Allocating, Detecting and Mining Sound Structures

Dörfler, Monika [1]
monika.doerfler@univie.ac.at

28th Sept., 2012

[1] Funded by Audio.Miner (WWTF Vienna)

The notion of sound object

AudioMiner: Vienna-based interdisciplinary project.
Exploring applications of modern mathematical methods in MIR.

The notion of sound object

AudioMiner: Vienna-based interdisciplinary project.
Exploring applications of modern mathematical methods in MIR.
Detection, classification and processing of sound objects

The notion of sound object

AudioMiner: Vienna-based interdisciplinary project.
Exploring applications of modern mathematical methods in MIR.
Detection, classification and processing of sound objects
Mathematics $\longleftrightarrow$ MIR - both ways!

The notion of sound object

AudioMiner: Vienna-based interdisciplinary project.
Exploring applications of modern mathematical methods in MIR.
Detection, classification and processing of sound objects
Mathematics $\longleftrightarrow$ MIR - both ways!
Particular focus on identification and extraction of sound objects from (polyphonic) audio or acousmatic sound signals

The notion of sound object

AudioMiner: Vienna-based interdisciplinary project.
Exploring applications of modern mathematical methods in MIR.
Detection, classification and processing of sound objects
Mathematics $\longleftrightarrow$ MIR - both ways!
Particular focus on identification and extraction of sound objects from (polyphonic) audio or acousmatic sound signals
$\longrightarrow$ but what is a sound object?

The notion of sound object

Sound object, sound structure or Gestalt?

- G. Monaci, P.Vandergheynst, "Audiovisual Gestalts," 2012
- I. Zannos (Ed.):Music and Signs – Semiotic and Cognitive Studies in Music, 1999 (M. Leman)
- V. Klien, T. Grill, and A. Flexer. On automated annotation of acousmatic music, 2012
- D. Smalley. Spectromorphology: explaining sound-shapes, 1997.

The notion of sound object

Sound object, sound structure or Gestalt?

- G. Monaci, P.Vandergheynst, "Audiovisual Gestalts," 2012
- I. Zannos (Ed.):Music and Signs – Semiotic and Cognitive Studies in Music, 1999 (M. Leman)
- V. Klien, T. Grill, and A. Flexer. On automated annotation of acousmatic music, 2012
- D. Smalley. Spectromorphology: explaining sound-shapes, 1997.

Object, Structure or Gestalt: interpretation depends on representation!

The notion of sound object

Object or Gestalt implies sparsity or compactness in representation.

The notion of sound object

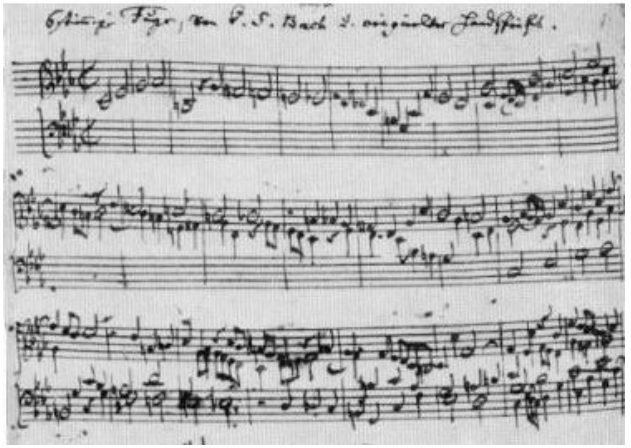
Object or Gestalt implies sparsity or compactness in representation.
Representation by *few* as opposed to *many* elements.

The notion of sound object

Object or Gestalt implies sparsity or compactness in representation.
Representation by *few* as opposed to *many* elements.
What do we mean by representation?

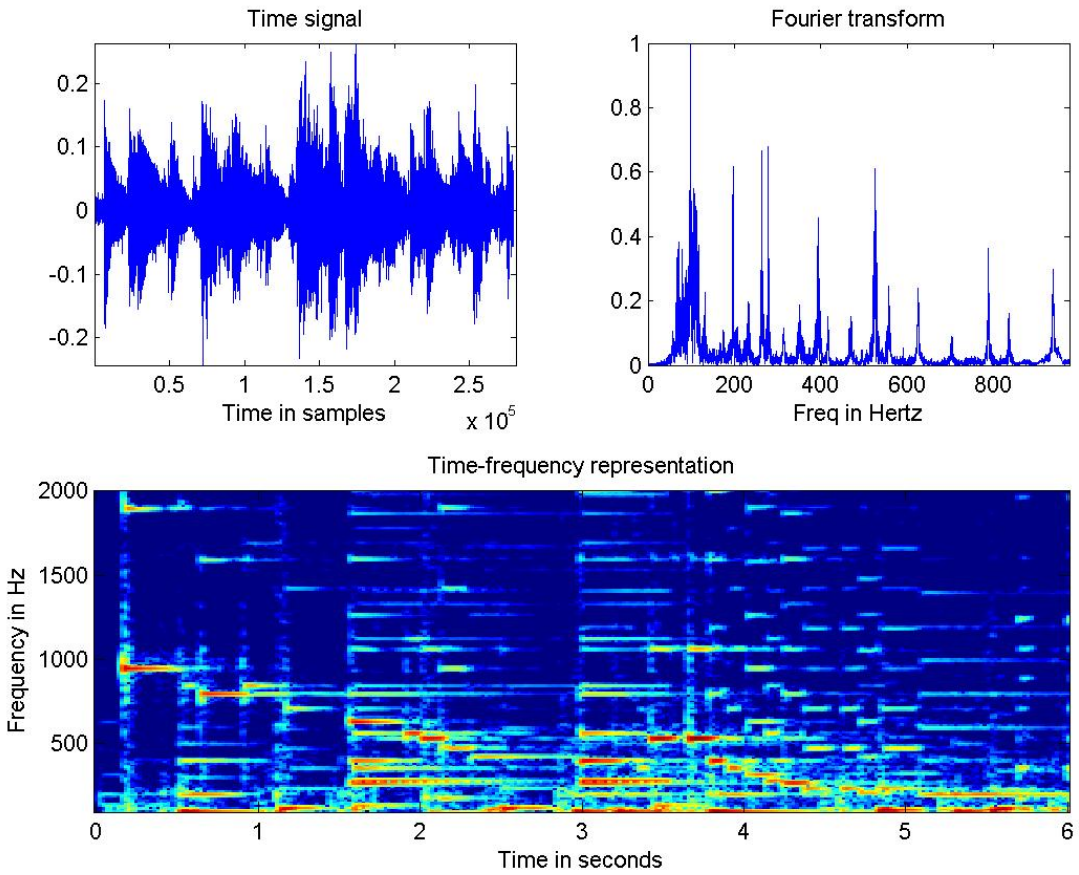
The notion of sound object

Object or Gestalt implies sparsity or compactness in representation.
Representation by *few* as opposed to *many* elements.
What do we mean by representation?



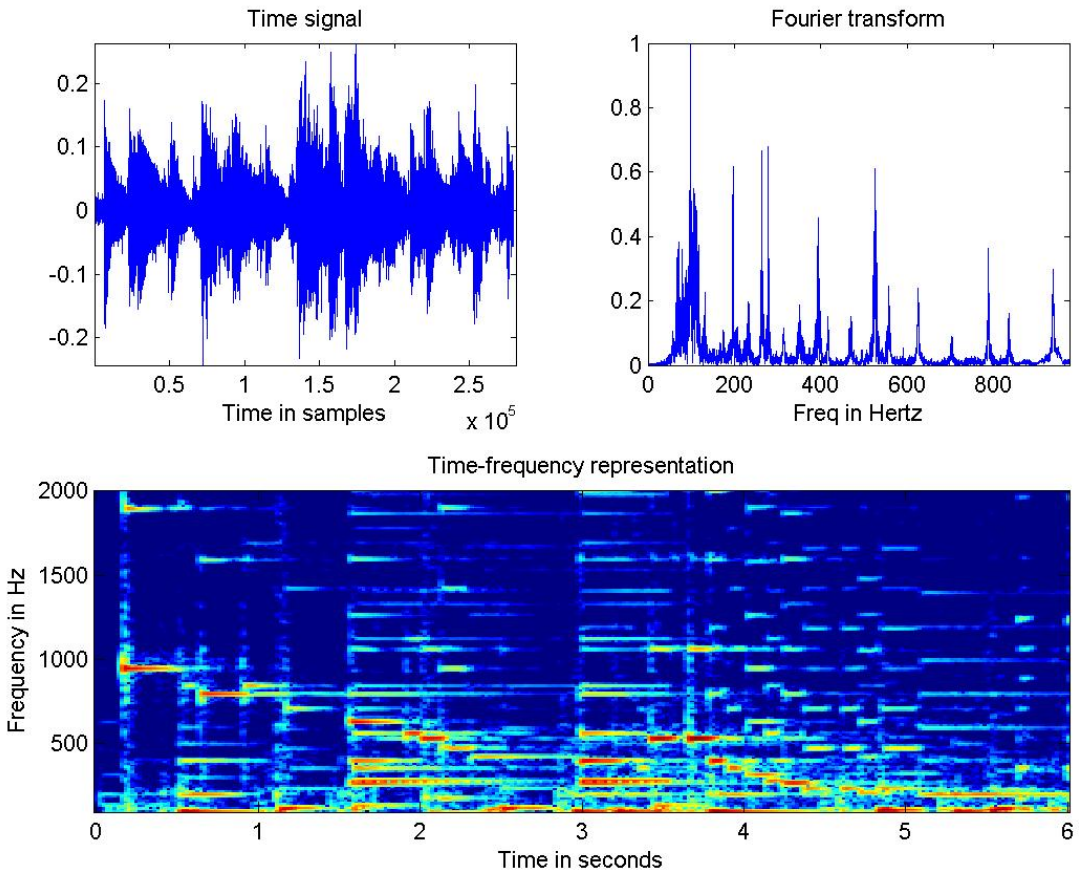
Time, Frequency, and Time-Frequency

$$f[n] = \sum_n c_n \delta_n = \sum_k \hat{c}_k (\cos 2\pi kn + \sin 2\pi kn)$$



Time, Frequency, and Time-Frequency

$$f[n] = \sum_n c_n \delta_n = \sum_k \hat{c}_k (\cos 2\pi kn + \sin 2\pi kn) = \sum_{n,k} c_{n,k} \varphi_{n,k}$$



Time, Frequency, and Time-Frequency

- Analysis tool should render representation for visual display, reflecting a user's acoustical impression.

Time, Frequency, and Time-Frequency

- Analysis tool should render representation for visual display, reflecting a user's acoustical impression.
- Sound objects should be visible as distinct from a more textural background.
- More sophisticated representations of sound signals than classical representations do exist now:
  - Adaptivity in the transformation parameters: sharpen the visual and assure a perfect connection between signal and representation (invertibility).
  - Bayesian coefficient priors highlight particular structures by means of informed analysis.

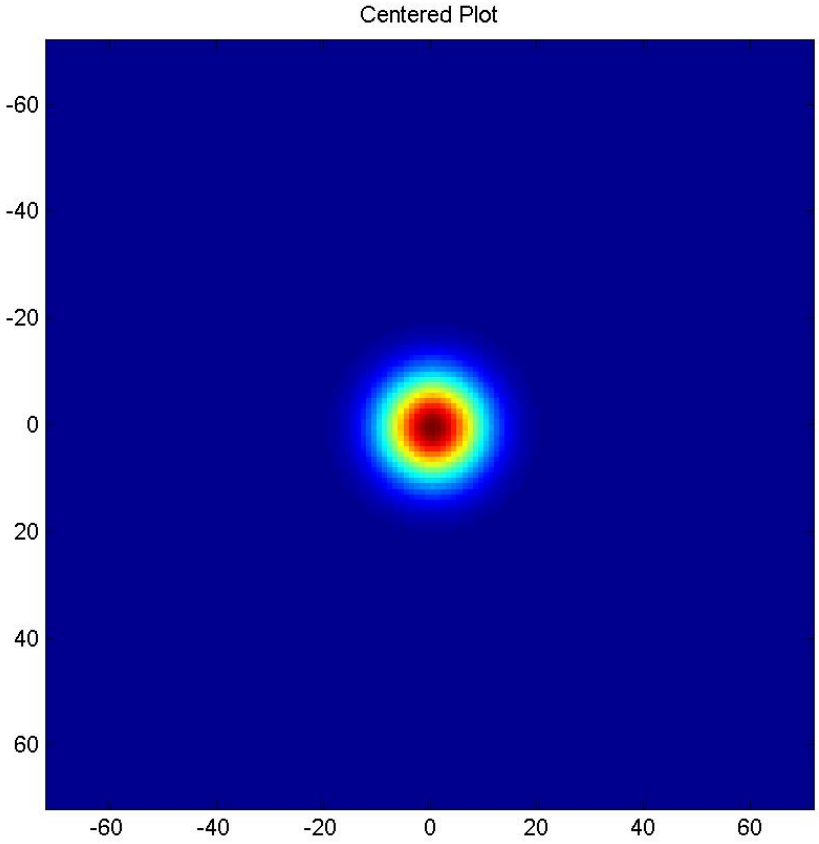
Time, Frequency, and Time-Frequency

- Analysis tool should render representation for visual display, reflecting a user's acoustical impression.
- Sound objects should be visible as distinct from a more textural background.
- More sophisticated representations of sound signals than classical representations do exist now:
  - Adaptivity in the transformation parameters: sharpen the visual and assure a perfect connection between signal and representation (invertibility).
  - Bayesian coefficient priors highlight particular structures by means of informed analysis.

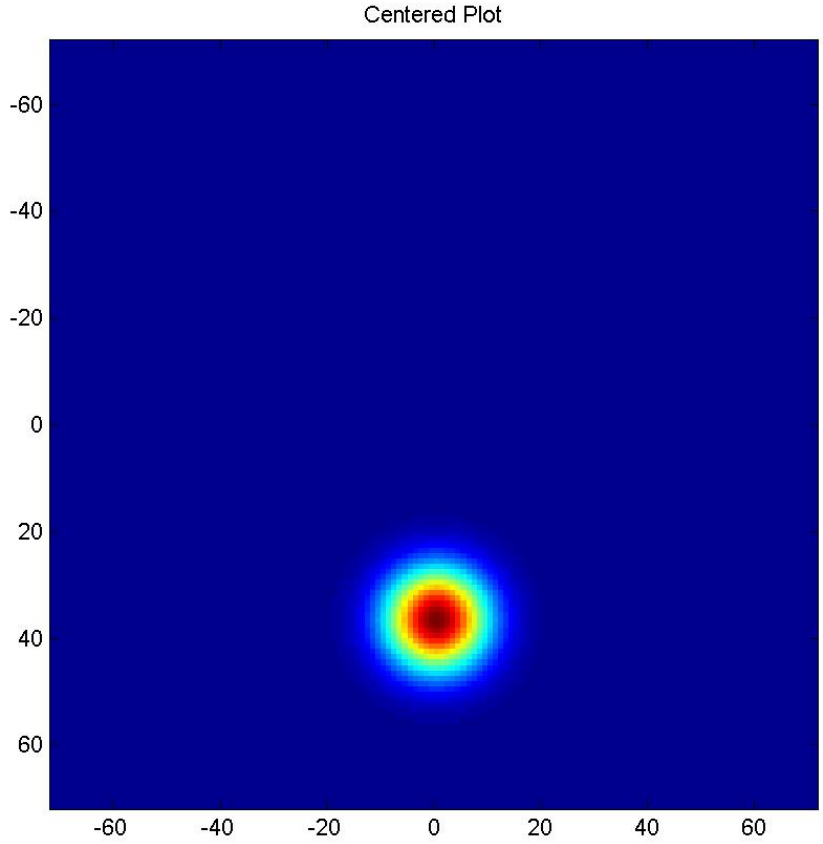
Time-Frequency analysis: Principles

- FFT-based time-variant analysis
- Mathematical frame theory always provides perfect reconstruction
- $f = \sum_{n,k} \langle f, \varphi^t_{n,k} \rangle \varphi^t_{n,k}$
- $\varphi_{n,k} := M_{kb} T_{na} \varphi$
- Choice of involved parameters (φ, a, b) is vital for the quality of representation

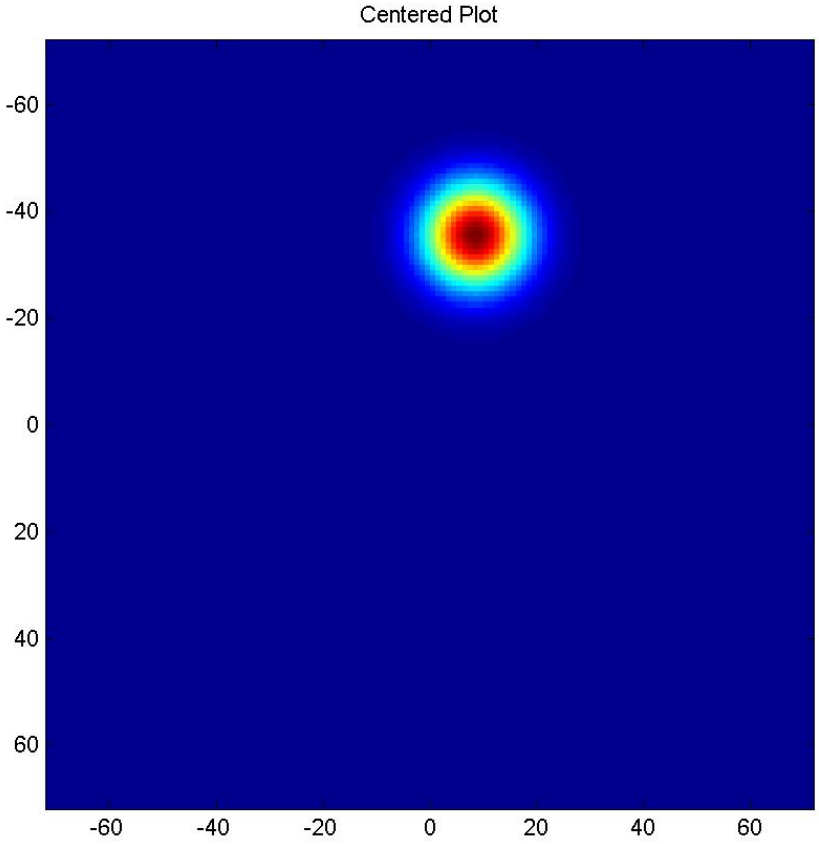
Time-Frequency analysis: Principles



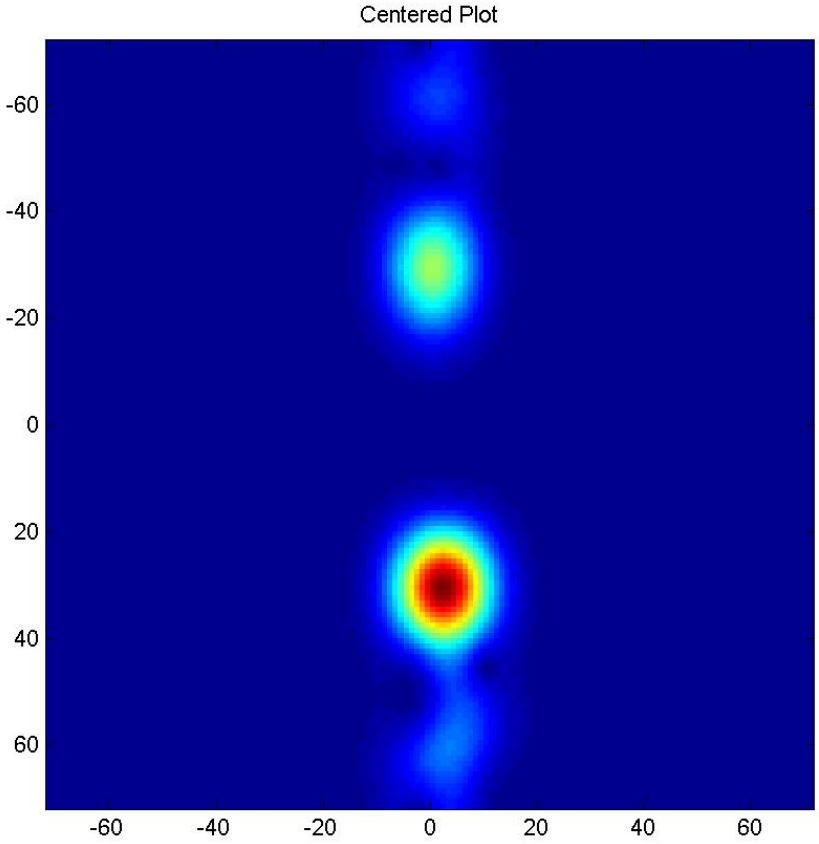
Time-Frequency analysis: Principles



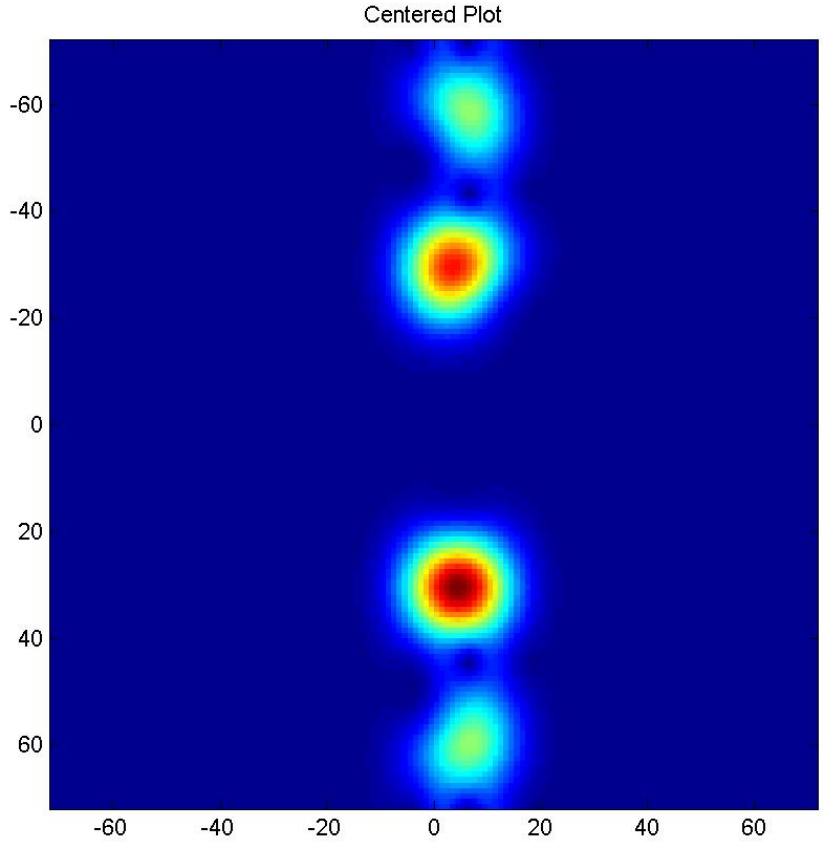
Time-Frequency analysis: Principles



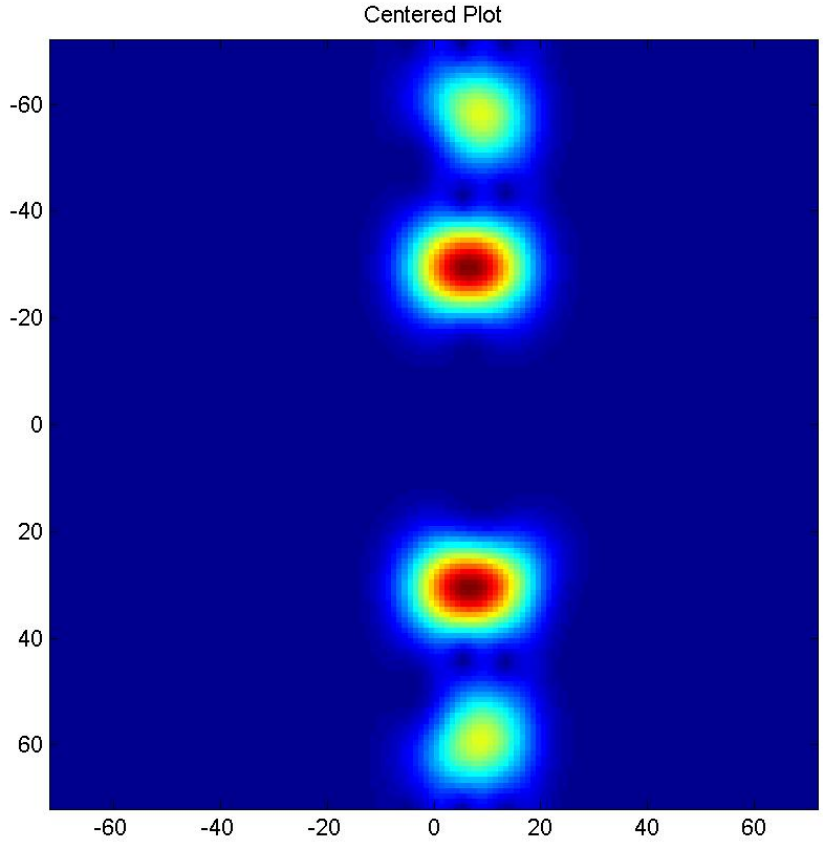
Time-Frequency analysis: Principles



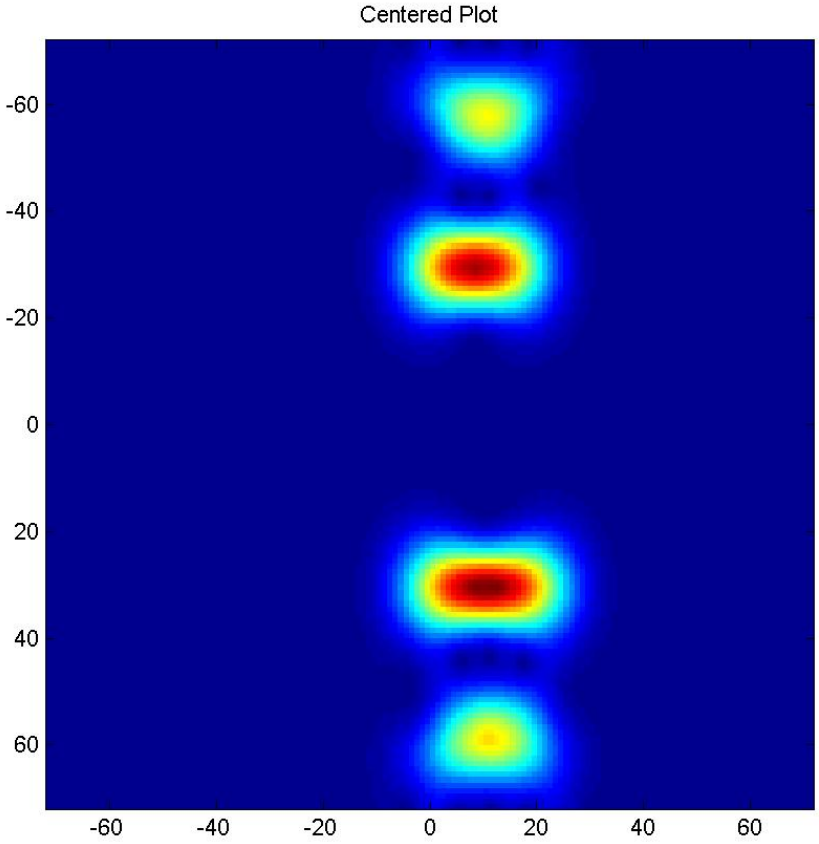
Time-Frequency analysis: Principles



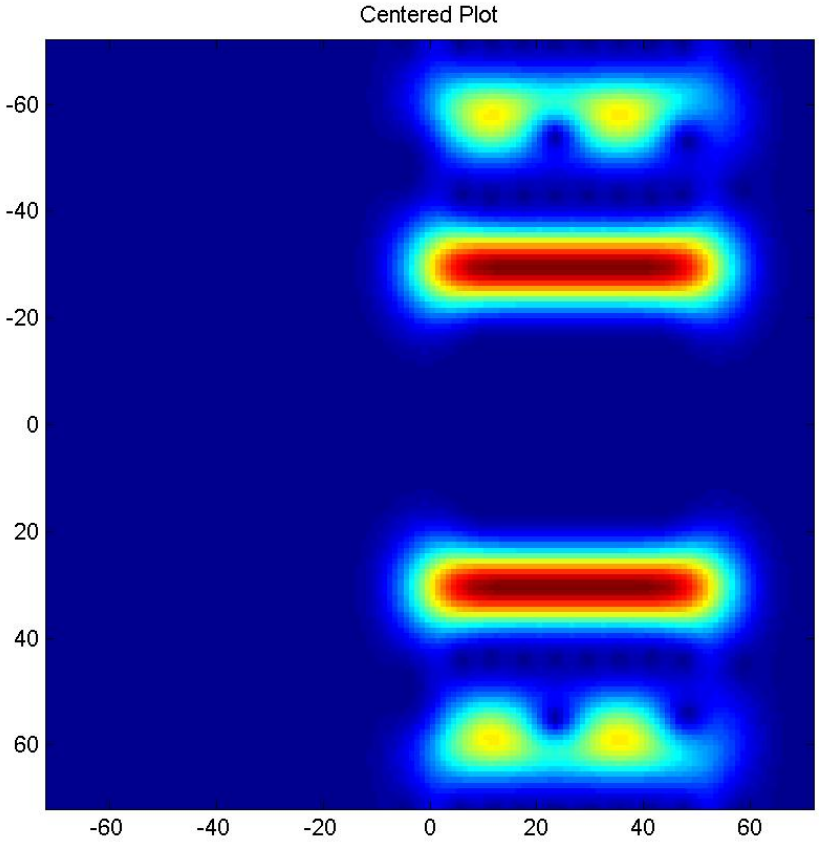
Time-Frequency analysis: Principles



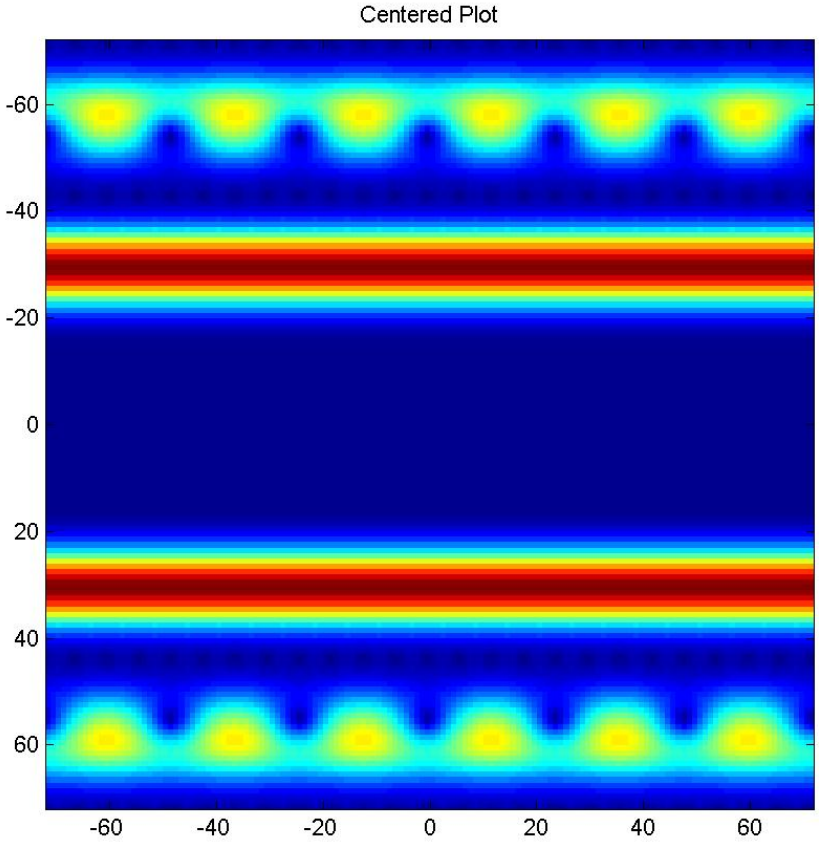
Time-Frequency analysis: Principles



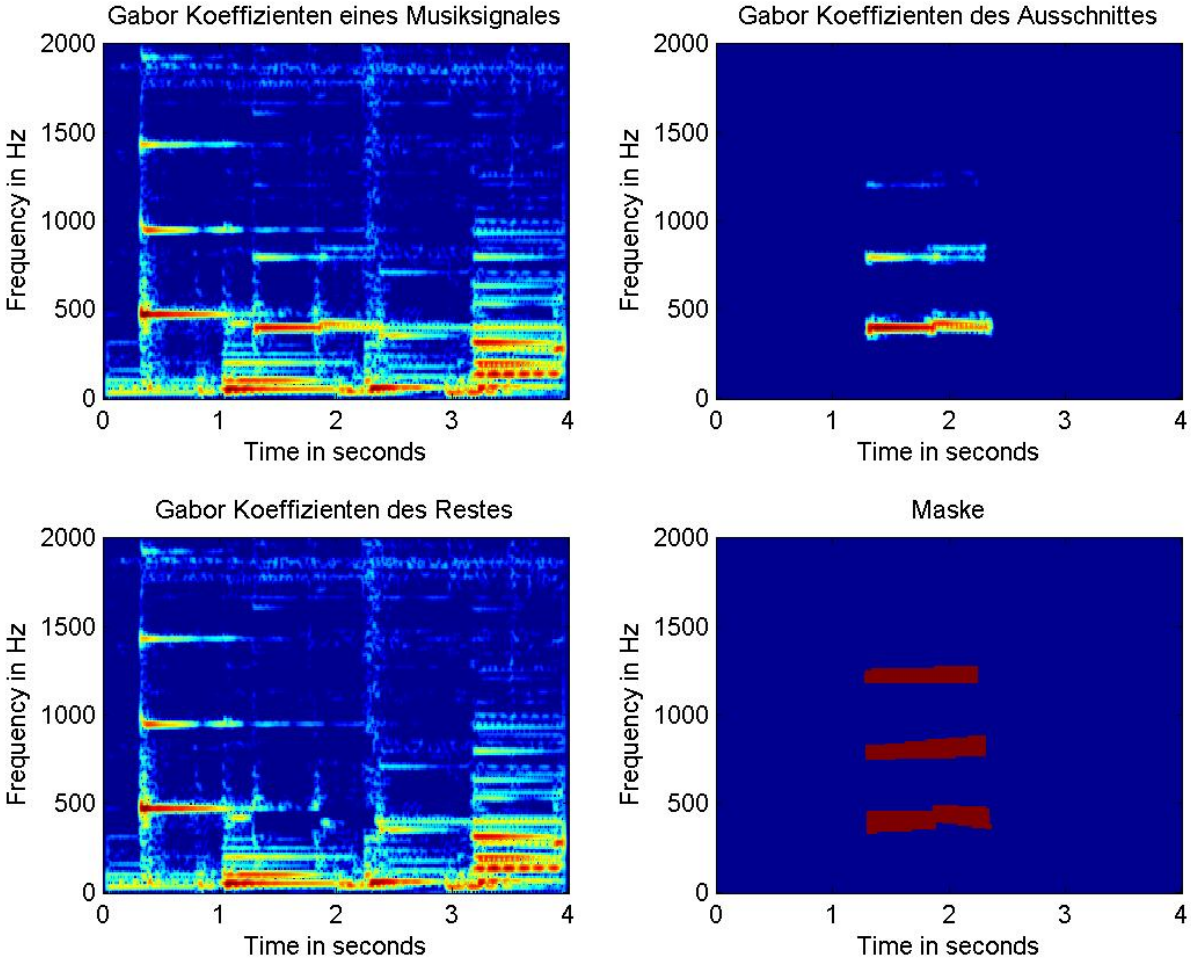
Time-Frequency analysis: Principles



Time-Frequency analysis: Principles



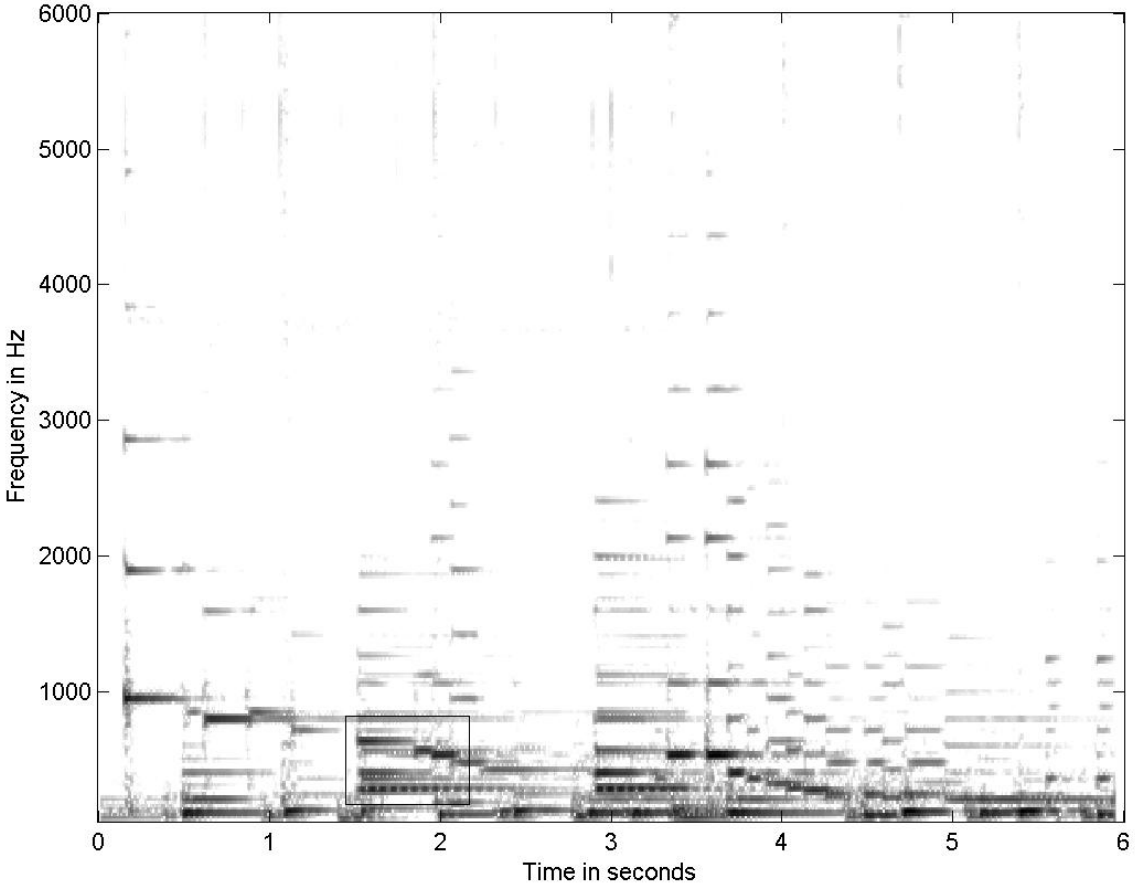
Time-frequency Multipliers



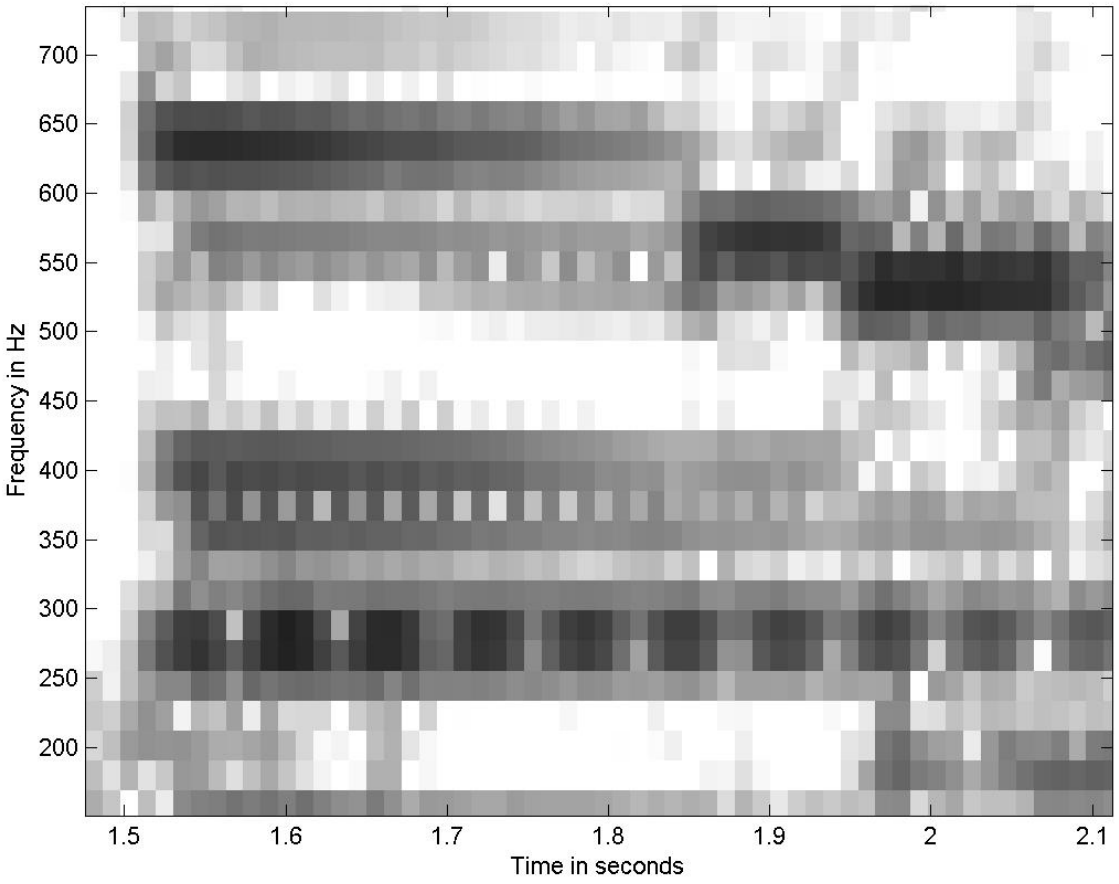
(*Listen!*)

Time-Frequency analysis: Principles

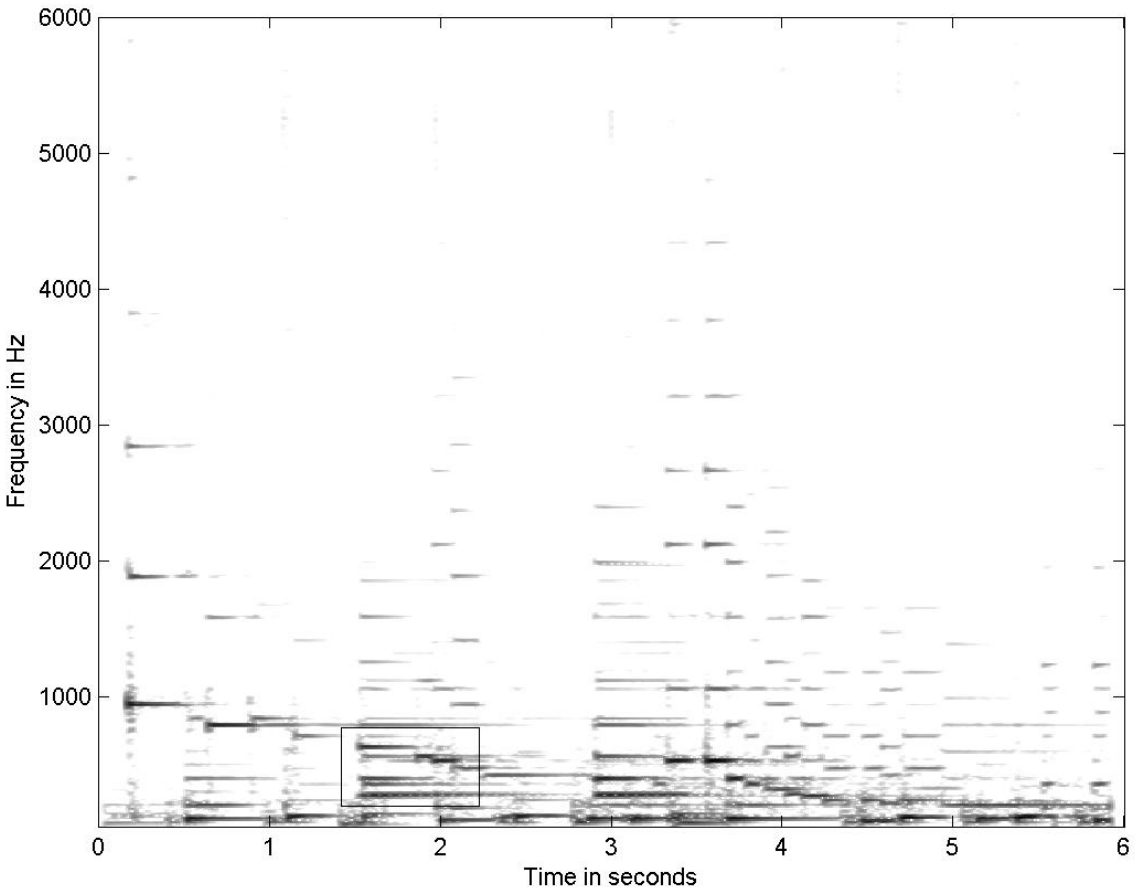
Time-Frequency analysis: Principles



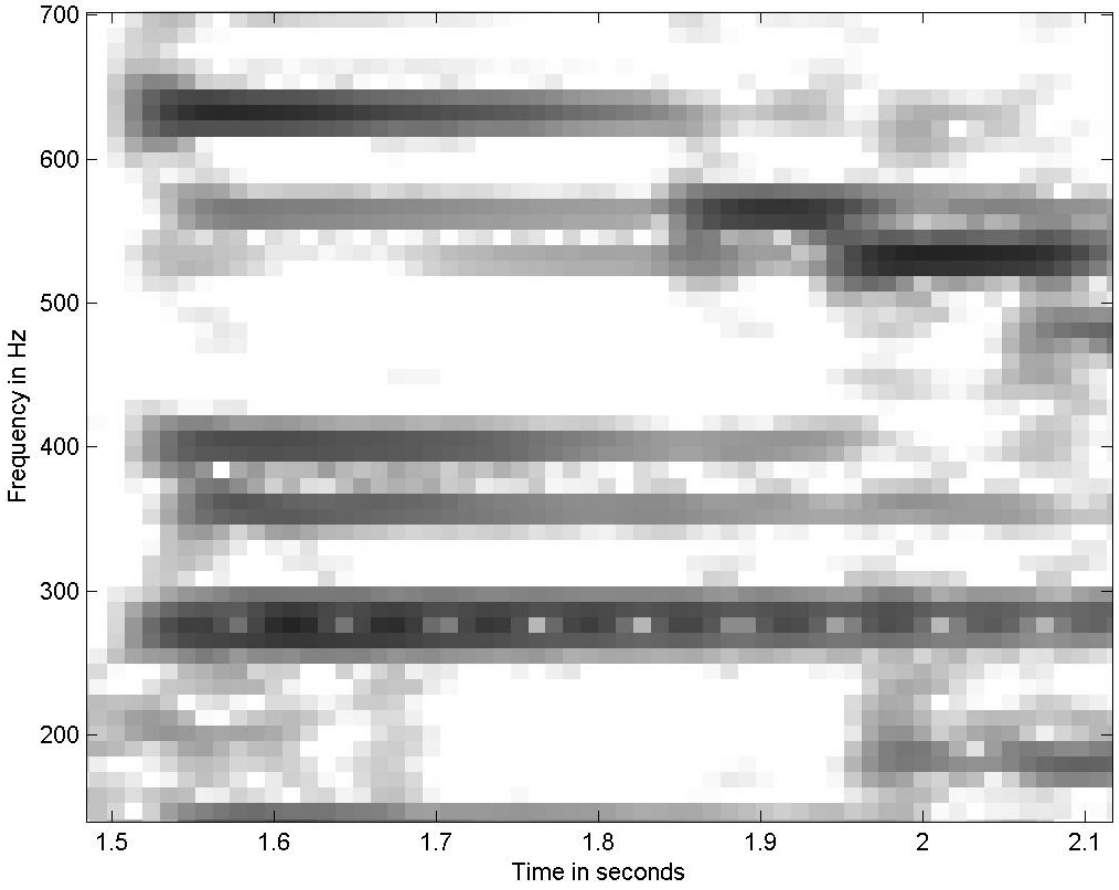
Time-Frequency analysis: Principles



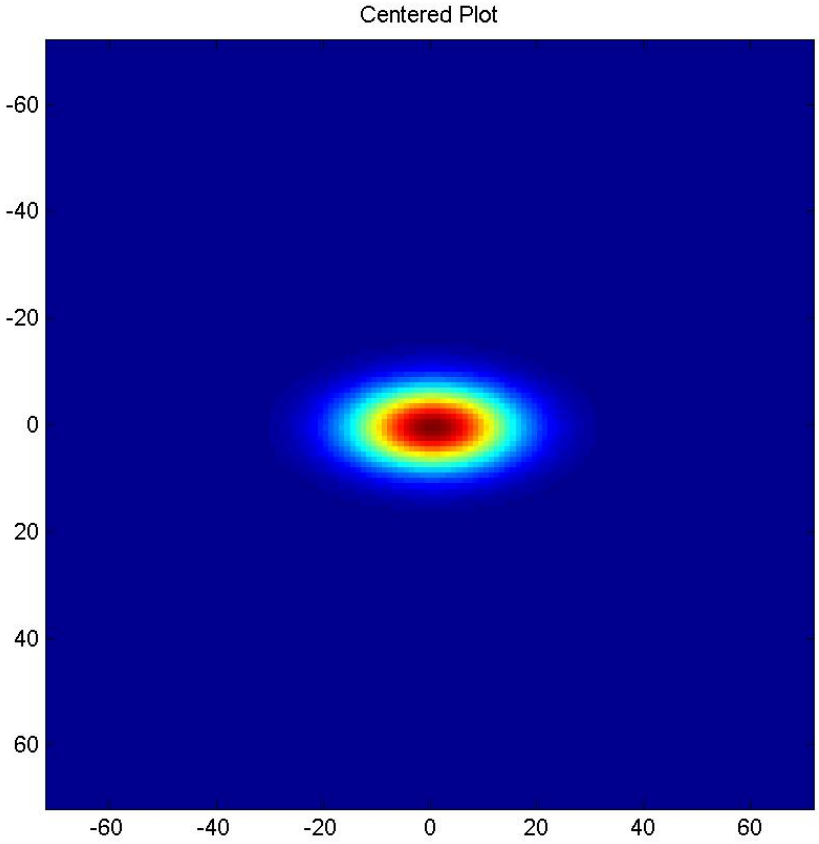
Time-Frequency analysis: Principles



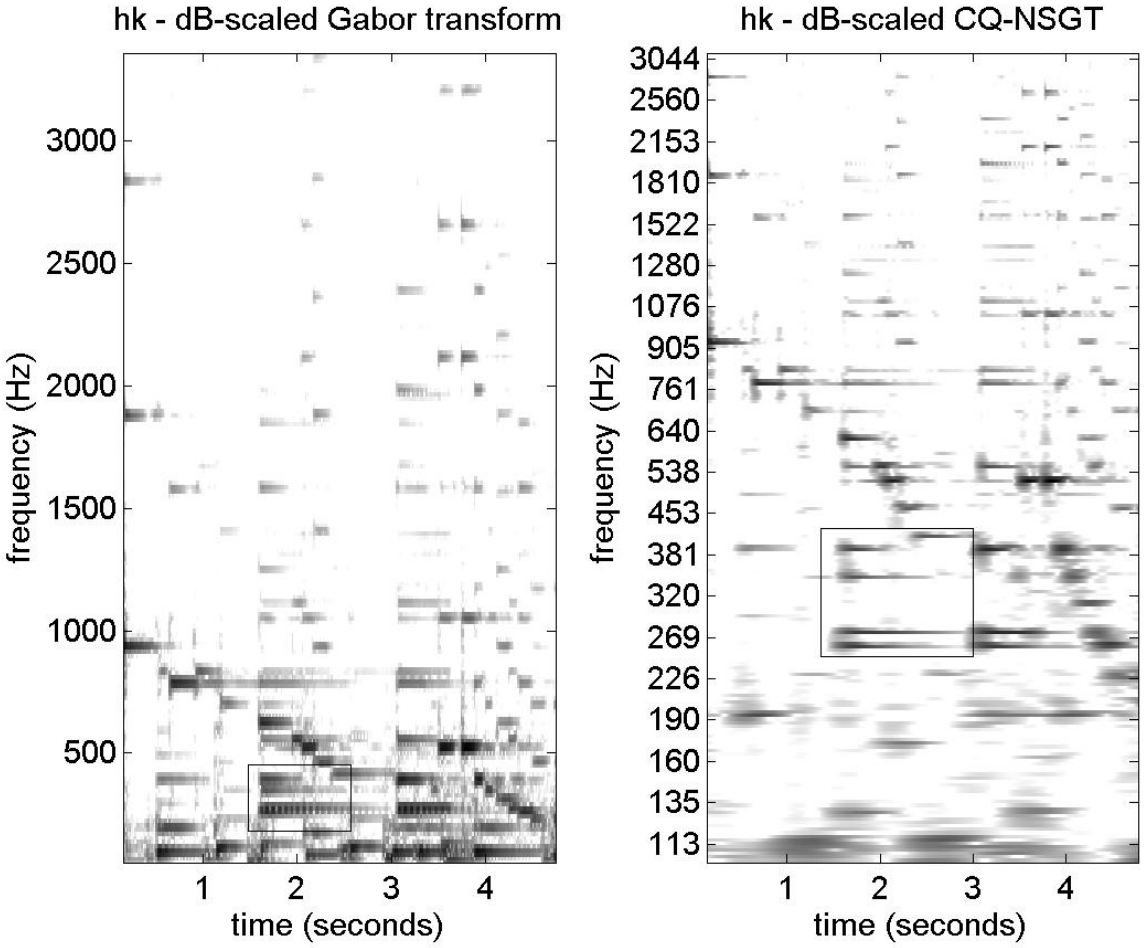
Time-Frequency analysis: Principles



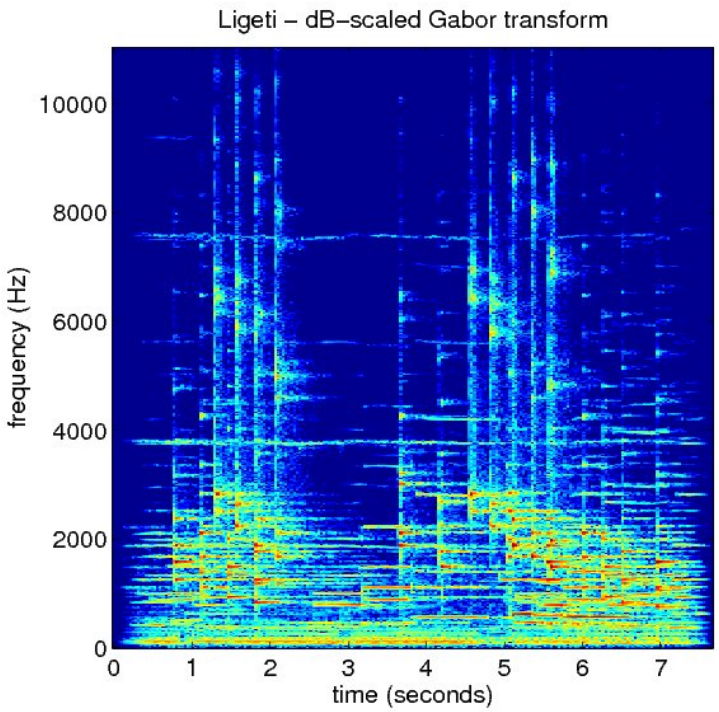
Time-Frequency analysis: Principles

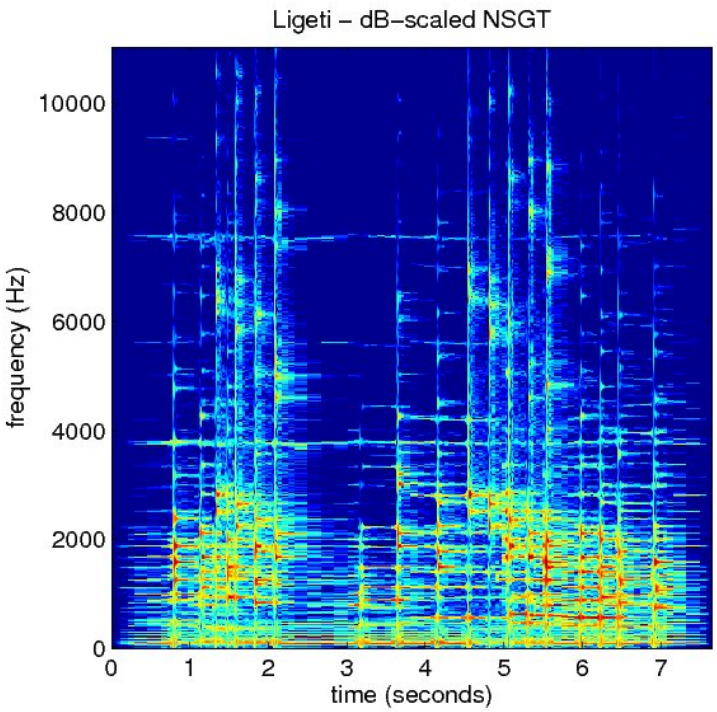


Adaptive signal representations

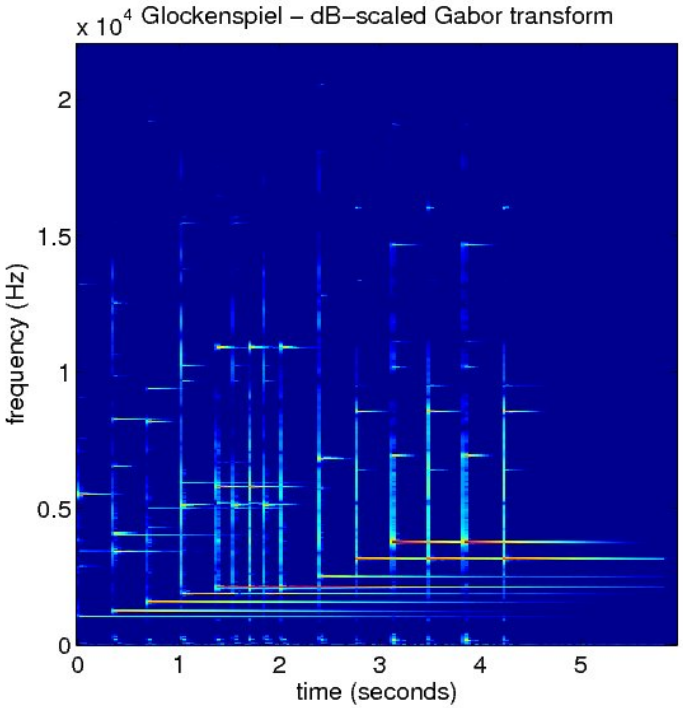


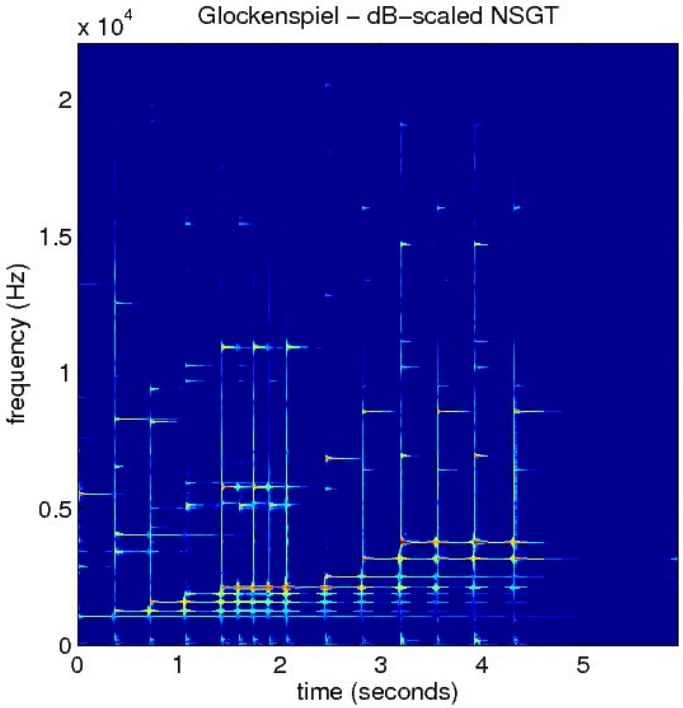
Adaptive signal representations





Adaptive signal representations

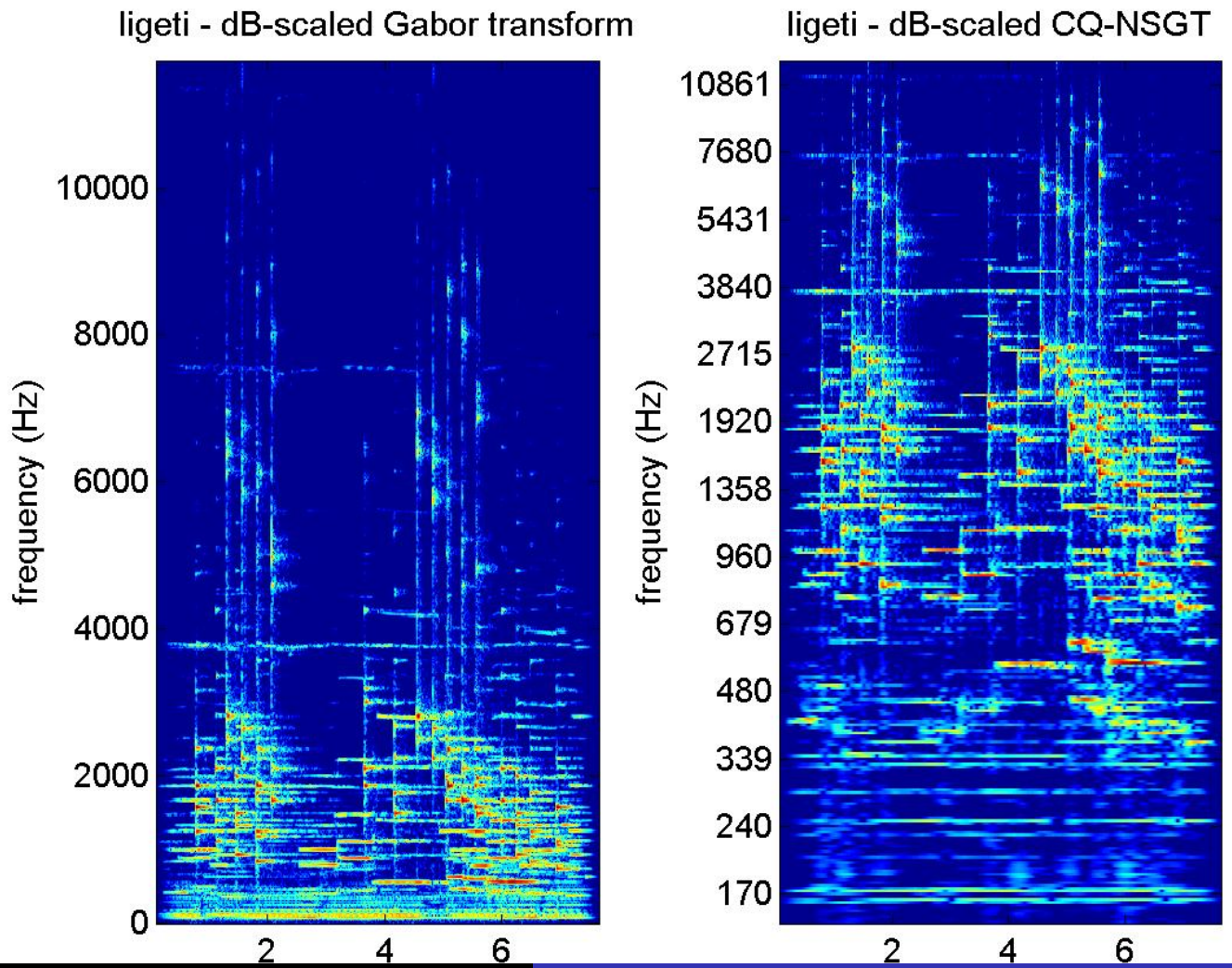




(*Listen!*) (*Listen!*)

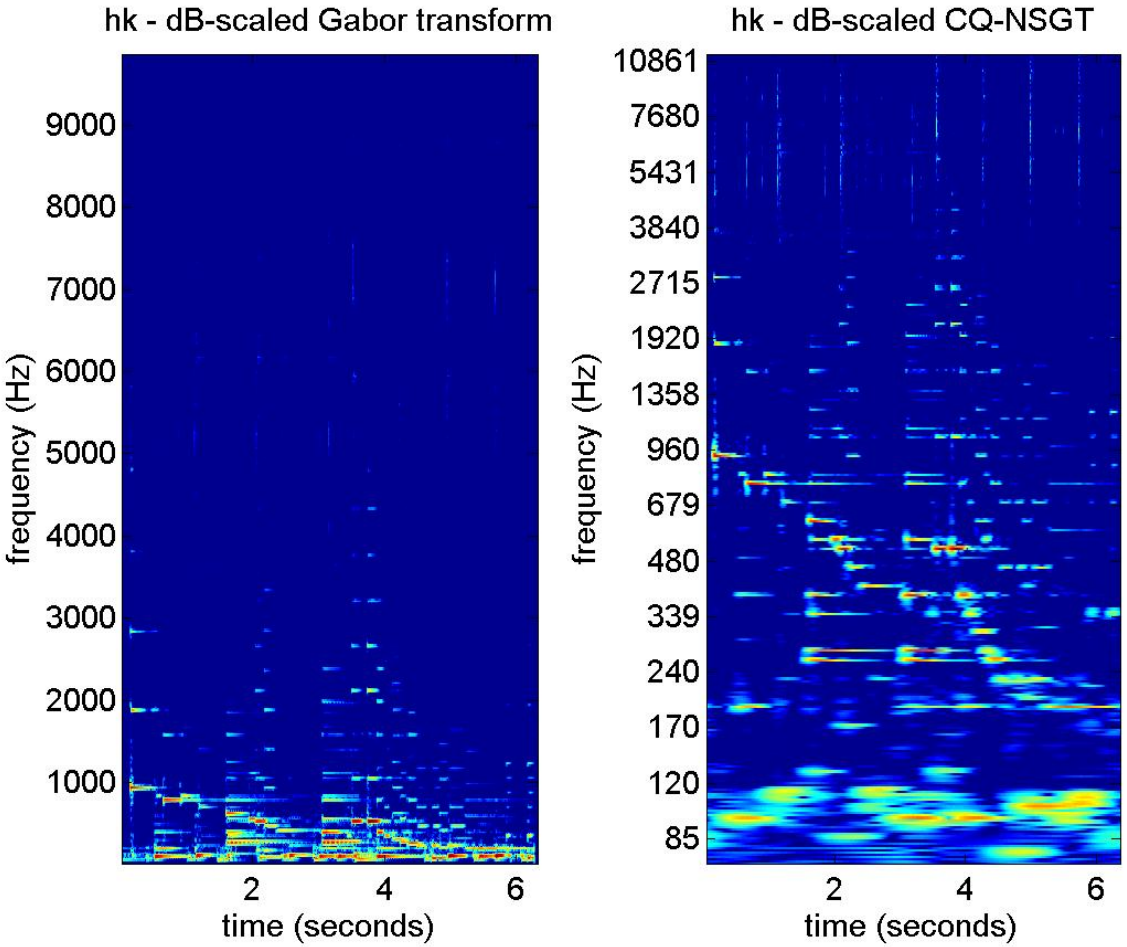
Adaptive signal representations

Exploring non-stationarity on the frequency side



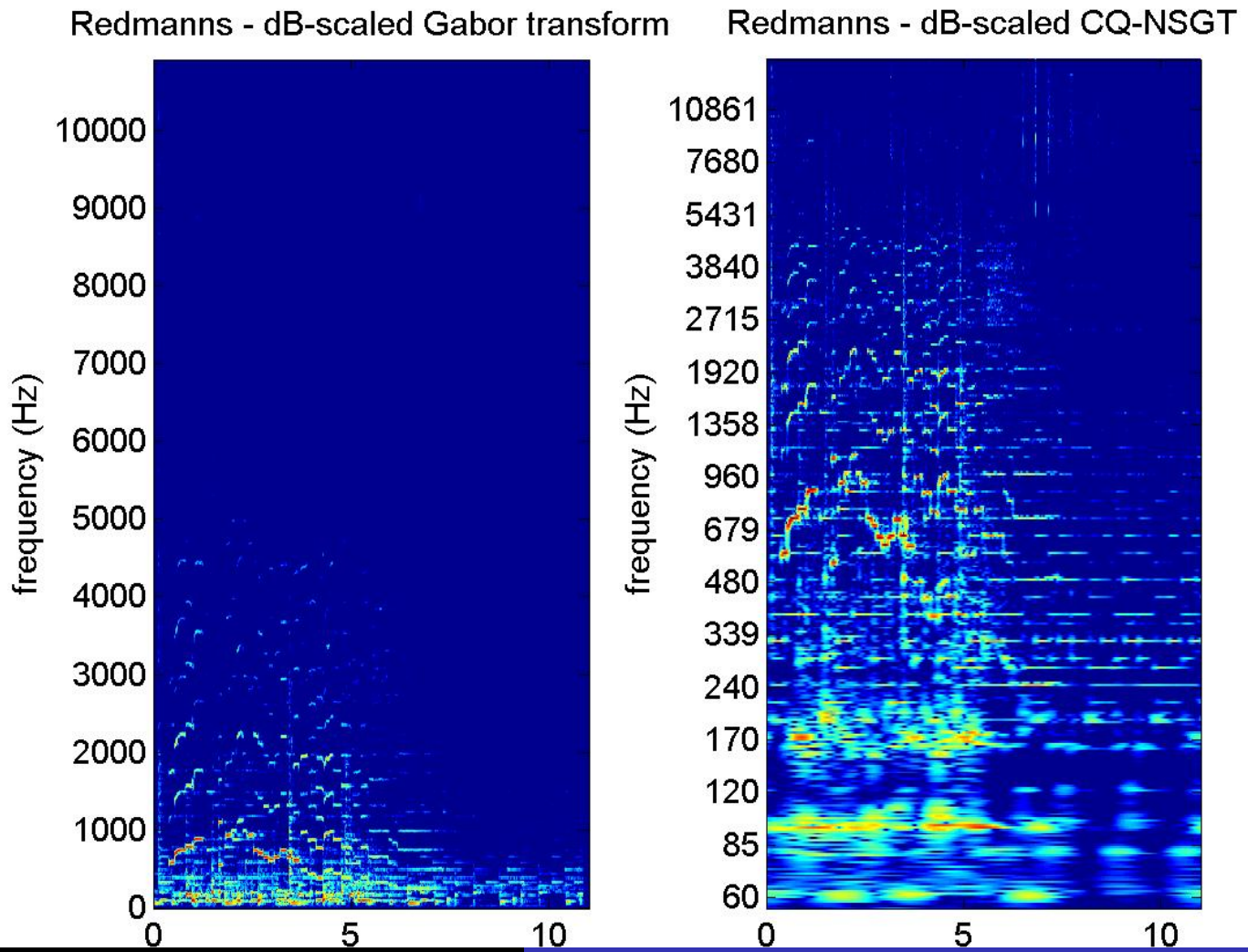
Adaptive signal representations

Exploring non-stationarity on the frequency side



Adaptive signal representations

Exploring non-stationarity on the frequency side



DISCUSSION

- The meaning of reconstruction.

DISCUSSION

- The meaning of reconstruction.
- The meaning of adaptivity: salience of sound objects/Gestalts.

Sparsity and structured representations

http://homepage.univie.ac.at/monika.doerfler/StrucAudio.html

- Regression problems with mixed-norm priors on time-frequency coefficients lead to structured, sparse representations of audio signals.

Sparsity and structured representations

http://homepage.univie.ac.at/monika.doerfler/StrucAudio.html

- Regression problems with mixed-norm priors on time-frequency coefficients lead to structured, sparse representations of audio signals.
- Systematic formulation of thresholding operators allows for weighting in the time-frequency domain

Sparsity and structured representations

http://homepage.univie.ac.at/monika.doerfler/StrucAudio.html

- Regression problems with mixed-norm priors on time-frequency coefficients lead to structured, sparse representations of audio signals.
- Systematic formulation of thresholding operators allows for weighting in the time-frequency domain
- Related iterative algorithms are evaluated on synthetic and real-life audio signals in the context of denoising and multi-layer decomposition.

Sparsity and structured representations

http://homepage.univie.ac.at/monika.doerfler/StrucAudio.html

- Regression problems with mixed-norm priors on time-frequency coefficients lead to structured, sparse representations of audio signals.
- Systematic formulation of thresholding operators allows for weighting in the time-frequency domain
- Related iterative algorithms are evaluated on synthetic and real-life audio signals in the context of denoising and multi-layer decomposition.
- Shape of the weighting masks influences results

Regression problems with mixed-norm priors

$$\Delta(f) = \| \sum_{n,k} c_{n,k} \varphi^t_{n,k} - \hat{f} \|_2^2 + \mu \| \mathbf{c} \|_{\ell^1}$$

Regression problems with mixed-norm priors

$$\Delta(f) = \| \sum_{n,k} c_{n,k} \varphi^t_{n,k} - \hat{f} \|_2^2 + \mu \|\mathbf{c}\|_{\ell^1}$$

...becomes...

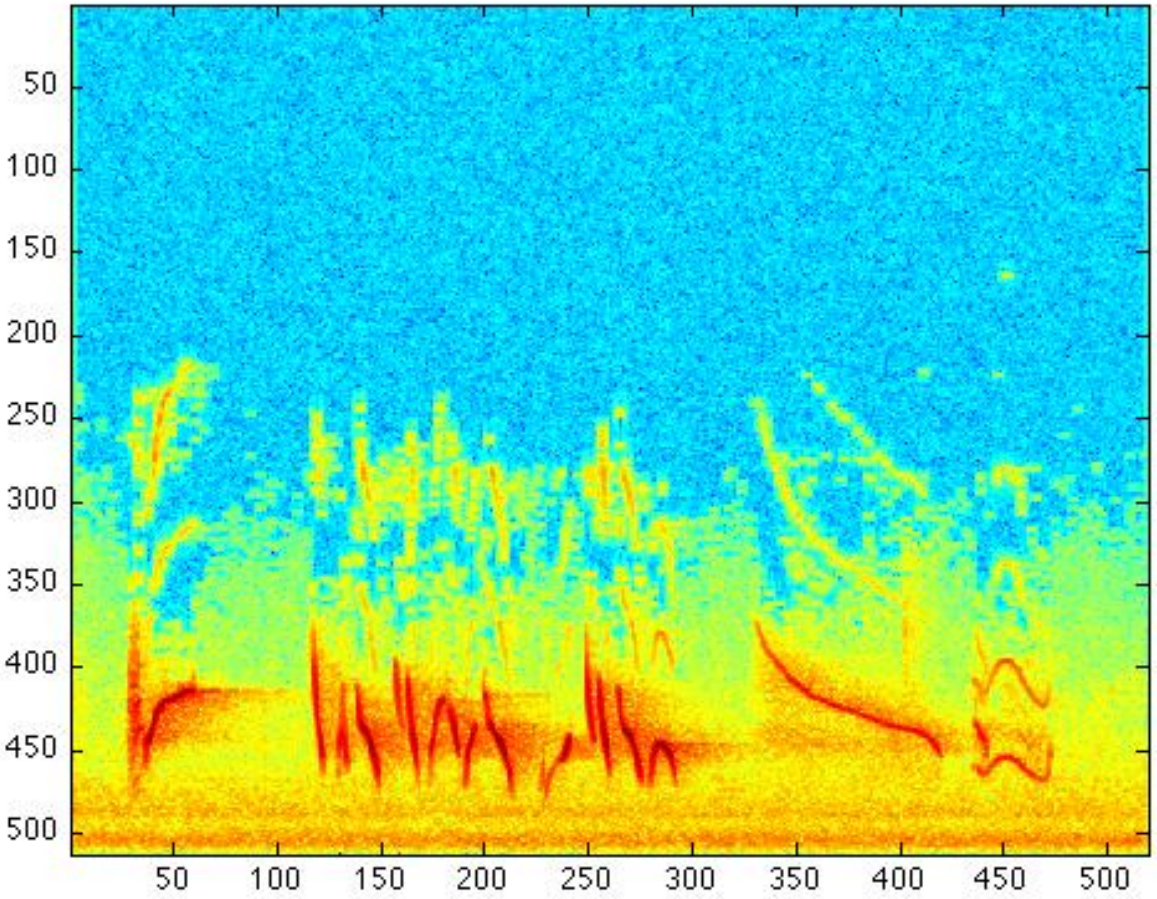
$$\Delta(f) = \| \sum_{n,k} c_{n,k} \varphi^t_{n,k} - \hat{f} \|_2^2 + \mu \|\mathbf{c}\|_{p,q}$$

algorithmic solution is iterative soft thresholding.

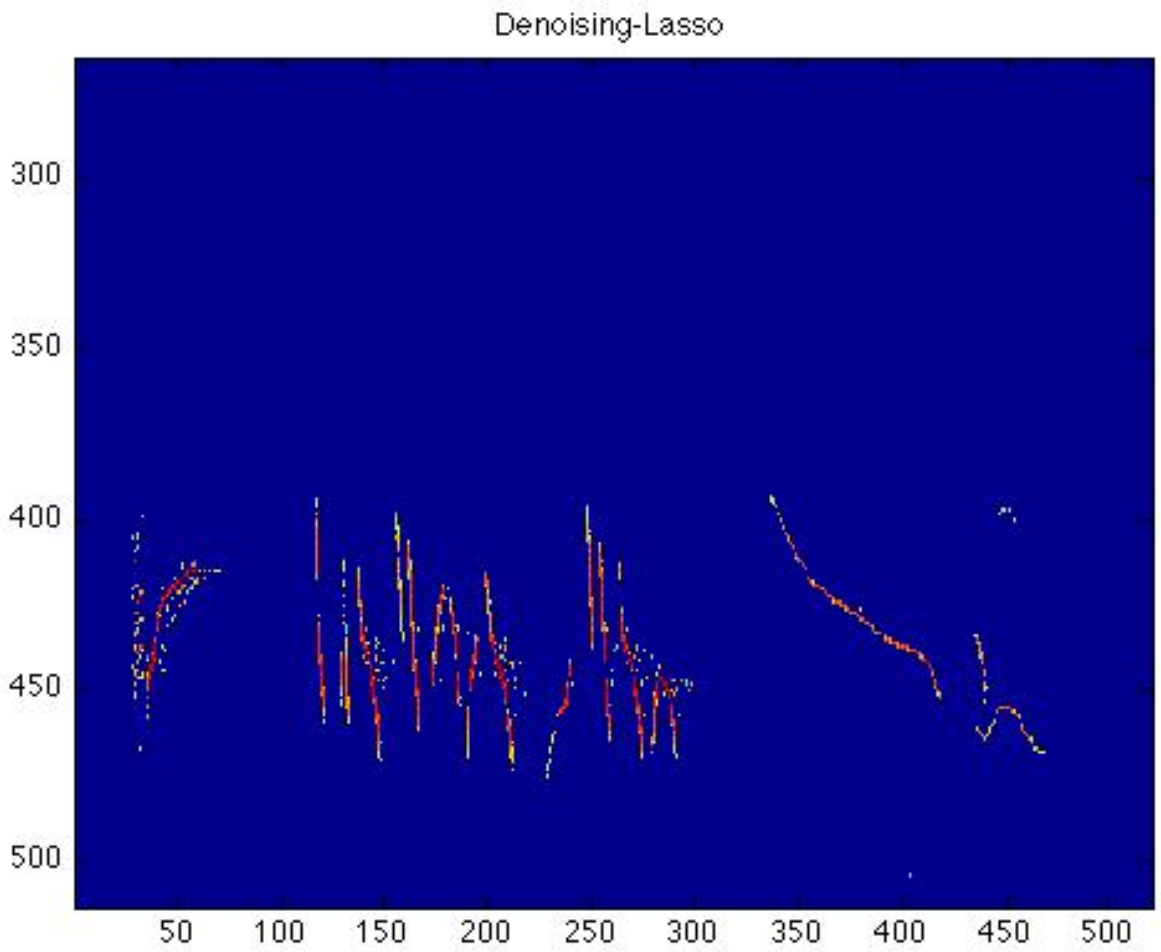
Thresholding operators with weighting in the time-frequency domain

$$\mathbb{S}_\omega(c_\gamma) = c_\gamma\left(1 - \frac{\mu}{\left(\sum_{\gamma' \in U_\gamma} \nu_\gamma(\gamma')|c_{\gamma'}|^2\right)^{1/2}}\right)^+$$

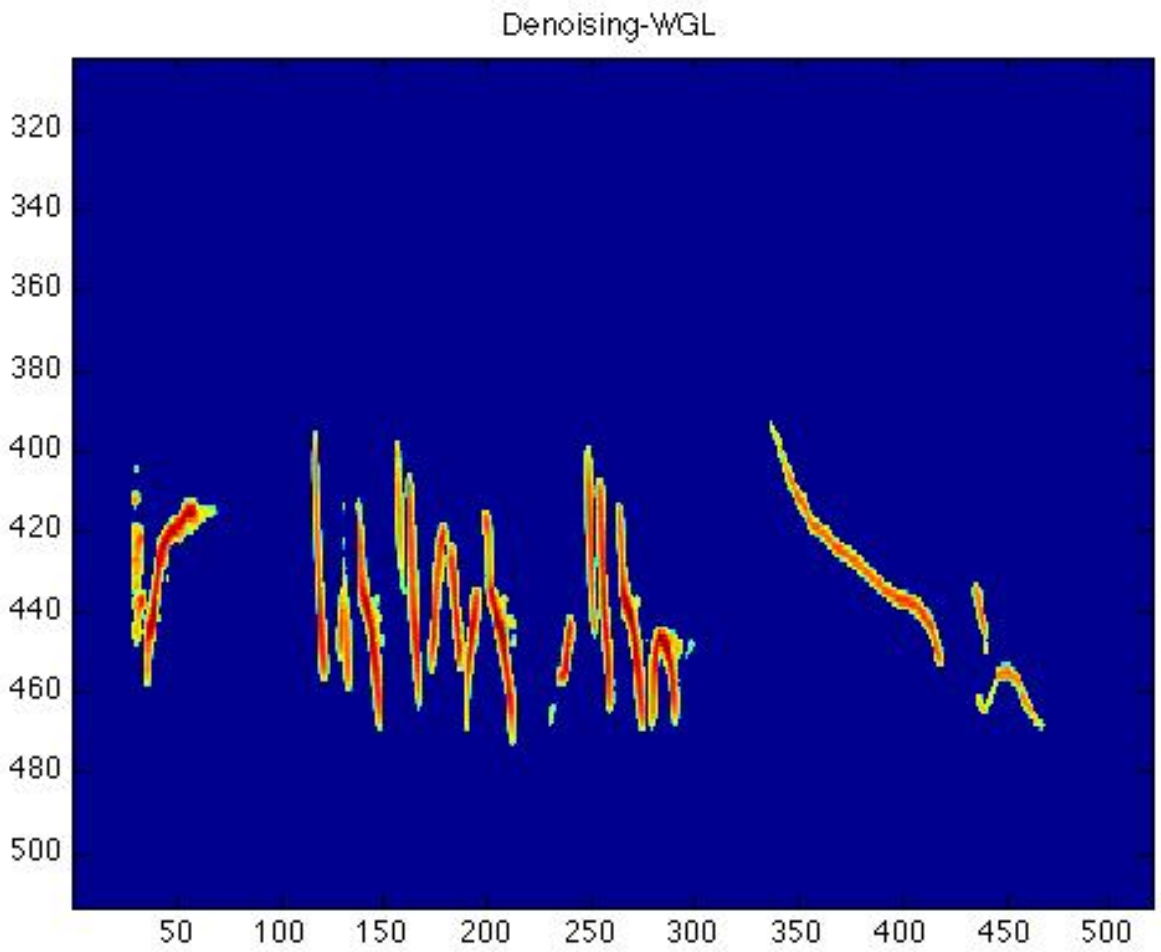
Application to denoising-Bird



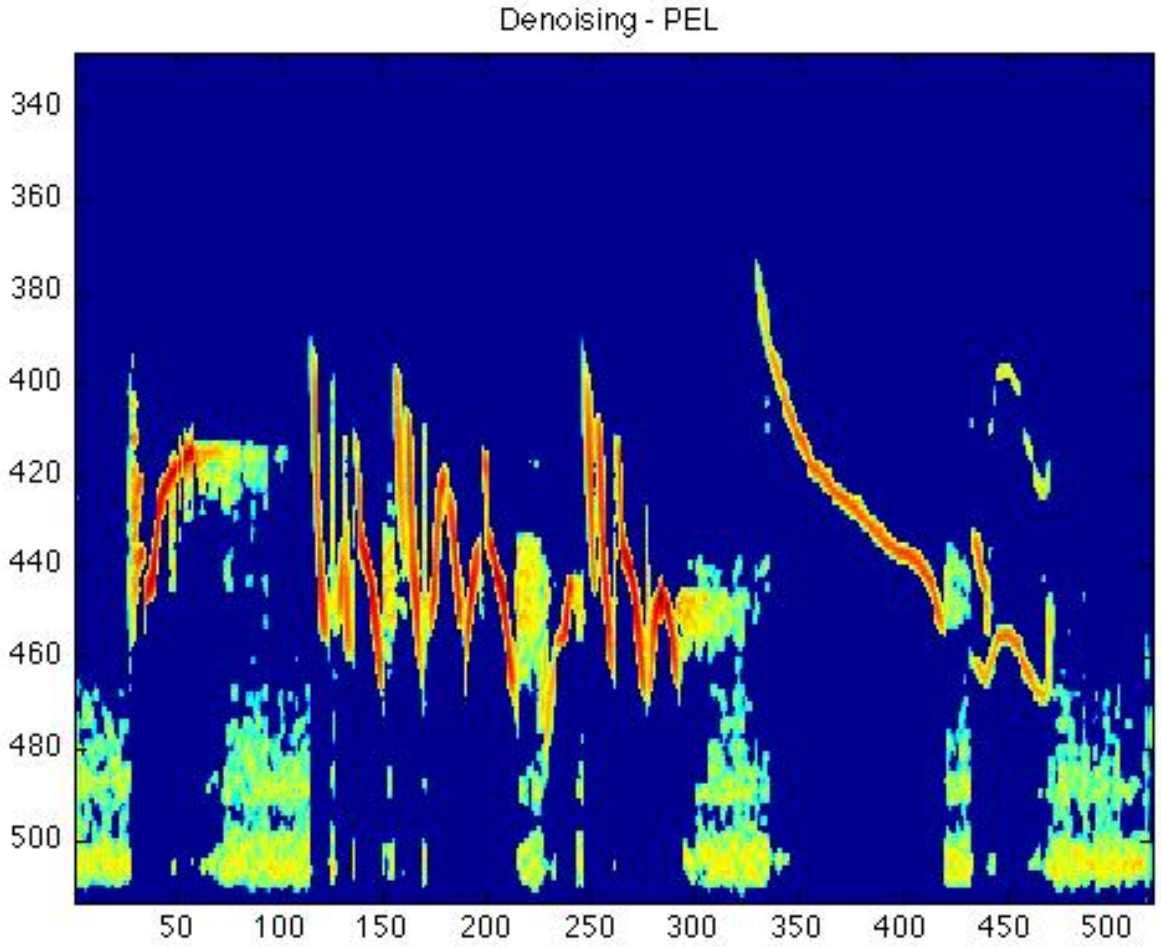
Application to denoising - Lasso, unstructured



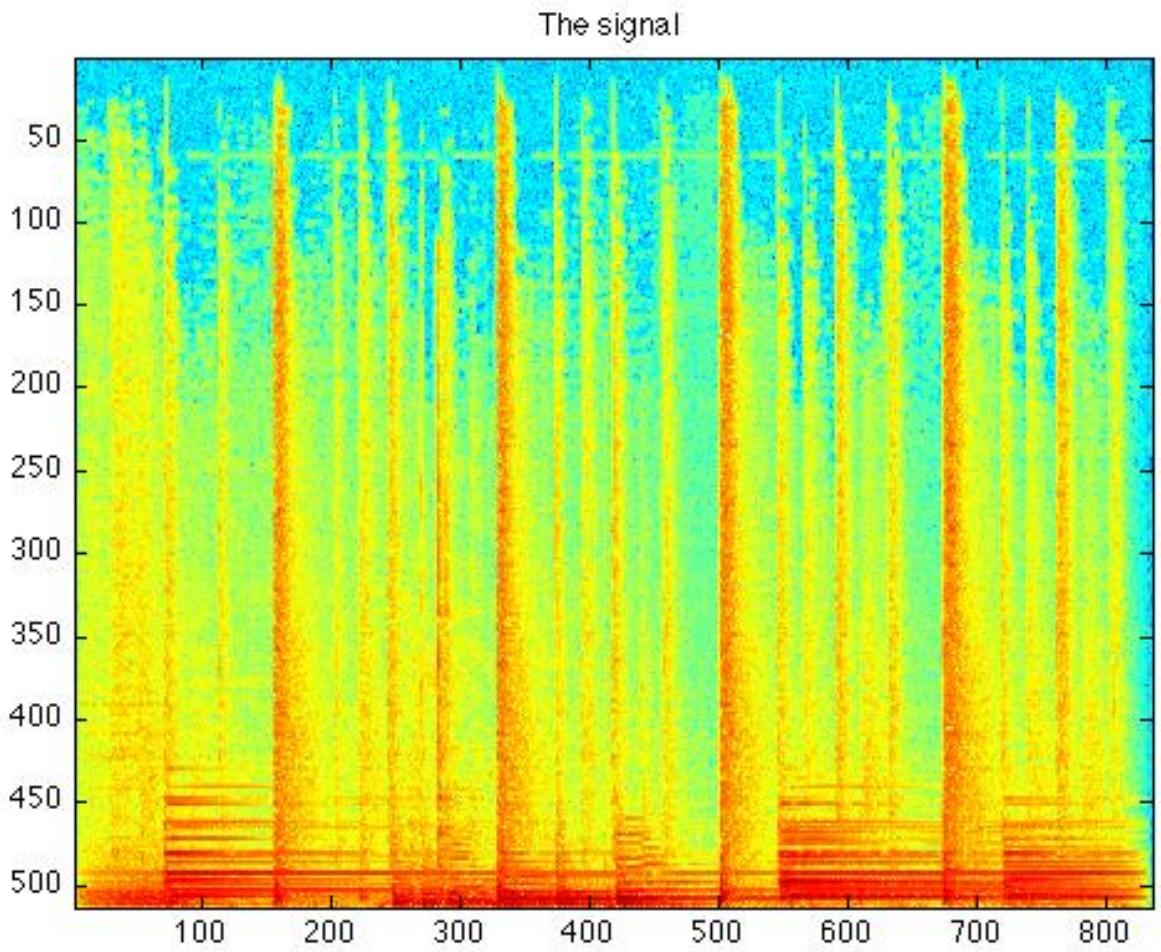
Application to denoising - Windowed Lasso



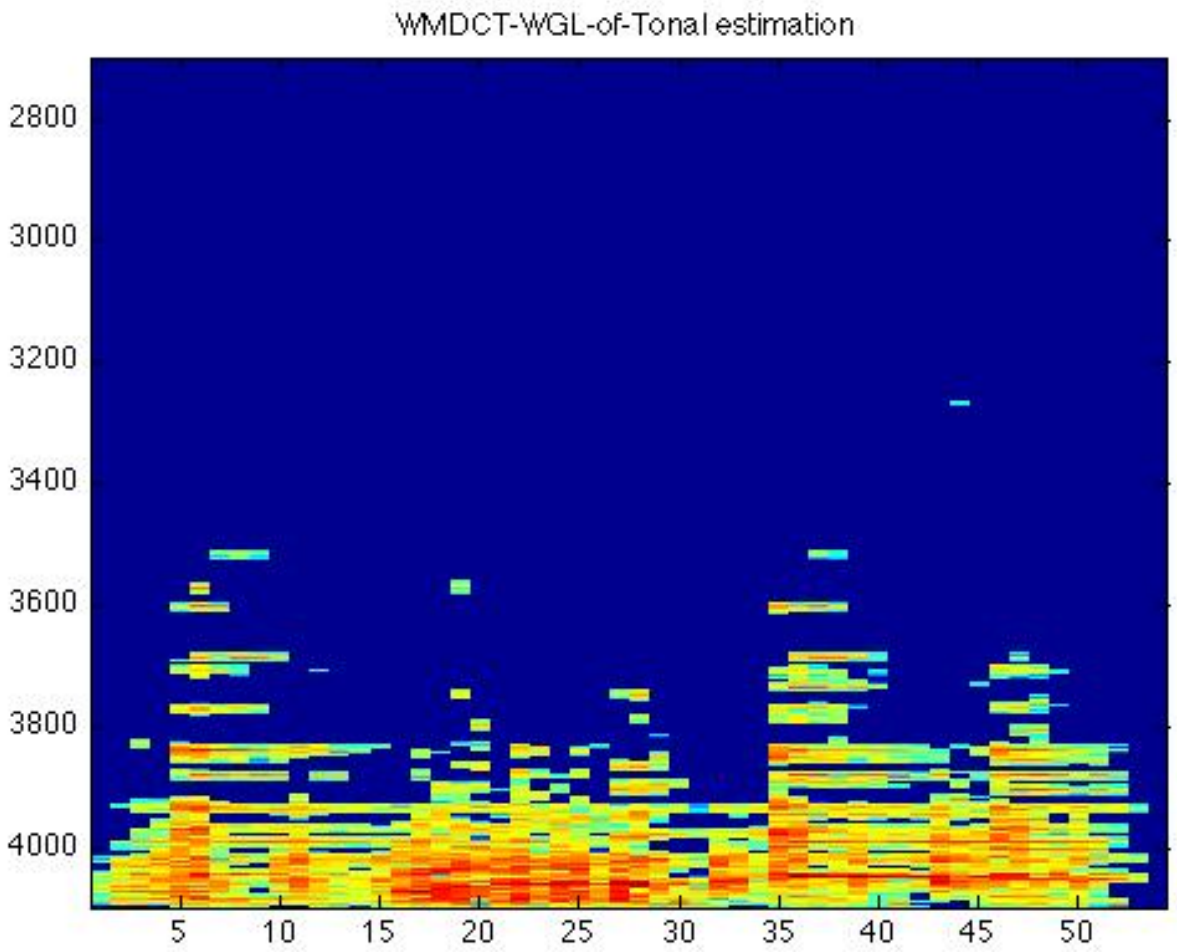
Application to denoising - Persistent Elitist Lasso



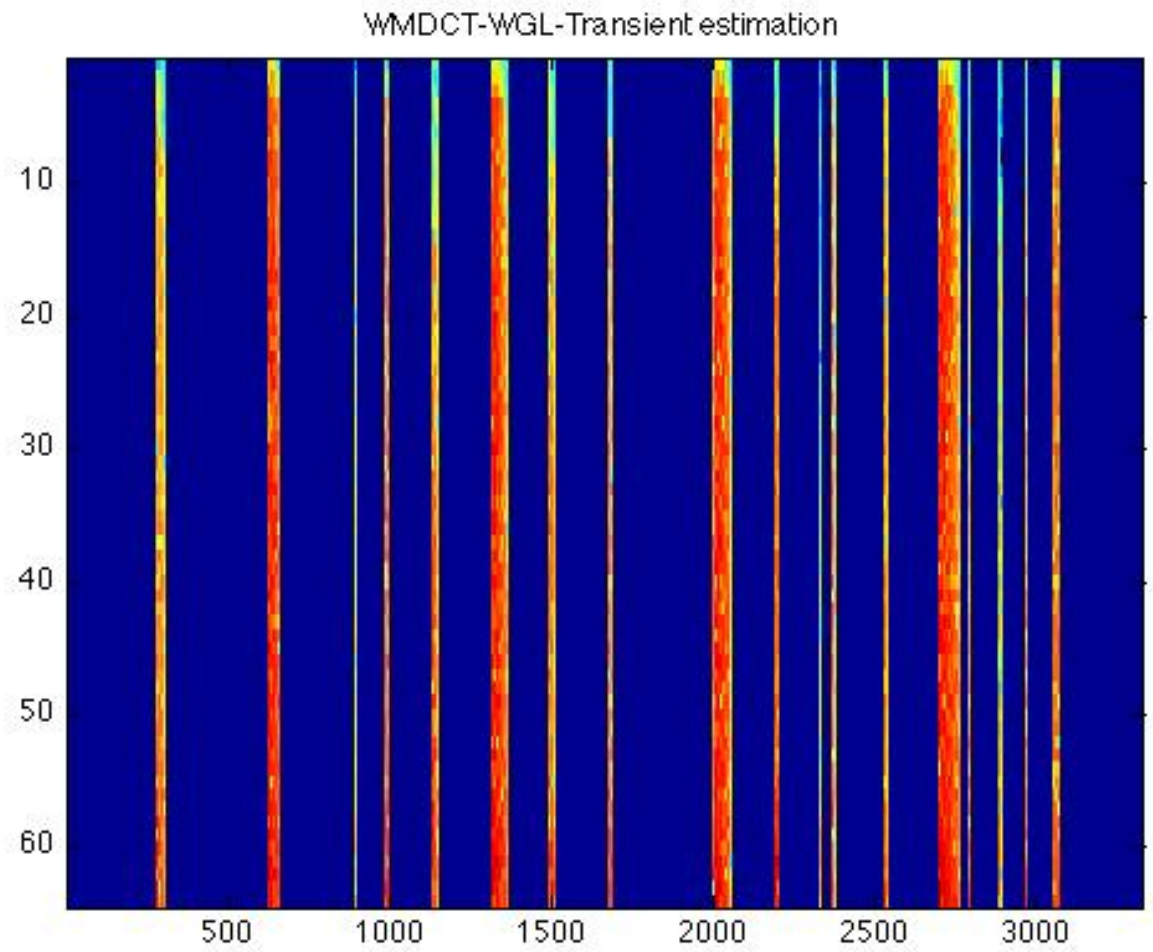
Application to multi-layer decomposition.



Application to multi-layer decomposition.

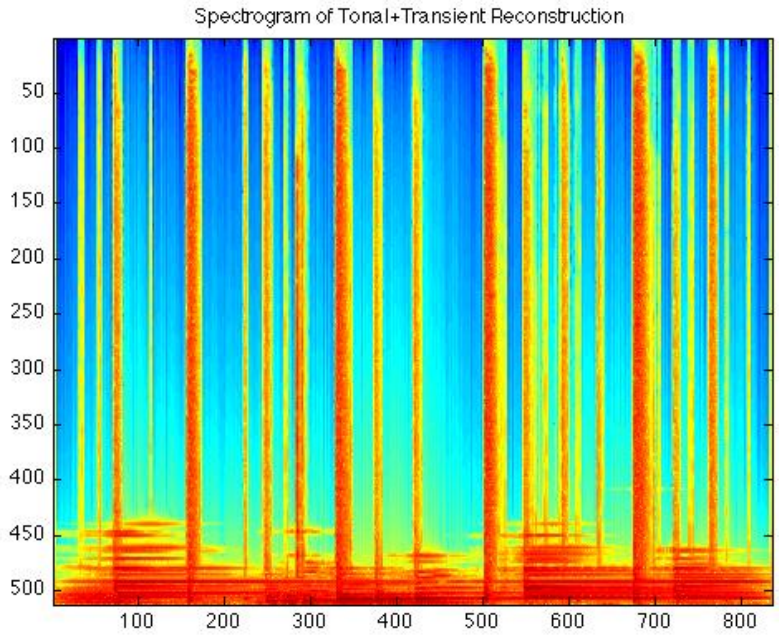


Application to multi-layer decomposition.



Application to multi-layer decomposition.

- Listen to residual!



(*Reconstruction Ton+Trans+Res*)

Influence of weighting mask shape

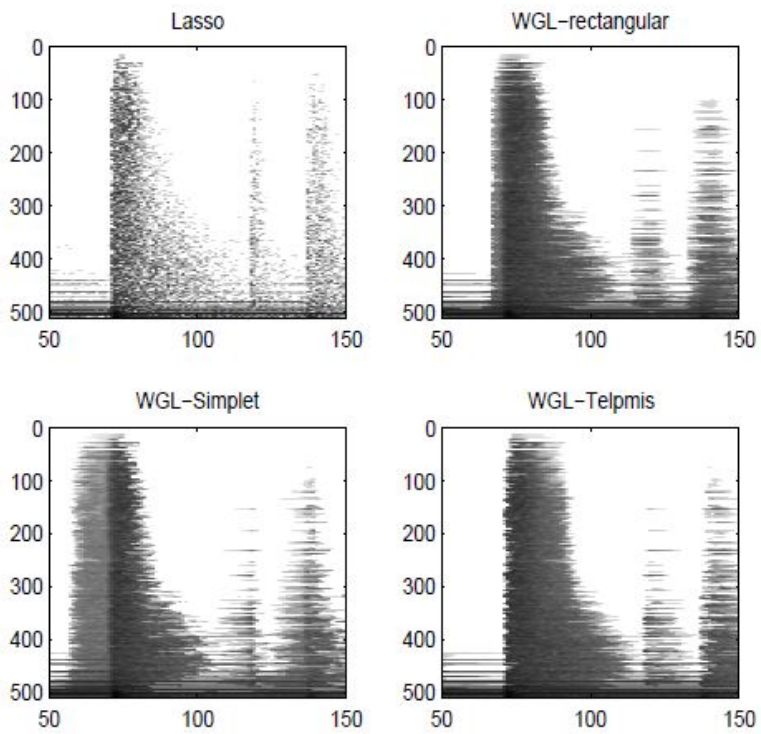


Figure: Iterated WGL shrinkage results for different shapes (i.e. weightings) of the neighborhood on an excerpt including piano, double-bass and drums and on an snare drum hit excerpt. The result of unsmoothed shrinkage, Rectangular, Simplet (= simple tent, starting at 1 and then linearly decaying to zero), Telpmis (= time-reversed simple tent)

Conclusion

- Perception of visual representation is highly dependent on analysis tools
- Also in higher level processing, this influence should be considered
- Prior knowledge can also be used in low level processing to extract meaningful information