CONSTRUCTING AN INVERTIBLE CONSTANT-Q TRANSFORM WITH NONSTATIONARY GABOR FRAMES

Gino Angelo Velasco[*†], *Nicki Holighaus*[*], *Monika Dörfler*[*], *Thomas Grill*[♯]

[*]NuHAG, Faculty of Mathematics, University of Vienna, Austria
[†]Institute of Mathematics, University of the Philippines, Diliman, Quezon City, Philippines
[♯]Austrian Research Institute for Artificial Intelligence (OFAI), Vienna, Austria
{gino.velasco,nicki.holighaus,monika.doerfler}@univie.ac.at,
thomas.grill@ofai.at

ABSTRACT

An efficient and perfectly invertible signal transform featuring a constant-Q frequency resolution is presented. The proposed approach is based on the idea of the recently introduced nonstationary Gabor frames. Exploiting the properties of the operator corresponding to a family of analysis atoms, this approach overcomes the problems of the classical implementations of constant-Q transforms, in particular, computational intensity and lack of invertibility. Perfect reconstruction is guaranteed by using an easy to calculate dual system in the synthesis step and computation time is kept low by applying FFT-based processing. The proposed method is applied to real-life signals and evaluated in comparison to a related approach, recently introduced specifically for audio signals.

1. INTRODUCTION

Many traditional signal transforms impose a regular spacing of frequency bins. In particular, Fourier transform based methods such as the *short-time Fourier transform* (STFT) lead to a frequency resolution that does not depend on frequency, but is constant over the whole frequency range. In contrast, the constant-Q transform (CQT), originally introduced by J. Brown [1, 2], features a frequency resolution dependent on the center frequencies of the windows used for each bin and the center frequencies of the frequency bins are not linearly, but geometrically spaced. In this sense, the principal idea of CQT is reminiscent of wavelet transforms, compare [3]: the Q-factor, i.e. the ratio of the center frequency to bandwidth is constant over all bins and thus the frequency resolution is better for low frequencies whereas time resolution improves with increasing frequency. However, the transform proposed in the original paper [1] is not invertible and does not rely on any concept of (orthonormal) bases. In fact, the number of bins used per octave is much higher than most traditional wavelet techniques would allow for. Furthermore, the computational efficiency of the original transform and its improved versions, [4], is insufficient.

CQTs rely on perception-based considerations, which is one of the reasons for their importance in the processing of speech and music signals. In these fields, the lack of invertibility of existing CQTs has become an important issue: for important applications such as masking of certain signal components or transposition of an entire signal or, again, some isolated signal components, the unbiased reconstruction from analysis coefficients is crucial. An interesting and promising approach to music processing with CQT was recently suggested in [5], also cf. references therein.

In the present contribution, we take a different point of view and consider both the implementation and inversion of a constant-Q transform in the context of the *nonstationary Gabor transform* (NSGT). Classical Gabor transform [6, 7] may be understood as STFT or sliding window transform. The generalization to NSGT was introduced in [8, 9] and allows for windows with flexible, adaptive bandwidths. Figure 1 shows an example of a classical sampled STFT (Gabor transform) and a CQ-NSGT of the same signal.

If the analysis windows are chosen appropriately, both analysis and reconstruction is realized efficiently with FFT-based methods. The original motivation for the introduction of NSGT was the desire to adapt both window size and sampling density in time, in order to resolve transient signal components more accurately. Here, we apply the same idea in frequency: we use windows with adaptive, compact bandwidth and choose the time-shift parameters dependent on the bandwidth of each window. The construction of the atoms, i.e. the shifted versions of the basic window functions used in the transform, is done directly in the frequency domain, see Section 2.2. This approach allows for efficient implementation using the FFT, as explained in Section 2.3. To exploit the efficiency of FFT, the signal of interest must be transformed into the frequency domain. For long real-life signals (e.g. signals longer than 10 seconds at a sampling rate of 44100Hz), processing is therefore done on consecutive time-slices, which is a natural processing step in real-time signal analysis [1]. The resolution of the proposed constant-Q type NSGT is identical to that of the CQT and perfect reconstruction is assured by relying on concepts from frame theory, which will be discussed next.

2. NONSTATIONARY GABOR FRAMES

Frames were first mentioned in [11], also see [12, 13]. Frames are a generalization of (orthonormal) bases and allow for redundancy and thus for much more flexibility in design of the signal representation. Thus, frames may be tailored to a specific application or certain requirements such as a constant-Q frequency resolution. Loosely speaking, we wish to expand, or represent, a given sig-

This work was supported by the Austrian Science Fund (FWF) project LOCATIF(T384-N13) and the WWTF projects MulAc (MA07-025) and Audio-Miner (MA09-024).

[1] If the time-slicing is done using smooth windows with an judiciously chosen amount of zero-padding, no undesired artifacts after modification of the analysis coefficients have to be expected. Mathematical details and error estimates will be given elsewhere, cf. [10]

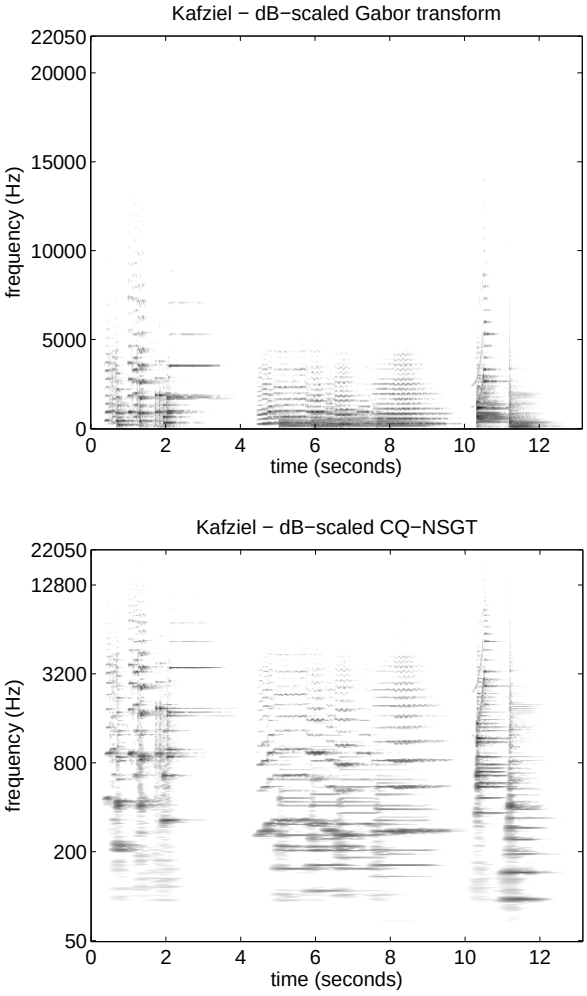

Figure 1: STFT and constant-Q nonstationary Gabor representations of a musical piece for violin and piano. The transform parameters were $B = 48$ and $\xi_{\text{min}} = 50$ Hz.

nal of interest as a linear combination of some building blocks or atoms $\varphi_{n,k}$, which are the members of our frame:

$$f = \sum_{n,k} c_{n,k}\varphi_{n,k} \tag{1}$$

for some coefficients $c_{n,k}$. The double indexes (n, k) allude to the fact that each atom has a certain location and concentration in time and frequency, compare Figure 2. Frame theory now allows us to determine, under which conditions an expansion (1) is possible and how coefficients leading to stable, perfect reconstruction may be determined.

We introduce the concept of frames for a Hilbert space $\mathcal{H}$. In a continuous setting, one may think of $\mathcal{H} = L^2(\mathbb{R})$, whereas we will choose $\mathcal{H} = \mathbb{C}^L$, L being the signal length, for describing the implementation.

2.1. Frames

Consider a collection of atoms $\varphi_{n,k} \in \mathcal{H}$ with $(n, k) \in \mathbb{Z} \times \mathbb{Z}$. Here, n may be thought of as a time index and k as an index related to frequency. We then define the frame operator S by

$$Sf = \sum_{n,k} \langle f, \varphi_{n,k}\rangle \varphi_{n,k},$$

for all $f \in \mathcal{H}$. Note that, if the set of functions $\{\varphi_{n,k}, (n, k) \in \mathbb{Z} \times \mathbb{Z}\}$ is an orthonormal basis, then S is the identity operator. If S is invertible on $\mathcal{H}$, then the collection $\{\varphi_{n,k}\}$, $(n, k) \in \mathbb{Z} \times \mathbb{Z}$ is a frame. In this case, we may define a *dual frame* by

$$\gamma_{n,k} = S^{-1}\varphi_{n,k}.$$

Then, reconstruction from the coefficients $c_{n,k} = \langle f, \varphi_{n,k}\rangle$ is possible:

$$f = S^{-1}Sf = \sum_{n,k} \langle f, \varphi_{n,k}\rangle S^{-1}\varphi_{n,k} = \sum_{n,k} c_{n,k}\gamma_{n,k}.$$

2.2. The Case of Painless Nonstationarity

In a general setting, the inversion of the operator S poses a problem in numerical realization of frame analysis. However, it was shown in [14], that under certain conditions, usually fulfilled in practical applications, S is diagonal. This situation of *painless non-orthogonal expansions* can now be generalized to allowing for adaptive resolution. Adaptive time-resolution was described in [8, 9], and here we turn to *adaptivity in frequency* in the same manner.

In the sequel, let T_x denote a time-shift by x, M_w denote a frequency shift (or modulation) by w and $\mathcal{F}f = \hat{f}$ the Fourier transform of f. Let φ_k, $k \in \mathbb{Z}$, be band-limited windows, well-localized in time, whose Fourier transforms $\psi_k = \hat{\varphi}_k$ are centered around possibly irregularly (or, e.g. geometrically) spaced frequency points ξ_k.

Then, we choose frequency dependent time-shift parameters (hop-sizes) a_k as follows: if the support of $\hat{\varphi}_k$ is contained in an interval of length $|\mathcal{I}_k|$, then we choose a_k such that

$$a_k \leq \frac{1}{|\mathcal{I}_k|} \quad \text{for all } k.$$

In other words, the time-sampling points have to be chosen dense enough to guarantee this condition. Finally, we obtain the frame members by setting

$$\varphi_{n,k} = T_{na_k}\varphi_k.$$

Under these conditions on the windows φ_k and the hop-sizes a_k, the frame operator is diagonal in the Fourier domain: since, by unitarity of the Fourier transform [15] and the Walnut representation of the frame operator [16], we have

$$\begin{aligned}\langle Sf, f\rangle =& \sum_{n,k} |\langle f, T_{na_k}\varphi_k\rangle|^2 = \sum_{n,k} |\langle \hat{f}, M_{-na_k}\hat{\varphi}_k\rangle|^2 \\ =& \langle \sum_k \frac{1}{a_k}|\hat{\varphi}_k|^2 \hat{f}, \hat{f}\rangle,\end{aligned}$$

the frame operator assumes the following form:

$$Sf = \mathcal{F}^{-1}\left(\sum_k \frac{1}{a_k}|\hat{\varphi}_k|^2 \hat{f}\right). \tag{2}$$

See [17, 14, 18] for detailed proofs of the diagonality of the frame operator in the described setting. From (2), it follows immediately that the frame operator is invertible whenever

$$0 < A \leq \sum_k \frac{1}{a_k} |\hat{\varphi}_k|^2 \leq B < \infty \tag{3}$$

almost everywhere. In this case, the dual frame is given by the elements

$$\gamma_{n,k} = T_{na_k} \big[\mathcal{F}^{-1} (\hat{\varphi}_k \big/ \sum_l \frac{1}{a_l} |\hat{\varphi}_l|^2) \big].$$

2.3. Realization in the Frequency domain

Based on the implementation of nonstationary Gabor frames performing adaptivity in the time domain [9], the above framework permits a fast realization by considering the Fourier transform of the input signal. The transform coefficients $c_{n,k} = \langle f, \varphi_{n,k} \rangle$ take the form

$$c_{n,k} = \langle f, T_{na_k} \varphi_k \rangle = \langle \hat{f}, M_{-na_k} \hat{\varphi}_k \rangle,$$

and can be calculated, for each k, with an FFT of length determined by the support of $\psi_k = \hat{\varphi}_k$. Similarly, reconstruction is realized by applying the dual windows $\hat{\gamma}_k = \hat{\varphi}_k \big/ \sum_l \frac{1}{a_l} |\hat{\varphi}_l|^2$ in a simple overlap-add process:

$$\hat{f} = \sum_{n,k} \langle \hat{f}, M_{-na_k} \hat{\varphi}_k \rangle M_{-na_k} \hat{\gamma}_k. \tag{4}$$

3. THE CQ-NSGT PARAMETERS: WINDOWS AND LATTICES

We will now describe in detail the parameters involved in the design of a nonstationary Gabor transform with constant-Q frequency resolution.

The CQT in [1] depends on the following parameters: the window functions, the number of frequency bins per octave, the minimum and maximum frequencies. These parameters determine the Q-factor, which is, as mentioned before, the ratio of the center frequency to the bandwidth. Here, the Q-factor is desired to be constant for all the relevant bins.

Let B and $\xi_{\min}$ denote the number of frequency bins per octave and the desired minimum frequency, respectively. For the proposed constant-Q nonstationary Gabor transform (CQ-NSGT), we consider band-limited window functions $\varphi_k \in \mathbb{C}^L$, $k = 1, \ldots, K$, with center frequencies ξ_k (in Hz) satisfying $\xi_k = \xi_{\min} 2^{\frac{k-1}{B}}$, as in the classical CQT. The maximum frequency $\xi_{\max}$ is restricted to be less than the Nyquist frequency $\xi_s/2$, where ξ_s denotes the sampling frequency. Further, we require the existence of an index K such that $\xi_{\max} \leq \xi_K < \xi_s/2$. We may set $K = \lceil B \log_2(\xi_{\max}/\xi_{\min}) + 1 \rceil$, with $\lceil z \rceil$ denoting the smallest integer greater than or equal to z.

Note that in the CQT, since the frequency spacing in the CQT is geometric, no 0-frequency is present and some high frequency content might not be represented. In the CQ-NSGT, however, there is freedom to use additional center frequencies, at negligible computational cost, to guarantee perfect reconstruction.

In our current implementation, tailored to (real) audio signals, we consider some symmetry in the frequency domain, and take the following values for the frequency-centers ξ_k:

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

The bandwidth Ω_k (the support of the window in frequency) of φ_k is set to be $\Omega_k = \xi_{k+1} - \xi_{k-1}$, for $k = 2, \ldots, K-1$, which leads to a constant Q-factor $Q = (2^{\frac{1}{B}} - 2^{-\frac{1}{B}})$. To obtain the same Q-factor on the relevant frequency bins, Ω_1 and Ω_K are therefore set to be ξ_1/Q and ξ_K/Q, respectively. Finally, we let $\Omega_0 = 2\xi_1 = 2\xi_{\min}$ and $\Omega_{K+1} = \xi_s - 2\xi_K$. In summary, we have the following values for Ω_k:

$$\Omega_k = \begin{cases} 2\xi_{\min}, & k = 0 \\ \xi_k/Q, & k = 1, \ldots, K \\ \xi_s - 2\xi_K, & k = K+1 \\ \xi_{2K+2-k}/Q, & k = K+2, \ldots, 2K+1. \end{cases}$$

3.1. Window Choice: Satisfying the Frame Conditions

We now give the details on the windows φ_k to be used such that (3) and hence the frame property is fulfilled.

We use a Hann window $\hat{h}$ that is zero outside $[-1/2, 1/2]$, i.e. a standard Hann window centered at 0 with support of length 1. We obtain the atoms φ_k by translation and dilation of h: $\widehat{\varphi_k}[j] = \hat{h}((j\xi_s/L - \xi_k)/\Omega_k)$, $k = 1, \ldots, K$, $K+2, \ldots, 2K+1$, $j = 0 \ldots, L-1$.

For the windows corresponding to the 0 and Nyquist frequencies, we use a plateau-like function $\hat{\psi}$, e.g. a Tukey window. We obtain φ_0 and φ_{K+1} by setting $\widehat{\varphi_k}[j] = \hat{\psi}((j\xi_s/L - \xi_k)/\Omega_k)$, $k = 0, K+1$.

Now, for the collection of time-shifts of the constructed windows a_k, we require $a_k \leq \xi_s/\Omega_k$ in order to satisfy (3). The $\varphi_{n,k}$ are then given by their Fourier transforms as:

$$\widehat{\varphi_{n,k}} = \mathbf{M}_{-na_k} \widehat{\varphi_k},\ n = 0, \ldots, \lfloor \tfrac{L}{a_k} \rfloor - 1,$$

where $\lfloor z \rfloor$ denotes the largest integer less than or equal to z. Figure 2 illustrates the time-frequency sampling grid of the set-up with the sampling points taken geometrically over frequency and linearly over time. Given these parameters, the coefficients of the CQ-NSGT are of the form $c_{n,k} = \langle f, \varphi_{n,k} \rangle = \langle \hat{f}, \widehat{\varphi_{n,k}} \rangle$, $f \in \mathbb{C}^L$.

From the given support condition, the system $\{\widehat{\varphi_k}\}_k$ has an overlap factor of around $1/2$. This implies that for the case where $a_k = \xi_s/\Omega_k$, the redundancy of the system is approximately 2.

By construction, the sum $\sigma = \sum_{m=0}^{2K+1} \frac{L}{a_k} \big|\widehat{\varphi_k}\big|^2$ is finite and bounded away from 0. From Sections 2.2 and 2.3, the frame operator is invertible. Hence, perfect reconstruction of the signal is obtained from the coefficients $c_{n,k}$ using the sum

$$f = \sum_{k=0}^{2K+1} \sum_{n=0}^{\lfloor \frac{L}{a_k} \rfloor - 1} c_{n,k} \gamma_{n,k},$$

where $\gamma_{n,k}$ denote the dual frame elements given by $\gamma_{n,k} = \mathbf{T}_{na_k} (\mathcal{F}^{-1}(\widehat{\varphi_k}/\sigma))$.

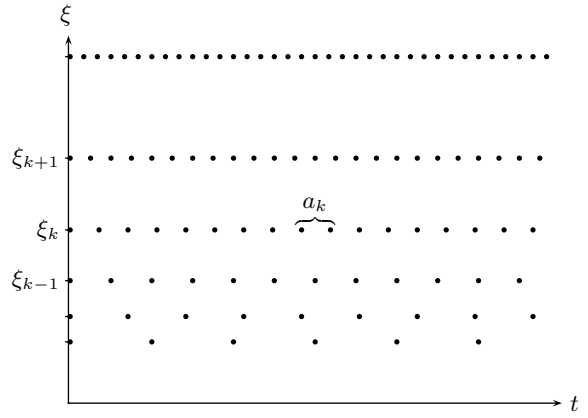

Figure 2: Examplary sampling grid of the time-frequency plane for a nonstationary Gabor system with resolution evolving over frequency.

signal length L	CQT mean time $\pm$ variance (in seconds)	CQ-NSGT mean time $\pm$ variance (in seconds)
262144	2.41 ± 0.03	0.64 ± 0.01
280789	2.42 ± 0.06	0.68 ± 0.06
579889	3.09 ± 0.06	1.28 ± 0.06
600569	3.13 ± 0.04	1.75 ± 0.04
805686	3.57 ± 0.09	1.51 ± 0.08

Table 1: Comparison of computation time between CQTs and CQ-NSGTs for signals of various lengths over 50 iterations. Parameters for all transforms were $B = 48$ and $\xi_{\text{min}} = 50$ Hz.

4. SIMULATIONS

We now present some experiments comparing the original CQT with the CQ-NSGT in terms of reconstruction error, computation time and (visual) representation of sound signals.

Technical framework: All 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 [5], available for free download at `http://www.elec.qmul.ac.uk/people/anssik/cqt/`. The CQ-NSGT algorithms are available at `http://univie.ac.at/nonstatgab/`.

For all experiments, ξ_{max} is taken to be $\xi_K = \xi_{\text{min}} 2^{\frac{K-1}{B}}$, where K is the largest integer such that $2\xi_K < \xi_s$.

4.1. Reconstruction Errors

The theoretical results, stating that the CQ-NSGT allows for perfect reconstruction, are confirmed by our experiments. For five test signals and various transform parameters, the relative reconstruction error

$$e_{\text{rec}} = \sqrt{\frac{\sum_{j=0}^{L-1} |f[j] - f_{\text{rec}}[j]|^2}{\sum_{j=0}^{L-1} |f[j]|^2}}$$

Bins per octave B	CQT mean time $\pm$ variance (in seconds)	CQ-NSGT mean time $\pm$ variance (in seconds)
12	0.95 ± 0.01	0.36 ± 0.00
24	1.44 ± 0.02	0.44 ± 0.00
48	2.42 ± 0.03	0.65 ± 0.00
96	4.50 ± 0.23	1.09 ± 0.15

Table 2: Comparison of computation time between CQTs and CQ-NSGTs of the Glockenspiel signals, varying the number of bins per octave. Values were obtained over 50 iterations. The minimum frequency ξ_{min} was chosen at 50 Hz.

was calculated. With ξ_{min} between 10 Hz and 130 Hz and B from 12 to 192, the largest reconstruction error of the CQ-NSGT algorithm was slightly smaller than $1.6 \cdot 10^{-15}$, perfect reconstruction up to numerical precision. For comparison, it was shown in [5] that a CQT with reasonable amounts of redundancy and bins per octave can be inverted with a relative error of 10^{-3}. This might not be enough for high-quality applications.

4.2. Computation Time and Computational Complexity

The required time for construction of the transform dictionary and computation of the corresponding coefficients was measured for audio signals of roughly 6 to 18 seconds length, at a sampling rate of 44.1 kHz. Each experiment was repeated 50 times, the results are listed in Table 1. We note that for all signals, the CQ-NSGT is faster than the CQT implementation proposed in [5] by a considerable factor.

Our approach is still of complexity $\mathcal{O}(L \log L)$,though, and the advantage over the CQT decreases for longer signals. Each frequency channel's time samples are acquired by means of sampled IFFT from the Fourier transform of the input signal, multiplied with the corresponding window. Therefore, a preliminary full length FFT is necessary. More explicitly, we assume $\widehat{\varphi_k}$ to have support of length M_k and we denote by N_k the corresponding IFFT length. Let $N = \max_k \{N_k\}$, i.e. the maximum IFFT-length, and we have $M_k \leq N_k \leq N$, since we only consider the painless case. Consequently, the number of operations is

1. FFT: $\mathcal{O}(L \cdot \log(L))$.
2. Windowing: M_k operations for the k-th window.
3. IFFT: $\mathcal{O}(N_k \cdot \log(N_k))$ for the k-th window.

The number of frequency channels $2K + 2$ is independent of L, since it is determined directly from the transform parameters. Thus, M_k and N_k are L-dependent and the computational complexity of the discrete CQ-NSGT is $\mathcal{O}(L \log L)$.

In applications, the dual windows are constructed directly on the frequency side and the painless case construction involves multiplication of the window functions by the inverse of a diagonal matrix, resulting in $\mathcal{O}(2\sum_{k=0}^{2K+1} M_k) = \mathcal{O}(L)$ operations. Finally, the inverse CQ-NSGT has numerical complexity $\mathcal{O}(L \cdot \log(L))$, since it entails computing for the FFT of each coefficient vector, multiplying with the corresponding dual windows and, after evaluating the sum, computing a length L IFFT.

We note that linear computation time may be achieved by processing the signal in a suitable piecewise manner. Some experiments on that matter have been conducted, but the details of this

procedure exceed the scope of this paper and are intended to be part of a future contribution [10].

In a second experiment, CQT and CQ-NSGT coefficients of the shortest sample, a Glockenspiel signal, were computed for several numbers of bins per octave. The results, listed in Table 2, illustrate that the advantage of the CQ-NSGT algorithm increases for large numbers of bins.

4.3. Visual Representation of Sound Signals

The spectral representation provided by CQT has several desirable properties, e.g. the logarithmic frequency scale resolves musical intervals in a similar way, independent of absolute frequencies. These properties are still present in the CQ-NSGT, in fact, its visual representation is practically identical to that of classical CQT as illustrated by Figure 3 for the exemplary case of the Glockenspiel signal. Figure 4 shows the CQ-NSGT of two additional music signals, further illustrating that even highly complex signals are nicely resolved by the proposed transform, similar to CQT.

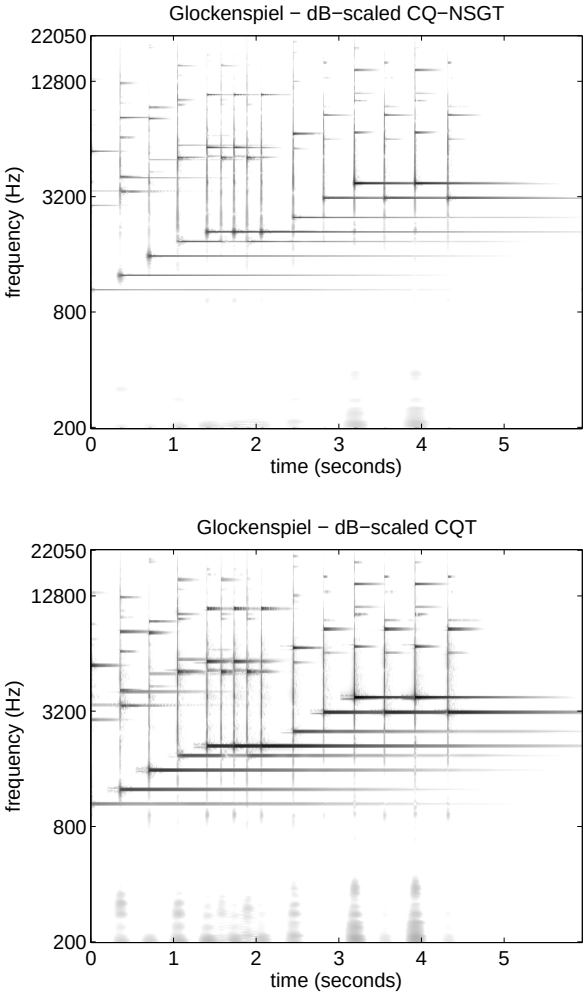

Figure 3: Nonstationary Gabor and constant-Q representations of the Glockenspiel signal. The transform parameters were $B = 48$ and $\xi_{\text{min}} = 200$ Hz.

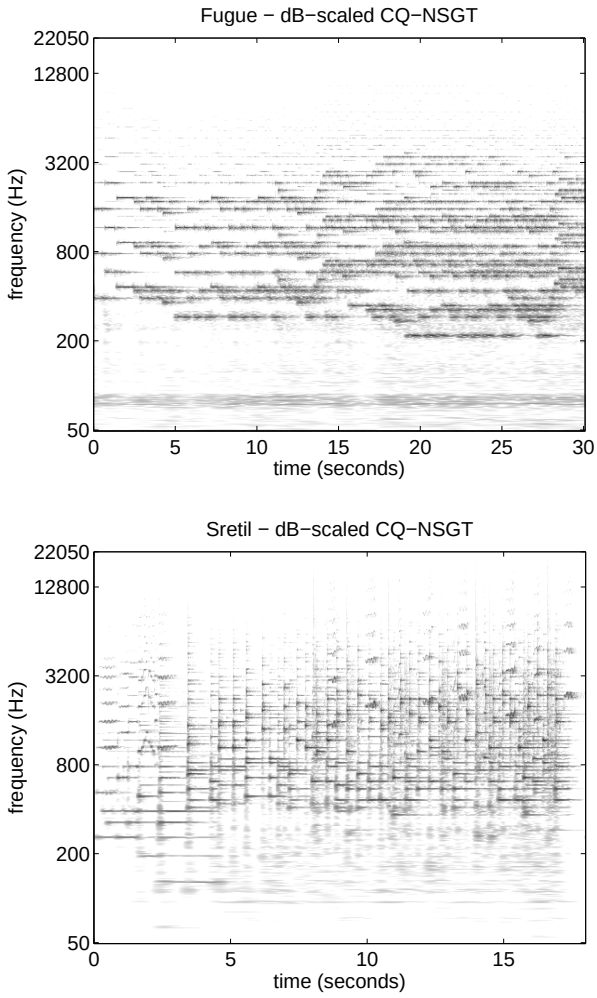

Figure 4: Nonstationary Gabor representations of ..... and a piano solo, respectively. The transform parameters were $B = 48$ and $\xi_{\text{min}} = 50$ Hz.

5. EXPERIMENTS ON APPLICATIONS

Our experiments show applications of the CQ-NSGT in musical contexts, where the property of a logarithmic frequency scale renders the method often superior to the traditional STFT. Corresponding sound examples can be found at `http://univie.ac.at/nonstatgab/cqt/`.

5.1. Masking

In the masking experiment, we show that the perfect reconstruction property of CQ-NSGT can be used to cut out components from a signal by directly modifying the time-frequency coefficients. The advantage of considerably higher spectral resolution at low frequencies (with a chosen application-specific temporal resolution at higher frequencies) compared to the STFT, makes the CQ-NSGT a very powerful, novel tool for masking or isolating time-frequency components of musical signals. Our example shows in Figure 5 a mask for extracting – or inversely, suppressing – a note from the Glockenspiel signal depicted in Figure 3. The mask was cre-

ated as a gray-scale bitmap using an ordinary image manipulation program and then resampled in order to conform to the irregular time-frequency grid of the CQ-NSGT. Figure 5 shows the mask spectrogram, along with the spectrograms of the synthesized, processed signal and remainder.

5.2. Transposition

A useful property of continuous constant-Q decompositions is the fact that the transposition of a harmonic structure, like a note including overtones, corresponds to a simple translation of the logarithmically scaled spectrum. Approximately, this is also the case for the finite, discrete CQ-NSGT. In this experiment, we transposed a piano chord simply by shifting the inner frequency bins accordingly. By inner frequency bins, we refer to all bins with constant Q-factor. This excludes the 0-frequency and Nyquist frequency bins. The onset portion of the signal has been damped, since inharmonic components, such as transients, produce audible artifacts when handled in this way. In Figure 6, we show spectrograms of the original and modified chords, shifted by 20 bins. This corresponds to a upwards transposition by 5 semitones.

6. SUMMARY AND PERSPECTIVES

We presented a constant-Q transform, based on nonstationary Gabor frames, that is computationally efficient and allows for perfect reconstruction. The described framework can easily be adapted to other perceptive frequency scales (e.g. mel or Bark scale) by choosing appropriate dictionaries.

The possibility of overcoming the difficulties that stem from dependence of the proposed transform on the signal length, e.g. by piecewise processing, is currently under investigation, cf. [10]. This will further reduce computational effort and enable the use of a single family of frame elements for signals of arbitrary length.

7. ACKNOWLEDGMENTS

Many thanks to the great number of anonymous reviewers!

8. REFERENCES

[1] Judith Brown, "Calculation of a constant Q spectral transform," *J. Acoust. Soc. Amer.*, vol. 89, no. 1, pp. 425–434, 1991.

[2] Stéphane Mallat, *A Wavelet Tour of Signal Processing: The Sparse Way*, Academic Press, 2009.

[3] İlker Bayram and Ivan W. Selesnick, "Frequency-domain design of overcomplete rational-dilation wavelet transforms," *IEEE Trans. Signal Process.*, vol. 57, no. 8, pp. 2957–2972, 2009.

[4] Judith C. Brown and Miller S. Puckette, "An efficient algorithm for the calculation of a constant Q transform," *J. Acoust. Soc. Am.*, vol. 92, no. 5, pp. 2698–2701, 1992.

[5] Christian Schörkhuber and Anssi Klapuri, "Constant-Q toolbox for music processing," in *Proceedings of SMC Conference 2010*, 2010.

[6] H. G. Feichtinger and T. Strohmer, *Gabor Analysis and Algorithms. Theory and Applications.*, Birkhäuser, 1998.

[7] Hans G. Feichtinger and Thomas Strohmer, *Advances in Gabor Analysis*, Birkhäuser, 2003.

[8] Florent Jaillet, *Représentation et traitement temps-fréquence des signaux audionumériques pour des applications de design sonore*, Ph.D. thesis, Université de Provence, 2005.

[9] F. Jaillet, P. Balazs, M. Dörfler, and N. Engelputzeder, "Nonstationary Gabor frames," in *SAMPTA'09, International Conference on SAMPling Theory and Applications*, 2009, pp. 227–230.

[10] M. Dörfler, N. Holighaus, and G.A. Velasco, "The sliced-Q transform," in preparation, 2011.

[11] R. J. Duffin and A. C. Schaeffer, "A class of nonharmonic Fourier series.," *Trans. Amer. Math. Soc.*, vol. 72, pp. 341–366, 1952.

[12] Amina Chebira and Jelena Kovacevic, "Life Beyond Bases: The Advent of Frames (Part I and II)," *IEEE Signal Processing Magazine*, vol. 24, no. 4 - 5, pp. 86–104, 115–125, 2007.

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

[14] Ingrid Daubechies, A. Grossmann, and Y. Meyer, "Painless nonorthogonal expansions," *J. Math. Phys.*, vol. 27, no. 5, pp. 1271–1283, May 1986.

[15] Walter Rudin, *Real and Complex Analysis (Third ed.)*, Singapore: McGraw Hill, 1987.

[16] D. F. Walnut, "Continuity properties of the Gabor frame operator," *J. Math. Anal. Appl.*, vol. 165, no. 2, pp. 479–504, 1992.

[17] M. Dörfler, "Time-frequency Analysis for Music Signals. A Mathematical Approach," *Journal of New Music Research*, vol. 30, no. 1, pp. 3–12, 2001.

[18] Peter Balazs, Monika Dörfler, Florent Jaillet, Nicki Holighaus, and Gino Angelo Velasco, "Theory, implementation and applications of nonstationary Gabor Frames," *preprint*, , no. submitted, 2011.

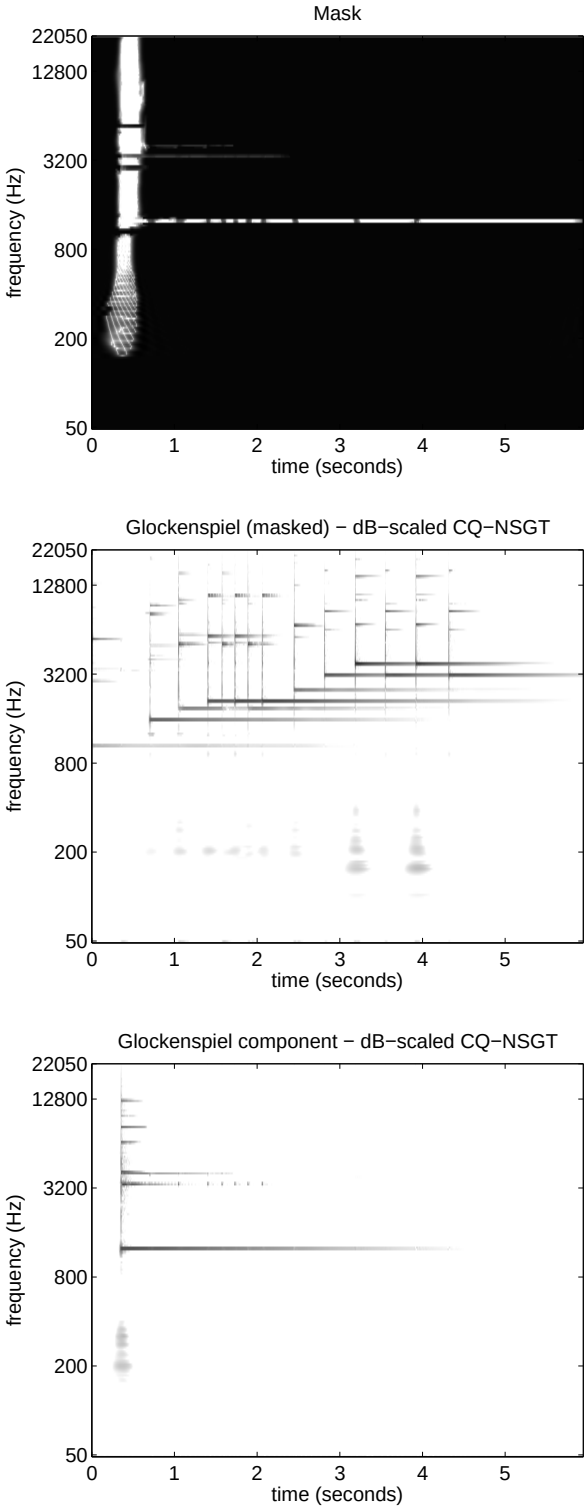

Figure 5: Note extraction from the Glockenspiel signal by masking. The CQ-NSGT coefficients of the Glockenspiel signal were weighted with the mask shown on top. The remaining signal and extracted component are depicted in the middle and bottom respectively. The transform parameters were $B = 24$ and $\xi_{\min} = 50$ Hz.

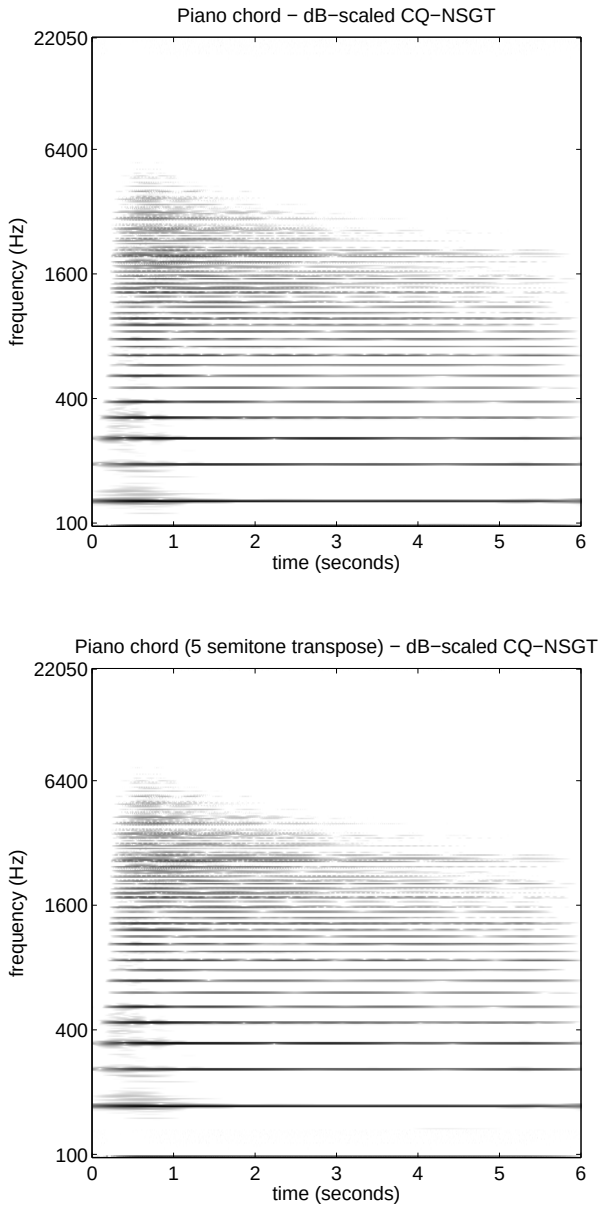

Figure 6: Piano chord signal and upwards transposition by 5 semitones, corresponding to a circular shift of the inner bins by 20. The transform parameters were $B = 48$ and $\xi_{\min} = 100$ Hz.