Nonstationary Gabor frames

Sevag Hanssian

MUMT 622, Winter 2021

February 25, 2021

2021-02-25

Nonstationary Gabor frames

Nonstationary Gabor frames

Sevag Hanssian

MUMT 622, Winter 2021

February 25, 2021

Nonstationary Gabor frames

2002 — Multiple Gabor systems for analysis of music[1]

2009 — Nonstationary Gabor frames[2]

2011 — **Theory, implementation and applications of non-stationary Gabor frames[3]**

[1] Monika Doerfler. "Gabor analysis for a class of signals called music". PhD thesis. Jan. 2002. URL: http://www.mathe.tu-freiberg.de/files/thesis/gamu_1.pdf

[2] Florent Jaillet, P. Balázs, and M. Dörfler. "Nonstationary Gabor Frames". In: *SAMPTA'09, International Conference on SAMPling Theory and Applications*. 2009. URL: https://github.com/ltfat/ltfat.github.io/blob/master/notes/ltfatnote010.pdf

[3] Peter Balazs et al. "Theory, implementation and applications of nonstationary Gabor frames". In: *Journal of computational and applied mathematics* 236 (Oct. 2011), pp. 1481–1496. DOI: 10.1016/j.cam.2011.09.011. URL: https://ltfat.github.io/notes/ltfatnote018.pdf

Nonstationary Gabor frames
2002 Multiple Gabor systems for analysis of music[1]
2009 Nonstationary Gabor frames[2]
2011 Theory, implementation and applications of non-stationary Gabor frames[3]

- Monika Dörfler's PhD dissertation on gabor, varying TF resolution, music

Nonstationary Gabor frames

The definition of multiple Gabor frames, which is comprehensively treated in [Dörfler 2002], provides Gabor frames with analysis techniques with multiple resolutions.

The nonstationary Gabor frames (see [Jaillet 2009], [Balazs 2011] for their definition and implementation) are a further development; they fully exploit theoretical properties [...] they provide for a class of FFT-based algorithms [...] together with perfect reconstruction formulas[4]

[4] Marco Liuni et al. "Automatic Adaptation of the Time-Frequency Resolution for Sound Analysis and Re-Synthesis". In: *IEEE Transactions on Audio Speech and Language Processing* 21 (May 2013). DOI: 10.1109/TASL.2013.2239989. URL: https://arpi.unipi.it/retrieve/handle/11568/159584/458549/tasl.pdf.

Nonstationary Gabor frames

The definition of multiple Gabor frames, which is comprehensively treated in [Dörfler 2002], provides Gabor frames with analysis techniques with multiple resolutions.

The nonstationary Gabor frames (see [Jaillet 2009], [Balazs 2011] for their definition and implementation) are a further development; they fully exploit theoretical properties [...] they provide for a class of FFT-based algorithms [...] together with perfect reconstruction formulas[4]

[4] Marco Liuni et al. "Automatic Adaptation of the Time-Frequency Resolution for Sound Analysis and Re-Synthesis". In: IEEE Transactions on Audio Speech and Language Processing 21 (May 2013). DOI: 10.1109/TASL.2013.2239989. URL: https://arpi.unipi.it/retrieve/handle/11568/159584/458549/tasl.pdf.

Invertible Constant-Q Transform

2011	Invertible constant-Q transform with nonstationary Gabor frames[5]
2012	Invertible, realtime CQT[6]
2013	Non-stationary Gabor systems, CQT added to LTFAT[7]
2018	CQT/ICQT added to MATLAB Wavelet Toolbox[8]

[5] Gino Angelo Velasco et al. *Constructing an invertible constant-Q transform with nonstationary Gabor frames*. Sept. 2011. URL: https://www.univie.ac.at/nonstatgab/pdf_files/dohogrve11_amsart.pdf

[6] Nicki Holighaus et al. "A Framework for Invertible, Real-Time Constant-Q Transforms". In: *IEEE Transactions on Audio, Speech and Language Processing* 21 (Sept. 2012). DOI: 10.1109/TASL.2012.2234114. URL: https://arxiv.org/pdf/1210.0084.pdf

[7] *ltfat/ChangeLog*. URL: https://github.com/ltfat/ltfat/blob/master/ChangeLog#L291-L301

[8] *Constant-Q nonstationary Gabor transform - MATLAB cqt*. 2018. URL: https://www.mathworks.com/help/wavelet/ref/cqt.html

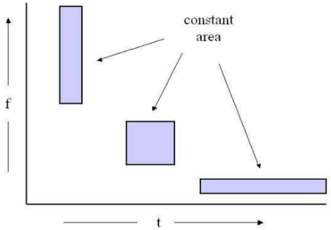
- frame theory led to a serious solution for a practical, invertible CQT

Review: Gabor frames

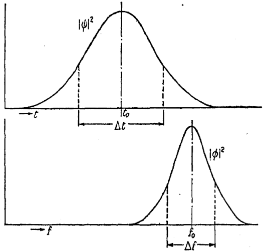
Gabor's 1946 "Theory of Communication"[9]:

1. First introduction of the time-frequency uncertainty principle
2. Proposed that any signal can be chopped up and windowed with Gaussian functions to minimize TF uncertainty

$$\sigma_t \sigma_f \geq \frac{1}{4\pi}, \qquad \Delta t \Delta f \geq 1$$

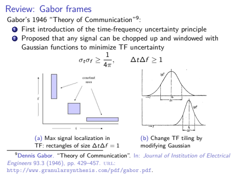


(a) Max signal localization in TF: rectangles of size $\Delta t \Delta f = 1$



(b) Change TF tiling by modifying Gaussian

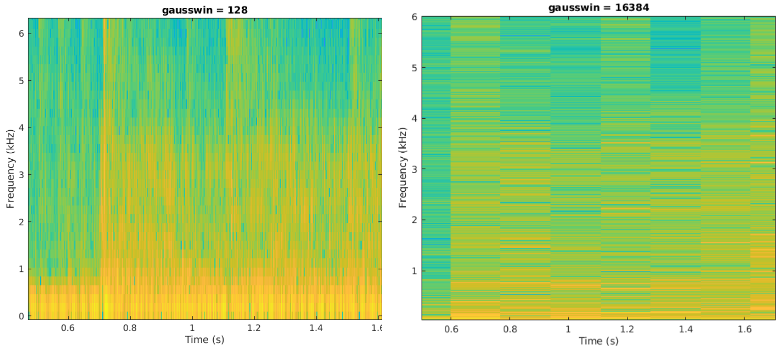
[9] Dennis Gabor. "Theory of Communication". In: *Journal of Institution of Electrical Engineers* 93.3 (1946), pp. 429–457. URL: http://www.granularsynthesis.com/pdf/gabor.pdf.



- this material builds on the material from my last presentation
- start with the refresher
- recall that we are constrained to rectangles of this area

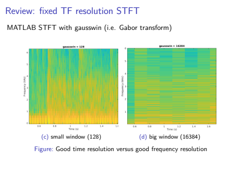
Review: fixed TF resolution STFT

MATLAB STFT with gausswin (i.e. Gabor transform)



(c) small window (128) (d) big window (16384)

Figure: Good time resolution versus good frequency resolution



Fixed TF resolution for music

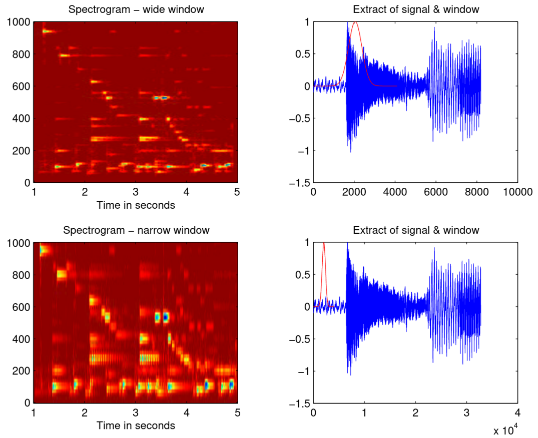
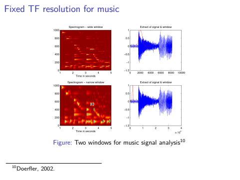


Figure: Two windows for music signal analysis[10]

[10] Doerfler, 2002.

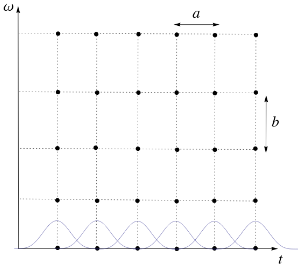


Stationary Gabor frames

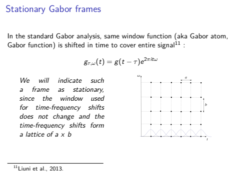
In the standard Gabor analysis, same window function (aka Gabor atom, Gabor function) is shifted in time to cover entire signal[11] :

$$g_{\tau,\omega}(t) = g(t-\tau)e^{2\pi it\omega}$$

We will indicate such a frame as stationary, since the window used for time-frequency shifts does not change and the time-frequency shifts form a lattice of a x b



[11] Liuni et al., 2013.

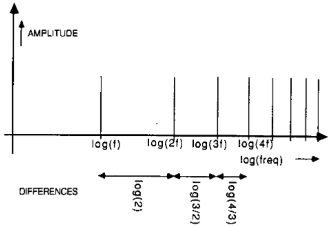


- Last presentation, I didn't call it stationary - only "gabor frame" alone
- until the non-stationary variant was created, there was no need to call the original stationary

Early CQT: musical motivation

Constant-Q transform for music analysis[12],[13]:

1. Harmonics of the fundamental have consistent spacing in the log scale – the constant pattern



2. Log-frequency spectra, demonstrating the constant pattern for harmonics, would be more useful in musical tasks

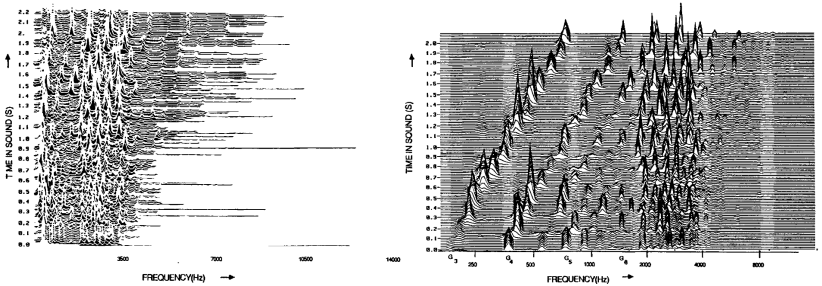
[12] J. Brown. "Calculation of a constant Q spectral transform". In: *Journal of the Acoustical Society of America* 89 (1991), pp. 425–434. URL: https://www.ee.columbia.edu/~dpwe/papers/Brown91-cqt.pdf.

[13] Judith Brown and Miller Puckette. ""An efficient algorithm for the calculation of a constant Q transform"". In: *Journal of the Acoustical Society of America* 92 (Nov. 1992), p. 2698. DOI: 10.1121/1.404385.

Early CQT: musical motivation
Constant-Q transform for music analysis[12],[13]:
Harmonics of the fundamental have consistent spacing in the log scale – the constant pattern
Log-frequency spectra, demonstrating the constant pattern for harmonics, would be more useful in musical tasks

- judith brown, MSP (of miller s puckette fame)
- constant-Q transform was known about before Gabor frames - like i discussed with prof
- tasks such as instrument identification by timbre, etc.
- also lines up with pitch perception as pattern recognition

Violin: DFT vs. CQT

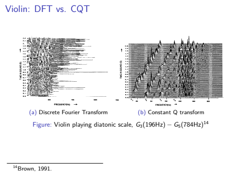


(a) Discrete Fourier Transform

(b) Constant Q transform

Figure: Violin playing diatonic scale, $G_3(196\text{Hz}) - G_5(784\text{Hz})$[14]

[14] Brown, 1991.



- Not explicitly named as an STFT but we know it is
- we can see note changes clearly, the fundamental, and even the formant in 3000hz region

Constant-Q Transform

"Constant ratio of frequency to frequency resolution": $\frac{f}{\delta f} = Q$

	Constant Q	DFT
Frequency	$(2^{1/24})^k \cdot f\min$ exponential in k	$k\,\Delta f$ linear in k
Window	variable $= \mathrm{N}[\mathrm{k}] = \frac{SR \cdot Q}{f_k}$	constant $= N$
Resolution Δf	variable $= f_k/Q$	constant $= \mathrm{SR}/\mathrm{N}$
$\frac{f_k}{\Delta f_k}$	constant $= Q$	variable $= k$
Cycles in Window	constant $= Q$	variable $= \mathrm{k}$

(a) Properties of DFT, CQT

Frequency (Hz)	Window (Samples)	(ms)
175	6231	195
208	5239	164
247	4406	138
294	3705	116
349	3115	97
415	2619	82
494	2203	69
587	1852	58
699	1557	49
831	1309	41
988	1101	34
1175	926	29

(b) Window sizes for CQT

Computed with windowed DFT where window changes with frequency to maintain Constant-Q[15] – **non-invertible!**:

$$\text{window len } N[k] = \frac{f_s}{f_k} Q, W[k,n] = \alpha + (1-\alpha)\cos\left(\frac{2\pi n}{N[k]}\right)$$

[15] Brown, 1991.

- meanwhile the linear spacing of conventional DFT leads to a pattern that varies with the harmonic – making it harder to identify
- in linear DFT, pattern of harmonic spacing changes with frequency

Multiple Gabor systems: musical motivation

In the case of music signals, for example, transients are important for several reasons. They give important cues for onset timing, and they carry information about instrument timbre. As another example, in low-frequency regions, very fine frequency resolution is required, because notes in this region lay the harmonic basis, musically speaking.

In order to achieve a setting adapted to music as discussed above, it will be necessary to use wider windows with good frequency concentration in low-frequency regions, whereas in the high-pass regions, where mainly transients and broadband signals components occur, rather short windows, which don't have to be very localised in frequency, will be of use [16]

[16] Doerfler, 2002.

Multiple Gabor systems: musical motivation

In the case of music signals, for example, transients are important for several reasons. They give important cues for onset timing, and they carry information about instrument timbre. As another example, in low-frequency regions, very fine frequency resolution is required, because notes in this region lay the harmonic basis, musically speaking.

In order to achieve a setting adapted to music as discussed above, it will be necessary to use wider windows with good frequency concentration in low-frequency regions, whereas in the high-pass regions, where mainly transients and broadband signals components occur, rather short windows, which don't have to be very localised in frequency, will be of use [16]

[16] Doerfler, 2002.

- similar motivation, except Dorfler et all are using this motivation to drive mathematical theory
- the lack of theory didn't affect the evolution, or desire, for the CQT

Multi-window Gabor dictionary

Stationary Gabor atom, where a and b are TF shift parameters:

$$g_{m,n}(t) = g(t - na)e^{j2\pi mbt}, \qquad m, n \in \mathbb{Z}$$

$$f(t) = \sum_{m,n \in \mathbb{Z}} c_{m,n} g_{m,n}(t)$$

Use R different windows, where a_r and b_r are TF shift parameters for each distinct window:

$$g^r_{m,n}(t) = g(t - na_r)e^{j2\pi mb_r t}, \qquad m, n \in \mathbb{Z}$$

$$f(t) = \sum_{r=0}^{R-1} \sum_{m,n \in \mathbb{Z}} c^r_{m,n} g^r_{m,n}(t)$$

Multi-window Gabor dictionary

Stationary Gabor atom, where a and b are TF shift parameters:

$$g_{m,n}(t) = g(t - na)e^{j2\pi mbt}, \qquad m, n \in \mathbb{Z}$$

$$f(t) = \sum_{m,n \in \mathbb{Z}} c_{m,n} g_{m,n}(t)$$

Use R different windows, where a_r and b_r are TF shift parameters for each distinct window:

$$g^r_{m,n}(t) = g(t - na_r)e^{j2\pi mb_r t}, \qquad m, n \in \mathbb{Z}$$

$$f(t) = \sum_{r=0}^{R-1} \sum_{m,n \in \mathbb{Z}} c^r_{m,n} g^r_{m,n}(t)$$

- f = signal, constructed from time-frequency shifts of gabor atoms

Example: multiple STFTs

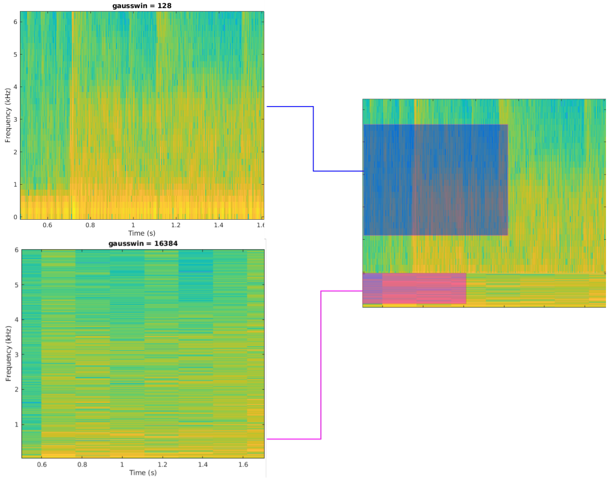
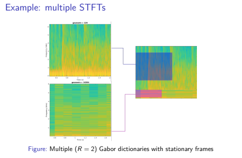


Figure: Multiple ($R = 2$) Gabor dictionaries with stationary frames



- fulfills musical requirements (good frequency resolution in low-frequency region, good time resolution in high-frequency region)
- highly redundant, overcomplete
- choose appropriate window STFT for each region of interest
- According to the multi-window approach, the dictionary will have high redundancy as it is basically the combination of R complete Gabor dictionaries. Due to the structure of audio signals, it is an appealing idea to reduce this highly redundant dictionary to fit to the special characteristics of these signals.

Nonstationary Gabor frame – resolution changing over time

Stationary Gabor atom, where a and b are TF shift parameters:

$$g_{m,n}(t) = g(t - na)e^{j2\pi mbt}, \qquad m, n \in \mathbb{Z}$$

Nonstationary Gabor atom from a set of functions $\{g_n\}$ and a fixed frequency sampling step b_n[17] :

$$g_{m,n}(t) = g_n(t)e^{j2\pi mb_n t}, \qquad m, n \in \mathbb{Z}$$

We get back classic nonstationary Gabor frame by setting:

$$g_n(t) = g(t - na) \text{ for a fixed time constant } a, b_n = b \forall n$$

[17] Balazs et al., 2011.

Nonstationary Gabor frame – resolution changing over time

Stationary Gabor atom, where a and b are TF shift parameters:

$$g_{m,n}(t) = g(t - na)e^{j2\pi mbt}, \qquad m, n \in \mathbb{Z}$$

Nonstationary Gabor atom from a set of functions $\{g_n\}$ and a fixed frequency sampling step b_n[17] :

$$g_{m,n}(t) = g_n(t)e^{j2\pi mb_n t}, \qquad m, n \in \mathbb{Z}$$

We get back classic nonstationary Gabor frame by setting:

$$g_n(t) = g(t - na) \text{ for a fixed time constant } a, b_n = b \forall n$$

[17] Balazs et al., 2011.

Nonstationary Gabor frame – resolution changing over time

> *[...] the functions $\{g_n\}$ are well-localized and centered around time-points a_n. This is similar to the standard Gabor scheme [...] with the possibility to vary the window g_n for each position a_n. Thus, sampling of the time-frequency plane is done on a grid which is irregular over time, but regular over frequency at each temporal position.*[18]

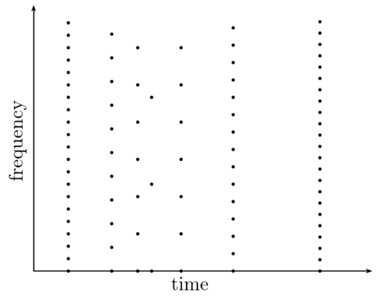
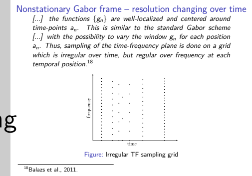


Figure: Irregular TF sampling grid

[18] Balazs et al., 2011.

└─Nonstationary Gabor frame – resolution changing over time



Nonstationary Gabor frame – resolution changing over frequency

Stationary Gabor atom, where a and b are TF shift parameters:

$$g_{m,n}(t) = g(t - na)e^{j2\pi mbt}, \qquad m, n \in \mathbb{Z}$$

Nonstationary Gabor atom from a family of functions $\{h_m\}$ and a fixed time sampling step a_m[19] :

$$h_{m,n}(t) = h_m(t - na_m), \qquad m, n \in \mathbb{Z}$$

We get back classic nonstationary Gabor frame by setting:

$h_m(t-na_m) = g(t-na)e^{j2\pi mbt}$ for a fixed frequency constant $b, a_m = a \forall m$

[19] Balazs et al., 2011.

Nonstationary Gabor frame – resolution changing over frequency

Stationary Gabor atom, where a and b are TF shift parameters:

$$g_{m,n}(t) = g(t - na)e^{j2\pi mbt}, \qquad m, n \in \mathbb{Z}$$

Nonstationary Gabor atom from a family of functions $\{h_m\}$ and a fixed time sampling step a_m[19] :

$$h_{m,n}(t) = h_m(t - na_m), \qquad m, n \in \mathbb{Z}$$

We get back classic nonstationary Gabor frame by setting:

$h_m(t-na_m) = g(t-na)e^{j2\pi mbt}$ for a fixed frequency constant $b, a_m = a \forall m$

[19] Balazs et al., 2011.

Nonstationary Gabor frame – resolution changing over frequency

In practice we will choose each function h_m as a well-localized band-pass function with center frequency b_n[20]

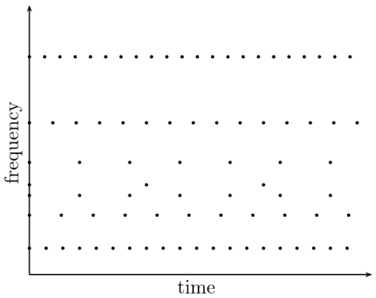
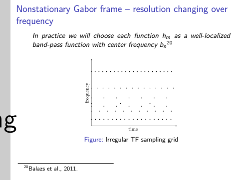


Figure: Irregular TF sampling grid

[20] Balazs et al., 2011.



Nonstationary Gabor frame construction

Construction of painless nonstationary Gabor frames relies on three properties of the windows and time-frequency shift parameters used[21] :

- The signal f of interest is localized at time- (or frequency-) positions n by means of multiplication with a compactly supported (or limited bandwidth, respectively) window function g_n
- The Fourier transform is applied on the localized pieces $f \cdot g_n$. The resulting spectra are sampled densely enough in order to perfectly re-construct $f \cdot g_n$ from these samples
- Adjacent windows overlap to avoid loss of information. At the same time, unnecessary overlap is undesirable. We assume that $0 < A \leq \sum_{n \in \mathbb{Z}} |g_n(t)|^2 \leq B < \infty$, a.e., for some positive A and B

These requirements lead to invertibility of the frame operator and therefore to perfect reconstruction.

[21] Balazs et al., 2011.

Nonstationary Gabor frame construction

Construction of painless nonstationary Gabor frames relies on three properties of the windows and time-frequency shift parameters used[21] :

- The signal f of interest is localized at time- (or frequency-) positions n by means of multiplication with a compactly supported (or limited bandwidth, respectively) window function g_n
- The Fourier transform is applied on the localized pieces $f \cdot g_n$. The resulting spectra are sampled densely enough in order to perfectly re-construct $f \cdot g_n$ from these samples
- Adjacent windows overlap to avoid loss of information. At the same time, unnecessary overlap is undesirable. We assume that $0 < A \leq \sum_{n \in \mathbb{Z}} |g_n(t)|^2 \leq B < \infty$, a.e., for some positive A and B

These requirements lead to invertibility of the frame operator and therefore to perfect reconstruction.

[21] Balazs et al., 2011.

- a.e. = almost everywhere
- Moreover, the frame operator is diagonal and its inversion is straight-forward. Further, the canonical dual frame has the same structure as the original one. Because of these pleasant consequences following from the three above-mentioned requirements, the frames satisfying all of them will be called painless nonstationary Gabor frames and we refer to this situation as the painless case

Discrete nonstationary Gabor frame numerical complexity

Discrete nonstationary Gabor frame in discrete time (analogous to continuous time)[22] :

$$g_{m,n}[k] = g_n[k] \cdot e^{\frac{j2\pi m b_n k}{L}} = g_n[k] \cdot W_L^{m b_n k}$$

$$n = 0, ..., N-1, m = 0, ..., M_n - 1, k = 0, ..., L = 1$$

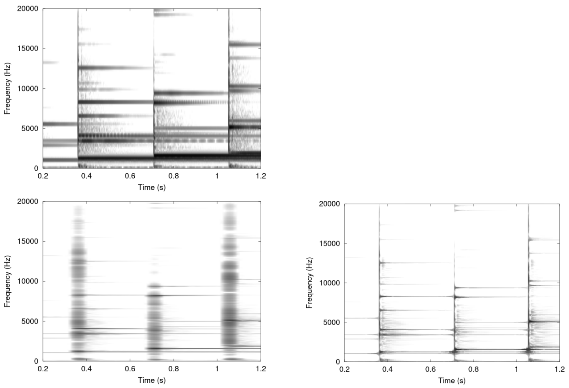
Nonstationary Gabor coefficients are given by an FFT of length M_n for each window g_n with a signal of length L_n

1. Windowing: L_n operations for the nth window
2. FFT: $O(M_n \cdot \log(M_n))$ for the nth window
3. Total: $O(N \cdot (M \log(M)))$ for N windows

[22] Balazs et al., 2011.

Discrete nonstationary Gabor frame numerical complexity
Discrete nonstationary Gabor frame in discrete time (analogous to continuous time)[22] :
$$g_{m,n}[k] = g_n[k] \cdot e^{\frac{j2\pi m b_n k}{L}} = g_n[k] \cdot W_L^{m b_n k}$$
$$n = 0, ..., N-1, m = 0, ..., M_n - 1, k = 0, ..., L = 1$$
Nonstationary Gabor coefficients are given by an FFT of length M_n for each window g_n with a signal of length L_n
1. Windowing: L_n operations for the nth window
2. FFT: $O(M_n \cdot \log(M_n))$ for the nth window
3. Total: $O(N \cdot (M \log(M)))$ for N windows
[22] Balazs et al., 2011.

Result of nonstationary Gabor decomposition

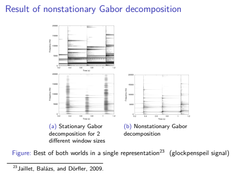


(a) Stationary Gabor decomposition for 2 different window sizes

(b) Nonstationary Gabor decomposition

Figure: Best of both worlds in a single representation[23] (glockenspeil signal)

[23] Jaillet, Balázs, and Dörfler, 2009.



- i skipped many details proving the invertibility, tightness, minimizing redundancy

Applications of NSGT: invertible CQT

- Original CQT[24] is non-invertible and computationally intensive
- Modification[25] improves computational efficiency
- Previous approach by Schörkhuber and Klapuri[26] has an RMS error of 10^{-3} from approximate reconstruction (used in librosa)

  > *The lack of perfect invertibility prevents the convenient modification of CQT coefficients with subsequent resynthesis required in complex music processing tasks such as masking or transposition.*

- NSGT CQT is faster and perfectly invertible; adaptive resolution in frequency results in desired constant Q-factor
- Drop-in replacement for STFT in music algorithms

---

[24] Brown, 1991.

[25] Brown and Puckette, 1992.

[26] C. Schörkhuber and A. Klapuri. *Constant-Q Transform Toolbox for Music Processing*. July 2010. DOI: 10.5281/zenodo.849741. URL: https://doi.org/10.5281/zenodo.849741.

Applications of NSGT: invertible CQT
- Original CQT[24] is non-invertible and computationally intensive
- Modification[25] improves computational efficiency
- Previous approach by Schörkhuber and Klapuri[26] has an RMS error of 10^{-3} from approximate reconstruction (used in librosa) The lack of perfect invertibility prevents the convenient modification of CQT coefficients with subsequent resynthesis required in complex music processing tasks such as masking or transposition.
- NSGT CQT is faster and perfectly invertible; adaptive resolution in frequency results in desired constant Q-factor
- Drop-in replacement for STFT in music algorithms
[24] Brown, 1991.
[25] Brown and Puckette, 1992.
[26] C. Schörkhuber and A. Klapuri. Constant-Q Transform Toolbox for Music Processing. July 2010. DOI: 10.5281/zenodo.849741. URL: https://doi.org/10.5281/zenodo.849741.

- so first its interesting that even today the outcomes of this paper are yet to spread - librosa currently has an unusable CQT!
- the invertible CQT is amazing because it's basically a drop-in replacement of the STFT - you can do time-frequency masking, magnitude/power "spectrogram" equivalent - benefit automatically from the musical considerations baked into the CQT
- talk about fitzgerald source separation