Theory, implementation and applications of nonstationary Gabor Frames

P. Balazs[a], M. Dörfler[b,*], F. Jaillet[c], N. Holighaus[b], G. Velasco[b,d]

[a] *Acoustics Research Institute, Austrian Academy of Sciences, Wohllebengasse 12-14, 1040 Wien, Austria*
[b] *Numerical Harmonic Analysis Group, Faculty of Mathematics, University of Vienna, Alserbachstraße 23, 1090 Wien, Austria*
[c] *Institut de Neurosciences Cognitives de la Méditerranée, UMR 6193 CNRS - Université de la Méditerranée, 31 chemin Joseph Aiguier, 13402 Marseille cedex 20, France*
[d] *Institute of Mathematics, University of the Philippines - Diliman, 1101 Quezon City, Philippines*

Abstract

Signal analysis with classical Gabor frames leads to a fixed time-frequency resolution over the whole time-frequency plane. To overcome the limitations imposed by this rigidity, we propose an extension of Gabor theory that leads to the construction of frames with time-frequency resolution changing over time or frequency. We describe the construction of the resulting *nonstationary Gabor frames* and give the explicit formula for the canonical dual frame for a particular case, the *painless case*. We show that wavelet transforms, constant-Q transforms and more general filter banks may be modeled in the framework of nonstationary Gabor frames. Further, we present the results in the finite-dimensional case, which provides a method for implementing the above-mentioned transforms with perfect reconstruction. Finally, we elaborate on two applications of nonstationary Gabor frames in audio signal processing, namely a method for automatic adaptation to transients and an algorithm for an invertible constant-Q transform.

Keywords: Time-frequency analysis, Adaptive representation, Constant-Q

*Corresponding Author

Email addresses: peter.balazs@oeaw.ac.at (P. Balazs), monika.doerfler@univie.ac.at (M. Dörfler), florent.jaillet@incm.cnrs-mrs.fr (F. Jaillet), nicki.holighaus@univie.ac.at (N. Holighaus), gino.velasco@univie.ac.at (G. Velasco)

transform, Invertibility

1. Introduction

Redundant short-time Fourier methods, also known as Gabor analysis [1], are widely used in signal processing applications. The basic idea is the analysis of a signal f by consideration of the projections $\langle f, g_{\tau,\omega} \rangle$ of f onto time-frequency atoms $g_{\tau,\omega}$. The $g_{\tau,\omega}$ are obtained by translation of a unique prototype function over time and frequency: $g_{\tau,\omega}(t) = g(t-\tau)e^{2\pi it\omega}$. This classical construction leads to a signal decomposition with fixed time-frequency resolution over the whole time-frequency plane. The restriction to a fixed resolution is often undesirable in processing signals with variable time-frequency characteristics. Alternative decompositions have been introduced to overcome this deficit, e.g. the wavelet transform [2], the constant-Q transform (CQT) [3] or decompositions using filter banks [4], in particular based on perceptive frequency scales [5]. Adaptation over time is considered in approaches such as modulated lapped transforms [6], adapted local trigonometric transforms [7] or (time-varying) wavelet packets [8].

Most of the cited work achieves flexible tilings of the time-frequency plane, but efficient reconstruction from signal-adaptive, overcomplete time-frequency transforms is rarely addressed. One exception is a recent approach in [9], which is in fact a special case of the more general model considered in the present paper. The wealth of existing approaches to fast adaptive transforms underlines the need for flexibility arising from many applications. On the other hand, the introduction of flexibility in a transform that is based on accurate mathematical modeling causes technical complications that are not always easy to overcome. We introduce an approach to fast adaptive time-frequency transforms, that is based on a generalization of *painless nonorthogonal expansions* [10]. It allows for adaptivity of the analysis windows *and* the sampling points. Since the resulting frames locally resemble classical Gabor frames and share some of their structure, they are called *nonstationary Gabor frames*. The corresponding transform is likewise referred to as *nonstationary Gabor transform* (NSGT).

The central feature of painless expansions is the diagonality of the frame operator associated with the proposed analysis system. This idea is used here to yield painless nonstationary Gabor frames and will allow for both mathematical accuracy in the sense of perfect reconstruction (the frame operator is invertible) and numerical feasibility by means of an FFT-based implementation. The construction of painless nonstationary Gabor frames relies on three intuitively accessible properties of the windows and time-frequency shift parameters used.

1. 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.
2. The Fourier transform is applied on the localized pieces $f \cdot g_n$. The resulting spectra are sampled densely enough in order to perfectly reconstruct $f \cdot g_n$ from these samples.
3. Adjacent windows overlap to avoid loss of information. At the same time, unnecessary overlap is undesirable. In other words, 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.

We will show that these requirements lead to invertibility of the frame operator and therefore to perfect reconstruction. 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*. Under application of a Fourier transform to the signal of interest, our approach leads to adaptivity in either time or frequency. The concept of this paper relies on ideas introduced in [11], and presented at [12]. In the present paper all formal proofs are given, the link to frame theory is provided, the possibility to represent other analysis/synthesis systems with this approach is established, the numerical issues are investigated and several applications are presented.

The rest of the article is organized as follows. We fix notation and review preliminary results from Gabor and frame theory in Section 2. Section 3 introduces the construction of (painless) nonstationary Gabor frames in detail and provides a proof for the frame property under the given conditions. The calculation of the dual or tight frames is also explicitly given for systems adaptive in time or frequency, respectively. Section 4 then establishes the details of implementation in a discrete and real-life setting and provides examples together with a comparison of numerical efficiency with existing approaches. We conclude, in Section 5 with a summary and a brief outlook on future work.

In the sense of reproducible research, [13], we provide all algorithms and scripts to reproduce the results in this paper at the webpage `http://univie.ac.at/nonstatgab/`. Please note that a nonstationary Gabor transform is also included in the Linear Time Frequency Analysis Toolbox (LTFAT) v.1.0 [14, 15], a Matlab/Octave toolbox, which is freely available at `http://ltfat.sourceforge.net/`.

2. Preliminaries

For an integrable function f, i.e. $f \in L^1(\mathbb{R})$, we denote its Fourier transform $\mathcal{F}f(\xi) = \hat{f}(\xi) = \int_{\mathbb{R}} f(t)\, e^{-2\pi i \xi t} dt$, with the usual extension to $L^2(\mathbb{R})$, the space of square-integrable functions from $\mathbb{R}$ to $\mathbb{C}$. The convolution of two functions $f, g \in L^1(\mathbb{R})$ is the function $f * g$ defined by $(f * g)(t) = \int_{\mathbb{R}} f(x) g(t-x)\, dx$, again with the usual extension to $L^2(\mathbb{R})$. It follows that $\mathcal{F}(f * g) = \hat{f} \cdot \hat{g}$. We use the notation $f(t) \simeq g(t)$ if there exist constants C_1, $C_2 > 0$, such that $C_1\, g(t) \leq f(t) \leq C_2\, g(t)$ for all t.

2.1. Frame Theory

We now give a short summary of frame theory on Hilbert spaces, first introduced in [16]. A thorough discussion can be found in [17] or [18].

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 (called lower and upper frame bounds, respectively) such that

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

i.e. $\sum_{l \in I} |\langle f, \psi_l \rangle|^2 \simeq \|f\|^2$. If $A = B$, then $(\psi_l)_{l \in I}$ is a *tight frame*. By $\mathbf{C} : \mathcal{H} \to \ell^2$, we denote the *analysis operator* defined by $(\mathbf{C}f)_l = \langle f, \psi_l \rangle$. The adjoint of $\mathbf{C}^*$ of $\mathbf{C}$ is the *synthesis operator* $\mathbf{C}^*(c_l) = \sum_l c_l \psi_l$. The *frame operator* is $\mathbf{S}f = \mathbf{C}^*\mathbf{C}f = \sum_l \langle f, \psi_l \rangle \psi_l$, hence $\langle \mathbf{S}f, f \rangle = \|\mathbf{C}f\|^2_{\ell^2}$.

The boundedness and invertibility of $\mathbf{S}$ is equivalent to the existence of frame bounds $0 < A, B < \infty$ in the frame inequality (1), as well as to the existence of *dual frames*, which can be used for reconstruction. In particular, the *canonical dual frame* $(\tilde{\psi_l})$, is found by applying the inverse of $\mathbf{S}$ to the original frame elements, i.e. $\tilde{\psi_l} = \mathbf{S}^{-1}\psi_l$ for all l. For all $f \in \mathcal{H}$ we then have the following reconstruction formulas:

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

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.

2.2. Gabor Theory

Recall that for any nonzero function $g \in L^2(\mathbb{R})$ (the *window*), the *short-time Fourier transform* (STFT) of a signal $f \in L^2(\mathbb{R})$ is defined as $\mathcal{V}_g(f)(\tau,\omega) = \langle f, \mathbf{M}_\omega \mathbf{T}_\tau g\rangle$, using the *translation operator* $\mathbf{T}_\tau f(t) = f(t-\tau)$ and the *modulation operator* $\mathbf{M}_\omega f(t) = f(t)\, e^{2\pi i\omega t}$. In $L^2(\mathbb{R})$, we have

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

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* [19]. Moreover, if $\mathcal{G}(g,a,b)$ is a frame, it is called a *Gabor frame* and the associated frame operator is denoted by $\mathbf{S}_{g,a,b}$. In the succeeding sections, where the dependence of the frame operator on the window g and the parameters a, b is clear, we simply denote the frame operator by $\mathbf{S}$. Note that the Gabor analysis coefficients are sampling points of the STFT of f with window g at the time-frequency points (an, bm), i.e. $\mathcal{V}_g(f)(an, bm) = \{\langle f, \mathbf{M}_{bm}\mathbf{T}_{an} g\rangle\}_{m,n}$.

A central property of Gabor frames is the fact that the 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 $\{(an, bm)\,|m, n \in \mathbb{Z}\}$. Note that the property that the dual system is again a system with the same structure, is a particular property of Gabor frames, shared by nonstationary Gabor frames in the painless setting considered in the present paper. For a more detailed introduction to Gabor analysis, see [1] or [20].

In the finite discrete case, we take the Hilbert space $\mathcal{H}$ to be $\mathbb{C}^L$. For a good introduction to Gabor analysis in this setting, see [21]. We shall restrict the lattice parameters a and b to factors of L such that the numbers $N = \frac{L}{a}$ and $M = \frac{L}{b}$ are integers. We regard all vectors as periodic, so discrete translation is a cyclic operator. Therefore the discretization of time-shift and modulation is given by

$$\mathbf{T}_n x = (x_{L-n}, x_{L-n+1}, \ldots, x_0, x_1, \ldots, x_{L-n-1})$$

and

$$\mathbf{M}_m x = \left(x_0 \cdot W_L^0, x_1 \cdot W_L^{1 \cdot m}, \ldots, x_{L-1} \cdot W_L^{(L-1)m}\right)$$

with $W_L = e^{\frac{2\pi i}{L}}$, respectively. We will consider the Gabor system

$$\mathcal{G}(g, a, b) = \{\mathbf{M}_{bm}\mathbf{T}_{an}g : n = 0, \ldots, N-1; m = 0, \ldots, M-1\},$$

which is a collection of $M \cdot N$ vectors in $\mathbb{C}^L$. Obviously, to fulfill the frame conditions (1), we need at least $M \cdot N \geq L$.

2.3. Wavelet Theory

As we will see below, nonstationary Gabor frames may be used to construct wavelet frames. We briefly sketch the continuous wavelet transform. Let $\psi \in L^2(\mathbb{R})$ and $(\alpha, \beta) \in \mathbb{R}_+^* \times \mathbb{R}$. We define the wavelet system by

$$\psi_{\alpha,\beta}(t) = \frac{1}{\sqrt{\alpha}}\psi\left(\frac{t-\beta}{\alpha}\right) = \mathbf{T}_\beta \mathbf{D}_\alpha \psi, \tag{2}$$

where $\mathbf{D}_\alpha$ denotes the dilation operator given by $\mathbf{D}_\alpha f(t) = \frac{1}{\sqrt{\alpha}} f(\frac{t}{\alpha})$.

The wavelet transform is then defined as

$$\mathbf{W}_\psi f(\alpha, \beta) = \langle f, \mathbf{T}_\beta \mathbf{D}_\alpha \psi \rangle = \left(f * \mathbf{D}_\alpha \overline{\mathcal{I}\psi}\right)(\beta), \tag{3}$$

where $\mathcal{I}$ denotes the involution $\mathcal{I}g(t) = g(-t)$.

If ψ is localized around τ_0, then $\psi_{\alpha,\beta}(t)$ is centered at $\alpha \cdot \tau_0 + \beta$. The frequency center is at η/α, where η is the center of $\hat{\psi}$.

3. Construction of nonstationary Gabor frames

3.1. Resolution changing over time

As opposed to standard Gabor analysis, where time translation is used to generate atoms, the setting of nonstationary Gabor frames allows for changing, hence adaptive, windows in different time positions. Then, for each time position, we build atoms by *regular* frequency modulation. Using a set of functions $\{g_n\}_{n\in\mathbb{Z}}$ in $L^2(\mathbb{R})$ and frequency sampling step b_n, for $m \in \mathbb{Z}$ and $n \in \mathbb{Z}$, we define atoms of the form:

$$g_{m,n}(t) = g_n(t)e^{2\pi i m b_n t} = \mathbf{M}_{mb_n} g_n(t),$$

implicitly assuming that the functions g_n are well-localized and centered around time-points a_n. This is similar to the standard Gabor scheme, however, 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.

Figure 1 shows an example of such a sampling grid. Note that some results exist in Gabor theory for semi-regular sampling grids, as for example in [22]. Our study uses a more general setting, as the sampling grid is in general not separable and, more importantly, the window can evolve over time. To get a first idea of the effect of nonstationary Gabor frames, the reader may take a look at Figure 2 and Figure 3, which show regular Gabor transforms and a nonstationary Gabor transform of the same signal. Note that the NSGT in Figure 3 was adapted to transients and the components are well-resolved.

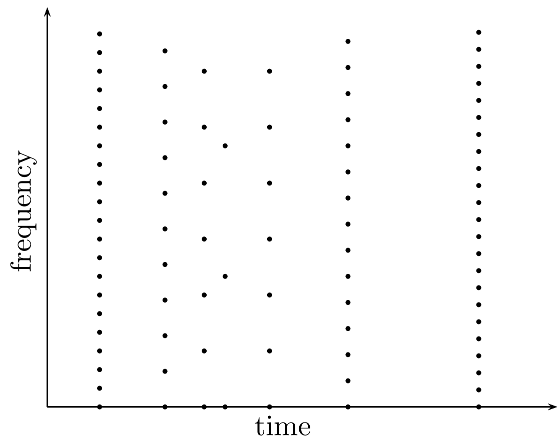

Figure 1: Example of a sampling grid of the time-frequency plane when building a decomposition with time-frequency resolution evolving over time

In the current situation, the analysis coefficients may be written as

$$c_{m,n} = \langle f, \mathbf{M}_{mb_n} g_n \rangle = \widehat{(f \cdot \overline{g_n})}(mb_n),\ m, n \in \mathbb{Z}.$$

Remark 1. If we set $g_n(t) = g(t - na)$ for a fixed time-constant a and $b_n = b$ for all n, we obtain the case of classical painless non-orthogonal expansions for regular Gabor systems introduced in [10].

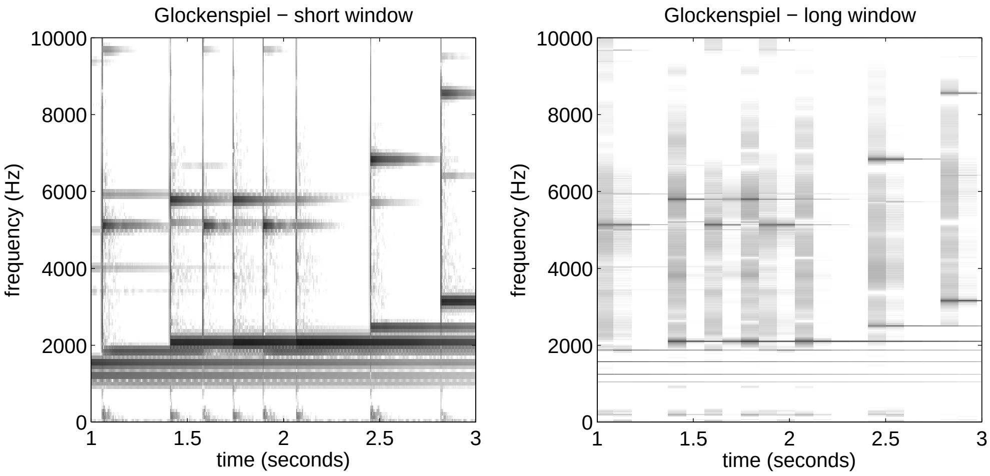

Figure 2: Glockenspiel (Example 1). Gabor representations with short window (11.6 ms), resp. long window (185.8 ms).

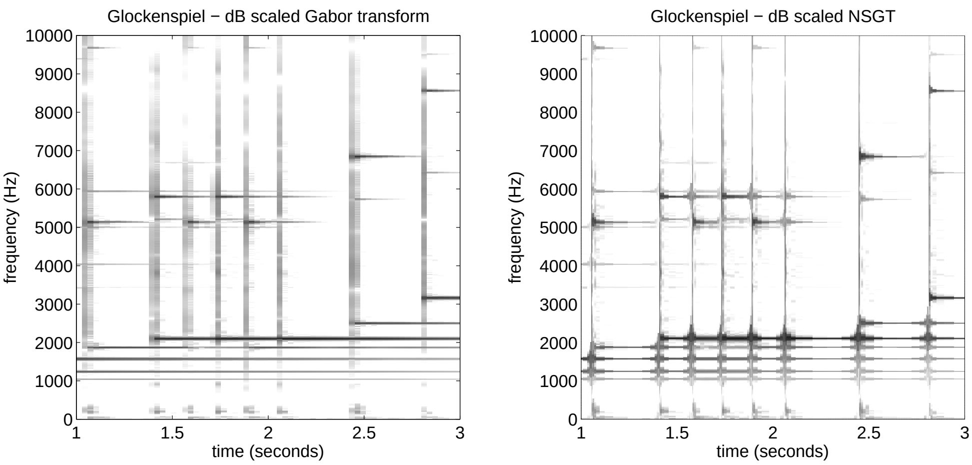

Figure 3: Glockenspiel (Example 1). Regular Gabor representation with a Hann window of 58 ms length and a nonstationary Gabor representation using Hann windows of varying length.

3.2. Resolution changing over frequency

An analog construction in the frequency domain leads to irregular sampling over frequency, together with windows featuring adaptive bandwidth. Then, sampling is regular over time. An example of the sampling grid in such a case is given in Figure 4.

In this case, we introduce a family of functions $\{h_m\}_{m\in\mathbb{Z}}$ of $L^2(\mathbb{R})$, and for $m \in \mathbb{Z}$ and $n \in \mathbb{Z}$, we define atoms of the form:

$$h_{m,n}(t) = h_m(t - na_m). \tag{4}$$

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

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

Hence, the situation is completely analog to the one described in the previous section, up to a Fourier transform.

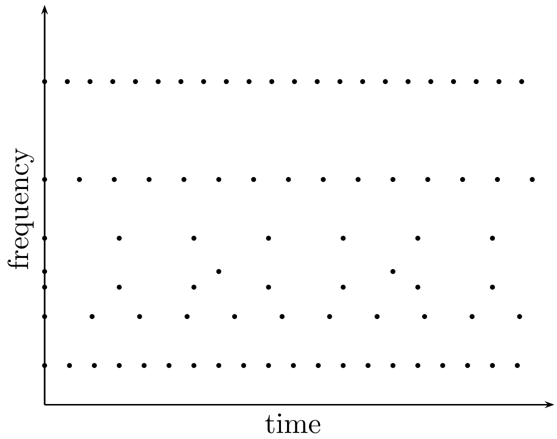

Figure 4: Example of a sampling grid of the time-frequency plane when building a decomposition with time-frequency resolution changing over frequency

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

3.2.1. Link between nonstationary Gabor frames, wavelet frames and filterbanks:

To obtain wavelet frames, the wavelet transform in (2) is sampled at sampling points (β_n, α_m). A typical discretization scheme [23] is $(n\beta_0, \alpha_0^m)$.

Then, the frame elements are $\psi_{m,n}(t) = \mathbf{T}_{n\beta_0}\mathbf{D}_{\alpha_0^m}\psi(t)$. Comparing this expression to (4) and setting $h_m = \mathbf{D}_{\alpha_0}^m\psi$ and $a_m = \beta_0$, we see that a wavelet frame with this discretization scheme corresponds to a nonstationary Gabor transform.

Another possibility for sampling the continuous wavelet transform [2] uses $\alpha = \alpha_0^m$ and $\beta = n\beta_0\alpha_0^m$. Again, we obtain a correspondence to nonstationary Gabor frames by setting $h_m = \mathbf{D}_{\alpha_0}^m\psi$ and $a_m = \beta_0 \cdot \alpha_0^m$.

Beyond the setting of wavelets, any *filter bank* [23], even with non-constant down-sampling factors D_m, can be written as a nonstationary Gabor frame. A filter bank is a set of time-invariant, linear filters $\mathfrak{h}_m$, i.e. Fourier multipliers. The response of a filter bank for the signal f and sampling period T_0 is given (in the continuous case) by

$$c_{m,n} = (f * \mathfrak{h}_m)(nD_mT_0) = \int_{\mathbb{R}} f(t)\mathfrak{h}_m(nD_mT_0 - t)\,dt = \langle f, h_{m,n}\rangle\,,$$

where $h_{m,n}(t) = \overline{\mathfrak{h}(nD_mT_0 - t)}$. Setting $h_m = \overline{\mathcal{I}\,\mathfrak{h}_m}$ and choosing $a_m = D_mT_0$ this construction is realized with nonstationary Gabor frames using (4). If the filters are band-limited and the down-sampling factors are small enough, then the conditions for the painless case are met and the corresponding reconstruction procedure can be applied.

3.3. Invertibility of the frame operator and reconstruction

In this central section we give the precise conditions under which painless nonstationary Gabor frames are constructed. The first two basic conditions, namely compactly supported windows and sufficiently dense frequency sampling points, lead to diagonality of the associated frame operator $\mathbf{S}$, as defined in Section 2.1. The third condition, the controlled overlap of adjacent windows, then leads to boundedness and invertibility of $\mathbf{S}$. The following theorem generalizes the results given for the classical case of painless non-orthogonal expansions [10, 20].

Theorem 1. *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

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

is given by a multiplication operator of the form

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

Proof. Note that,

$$\begin{aligned}
\langle \mathbf{S}f, f \rangle &= \sum_n \sum_m \Big| \int_{\mathbb{R}} f(t)\, \overline{g_n(t)}\, e^{-2\pi i m b_n t} dt \Big|^2 \\
&= \sum_n \sum_m \Big| \int_{c_n}^{d_n} f(t)\, \overline{g_n(t)}\, e^{-2\pi i m b_n t} dt \Big|^2,
\end{aligned}$$

due to the compact support property of the g_n. Let $I_n = [c_n, c_n + b_n^{-1}]$ for all n and χ_I denote the characteristic function of the interval I. Taking into account the compact support of g_n again, it is obvious that

$$f\, \overline{g_n} = \chi_{I_n} \sum_l \mathbf{T}_{l b_n^{-1}} (f \overline{g_n}),$$

with the b_n^{-1}-periodic function $\sum_l \mathbf{T}_{l b_n^{-1}} (f\, \overline{g_n})$. Hence, with $W_{m,n}(t) = e^{-2\pi i m b_n t}$,

$$\begin{aligned}
\Big| \int_{c_n}^{d_n} f(t)\, \overline{g_n(t)}\, W_{m,n}(t)\, dt \Big|^2 &= \Big| \int_{I_n} f(t)\, \overline{g_n(t)}\, W_{m,n}(t)\, dt \Big|^2, \\
&= \Big| \langle f\, \overline{g_n}, W_{m,n} \rangle_{L^2(I_n)} \Big|^2
\end{aligned}$$

and applying Parseval's identity to the sum over m yields

$$\begin{aligned}
\langle \mathbf{S}f, f \rangle &= \sum_n \sum_m |\langle f\, \overline{g_n}, W_{m,n} \rangle_{L^2(I_n)}|^2 \\
&= \sum_n \frac{1}{b_n} \|f\, \overline{g_n}\|^2 = \Big\langle \sum_n \frac{1}{b_n} |g_n|^2 f, f \Big\rangle.
\end{aligned}$$

□

While in general, the inversion of $\mathbf{S}$ can be numerically unfeasible, in the special case described in Theorem 1, the invertibility of the frame operator is easy to check and inversion is a simple multiplication.

Corollary 1. *Under the conditions given in Theorem 1, the system of functions $g_{m,n}$ forms a frame for $L^2(\mathbb{R})$ if and only if $\sum_n \frac{1}{b_n}|g_n(t)|^2 \simeq 1$. In this case, the canonical dual frame elements are given by:*

$$\tilde{g}_{m,n}(t) = \frac{g_n(t)}{\sum_l \frac{1}{b_l}|g_l(t)|^2} e^{2\pi i m b_n t}, \tag{5}$$

and the associated canonical tight frame elements can be calculated as:

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

Remark 2. The optimal lower and upper frame bounds are explicitly given by $A_{opt} = \operatorname{essinf} \sum_n \frac{1}{b_n}|g_n(t)|^2$ and $B_{opt} = \operatorname{esssup} \sum_n \frac{1}{b_n}|g_n(t)|^2$.

We next state the results of Theorem 1 and Corollary 1 in the Fourier domain. This is the basis for adaptation over frequency.

Corollary 2. *For every $m \in \mathbb{Z}$, let the function h_m be band-limited to $\operatorname{supp}(\widehat{h_m}) = [c_m, d_m]$ and let a_m be chosen such that $d_n - c_n \leq \frac{1}{a_m}$. Then the frame operator of the system*

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

is given by a convolution operator of the form

$$\langle \mathbf{S}f, f\rangle = \langle \mathcal{F}^{-1}\big(\sum_m \frac{1}{a_m}|\widehat{h_m}|^2\big) * f, f\rangle \tag{6}$$

for $f \in L^2(\mathbb{R})$. Hence, the system of functions $h_{m,n}$ forms a frame of $L^2(\mathbb{R})$ if and only if $\forall \nu \in \mathbb{R}$, $\sum_m \frac{1}{a_m}|\widehat{h_m}(\nu)|^2 \simeq 1$. The elements of the canonical dual frame are given by

$$\tilde{h}_{m,n}(t) = \mathbf{T}_{na_m}\mathcal{F}^{-1}\left(\frac{\widehat{h_m}}{\sum_l \frac{1}{a_l}|\widehat{h_l}|^2}\right)(t) \tag{7}$$

and the canonical tight frame is given by

$$\mathring{h}_{m,n}(t) = \mathbf{T}_{na_m}\mathcal{F}^{-1}\left(\frac{\widehat{h_m}}{\sqrt{\sum_l \frac{1}{a_l}|\widehat{h_l}|^2}}\right)(t). \tag{8}$$

Proof. We deduce the form of the frame operator in the current setting from the proof of Theorem 1 by setting

$$\langle \mathbf{S}f, f\rangle = \langle \widehat{\mathbf{S}f}, \hat{f}\rangle = \sum_{m,n} |\langle \hat{f}, \widehat{h_{m,n}}\rangle|^2$$

and the rest of the corollary is equivalent to Corollary 1. □

Remark 3. Classical Gabor frames are intimately related to modulation spaces, see [24] for an extensive discussion and relevant references. The characterization of modulation spaces depends on the joint time-frequency localization of the analysis window. Painless nonstationary Gabor frames characterize modulation spaces, if, in addition to compactness in one domain (time or frequency), the windows g_k exhibit a uniform decay in the sense of *time-frequency molecules*, see [25, Theorem 22], i.e., letting $\xi = (a_k, l/b_k)$, $k, l \in \mathbb{Z}$, we require $|\mathcal{V}_\varphi g_k(z)| \leq C(1 + |z - \xi|)^{-r}$ for some $r > 2$. Then, the corresponding frame operator is invertible on all modulation spaces M^p, $1 \leq p \leq \infty$, and the ℓ^p-norm of the corresponding coefficient sequence is equivalent to the modulation space norm.

Remark 4. As mentioned in Section 3.2.1 the NSGT is linked to wavelet frames. In the painless case it is possible to construct a dual sequence which has the same structure. For wavelets this is also possible, see e.g. [26, 27], where non-canonical duals are constructed.

In a similar way as modulation spaces are linked to the Gabor transform, Besov spaces are related to wavelet systems, see e.g. [28]. Also, Sobolev spaces can be linked to the wavelet transform [29]. Nonstationary Gabor frames can also be used to characterize Besov and Sobolev spaces, with some additional assumptions. Details will be reported elsewhere.

4. Discrete Finite Nonstationary Gabor Frames

4.1. Discrete, time-adaptive Gabor transform

For the practical implementation, the equivalent theory may be developed in a finite discrete setting using the Hilbert space $\mathbb{C}^L$. Since this is largely straight-forward from simple matrix multiplication, we only state the main result. Given a set of functions $\{g_n\}_{n\in\{0,\ldots,N-1\}}$, a set of integers (number of frequency samples for each time position) $\{M_n\}_{n\in\{0,\ldots,N-1\}}$ associated with the set of real values $\{b_n = \frac{L}{M_n}\}_{n\in\{0,\ldots,N-1\}}$, the discrete, nonstationary Gabor system is given by

$$g_{m,n}[k] = g_n[k] \cdot e^{\frac{2\pi i m b_n k}{L}} = g_n[k] \cdot W_L^{m b_n k}.$$

for $n = 0, \ldots, N-1$, $m = 0, \ldots, M_n - 1$ and all $k = 0, \ldots L-1$. Note that in practice, $g_{m,n}[k]$ will have zero-values for most k, allowing for efficient FFT-implementation: since $M_n = \frac{L}{b_n}$, we have $g_{m,n}[k] = g_n[k] \cdot e^{\frac{2\pi i m k}{M_n}}$ and the nonstationary Gabor coefficients are given by an FFT of length M_n for each g_n.

The number of elements of $\{g_{m,n}\}$ is $P = \sum_{n=0}^{N-1} M_n$. Let $\mathbf{G}$ be the $L \times P$ matrix such that its p-th column is $g_{m,n}$, for $p = m + \sum_{k=0}^{n-1} M_k$.

Corollary 3. *The frame operator $\mathbf{S} = \mathbf{G} \cdot \mathbf{G}^*$ is an $L \times L$ matrix with entries:*

$$\mathbf{S}_{k,j} = \sum_{n \in \mathcal{N}_{(k-j)}} M_n \, g_n[k] \, \overline{g_n[j]}$$

where $\mathcal{N}_p = \{n \in [0, N-1] \,|\, p = 0 \mod M_n\}$ for $p \in [-L, L]$. Therefore, if appropriate support conditions are met, $\mathbf{S}$ is a diagonal matrix.

4.1.1. Numerical complexity

Assuming that the windows g_n have support of length L_n, let $M = \max_n \{M_n\}$ be the maximum FFT-length. We consider the painless case where $L_n \leq M_n \leq M$. The number of operations is

1. Windowing: L_n operations for the n-th window.
2. FFT: $\mathcal{O}(M_n \cdot \log(M_n))$ for the n-th window.

Then the number of operations for the discrete NSGT is

$$\begin{aligned}\mathcal{O}\left(\sum_{n=0}^{N-1} M_n \cdot \log(M_n) + L_n\right) &= \mathcal{O}(N \cdot (M \log(M) + M)) \\ &= \mathcal{O}(N \cdot (M \log(M)))\end{aligned}$$

Similar to the regular Gabor case, the number of windows N will usually depend linearly on the signal length L while the maximum FFT-length M is assumed to be independent of L. In that case, the discrete NSGT is a linear cost algorithm.

For the construction of the dual windows in the painless case, the computation involves multiplication of the window functions by the inverse of the diagonal matrix $\mathbf{S}$ and results in $\mathcal{O}(2\sum_{n=0}^{N-1} L_n) = \mathcal{O}(N \cdot M)$ operations. Lastly, the inverse NSGT has numerical complexity $\mathcal{O}(N \cdot (M \log(M)))$, as in the NSGT, since it entails computing the IFFT of each coefficient vector, multiplying with the corresponding dual windows and evaluating the sum.

Technical framework: All subsequently presented simulations were done in MATLAB R2009b on a 2 Gigahertz Intel Core 2 Duo machine with 2 Gigabytes of RAM running Kubuntu 9.04. The CQTs were computed using the code published with [30], available for free download at `http://www.elec.qmul.ac.uk/people/anssik/cqt/`. The constant-Q nonstationary Gabor transform (CQ-NSGT) algorithms are available at `http://univie.ac.at/nonstatgab/`.

4.1.2. Application: automatic adaptation to transients

In real-life applications, NSGT has the potential to represent local signal characteristics, e.g. transient sound events, in a more appropriate way than pre-determined, regular transform schemes. Since the appropriateness of a representation depends on the specific application, any adaptation procedure must be designed specifically. For the implementation itself, however, two observations generally remain true: First, the general nonstationary framework needs to be restricted to a well defined set of choices. Second, some measure is needed to determine the most suitable of the possible choices. For example, in the case of a sparsity measure, the most sparse representation will be chosen. To show that good results are achieved even when using quite simple adaptation methods, we describe a procedure suitable for signals consisting mainly of transient and sinusoidal components. The adaptation measure proposed is based on onset detection, i.e. estimating where transients occur in the signal. The transform setting is what we call *scale frames*: the analysis procedure uses a single window prototype and a countable set of dilations thereof.

For evaluation, the representation quality is measured by comparison of the number of representation coefficients leading to certain root mean square (RMS) reconstruction errors, for both NSGT and regular Gabor transforms. The results are especially convincing for sparse music signals with high energy transient components. Other possible adaptation methods might be based on time-frequency concentration, sparsity or entropy measures [9],[31],[32].

Scale frames: In the following paragraphs, we propose a family of nonstationary Gabor frames that allows for exponential changes in time-frequency resolution along time positions. To avoid heavy notation and since the formalism necessary for the discrete, finite case could obscure the principal idea, we describe the continuous case construction. Suitable standard sampling then yields discrete, finite frames with equivalent characteristics.

The basic idea is to build a sequence of windows g_n from a single, continuous window prototype g with support on an interval of length 1 in such a way that the resulting g_n satisfy Corollary 1. The window sequence will be unambiguously determined by a sequence of scales. Once this *scale sequence* is known, it is a simple task to choose modulation parameters b_n satisfying the necessary conditions.

As a scale sequence, we allow any integer-valued sequence $\{s_n\}_{n\in\mathbb{Z}}$ such that $|s_n - s_{n-1}| \in \{0, 1\}$, where the latter restriction is set in order to avoid sudden changes of window length. Then, g_n is, up to translation, given by a dilation of the prototype g:

$$\mathbf{D}_{2^{s_n}}(g)(t) = \sqrt{2^{-s_n}}g(2^{-s_n}t)$$

This implies that a change of scale from one time step to the next corresponds to the use of a window either half or twice as long. More precisely, for every time step n, set $s = \min\{s_{n-1}, s_n\}$ and fix an overlap of $2/3 \cdot 2^s$, if $s_n \neq s_{n-1}$ and $1/3 \cdot 2^s$, if $s_n = s_{n-1}$. Explicitly,

$$g_n = \mathcal{T}_n \mathbf{D}_{2^{s_n}}(g),$$

with recursively defined time shift operators $\mathcal{T}_n$ given by

$$\mathcal{T}_0 = \mathbf{T}_0, \mathcal{T}_n = \begin{cases} \mathbf{T}_{2^s 5/6}\mathcal{T}_{n-1}, & \text{if } s_n \neq s_{n-1} \\ \mathbf{T}_{2^{s+1}/3}\mathcal{T}_{n-1}, & \text{else.} \end{cases}$$

Defining the time shifts in this manner, we achieve exactly the desired overlap as illustrated in Figure 5.

By construction, each g_n has non-zero overlap with its neighbors g_{n-1} and g_{n+1} and at any point on the real line, at most two windows are non-zero. After performing a preliminary transient detection step, as explained before, the construction of the adapted frame reduces to the determination of a scale sequence.

In the subsequent figures and experiments we used the Hann window as prototype, but other window choices are possible. The described concept can

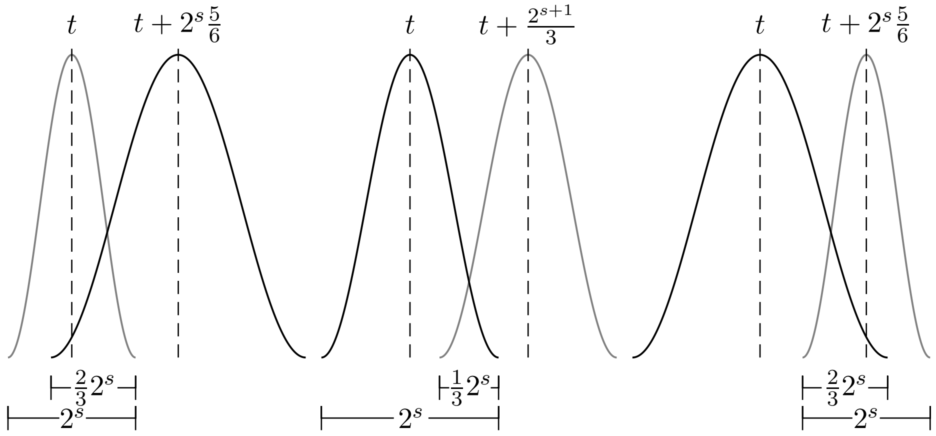

Figure 5: Illustration of scale frame overlaps and time shifts.

easily be generalized by admitting other overlap factors and scaling ratio than the ones specified above. The parameters have to be chosen with some care, though. Otherwise the resulting frames might be badly conditioned, with a big or even infinite condition number $\frac{B}{A}$, caused by accumulation points for the time shifts or gaps between windows. A more detailed description of general and discrete scale frames is beyond the scope of this article and will be part of a future contribution.

Frame construction from a sequence of onsets: In this paragraph, we assume that the signals of interest are mainly comprised of transient and sinusoidal components, an assumption met, e.g. by piano music. The instant a piano key is hit corresponds to a percussive, transient sound event, directly followed by harmonic components, concentrated in frequency. An intuitive adaptation to signals of this type would use high time resolution at the positions of transients. This corresponds to applying minimal scale at the transients and steadily increasing the scale with the distance from the closest transient. The transients' positions can be determined, e.g. by so-called onset detection procedures [33] which, if used carefully, work to a high degree of accuracy. Once the transient positions are known, the construction of a corresponding scale frame yields good nonstationary representations for sufficiently sparse signals.

Application of onset-based scale frames: We applied the procedure proposed above to various signals, mainly piano music. For this presentation, we selected three examples, all of them sampled at 44.1 kHz and consisting of a single channel. Some more examples and corresponding results as well as the source sound files can be found on the associated web-page `http://univie.ac.at/nonstatgab/`.

- Example 1: The widely used Glockenspiel signal shown in Figure 3.
- Example 2: An excerpt from a solo jazz piano piece performed by Herbie Hancock, characterized by its calmness and varied rhythmical pattern, resulting in irregularly spaced low-energy transients. See Figure 6.
- Example 3: A short excerpt of György Ligeti's piano concert. With highly percussive onsets in the piano and Glockenspiel voices and some orchestral background, this is the most polyphonic of our examples. See Figure 7.

For comparison, the plots in Figures 3, 6 and 7 also show standard Gabor coefficients with comparable (average) window overlap. A Hann window of 2560 samples length was chosen for the computation of regular Gabor transforms. The comparison shows that for the three signals, the NSGT features a better concentration of transient energy than a regular Gabor transform, while keeping, or even improving, frequency resolution.

Efficiency in sparse reconstruction: The onset detection procedure and a subsequent scale frame analysis were applied, along with a regular Gabor decomposition, to the Glockenspiel and Ligeti signals. As a test of the representations' sparsity, the signals were synthesized from their corresponding coefficients, modified by hard thresholding followed by reconstruction using the canonical dual frame. Then the numbers of largest magnitude coefficients needed for a certain relative root mean square (RMS) reconstruction error for each representation were compared. The RMS error of a vector f and its reconstruction f_{rec} is given by

$$RMS(f, f_{rec}) = \sqrt{\frac{\sum_{k=0}^{L-1} |f[k] - f_{rec}[k]|^2}{\sum_{k=0}^{L-1} |f[k]|^2}}.$$

All transforms are of redundancy about $\frac{5}{3}$. The results for NSGT and different regular Gabor transform schemes are listed in Figure 8. On the Glockenspiel signal the NSGT method performs vastly better than the ordinary

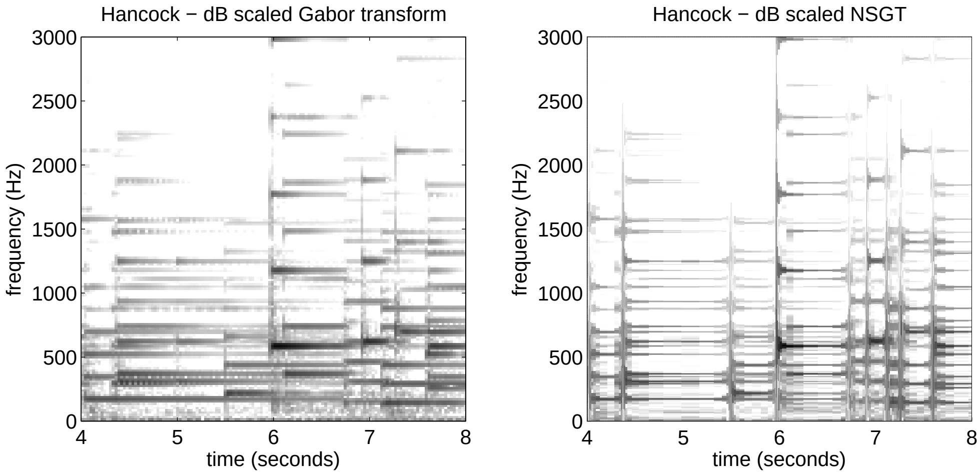

Figure 6: Hancock (Example 2). Regular and nonstationary Gabor representations.

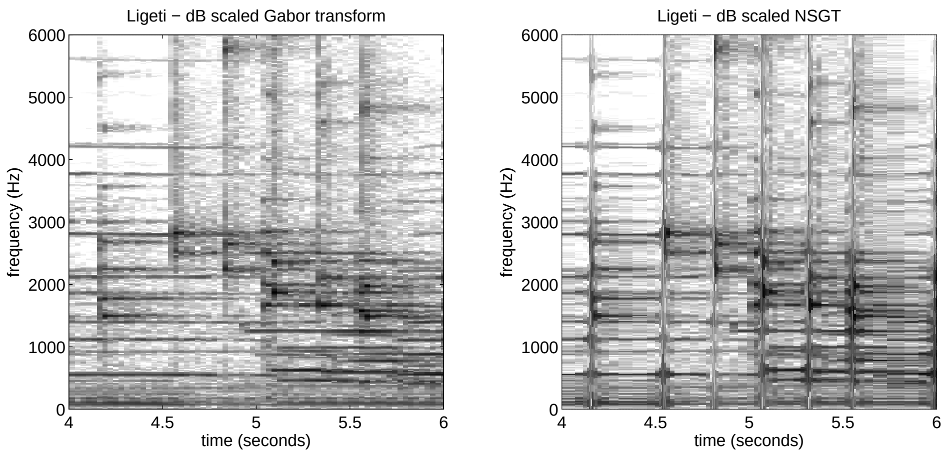

Figure 7: Ligeti (Example 3). Regular and nonstationary Gabor representations.

Gabor transform. For Ligeti, the differences are not as significant, but still the NSGT-based procedure shows better overall results.

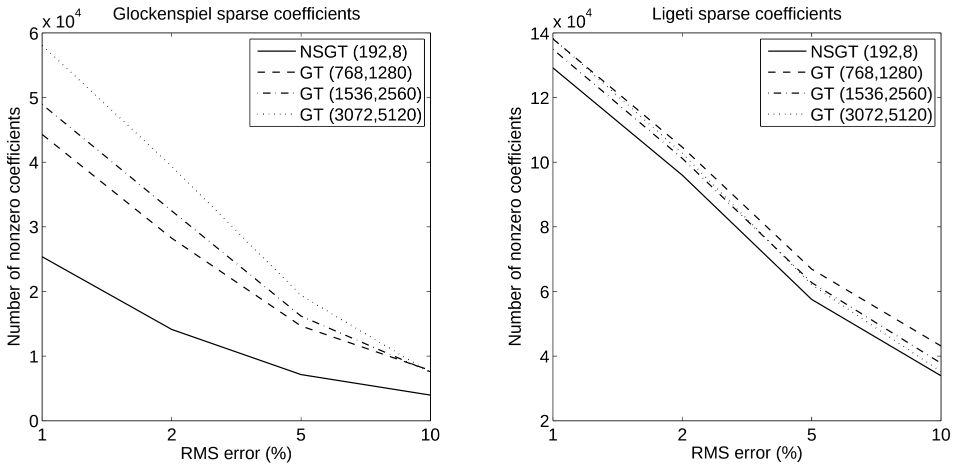

Figure 8: RMS error in sparse representations of Example 1 and Example 3. Parameters (in parentheses) are hop size and window length in the regular case (GT) or shortest window length and number of scales for the nonstationary case (NSGT). The values are estimated to be the optimal numbers of coefficients necessary to achieve reconstruction with less than the respective error.

Further experiments and a more exhaustive discussion of the parameters used in the experiments, can be found on the web-page `http://univie.ac.at/nonstatgab/`. Along them, examples of regular and nonstationary reconstructions from a specified amount of coefficients can be found, so the reader might get a subjective impression of perceptive reconstruction quality. In conclusion, the experiments show that for real music signals, NSGT can provide a sparser representation than regular Gabor transforms, admitting reasonable reconstruction error.

4.2. Implementation of a discrete, frequency-adaptive Gabor Transform

Since our construction of Gabor frames with adaptivity in the frequency domain relies on the fact that analysis windows h_m possess compact bandwidth, an FFT-based implementation is highly efficient. We take the input signal's Fourier transform and treat the procedure in complete analogy to the situation developed for time-adaptive transforms, i.e. $h_{m,n}[k] = \mathbf{T}_{na_m} h_m[k]$ and $\widehat{h_{m,n}}[j] = \mathbf{M}_{-na_m} \widehat{h_m}[j]$.

As observed in Section 3.2.1, we are able to obtain wavelet frames using Gabor frames that exhibit nonstationarity in the frequency domain. Moreover, we may design general transforms with flexible frequency resolution, such as a constant-Q transform. While various other adjustments (e.g. Mel- or Bark-scaled transforms) are feasible, we will focus our discussion on the constant-Q case. To the best knowledge of the authors, the approach to implement the constant-Q transform directly in the frequency domain by means of FFT is new in audio processing.

Remark 5. Note that for real-valued signals the symmetry of their FFT can be exploited to further reduce the computational effort. We particularly refer to the LTFAT routines `filterbankrealdual.m` and `filterbankrealtight.m`.

4.2.1. Application: an invertible constant-Q transform

The constant-Q transform (CQT), introduced by Brown [34], transforms a time signal into the time-frequency domain, where the center frequencies of the frequency bins are geometrically spaced. Since the *Q-factor* (the ratio of the center frequencies to the window's bandwidth) is constant, the representation allows for a better frequency resolution at lower frequencies and a better time resolution at the higher frequencies. This is sometimes preferable to the fixed resolution of the standard Gabor transform, for which the frequency bins are linearly spaced. In particular, this kind of resolution is often desired in the analysis of musical signals, since the transform can be set to coincide the temperament, e.g. semitone or quarter tone, used in Western music.

The originally introduced constant-Q transform, however, is not invertible and is computationally more intensive than the DFT. A computationally more efficient approach was presented in the sequel [3]: for the nth time slice of the signal f, the coefficient vector $c_{m,n}$, equal to inner product of the signal f with the *time*-limited window $h_{m,n}$ is computed in the Fourier side via $\langle \hat{f}, \widehat{h_{m,n}} \rangle$. This approximate computation takes advantage of the sparsity of the frequency domain kernel or *spectral kernel*. In contrast, we compute the coefficient vector for each frequency bin, making use of *band*-limited window functions.

Perfect reconstruction wavelet transforms with rational dilation factors were proposed in [35]. Since they are based on iterated filter banks, these methods are computationally too expensive for long, real-life signals, when high Q-factors, such as 12-96 bins per octave, are required.

In [30], Klapuri and Schörkhuber presented a computation of the CQT that shows improved efficiency and flexibility compared to the method proposed in [3], among others. However, the approximate inversion introduced in [30] still gives an RMS error of around 10^{-3}. 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. By allowing adaptive resolution in frequency, we can construct an invertible nonstationary Gabor transform with a constant Q-factor on the relevant frequency bins.

Setting: For the frame elements in the transform, we consider functions $h_m \in \mathbb{C}^L$, $m = 1, \ldots, M$ having center frequencies (in Hz) at $\xi_m = \xi_{\min} 2^{\frac{m-1}{B}}$, as in the CQT. Here, B is the number of frequency bins per octave, and $\xi_{\min}$ and $\xi_{\max}$ are the desired minimum and maximum frequencies, respectively. In the experiments, we restrict $\xi_{\max}$ to be less than the Nyquist frequency and there should exist an $M \in \mathbb{N}$ satisfying $\xi_{\max} \leq \xi_{\min} 2^{\frac{M-1}{B}} < \xi_s/2$, where ξ_s denotes the sampling frequency. In this case, we take $M = \lceil B \log_2(\xi_{\max}/\xi_{\min}) + 1 \rceil$, where $\lceil z \rceil$ is the smallest integer greater than or equal to z. While in the CQT no 0-frequency is present, the NSGT provides all necessary freedom to use additional center frequencies. Since the signals of interest are real-valued, we put filters at center frequencies beyond the Nyquist frequency in a symmetric manner. This results in the following values for the center frequencies:

$$\xi_m = \begin{cases} 0, & m = 0 \\ \xi_{\min} 2^{\frac{m-1}{B}}, & m = 1, \ldots, M \\ \xi_s/2, & m = M+1 \\ \xi_s - \xi_{2M+2-m}, & m = M+2, \ldots, 2M+1. \end{cases}$$

For the corresponding bandwidth Ω_m of h_m, we set $\Omega_m = \xi_{m+1} - \xi_{m-1}$, for $m = 1, \ldots, M$, and $\Omega_0 = 2\xi_1 = 2\xi_{\min}$. By construction, these result in a constant Q-factor $Q = (2^{\frac{1}{B}} - 2^{-\frac{1}{B}})^{-1}$ for $m = 2, \ldots, M-1$. And we can write each Ω_m as follows:

$$\Omega_m = \begin{cases} 2\xi_{\min}, & m = 0 \\ \xi_2, & m = 1, 2M+1 \\ \xi_m/Q, & m = 2, \dots, M-1 \\ (\xi_s - 2\xi_{M-1})/2, & m = M, M+2 \\ \xi_s - 2\xi_M, & m = M+1 \\ \xi_{2M+2-m}/Q, & m = M+3, \dots, 2M. \end{cases}$$

If we use a Hann window $\hat{h}$, supported on $[-1/2, 1/2]$, then we can obtain each h_m via $\widehat{h_m}[j] = \hat{h}((j\frac{\xi_s}{L} - \xi_m)/\Omega_m)$, where $j = 0, \dots, L-1$. Letting $a_m \leq \frac{\xi_s}{\Omega_m}$, we define $h_{m,n}$ by their Fourier transform $\widehat{h_{m,n}} = \mathbf{M}_{-na_m}\widehat{h_m}$, $n = 0, \dots, \lfloor \frac{L}{a_m} \rfloor - 1$. Figure 9 illustrates the time-frequency sampling grid of the set-up, where the center frequencies are geometrically spaced and sampling points regularly spaced.

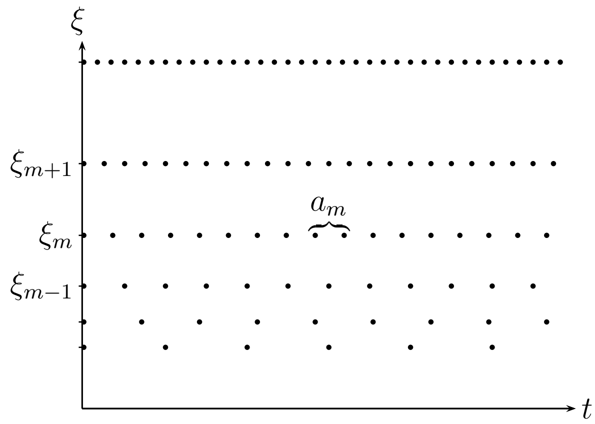

Figure 9: Exemplary sampling grid of the time-frequency plane for a constant-Q nonstationary Gabor system.

The support conditions on $\widehat{h_m}$ imply that the sum $\sigma = \sum_{m=0}^{2M+1} \frac{L}{a_m} \left|\widehat{h_m}\right|^2$ is finite and bounded away from 0. From Section 3.3, the frame operator is therefore invertible and we can apply Corollary 2.

Note that we consider the bandwidth to be the support of the window in frequency. This makes sense in the considered painless case. Very often, see e.g. [30], the bandwidth is taken as the width between the points, where the filter response drops to half of the maximum, i.e. the -3dB -bandwidth. This definition would also make sense in a non-compactly supported case. For the chosen filters, Hann windows, the Q-factor considering the -3dB -bandwidth is just double of the one considered above.

We see in Figure 10 the standard Gabor transform spectrogram and the constant-Q NSGT spectrogram of the Glockenspiel signal, the latter being very similar to the CQT spectrogram obtained from the original algorithm [34] but with the additional property that the signal can be perfectly reconstructed from the coefficients. Figures 11 and 12 compare the standard Gabor transform spectrogram and the constant-Q NSGT spectrogram of two additional test signals, both sampled at 44.1 kHz:

- Example 4: A recording of Bach's Little Fugue in G Minor, BWV578 performed by Christopher Herrick on a pipe organ. Low frequency noise and the characteristic structure of pipe organ notes are resolved very well by a CQT. See Figure 11.

- Example 5: An excerpt from a duet between violin and piano. Written by John Zorn and performed by Sylvie Courvoisier and Mark Feldman, the sample is made up of three short segments: A frantic sequence of violin and piano notes, a slow violin melody with piano backing and an inharmonic part with chirp component. See Figure 12.

Efficiency: The computation time of the nonstationary Gabor transform was found to be better than a recent fast CQT implementation [30], as seen in Figure 13. The two plots show mean values for computation time in seconds and the corresponding variance over 50 iterations, with varying window lengths and number of frequency bins, respectively. The outlier, drawn in gray, in Figure 13 (left) at the prime number 600569 illustrates dependence of the current CQ-NSGT implementation on the signal length's prime factor structure, analogous to FFT.

It is again reasonable to assume that the number of filters is bounded, independently of L, while the number of temporal points depend on L. As the role of M and N is switched in the assumption in Section 4.1.1 for the complexity, we arrive at a complexity of $\mathcal{O}(L \log L)$. This is also the complexity of the FFT of the whole signal. So the overall complexity of the frequency-dependent nonstationary Gabor transform is $\mathcal{O}(L \log L)$. The advantage of the method in terms of computational efficiency thus decreases as longer signals are considered.

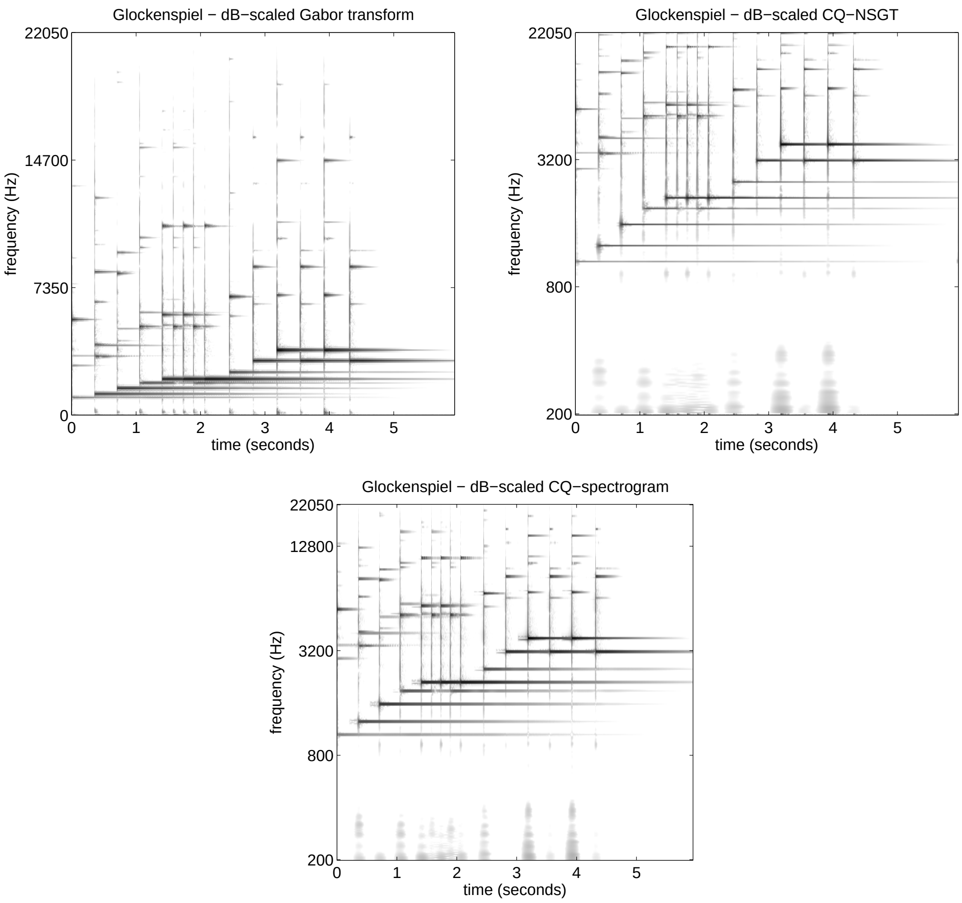

Figure 10: Glockenspiel (Example 1). Regular Gabor, constant-Q nonstationary Gabor and constant-Q representations of the signal. The transform parameters were $B = 48$ and $\xi_{\min} = 200$ Hz.

We note that at this point, since the windows used are band-limited, the current procedure is not suitable for real-time processing, despite its efficiency. The next step would be to process the incoming samples in a piecewise manner, using only a single family of frame elements for signals of arbitrary length. This entails working on finite, discrete parts of the given signal, thus considering the Fourier-transformed versions of vectors $f \cdot h \in \mathbb{C}^{\mathcal{L}}$, where h denotes some function of length $\mathcal{L} \ll L$. This window, together with the frame elements, will be designed to minimize undesired effects that stem from the cutting of the signal. Details of this piecewise processing, as well

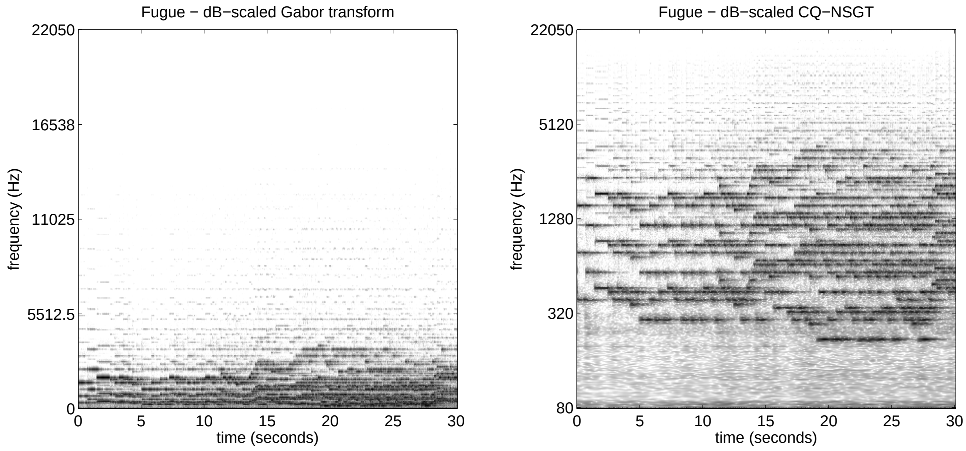

Figure 11: Bach's Little Fugue (Example 4). Regular and constant-Q nonstationary Gabor representations of the signal. The transform parameters were $B = 48$ and $\xi_{\min} = 75$ Hz.

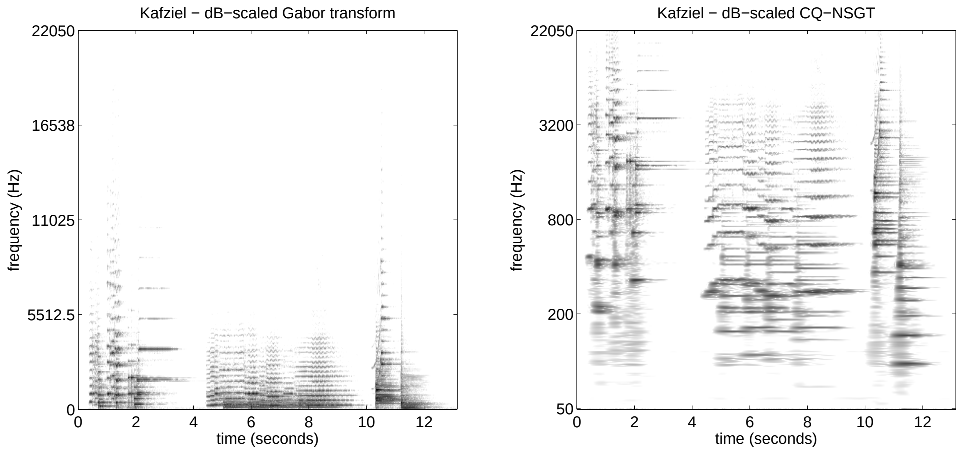

Figure 12: Violin and piano duet (Example 5). Regular and constant-Q nonstationary Gabor representations of the signal. The transform parameters were $B = 48$ and $\xi_{\min} = 50$ Hz.

as a proposed *variable-Q* transform, will be further discussed in a future contribution.

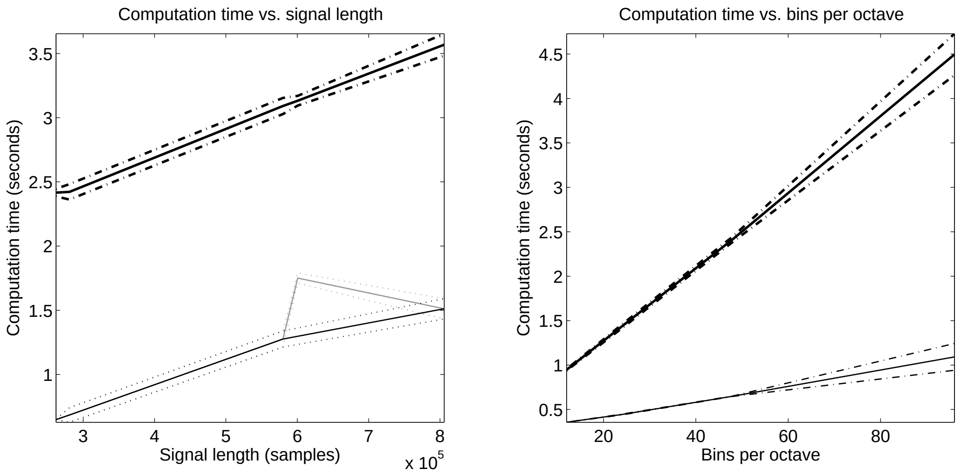

Figure 13: Comparison of computation time of CQT (top curves) and NSGT (bottom curves). The figure on the left shows the computation times for signals of various lengths with the number of bins per octave fixed at $B = 48$, while the figure on the right shows the computation times for the Glockenspiel signal, varying the number of bins per octave. In both figures, the solid lines represent the mean time (in seconds) and the dashed or dotted lines signify the mean time with corresponding variance. The lower left curve also shows gray solid lines indicating an outlier. The minimum frequency for all cases $\xi_{\min}$ was chosen at 50 Hz.

5. Conclusion and perspectives

Our approach enables the construction of frames with flexible evolution of time-frequency resolution over time or frequency. The resulting frames are well suited for applications as they can be implemented using fast algorithms, at a computational cost close to standard Gabor frames.

Exploiting evolution of resolution over time, the proposed approach can be of particular interest for applications where the frequency characteristics of the signal are known to evolve significantly with time. Order analysis [36], in which the signal analyzed is produced by a rotating machine having changing rotating speed, is an example of such an application.

Exploiting evolution of resolution over frequency, the presented approach is valuable for applications requiring the use of a tailored non uniform filter bank. In particular, it can be used to build filter banks following some perceptive frequency scale, see e.g. [5]. In the present contribution, we described in detail an invertible constant-Q transform.

One difficulty when using our approach is to adapt the time-frequency resolution to the evolution of the signal characteristics. If prior knowledge is available, this can be done by hand. An automatic adaptation algorithm based on onset detection was described in Section 4.1.2. A different approach will involve the investigation of sparsity criteria as proposed in [31]. Finally, future work will lead to adaptability in both time and frequency leading to *quilted frames* as introduced in [37].

Acknowledgment

This work was supported by the WWTF projects *Audiominer* (MA09-24) and *MulAc* (MA07-025), the Austrian Science Fund (FWF):[T384-N13] and [S10602-N13]. We thank two anonymous reviewers for their valuable comments and suggestions.

References

[1] H. G. Feichtinger, T. Strohmer, Gabor Analysis and Algorithms - Theory and Applications, Birkhäuser Boston, 1998.

[2] I. Daubechies, Ten Lectures On Wavelets, CBMS-NSF Regional Conference Series in Applied Mathematics, SIAM Philadelphia, 1992.

[3] J. C. Brown, M. S. Puckette, An efficient algorithm for the calculation of a constant Q transform, J. Acoust. Soc. Am. 92 (1992) 2698–2701.

[4] H. Bölcskei, F. Hlawatsch, H. G. Feichtinger, Frame-theoretic analysis of oversampled filter banks, IEEE Trans. Signal Process. 46 (1998) 3256–3268.

[5] W. M. Hartmann, Signals, Sounds, and Sensation, Springer, 1998.

[6] H. Malvar, Signal Processing with Lapped Transforms, Boston, MA: Artech House. xvi, 1992.

[7] E. Wesfreid, M. V. Wickerhauser, Adapted local trigonometric transforms and speech processing, IEEE Trans. Signal Process. 41 (1993) 3596–3600.

[8] K. Ramchandran, Z. Xiong, C. Herley, M. Orchard, Flexible tree-structured signal expansions using time-varying wavelet packets, IEEE Trans. Signal Process. 45 (1997) 233–245.

[9] D. Rudoy, B. Prabahan, P. Wolfe, Superposition frames for adaptive time-frequency analysis and fast reconstruction, IEEE Trans. Signal Process. 58 (2010) 2581–2596.

[10] I. Daubechies, A. Grossmann, Y. Meyer, Painless nonorthogonal expansions, J. Math. Phys. 27 (1986) 1271–1283.

[11] F. Jaillet, Représentation et traitement temps-fréquence des signaux audionumériques pour des applications de design sonore, Ph.D. thesis, Université de la Méditerranée Aix-Marseille II, 2005.

[12] F. Jaillet, P. Balazs, M. Dörfler, N. Engelputzeder, Nonstationary Gabor frames, in: Proceedings of the 8th International Conference on Sampling Theory and Applications (SampTA), 2009, pp. 227–230.

[13] P. Vandewalle, J. Kovačević, M. Vetterli, Reproducible research in signal processing, IEEE Signal Process. Mag. 26 (2009) 37–47.

[14] P. L. Søndergaard, Finite Discrete Gabor Analysis, Ph.D. thesis, Technical University of Denmark, Lyngby, 2007.

[15] P. L. Søndergaard, B. Torrésani, P. Balazs, The linear time frequency analysis toolbox, Int. J. Wavelets Multiresolut. Inf. Process. to appear (2011) accepted.

[16] R. J. Duffin, A. C. Schaeffer, A class of nonharmonic Fourier series, Trans. Amer. Math. Soc. 72 (1952) 341–366.

[17] P. G. Casazza, The art of frame theory, Taiwanese J. Math. 4 (2000) 129–202.

[18] O. Christensen, An Introduction To Frames And Riesz Bases, Birkhäuser, 2003.

[19] D. Gabor, Theory of communications, J. IEE III (1946) 429–457.

[20] K. Gröchenig, Foundations of Time-Frequency Analysis, Birkhäuser Boston, 2001.

[21] T. Strohmer, Numerical algorithms for discrete Gabor expansions, in: [1], pp. 267–294.

[22] P. G. Casazza, O. Christensen, Gabor frames over irregular lattices, Adv. Comput. Math. 18 (2003) 329–344.

[23] S. Mallat, A Wavelet Tour of Signal Processing, Academic Press London, 1998.

[24] K. Gröchenig, Foundations of Time-Frequency Analysis, Appl. Numer. Harmon. Anal., Birkhäuser Boston, 2001.

[25] K. Gröchenig, Localization of frames, Banach frames, and the invertibility of the frame operator, J. Fourier Anal. Appl. 10 (2004) 105–132.

[26] M. Ehler, On multivariate compactly supported bi-frames, J. Fourier Anal. Appl. 13 (2007) 511–532.

[27] M. Ehler, B. Han, Wavelet bi-frames with few generators from multivariate refinable functions, Appl. Comput. Harmon. Anal. 25 (2008) 407–414.

[28] M. Ehler, M. ehler: Nonlinear approximation associated with nonseparable wavelet bi-frames, J. Approx. Theory. 161 (2009) 292–313.

[29] M. Ehler, The multiresolution structure of pairs of dual wavelet frames for a pair of sobolev spaces, Jaen J. Approx. 2 (2010) 193–214.

[30] C. Schörkhuber, A. Klapuri, Constant-Q toolbox for music processing, in: Proceedings of the 7th Sound and Music Computing Conference (SMC), 2010.

[31] F. Jaillet, B. Torrésani, Time-frequency jigsaw puzzle: adaptive multiwindow and multilayered gabor representations, Int. J. Wavelets Multiresolut. Inf. Process. 5 (2007) 293–316.

[32] M. Liuni, A. Röbel, M. Romito, X. Rodet, Rényi information measures for spectral change detection, in: Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2011, pp. 3824 – 3827.

[33] S. Dixon, Onset Detection Revisited, in: Proceedings of the 9th International Conference on Digital Audio Effects (DAFx), 2006, pp. 133–137.

[34] J. Brown, Calculation of a constant Q spectral transform, J. Acoust. Soc. Amer. 89 (1991) 425–434.

[35] İ. Bayram, I. W. Selesnick, Frequency-domain design of overcomplete rational-dilation wavelet transforms, IEEE Trans. Signal Process. 57 (2009) 2957–2972.

[36] H. Shao, W. Jin, S. Qian, Order tracking by discrete Gabor expansion, IEEE Trans. Inst. Meas. 52 (2003) 754–761.

[37] M. Dörfler, Quilted Gabor frames - A new concept for adaptive time-frequency representation, Adv. Appl. Math. In Press, Corrected Proof, Doi: 10.1016/j.aam.2011.02.0 (2011).