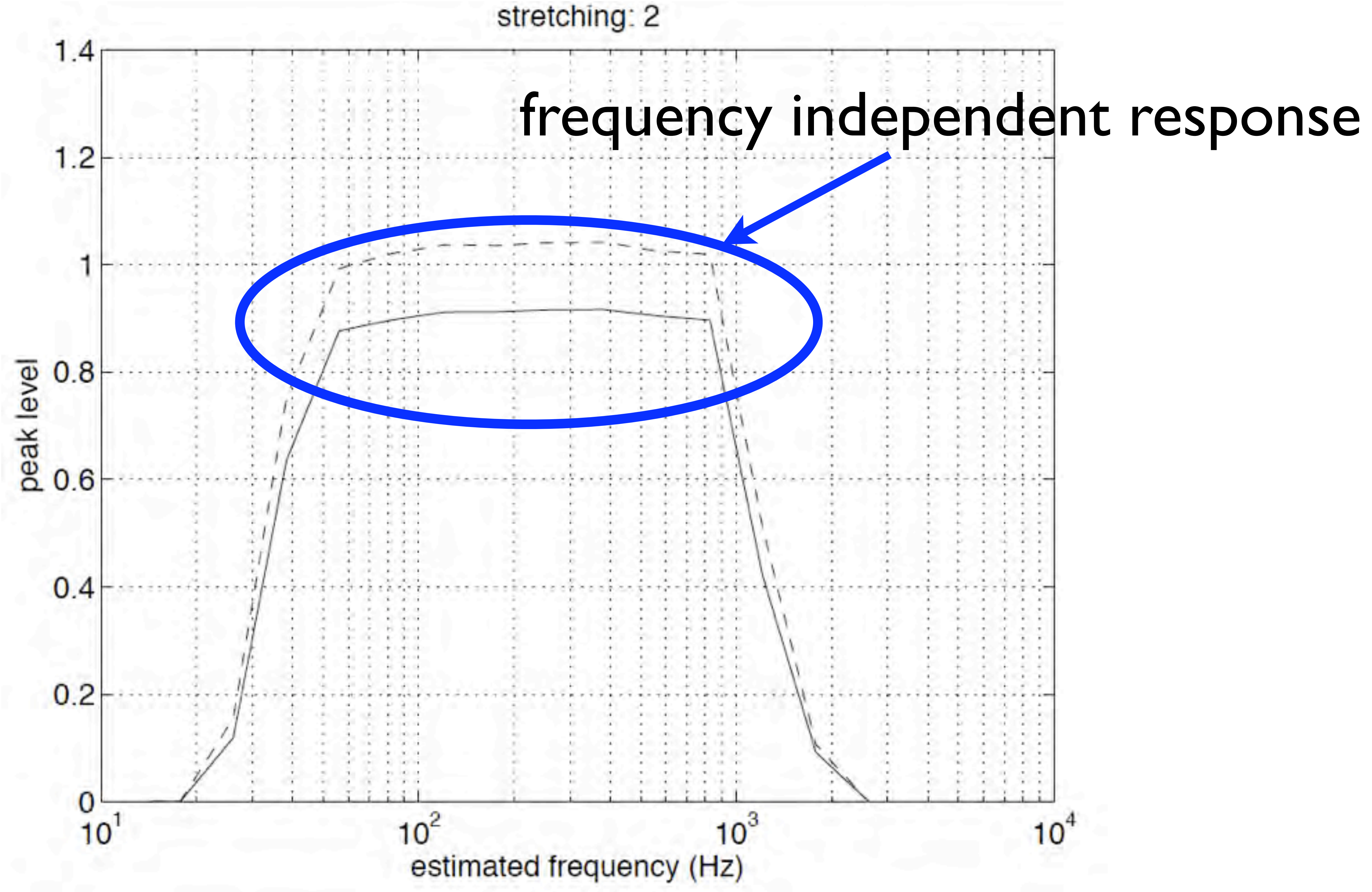
$$\bar{A}(\tau) = \frac{1}{M}\sum_{k=1}^{M} w_{LAG}(\tau; T_L 2^{\frac{1-k}{L}}) A\left(\tau; T_L 2^{\frac{1-k}{L}}\right)$$



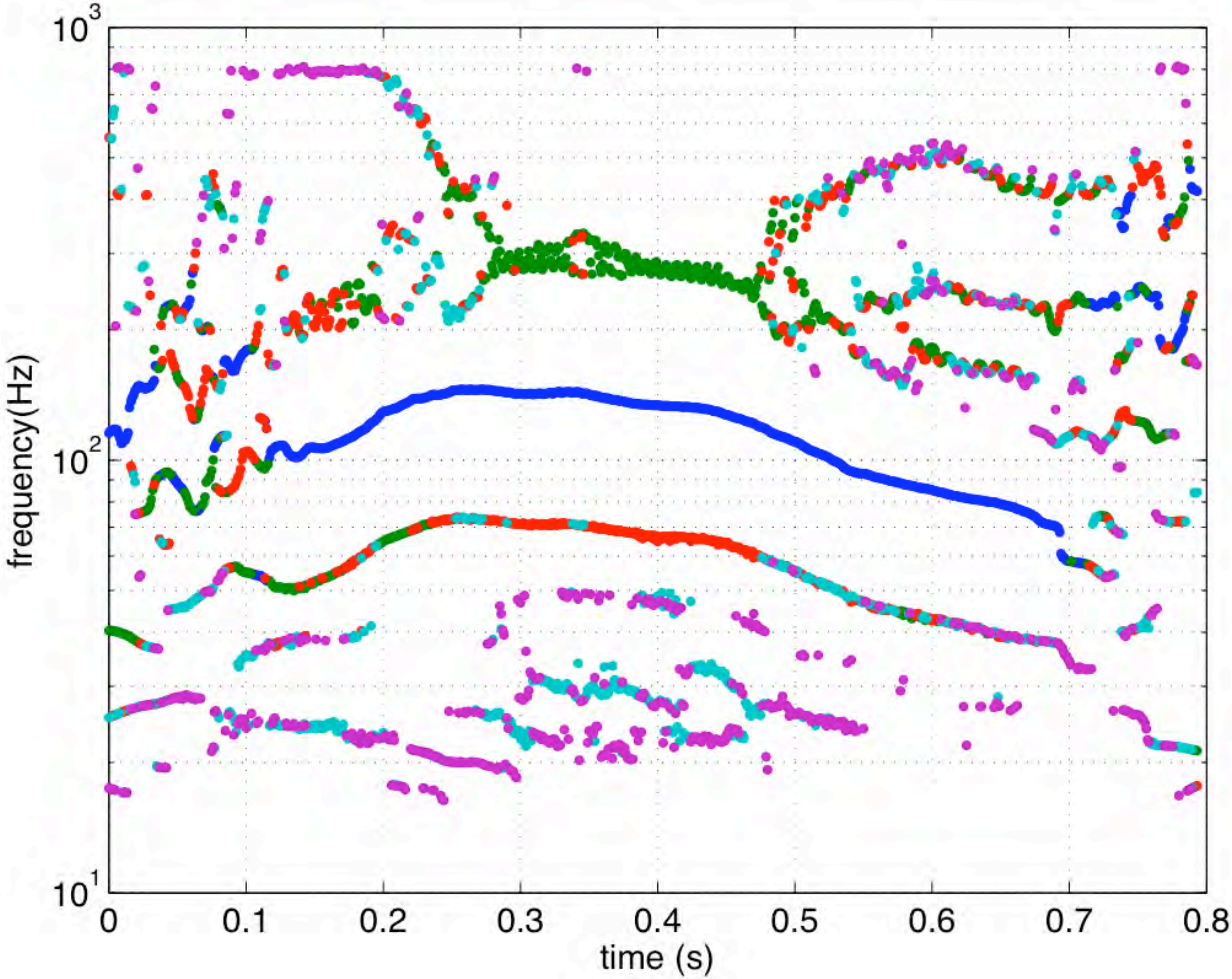
Combined detectors



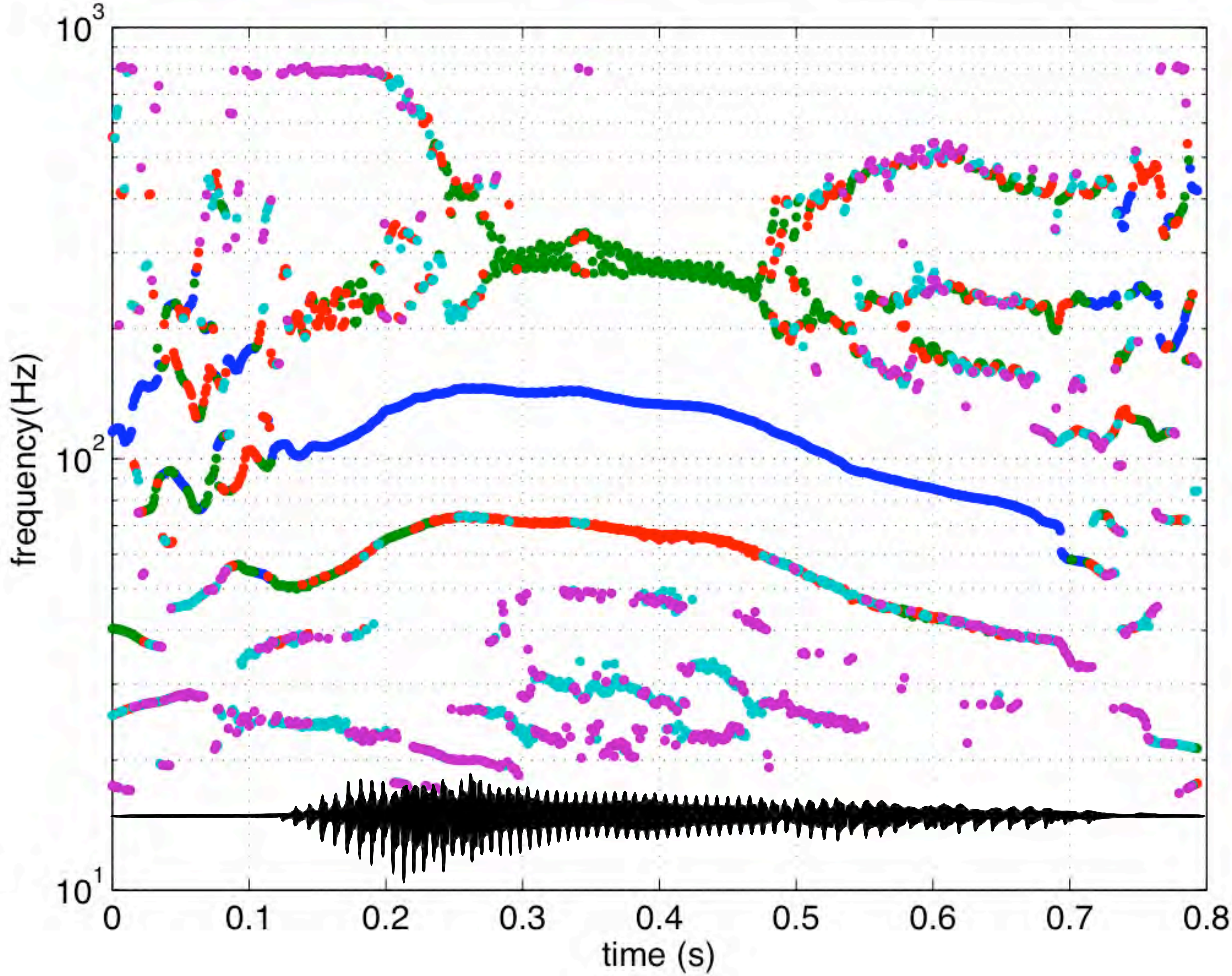
Combined detectors



Example

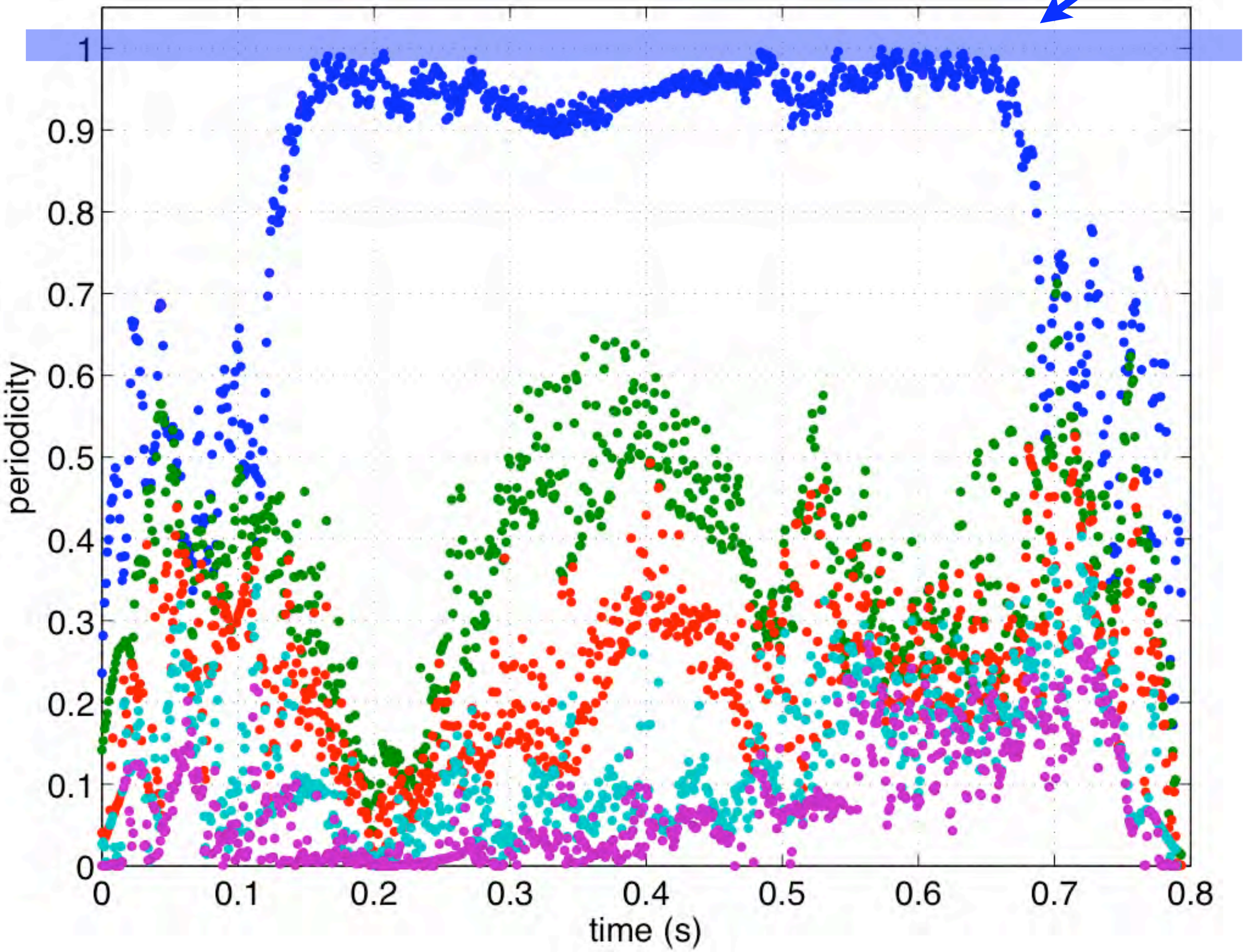


Example

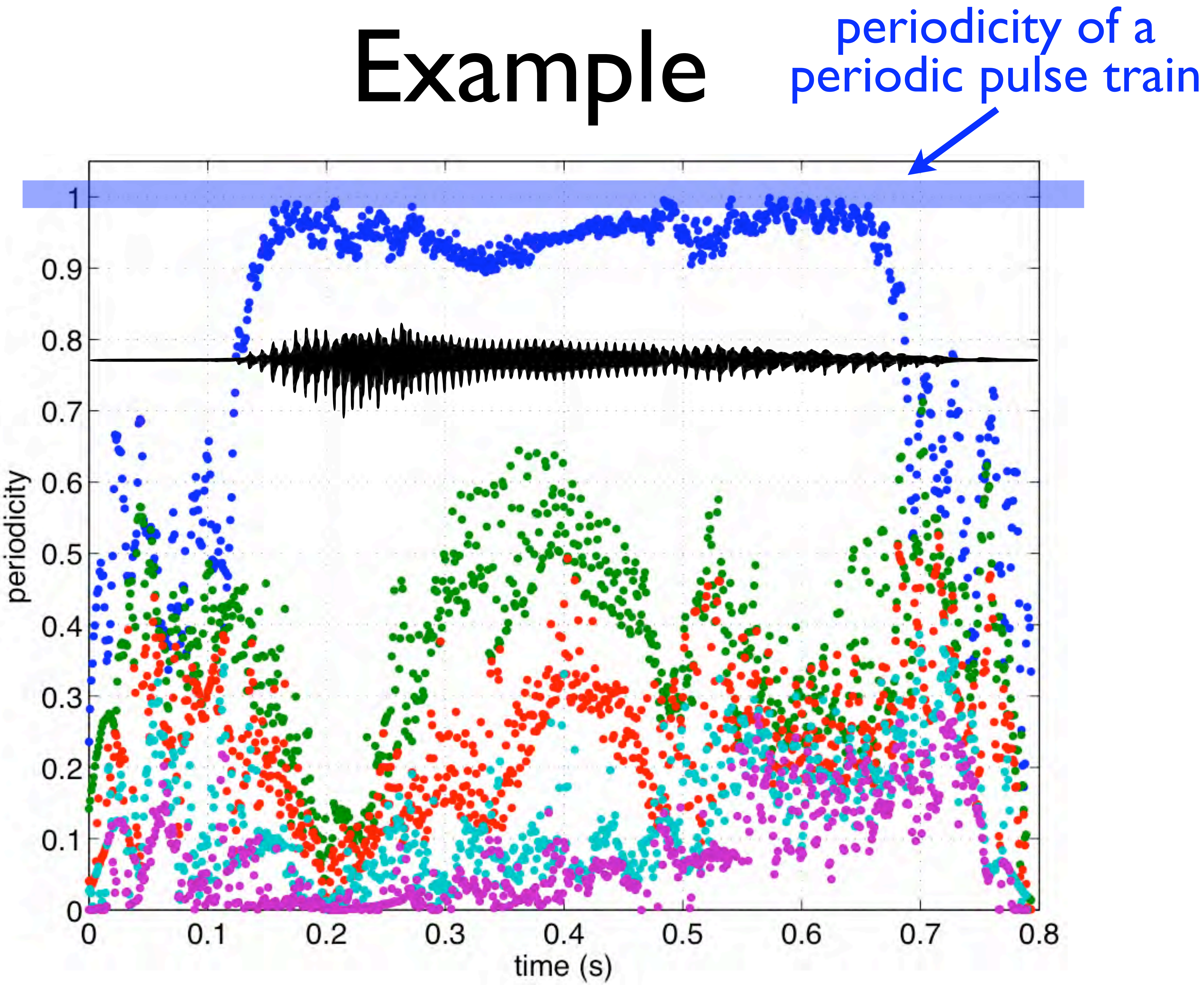


Example

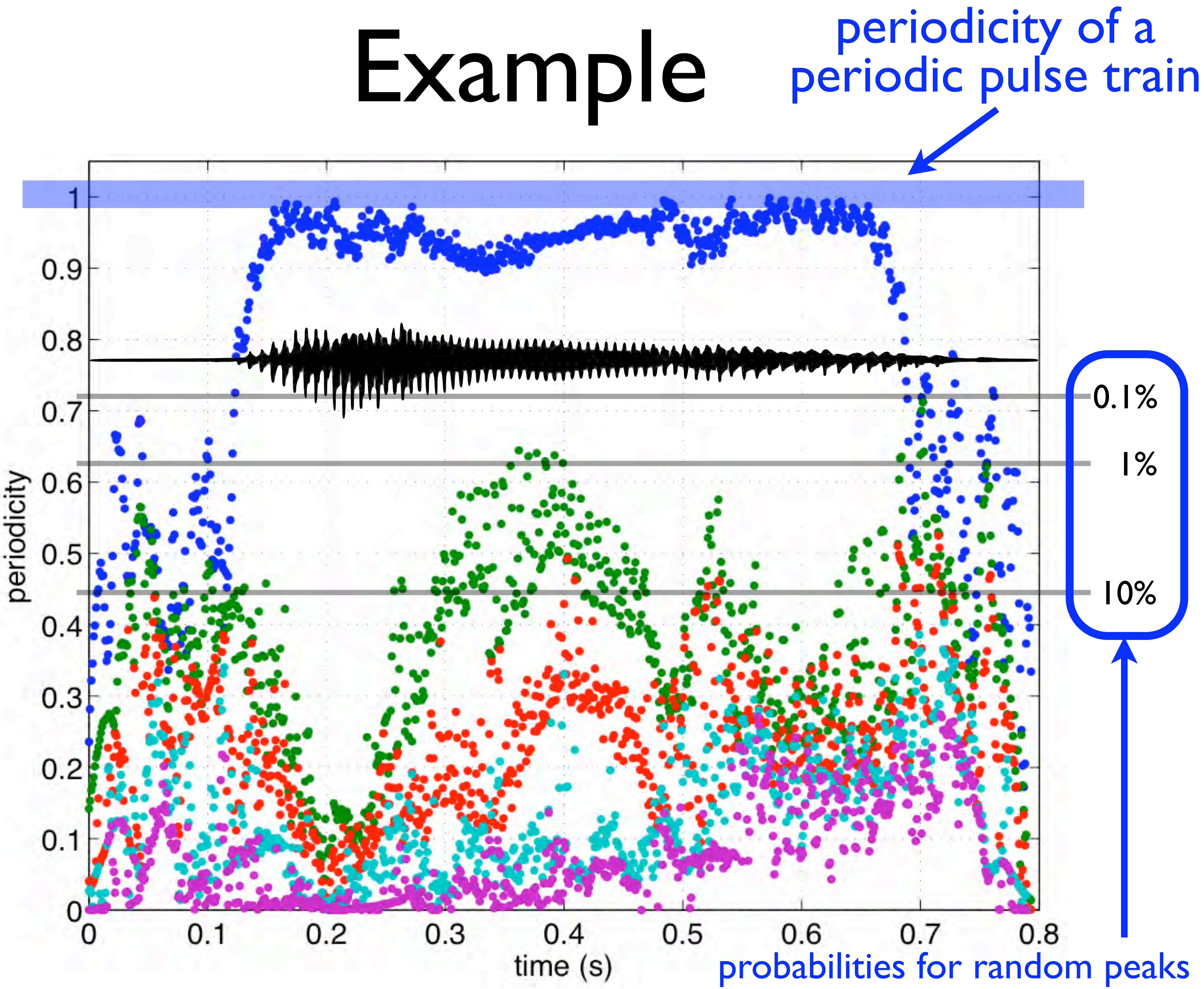
periodicity of a
periodic pulse train



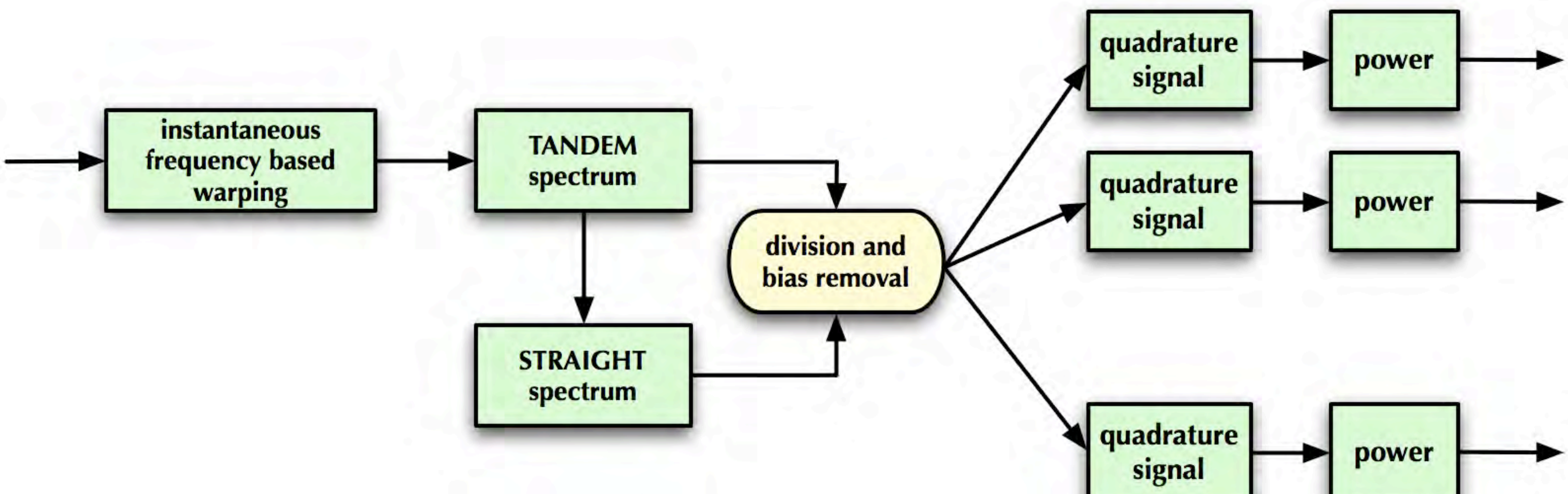
Example


periodicity of a periodic pulse train
periodicity
time (s)
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8

Example


periodicity of a
periodic pulse train
0.1%
1%
10%
probabilities for random peaks
periodicity
time (s)
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1

Periodicity spectrogram



Periodicity spectrogram

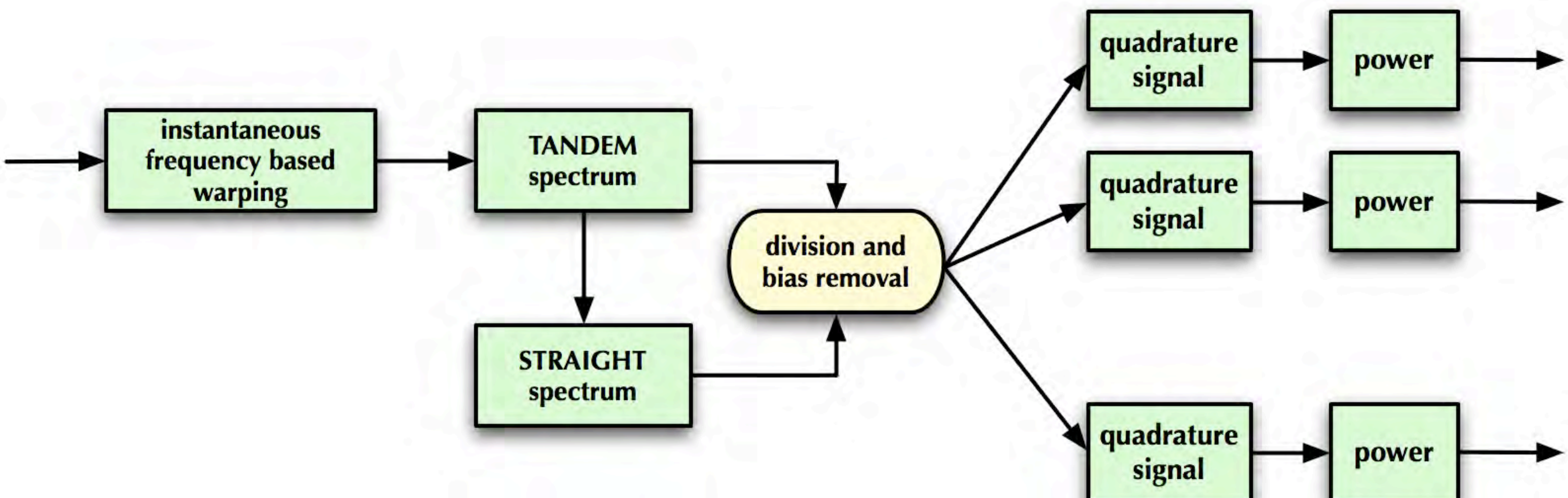
$$Q_C^2(\omega; T_C) = \left| \int_{-\infty}^{\infty} h_N(\lambda; T_C) P_C(\omega - \lambda; T_C) d\lambda \right|^2$$

quadrature signal to select
F0 related variation

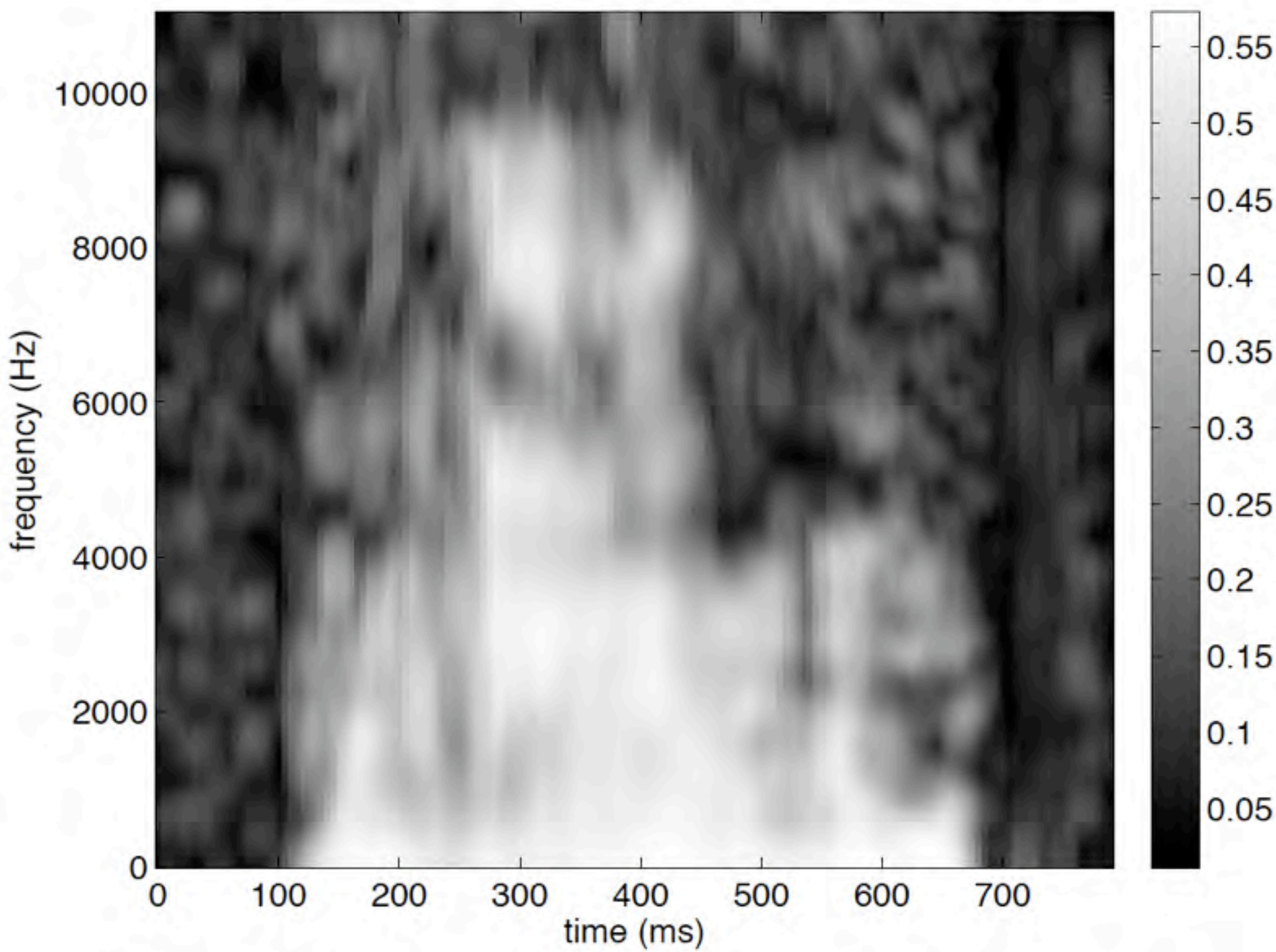
$$h_N(\omega) = w_{\omega 0, N}(\omega) \exp\left(2\pi j \omega / \omega_0\right)$$

a practical implementation
of envelope

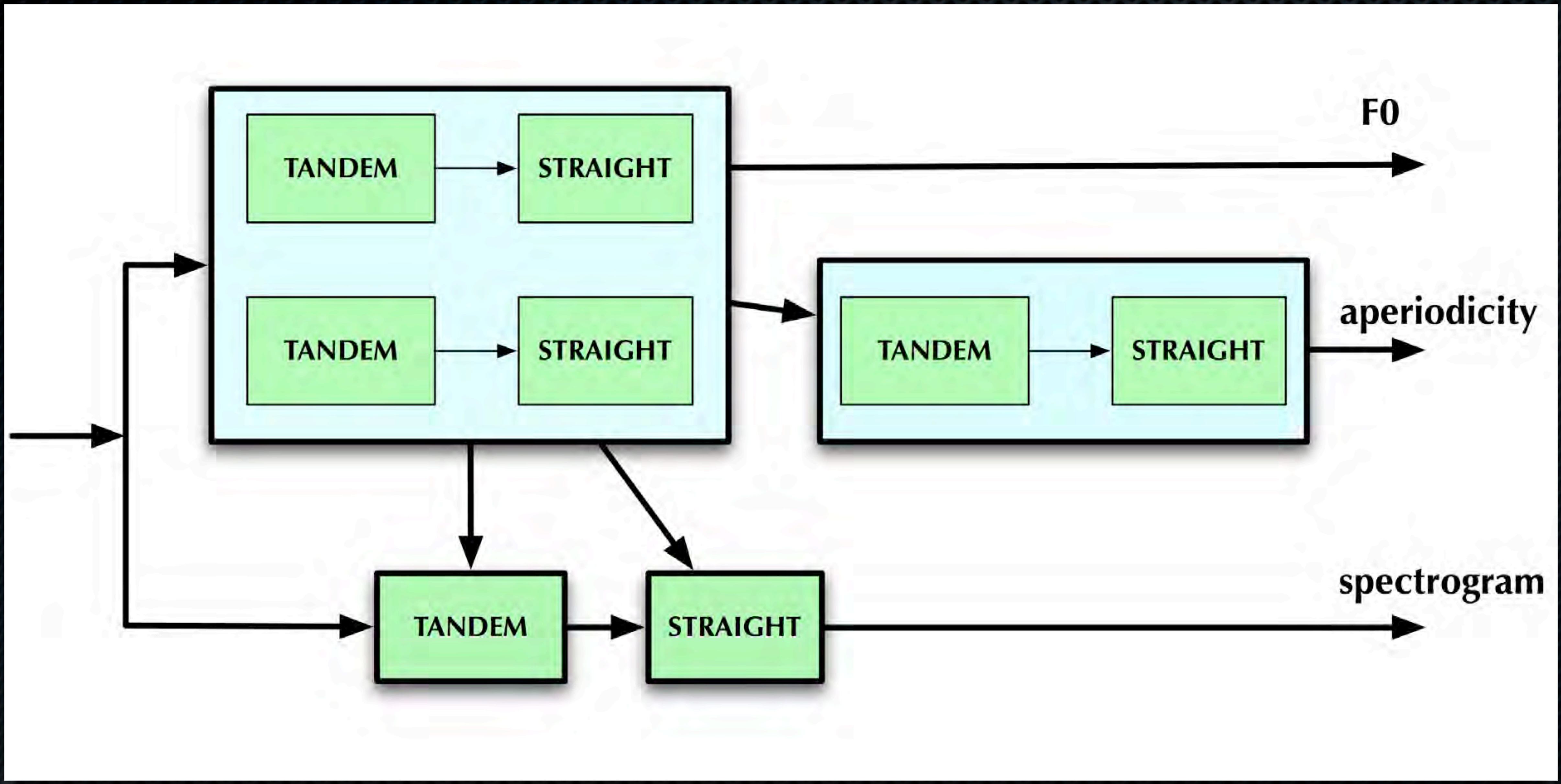
$$w_{\omega_C, N}(\omega) = c_0 \left(1 + \cos\left(\pi \omega / N \omega_C\right)\right)$$



Example

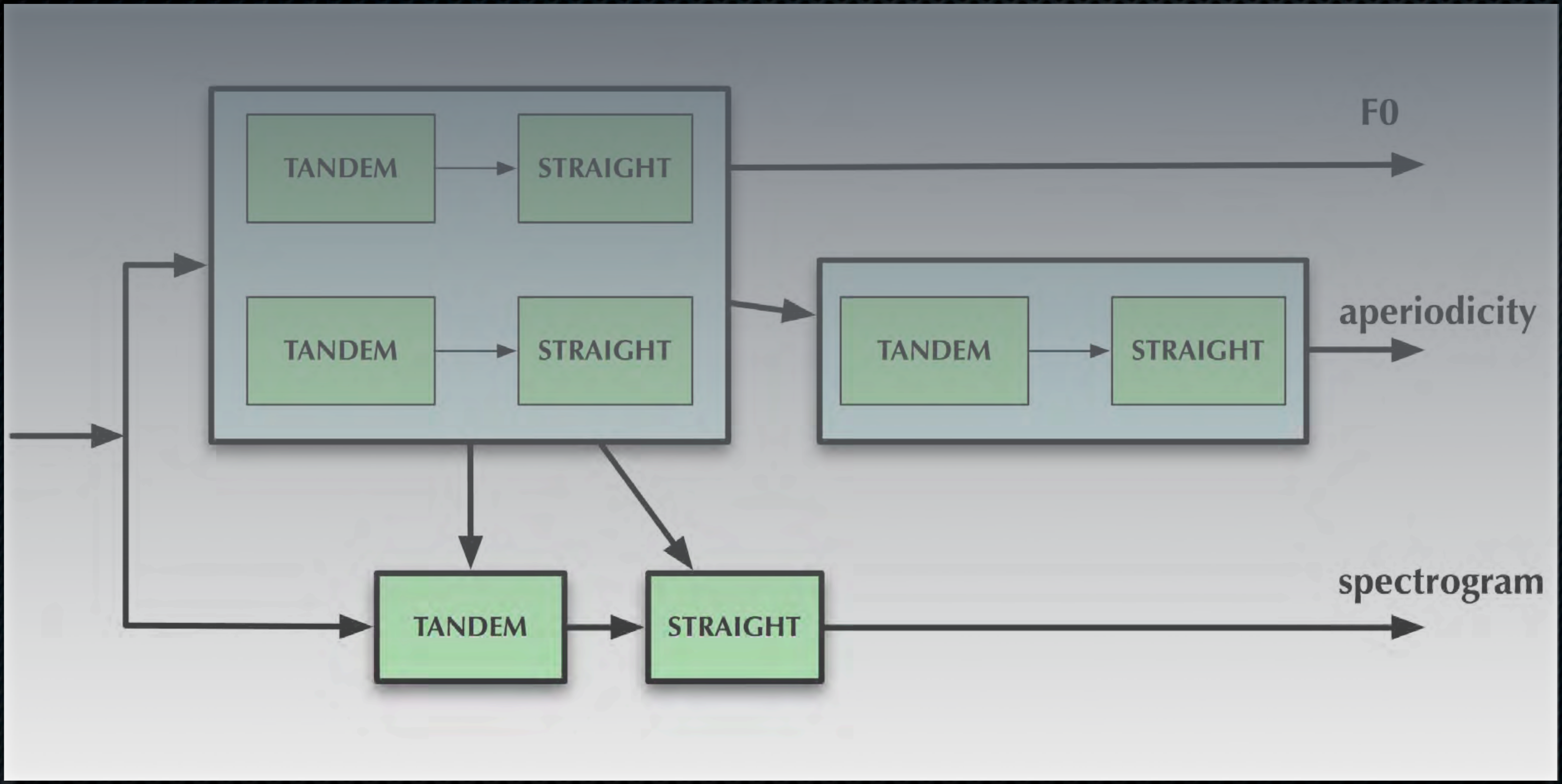


10000
8000
6000
4000
2000
0
frequency (Hz)
0
100
200
300
400
500
600
700
time (ms)
0.55
0.5
0.45
0.4
0.35
0.3
0.25
0.2
0.15
0.1
0.05



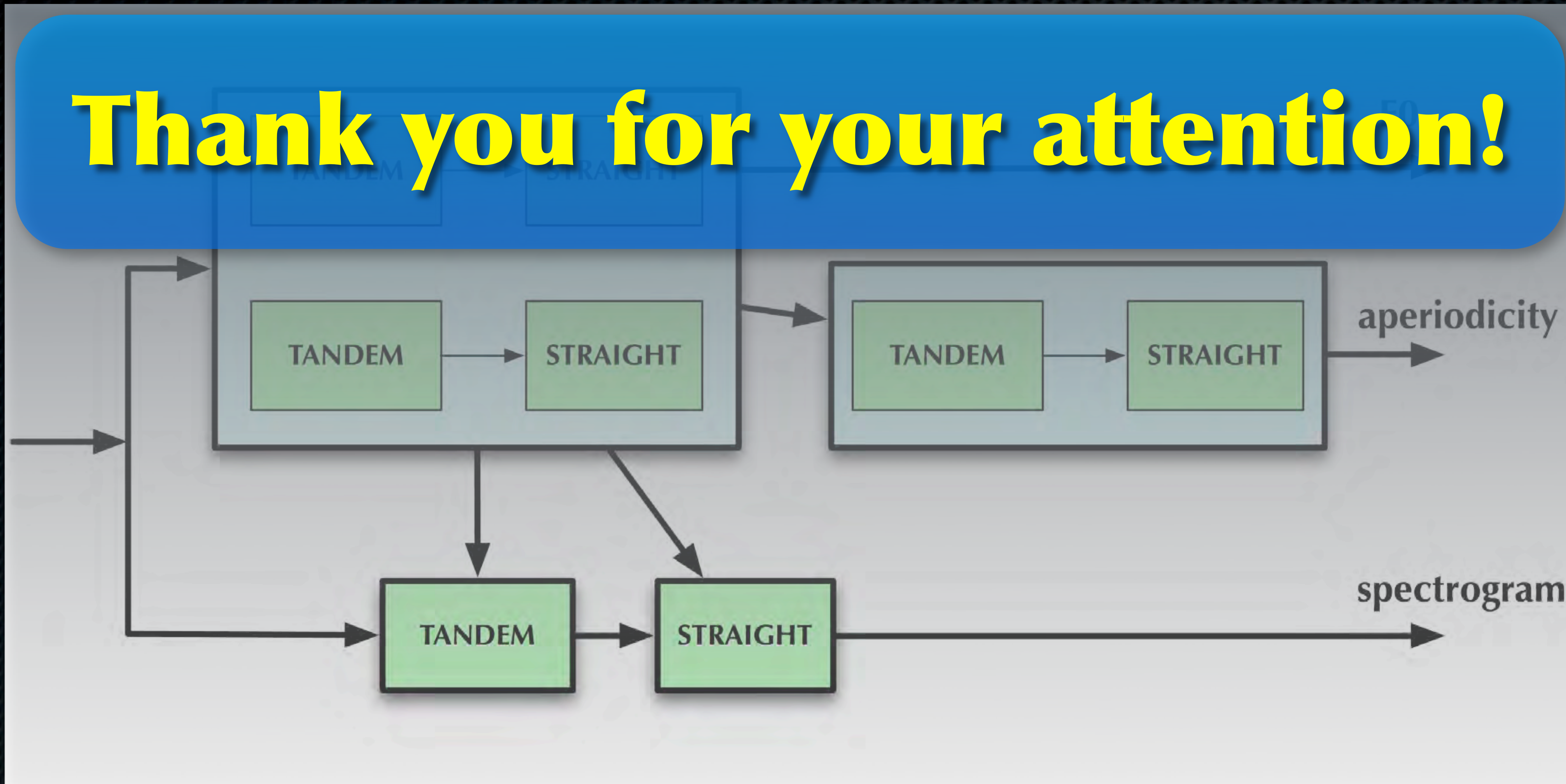
Concluding remarks

TANDEM-STRAIGHT is theoretically tractable



Concluding remarks

TANDEM-STRAIGHT is theoretically tractable

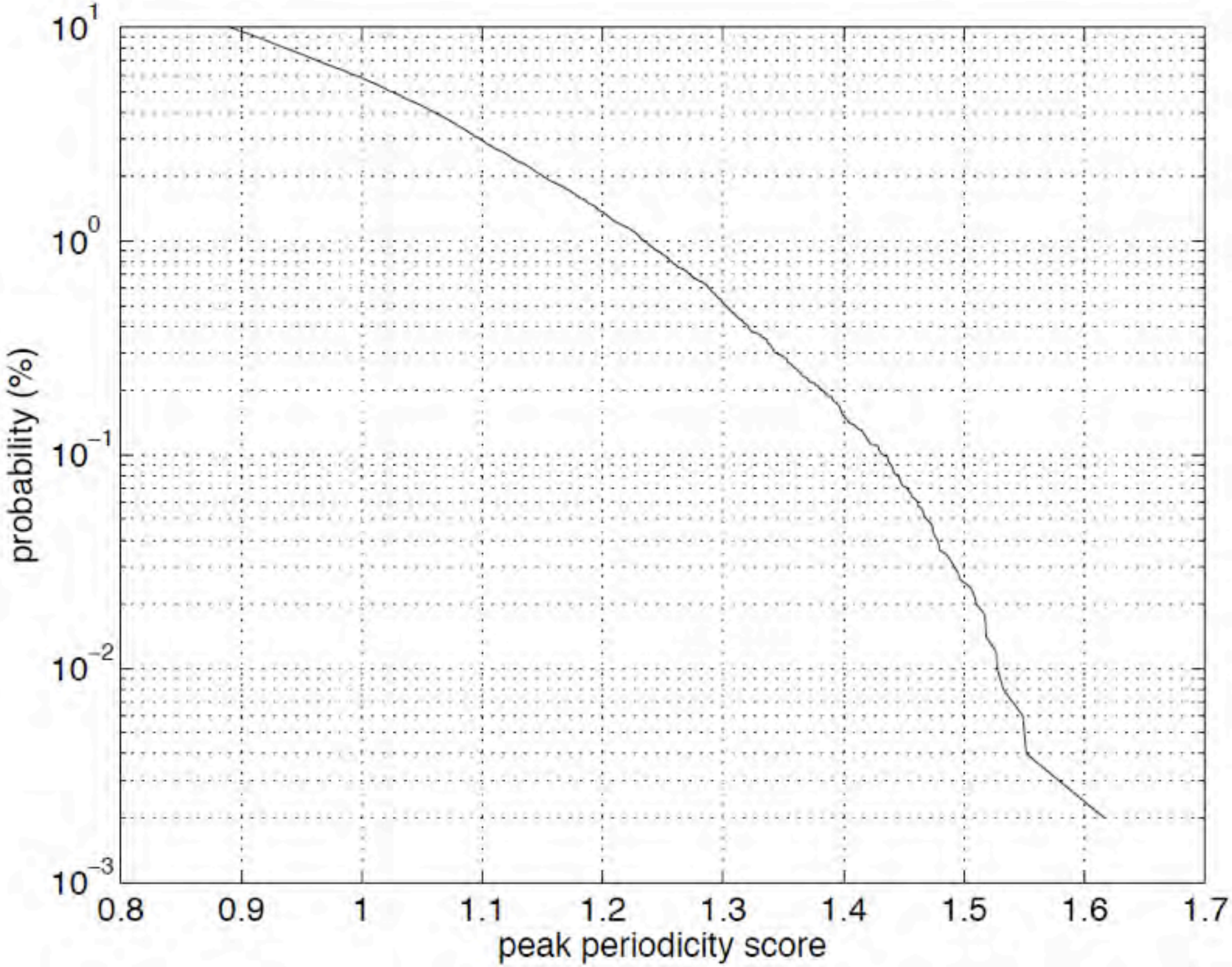


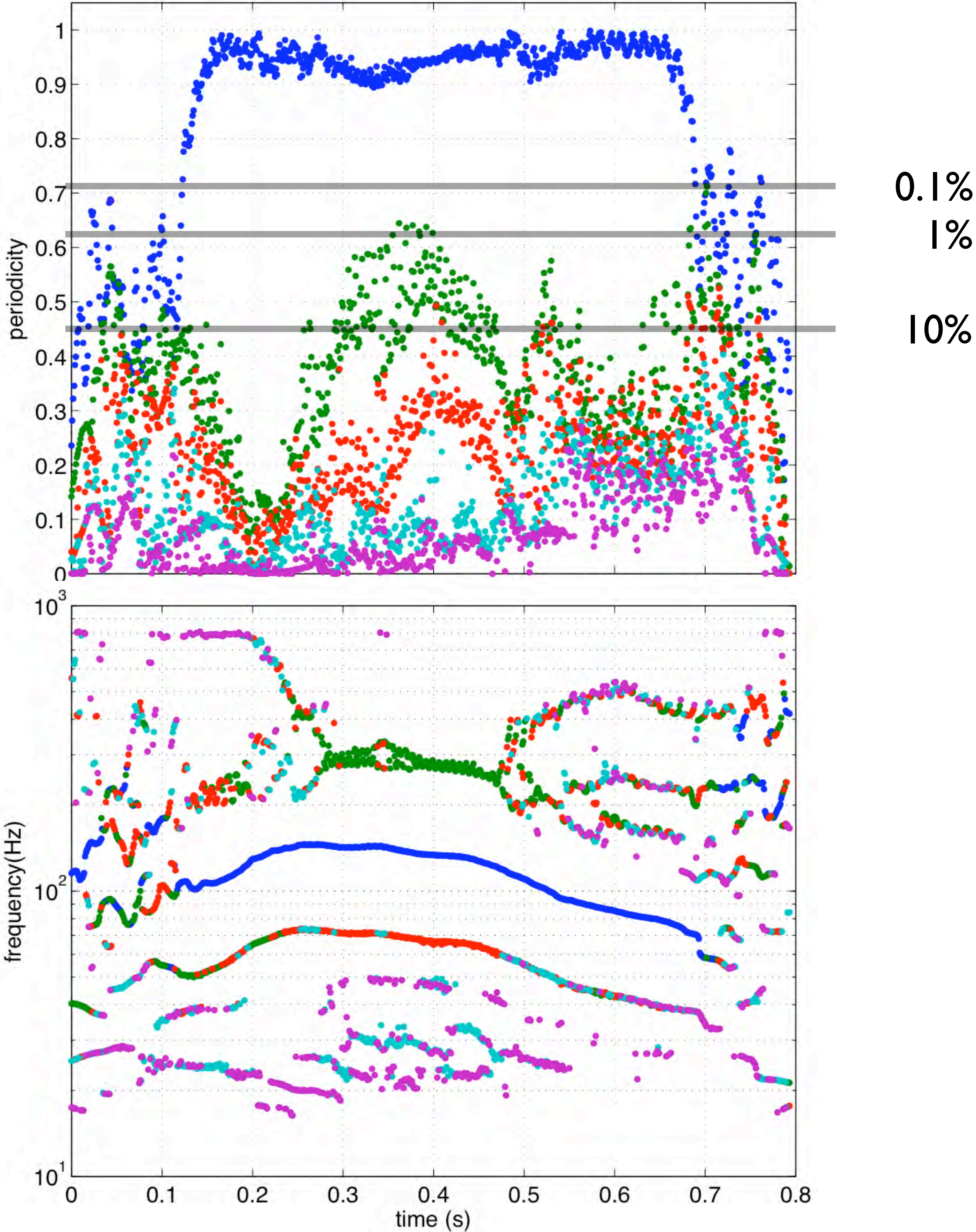
Concluding remarks

TANDEM-STRAIGHT is theoretically tractable

Appendix

Combined F0 detectors





periodicity
0.1%
1%
10%
frequency(Hz)
time (s)