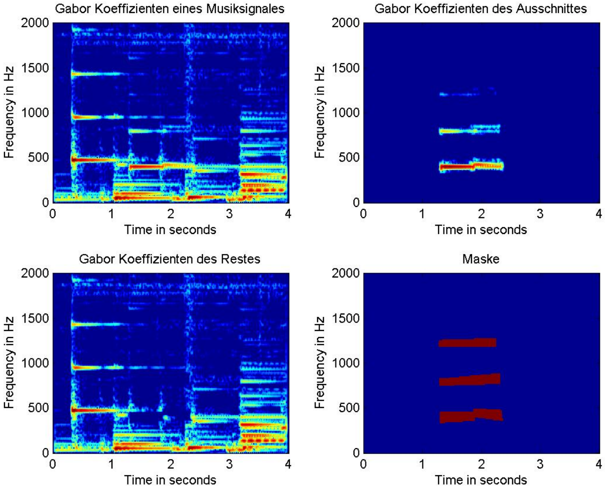
Nonstationary Gabor frames in audio signal processing

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

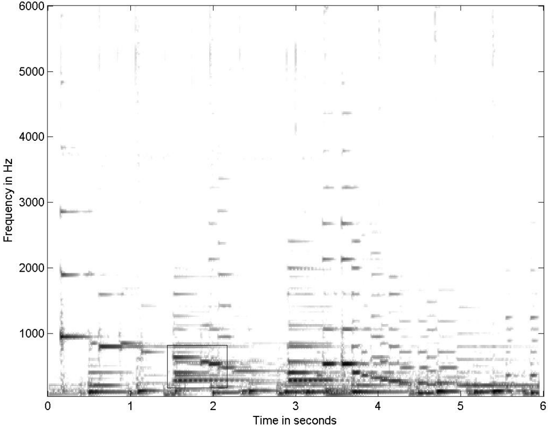
Brno, 28th Oct, 2011

[1] Funded by Locatif (FWF Austria) and Audio.Miner (WWTF Vienna)

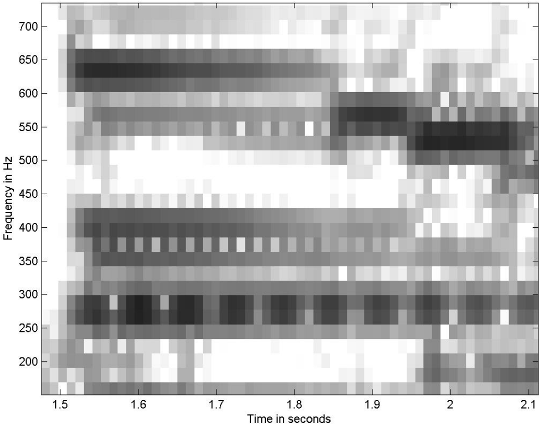
Combining mathematics and music



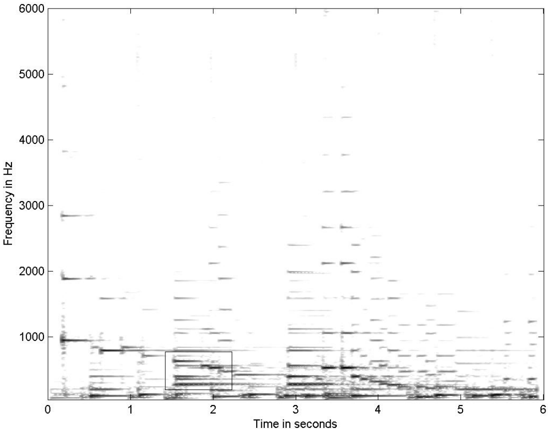
Combining mathematics and music



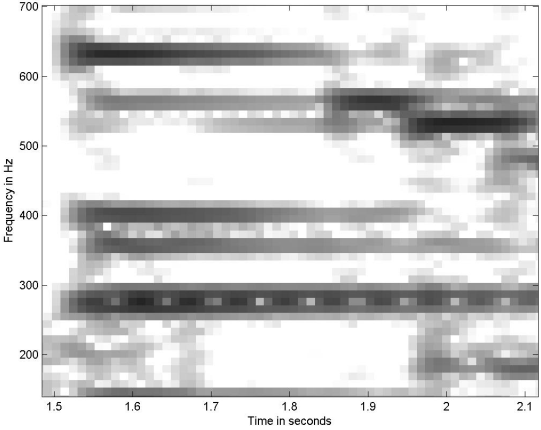
Combining mathematics and music



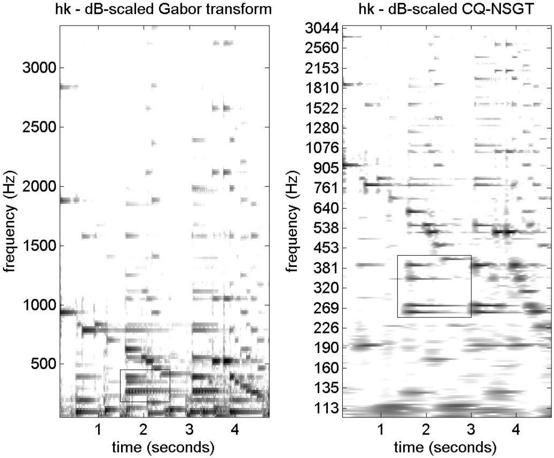
Combining mathematics and music



Combining mathematics and music



Combining mathematics and music



OUTLINE

- Gabor frames
- Frame operator, Walnut representation, painless case
- Nonstationary in Gabor frames
- Example applications
- Perspectives: Quilted frames

Gabor Frames

A sequence $(\psi_l)_{l\in I}$ in the Hilbert space $\mathcal{H}$ is called a *frame*, if there exist positive constants A and B:

$$A\|f\|^2 \leq \sum_{l\in I} |\langle f, \psi_l\rangle|^2 \leq B\|f\|^2 \quad \forall f \in \mathcal{H}, \tag{1}$$

Gabor Frames

A sequence $(\psi_l)_{l\in I}$ in the Hilbert space $\mathcal{H}$ is called a *frame*, if there exist positive constants A and B:

$$A\|f\|^2 \leq \sum_{l\in I} |\langle f, \psi_l\rangle|^2 \leq B\|f\|^2 \quad \forall f \in \mathcal{H}, \tag{1}$$

$\mathbf{C} : \mathcal{H} \to \ell^2$ is the *analysis operator* defined by $(\mathbf{C}f)_l = \langle f, \psi_l\rangle$

Gabor Frames

A sequence $(\psi_l)_{l \in I}$ in the Hilbert space $\mathcal{H}$ is called a *frame*, if there exist positive constants A and B:

$$A\|f\|^2 \leq \sum_{l \in I} |\langle f, \psi_l \rangle|^2 \leq B\|f\|^2 \quad \forall f \in \mathcal{H}, \tag{1}$$

$\mathbf{C} : \mathcal{H} \to \ell^2$ is the *analysis operator* defined by $(\mathbf{C}f)_l = \langle f, \psi_l \rangle$
The adjoint $\mathbf{C}$ is the *synthesis operator* $\mathbf{C}^*(c_l) = \sum_l c_l \psi_l$.

Gabor Frames

A sequence $(\psi_l)_{l \in I}$ in the Hilbert space $\mathcal{H}$ is called a *frame*, if there exist positive constants A and B:

$$A\|f\|^2 \leq \sum_{l \in I} |\langle f, \psi_l \rangle|^2 \leq B\|f\|^2 \quad \forall f \in \mathcal{H}, \tag{1}$$

$\mathbf{C} : \mathcal{H} \to \ell^2$ is the *analysis operator* defined by $(\mathbf{C}f)_l = \langle f, \psi_l \rangle$
The adjoint $\mathbf{C}$ is the *synthesis operator* $\mathbf{C}^*(c_l) = \sum_l c_l \psi_l$.
Frame operator is $\mathbf{S}f = \mathbf{C}^*\mathbf{C}f = \sum_l \langle f, \psi_l \rangle \psi_l$.

Invertibility of $\mathbf{S}$ leads to existence of *dual frames* $\rightarrow$ reconstruction.

Gabor Frames

Invertibility of $\mathbf{S}$ leads to existence of *dual frames* $\rightarrow$ reconstruction.
Canonical dual frame: $\tilde{\psi}_l = \mathbf{S}^{-1}\psi_l$ for all l.

Gabor Frames

Invertibility of $\mathbf{S}$ leads to existence of *dual frames* $\rightarrow$ reconstruction.
Canonical dual frame: $\tilde{\psi}_l = \mathbf{S}^{-1}\psi_l$ for all l.
Reconstruction:

$$f = \sum_l \langle f, \psi_l \rangle \tilde{\psi}_l = \sum_l \langle f, \tilde{\psi}_l \rangle \psi_l .$$

For tight frames, the frame operator reduces to $\mathbf{S} = A\mathbf{I}$, where $\mathbf{I}$ denotes the identity operator, and therefore $\mathbf{S}^{-1} = \frac{1}{A}\mathbf{I}$. The *canonical tight frame* ($\mathring{\psi}_l$) is obtained by applying $\mathbf{S}^{-\frac{1}{2}}$ to the frame elements, i.e. $\mathring{\psi}_l = \mathbf{S}^{-\frac{1}{2}}\psi_l$ for all l.

Gabor Frames

For $g \in L^2(\mathbb{R})$ (the *window*), the *short-time Fourier transform* (STFT) of $f \in L^2(\mathbb{R})$ is defined as

$$\mathcal{V}_g\left(f\right)(\tau,\omega) = \langle f, \mathbf{M}_\omega \mathbf{T}_\tau g\rangle$$

translation operator $\mathbf{T}_\tau f\left(t\right) = f\left(t-\tau\right)$
modulation operator $\mathbf{M}_\omega f\left(t\right) = f\left(t\right)e^{2\pi i\omega t}$. In other words:

$$\mathcal{V}_g\left(f\right)(\tau,\omega) = \int_{\mathbb{R}} f(t)\,\overline{g(t-\tau)}\,e^{-2\pi i\omega t}dt.$$

Gabor Frames

For a (non-zero) window function g and parameters $a, b > 0$, the set of time-frequency shifts of g

$$\mathcal{G}(g, a, b) = \{\mathbf{M}_{bm}\mathbf{T}_{an}g : m, n \in \mathbb{Z}\}$$

is called a *Gabor system*, if $\mathcal{G}(g, a, b)$ is a frame, it is called a *Gabor frame*.

Gabor Frames

For a (non-zero) window function g and parameters $a, b > 0$, the set of time-frequency shifts of g

$$\mathcal{G}(g, a, b) = \{\mathbf{M}_{bm}\mathbf{T}_{an}g : m, n \in \mathbb{Z}\}$$

is called a *Gabor system*, if $\mathcal{G}(g, a, b)$ is a frame, it is called a *Gabor frame*.
Gabor analysis coefficients are sampling points of the STFT of f with window g at the points (an, bm).

Gabor Frames

For a (non-zero) window function g and parameters $a, b > 0$, the set of time-frequency shifts of g

$$\mathcal{G}(g, a, b) = \{\mathbf{M}_{bm}\mathbf{T}_{an}g : m, n \in \mathbb{Z}\}$$

is called a *Gabor system*, if $\mathcal{G}(g, a, b)$ is a frame, it is called a *Gabor frame*.

Gabor analysis coefficients are sampling points of the STFT of f with window g at the points (an, bm).

(Canonical) dual frame of a Gabor frame is again a Gabor frame: generated by the *dual window* $\tilde{g} = \mathbf{S}^{-1}g$ and the same *lattice*, i.e. the set of time-frequency points $\{(an, bm) \,|\, m, n \in \mathbb{Z}\}$.

Rewriting the frame operator

If the analysis window g is (at least) in the Wiener space $W(\mathbb{R})$, then the frame operator

$$\mathbf{S}f = \sum_{m,n} \langle f, \mathbf{M}_{bm}\mathbf{T}_{an}g \rangle \mathbf{M}_{bm}\mathbf{T}_{an}g$$

can be written as

$$\mathbf{S}f(x) = \frac{1}{b}\sum_{k,n} \bar{g}(x - \frac{n}{b} - ak)g(x - ak)\mathbf{T}_{\frac{n}{b}}f(x).$$

Rewriting the frame operator

If the analysis window g is (at least) in the Wiener space $W(\mathbb{R})$, then the frame operator

$$\mathbf{S}f = \sum_{m,n} \langle f, \mathbf{M}_{bm}\mathbf{T}_{an}g \rangle \mathbf{M}_{bm}\mathbf{T}_{an}g$$

can be written as

$$\mathbf{S}f(x) = \frac{1}{b}\sum_{k,n} \bar{g}(x - \frac{n}{b} - ak)g(x - ak)\mathbf{T}_{\frac{n}{b}}f(x).$$

Immediately leads to a general existence result (given by Walnut in 1992).

Rewriting the frame operator

If the analysis window g is (at least) in the Wiener space $W(\mathbb{R})$, then the frame operator

$$\mathbf{S}f = \sum_{m,n} \langle f, \mathbf{M}_{bm}\mathbf{T}_{an}g \rangle \mathbf{M}_{bm}\mathbf{T}_{an}g$$

can be written as

$$\mathbf{S}f(x) = \frac{1}{b} \sum_{k,n} \bar{g}(x - \frac{n}{b} - ak) g(x - ak) \mathbf{T}_{\frac{n}{b}} f(x).$$

Immediately leads to a general existence result (given by Walnut in 1992).
If g is compactly supported, and b is small enough, then $\mathbf{S}$ is diagonal! (Daubechies et al., 1988)

Frames are better than bases, but . . .

Advantages of (Gabor) frames in Music Signal Processing:

- Good time-frequency resolution is possible
- Fast processing (FFT based)
- Redundancy offers sparse representations
- Reconstruction straight-forward in painless case

but . . . what about the changing structures of signals?

Frames are better than bases, but ...

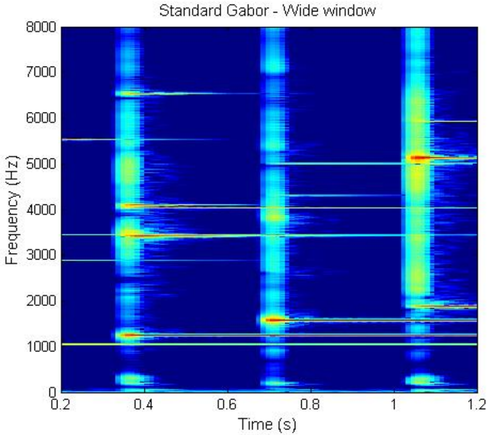


Figure: *STFT with wide window*

Frames are better than bases, but ...

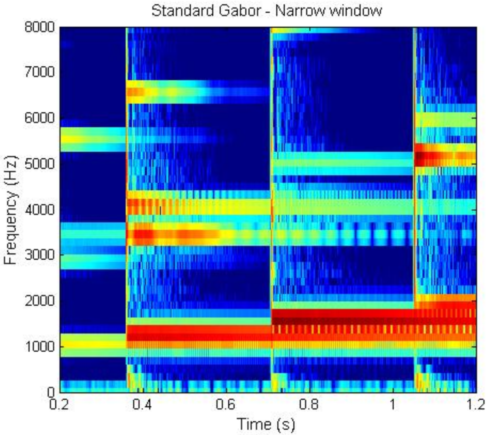


Figure: *STFT with narrow window*

The "painless" case is adaptable!

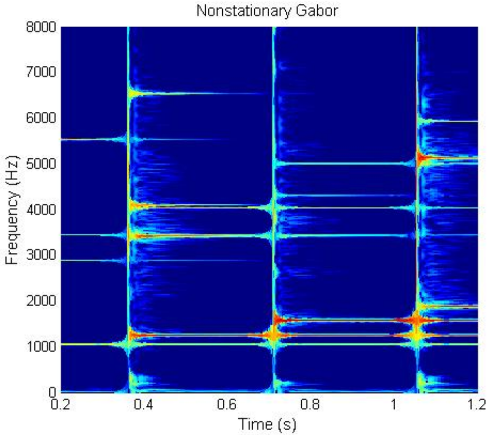


Figure: *STFT with adapted window*

The "painless" case is adaptable!

Basic technique: in analogy to the classical painless situation as suggested by Daubechies et al. (1988).

The "painless" case is adaptable!

Basic technique: in analogy to the classical painless situation as suggested by Daubechies et al. (1988).
Essential ingredients:

1. Signal is localized at time- (or frequency-)positions n by multiplication with a compactly supported (or limited bandwidth, respectively) window functions g_n.
2. The Fourier transform of the localized pieces is sampled densely enough.
3. Adjacent windows overlap to avoid loss of information.
4. http://www.univie.ac.at/nonstatgab/

The "painless" case is adaptable!

For every $n \in \mathbb{Z}$, let the function $g_n \in L^2(\mathbb{R})$ be compactly supported with $\operatorname{supp}(g_n) \subseteq [c_n, d_n]$ and let b_n be chosen such that $d_n - c_n \leq \frac{1}{b_n}$. Then the frame operator

$$\mathbf{S} : f \mapsto \sum_{m,n} \langle f, g_{m,n} \rangle g_{m,n}$$

of the system

$$g_{m,n}(t) = g_n(t)\, e^{2\pi i m b_n t}, \; m, n \in \mathbb{Z},$$

is given by a multiplication operator of the form

$$\mathbf{S} f(t) = \left(\sum_n \frac{1}{b_n} |g_n(t)|^2 \right) f(t).$$

The "painless" case is adaptable!

Analog construction in the frequency domain leads to irregular sampling over frequency and windows with adaptive bandwidth.

The "painless" case is adaptable!

Analog construction in the frequency domain leads to irregular sampling over frequency and windows with adaptive bandwidth. For a family of functions $\{h_m\}_{m\in\mathbb{Z}}$ define atoms of the form:

$$h_{m,n}(t) = h_m(t - na_m). \tag{2}$$

Therefore $\widehat{h_{m,n}}(\nu) = \widehat{h_m}(\nu) \cdot e^{-2\pi i n a_m \nu}$ and the analysis coefficients may be written as

$$c_{m,n} = \langle f, h_{m,n}\rangle = \langle \hat{f}, \mathcal{F}(\mathbf{T}_{na_m} h_m)\rangle = \mathcal{F}^{-1}(\hat{f} \cdot \overline{\widehat{h_m}})(na_m).$$

The "painless" case is adaptable!

Analog construction in the frequency domain leads to irregular sampling over frequency and windows with adaptive bandwidth. For a family of functions $\{h_m\}_{m\in\mathbb{Z}}$ define atoms of the form:

$$h_{m,n}(t) = h_m(t - na_m). \qquad (2)$$

Therefore $\widehat{h_{m,n}}(\nu) = \widehat{h_m}(\nu) \cdot e^{-2\pi i n a_m \nu}$ and the analysis coefficients may be written as

$$c_{m,n} = \langle f, h_{m,n}\rangle = \langle \hat{f}, \mathcal{F}(\mathbf{T}_{na_m} h_m)\rangle = \mathcal{F}^{-1}\left(\hat{f} \cdot \overline{\widehat{h_m}}\right)(na_m).$$

$\rightarrow$ situation analog to before, up to a Fourier transform.

The "painless" case is adaptable!

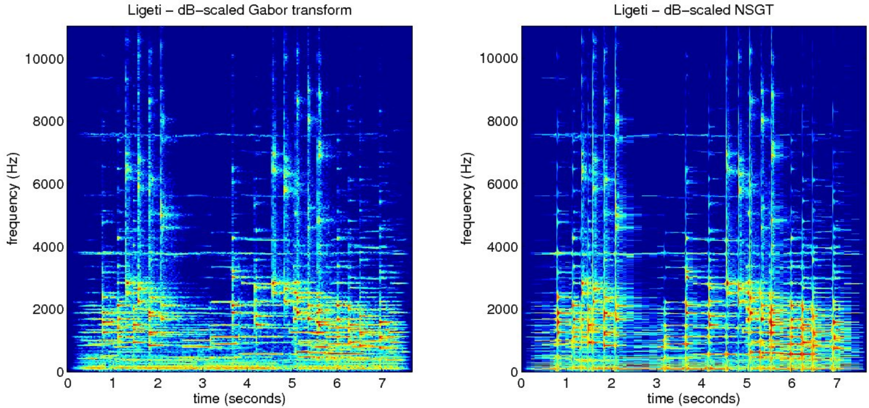
In general, the inversion of **S** can be numerically unfeasible, in the special "painless" case, the invertibility of the frame operator is easy to check and inversion is a simple multiplication:
$g_{m,n}$ forms a frame for $L^2(\mathbb{R})$ if and only if $\sum_n \frac{1}{b_n}|g_n(t)|^2 \simeq 1$.

The "painless" case is adaptable!

In general, the inversion of **S** can be numerically unfeasible, in the special "painless" case, the invertibility of the frame operator is easy to check and inversion is a simple multiplication:
$g_{m,n}$ forms a frame for $L^2(\mathbb{R})$ if and only if $\sum_n \frac{1}{b_n}|g_n(t)|^2 \simeq 1$.
Canonical dual frame elements are given by:

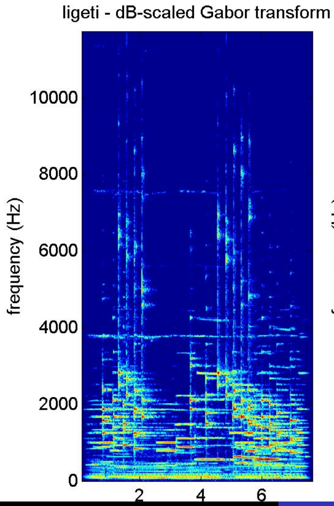
$$\tilde{g}_{m,n}(t) = \frac{g_n(t)}{\sum_l \frac{1}{b_l}|g_l(t)|^2} e^{2\pi i m b_n t}, \tag{3}$$

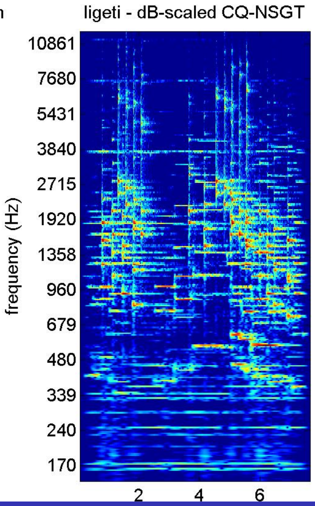
The "painless" case is adaptable!



The "painless" case is adaptable!

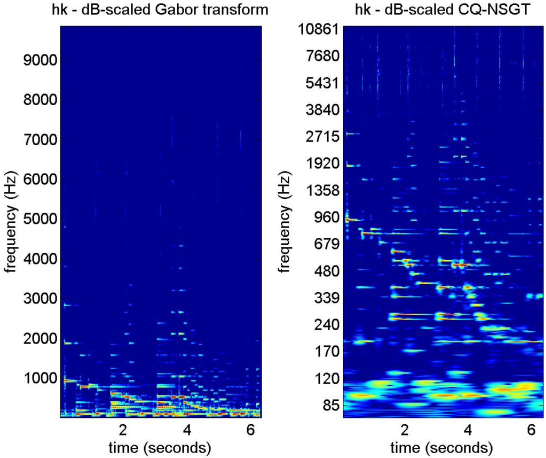
Exploring non-stationarity on the frequency side





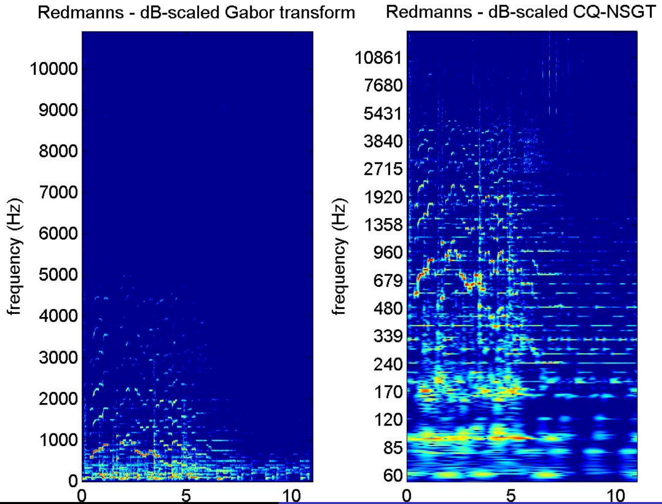
The "painless" case is adaptable!

Exploring non-stationarity on the frequency side



The "painless" case is adaptable!

Exploring non-stationarity on the frequency side



Transposition

Combining mathematics and music - the starting point

Time-frequency analysis

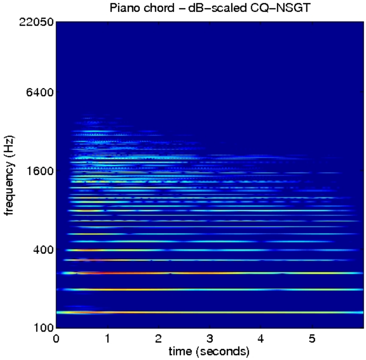
Gabor frames

The Walnut representation and the painless case

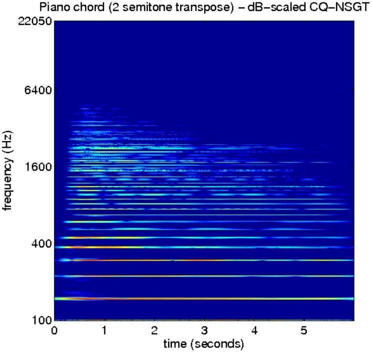
Nonstationarity in Gabor frames

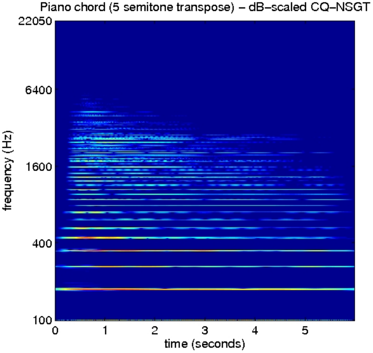
Example applications

Quilted frames and local adaptation

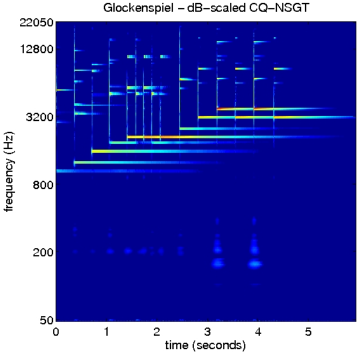


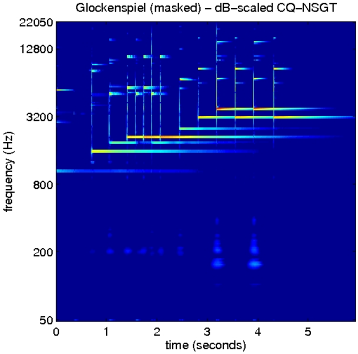
Transposition



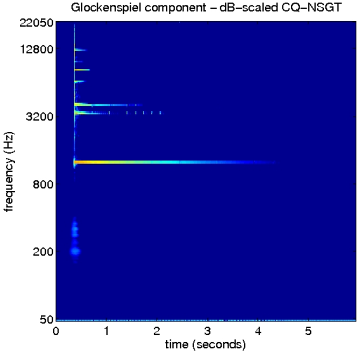


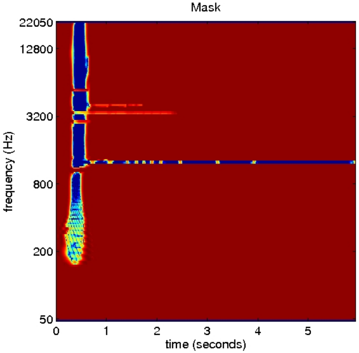
Masking


Glockenspiel – dB–scaled CQ–NSGT
frequency (Hz)
22050
12800
3200
800
200
50
0 1 2 3 4 5
time (seconds)


Glockenspiel (masked) – dB–scaled CQ–NSGT
frequency (Hz)
22050
12800
3200
800
200
50
0 1 2 3 4 5
time (seconds)

Masking


Glockenspiel component – dB–scaled CQ–NSGT
22050
12800
3200
800
200
50
frequency (Hz)
0
1
2
3
4
5
time (seconds)


Mask
22050
12800
3200
800
200
50
frequency (Hz)
0
1
2
3
4
5
time (seconds)

The "painless" case is adaptable!

References:

- P. Balazs, M. Dörfler, N. Holighaus, F. Jaillet and G. Velasco, *Theory, Implementation and Application of Nonstationary Gabor Frames* J. Comput. Appl. Math., In Press. (2011)
- Velasco, G., Holighaus, N.,Dörfler, M.,Grill, T., *Constructing an invertible constant-Q transform with non-stationary Gabor frames* Proceedings of DAFX11, Paris (2011)
  Note: constant-Q & sound patterns
- A. Holzapfel, G. A. Velasco, N. Holighaus, M. Dörfler, A. Flexer, *Advantages of nonstationary Gabor transforms in beat* tracking, MIRUM11, Arizona, USA (2011).
- `http://www.univie.ac.at/nonstatgab/`

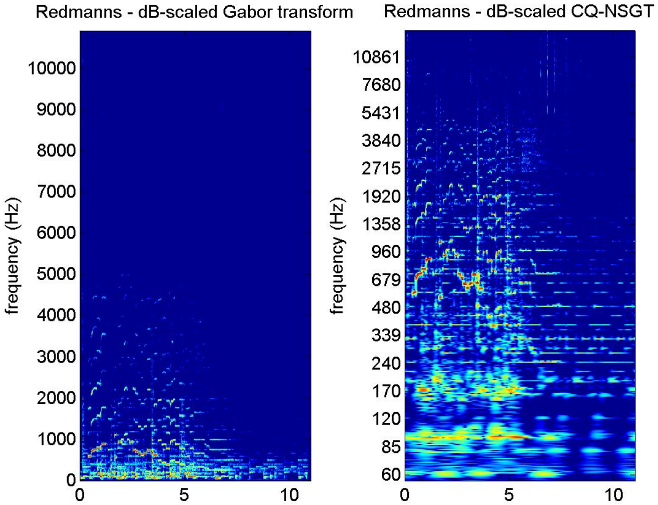
The "painless" case is adaptable!

References:

- P. Balazs, M. Dörfler, N. Holighaus, F. Jaillet and G. Velasco, *Theory, Implementation and Application of Nonstationary Gabor Frames* J. Comput. Appl. Math., In Press. (2011)
- Velasco, G., Holighaus, N.,Dörfler, M.,Grill, T., *Constructing an invertible constant-Q transform with non-stationary Gabor frames* Proceedings of DAFX11, Paris (2011)
  Note: constant-Q & sound patterns
- A. Holzapfel, G. A. Velasco, N. Holighaus, M. Dörfler, A. Flexer, *Advantages of nonstationary Gabor transforms in beat* tracking, MIRUM11, Arizona, USA (2011).
- http://www.univie.ac.at/nonstatgab/

So , are we happy yet?

The "painless" case is adaptable!



The "painless" case is adaptable!

$\rightarrow$ adaptivity in BOTH domains would be nice.
Problems: missing compactness, interpretation,..
Encouraging result: characterization by TF-shifts of operators, but NO strict locality!
Discrete version, new concept: Quilted frames

The idea and structure of quilted frames

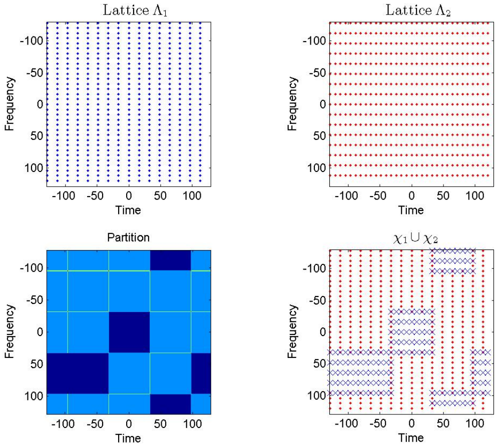


Figure: *Partition in time-frequency and resulting quilted lattice*

The idea and structure of quilted frames

Perspecitves:

- Implementation of quilted frames via ”sliced nonstationary Gabor frames”
- Adaptation via (structured) sparsity constraints
- Dictionary learning methods

The idea and structure of quilted frames

Perspecitves:

- Implementation of quilted frames via "sliced nonstationary Gabor frames"
- Adaptation via (structured) sparsity constraints
- Dictionary learning methods

Thanks for listening!