APPROXIMATIONS FOR ONLINE COMPUTATION OF REDRESSED FREQUENCY WARPED VOCODERS

Gianpaolo Evangelista

Institute for Composition and Electroacoustics
MDW University of Music and Performing Arts
Vienna, Austria
evangelista@mdw.ac.at

ABSTRACT

In recent work, the construction of non-uniform generalized Gabor frames for the time-frequency analysis of signals has been introduced. In particular, while preserving perfect reconstruction, these frames allow for tilings of the time-frequency plane with arbitrary allocation of partially overlapping frequency bands or time intervals.

In a recent paper, the author demonstrated that the construction of such frames can be entirely based on warping operators, which are specified by the required frequency or time warping maps, which, in turn, interpolate the desired frequency or time intervals edges. However, while the online computation of Gabor expansions on non-uniform time intervals presents little or no problem, the computation of Gabor expansions on non-uniform frequency bands requires knowledge of the Fourier transform of the entire signal, which precludes online computation.

In this paper we introduce approximations and ideas for the design of nearly perfect reconstruction analysis and synthesis atoms, which allow for the online computation of time-frequency representations on non-uniform frequency bands.

1. INTRODUCTION

Adapting time-frequency representations, such as the phase vocoder or Short-Time Fourier Transform (STFT), to features of the sound signals or to characteristics of perception, such as glissando, vibrato and 12-tone note system, is a desired goal in the analysis, synthesis and processing and in several contexts ranging from music information retrieval to transformations and special effects.

The STFT's uniform frequency bands can be transformed into non-uniform frequency bands by means of a frequency map, i.e. a monotonically increasing function remapping the frequency axis, as shown in Fig. 1 for adaptation to an equally tempered scale with a constant Q 1/3 octave band splitting.

In a similar way, non-uniform analysis time intervals can be allocated by remapping the time axis of the signal prior to performing uniform time-frequency analysis. The uniform analysis of the time warped signal achieves non-uniform time resolution.

Warping the signal prior to STFT analysis is equivalent to inverse warping the representative elements, i.e. the atoms of the representation.

However, being a time-shift dependent operation, frequency warping disrupts the time organization of signals. Uniform time-frequency analysis of the frequency warped signal results in a frequency dependent distortion of the time axis in the warped time-frequency representation. Similarly, the time warped time-frequency representation shows time dependent distortion of the frequency axis. Thus, warping one variable prior to uniform time-frequency analysis affects the conjugate variable in the representation plane.

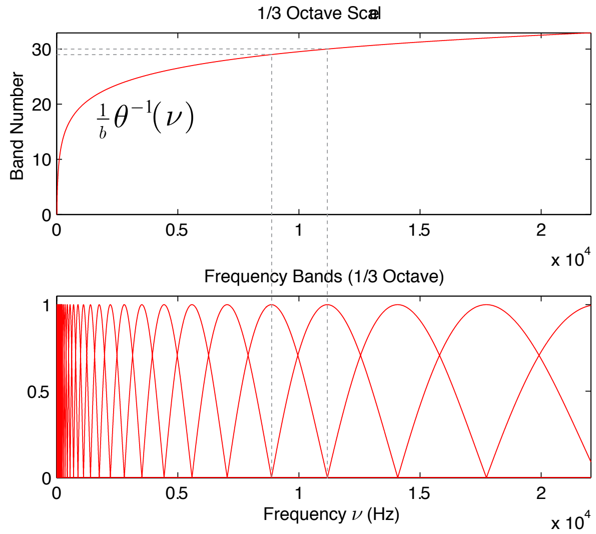

Figure 1: Frequency warping uniform frequency bands according to a $1/3$ of octave scale (top); resulting frequency band characteristics (bottom). Here b is the frequency shift in Hz of the original uniform bands.

In recent work [1, 2, 3], the problem of the construction of flexible frames that allow for arbitrary selection of the frequency bands of their atoms was addressed. In [3] it is shown that the required allocation of generalized Gabor atoms can be specified according to a frequency or time warping map. In [4] the STFT redressing method is introduced, which, with the use of additional warping in time-frequency, shows under which conditions one can have generalized Gabor frames. These conditions are dictated by the interaction of sampling in time-frequency and frequency or time warping operators.

The results in the previously cited work show that arbitrary allocation of the atoms is generally possible in the so called *painless case*, i.e. in the case of finite time support of the windows for arbitrary time interval allocation and of finite frequency support of the windows for arbitrary frequency band allocation.

Since online computation of the generalized Gabor analysis / synthesis is only possible with finite duration windows, the arbitrary frequency band allocation is not exactly feasible in applications that require real-time, while the arbitrary time interval allocation presents little or no problem.

In this paper, we address the problem of online computation of generalized Gabor analysis with arbitrary frequency band allocation, resorting to approximations that lead to near perfect reconstruction methods.

The paper is organized as follows. In Section 2 we review the concept of applying time and frequency warping to time-frequency representations derived from the continuous time Short-Time Fourier Transform, pointing out the problems introduced by dispersion and resolving them with the redressing method, which involves a further warping operations in the time-frequency domain. In Section 3 we apply the redressing method to frames, which allow for sampled time-frequency analysis and synthesis, and we provide the conditions by which the redressing of dispersion is exact. In Section 4 we introduce approximations suitable for the online computation of redressed frame expansions. In Section 5 we draw our conclusions.

2. REDRESSED WARPED TIME-FREQUENCY

In this section we review concepts that lead to the redressing method for the time alignment of the frequency warped Short-Time Fourier Transform (STFT).

In order to set the notation, the uniform STFT is obtained by applying the operator $\mathcal{S}$ to the signal s:

$$\begin{gathered}\left[\mathcal{S}s\right](\tau,\nu) = \langle s, h_{\tau,\nu}\rangle = \langle s, \mathbf{T}_\tau \mathbf{M}_\nu h_{0,0}\rangle = \\ \int_{-\infty}^{+\infty} s(t)\overline{h_{0,0}(t-\tau)}e^{-j2\pi\nu(t-\tau)}dt,\end{gathered} \tag{1}$$

where $\mathbf{T}_\tau s(t) = s(t-\tau)$ is the time-shift operator, $\mathbf{M}_\nu s(t) = e^{j2\pi\nu t}s(t)$ is the modulation operator and the overbar denotes complex conjugation. The operator $\mathcal{S}$ acts over time signals and the frequency ν is considered as a parameter. In (1), the analysis windows

$$h_{\tau,\nu}(t) = \left[\mathbf{T}_\tau \mathbf{M}_\nu h_{0,0}\right](t) = h_{0,0}(t-\tau)e^{j2\pi\nu(t-\tau)} \tag{2}$$

are modulated and shifted versions of a unique time window $h_{0,0}$. Their Fourier transforms are related to the Fourier transform of the original window $\hat{h}_{0,0}$ as follows:

$$\hat{h}_{\tau,\nu}(f) = \hat{h}_{0,0}(f-\nu)e^{-j2\pi f\tau}, \tag{3}$$

which are frequency shifted and modulated versions of the Fourier transform of the window $h_{0,0}$.

Since $\left[\mathcal{S}s\right](\tau,\nu) = s(\tau) * \overline{h_{0,\nu}(-\tau)}$, where the symbol $*$ denotes convolution, one can rewrite (1) in the frequency domain w.r.t. τ as follows:

$$\left[\widehat{\mathcal{S}s}\right](f,\nu) = \overline{\hat{h}_{0,\nu}(f)}\hat{s}(f) = \overline{\hat{h}_{0,0}(f-\nu)}\hat{s}(f). \tag{4}$$

Non-uniform time-frequency representations can be obtained from uniform ones via time and / or frequency warping, as discussed in Section 2.2, after we formally introduce warping operators in the next section.

2.1. Warped STFT

The warped STFT can be obtained by warping the signal prior to applying the STFT operator. The most general warping operator involves combined time-frequency warping, i.e. time dependent frequency warping or, equivalently, frequency dependent time warping. For the purpose of this paper we consider separable warping, which can be computed by cascading time invariant frequency warping with frequency independent time warping. We mostly focus on pure frequency warping.

A 1D warping operator performs a remapping of the abscissae, as obtained through function composition. A time warping operator $\mathbf{W}_\gamma$ is completely characterized by a function composition operator in the time domain:

$$s_{tw} = \mathbf{W}_\gamma s = s \circ \gamma, \tag{5}$$

where γ is the time warping map and s_{tw} is the time-warped signal. Similarly, a frequency warping operator $\mathbf{W}_{\tilde{\theta}}$ is completely characterized by a function composition operator $\mathbf{W}_\theta$ in the frequency domain:

$$\hat{s}_{fw} = \widehat{\mathbf{W}_{\tilde{\theta}}s} = \hat{\mathbf{W}}_{\tilde{\theta}}\hat{s} = \mathbf{W}_\theta \hat{s} = \hat{s}\circ\theta, \tag{6}$$

where θ is the frequency warping map, which transforms the Fourier transform $\hat{s} = \mathcal{F}s$ of a signal s into the Fourier transform $\hat{s}_{fw} = \mathcal{F}s_{fw}$ of another signal s_{fw}, where $\mathcal{F}$ is the Fourier transform operator and the hat over a symbol denotes the Fourier transformed quantity (signal or operator). We affix the $\tilde{}$ symbol over the map θ as a reminder that the map operates in the frequency domain. Accordingly, we have $\mathbf{W}_{\tilde{\theta}} = \mathcal{F}^{-1}\hat{\mathbf{W}}_{\tilde{\theta}}\mathcal{F} = \mathcal{F}^{-1}\mathbf{W}_\theta\mathcal{F}$.

If the warping map is one-to-one and almost everywhere differentiable then a unitary form of the warping operator can be defined by amplitude scaling, as given by the square root of the derivative of the map (dilation function). For example, a unitary frequency warping operator $\mathbf{U}_{\tilde{\theta}}$ has frequency domain action

$$\hat{s}_{fw}(\nu) = \left[\widehat{\mathbf{U}_{\tilde{\theta}}s}\right](\nu) = \sqrt{\left|\frac{d\theta}{d\nu}\right|}\hat{s}(\theta(\nu)), \tag{7}$$

where ν denotes frequency. We assume henceforth that all warping maps are almost everywhere increasing so that the magnitude sign can be dropped from the derivative under the square root.

2.2. Warped Time-Frequency Representations

Remapping signals prior to STFT allows for a reinterpretation of the representation elements: while the organization of the representation (tiling) remains the same, the elements capture different components of the signal. Time warping dilates / shrinks and displaces the characteristic analysis time intervals (resolution and centers) w.r.t. signals. Frequency warping remaps the characteristic analysis frequency bands w.r.t. signals (bandwidths and centers).

Given a frequency warping operator $\mathbf{W}_{\tilde{\theta}}$, the warped STFT is defined through the operator $\mathcal{S}_{\tilde{\theta}}$ as follows

$$\begin{gathered}\left[\mathcal{S}_{\tilde{\theta}}s\right](\tau,\nu) = \left[\mathcal{S}\mathbf{W}_{\tilde{\theta}}s\right](\tau,\nu) = \\ \langle \mathbf{W}_{\tilde{\theta}}s, h_{\tau,\nu}\rangle = \left\langle s, \mathbf{W}_{\tilde{\theta}}^\dagger h_{\tau,\nu}\right\rangle,\end{gathered} \tag{8}$$

which is indeed a warped version of (1), where $\mathbf{W}_{\tilde{\theta}}^\dagger$ is the adjoint of the warping operator. If the warping operator is unitary then we have $\mathbf{W}_{\tilde{\theta}}^\dagger = \mathbf{W}_{\tilde{\theta}}^{-1} = \mathbf{W}_{\tilde{\theta}^{-1}}$. In that case, warping the signal prior to STFT is perfectly equivalent to perform STFT analysis with inversely frequency warped windows. The warped STFT is unitarily equivalent to the STFT so that a number of properties concerning conditioning and reconstruction hold [5].

The Fourier transforms of the frequency warped STFT analysis elements are

$$\begin{gathered}\hat{\tilde{h}}_{\tau,\nu}(f) = \left[\mathbf{W}_{\tilde{\theta}^{-1}}\widehat{h_{\tau,\nu}}\right](f) = \\ \sqrt{\tfrac{d\theta^{-1}}{df}}\,\hat{h}_{0,0}(\theta^{-1}(f)-\nu)e^{-j2\pi\theta^{-1}(f)\tau},\end{gathered} \tag{9}$$

which shows how the analysis elements are obtained from frequency warped modulated windows centered at frequencies $f = \theta(\nu)$. The windows are time-shifted with dispersive delay, where the group delay is $\tau\frac{d\theta^{-1}}{df}$.

Frequency warping generally disrupts the time organization of signals. Indeed, the time-shift operator $\mathbf{T}_\tau$ does not commute with the frequency warping operator:

$$\left[\widehat{\mathbf{W}_{\tilde{\theta}}\mathbf{T}_\tau s}\right](\nu) = \left[\mathbf{W}_\theta\widehat{\mathbf{T}_\tau s}\right](\nu) = e^{-j2\pi\theta(\nu)\tau}\hat{s}(\theta(\nu)), \tag{10}$$

which is different from $\left[\widehat{\mathbf{T}_\tau\mathbf{W}_{\tilde{\theta}}s}\right](\nu) = e^{-j2\pi\nu\tau}\hat{s}(\theta(\nu))$, unless the map θ is the identity map. Thus, an event that starts at time T in the original signal, is dispersed into events starting at times $\phi_d(\nu)T$, where $\phi_d(\nu) = \theta(\nu)/\nu$ is the phase delay of the warping map, which depends on frequency unless the map is linear.

In the applications we would like to produce spectrograms with non-uniform time or frequency resolution but the dispersion introduced by warping results in misalignment and spreading of the time-frequency components in the conjugate variable of the warped one. In the next section we will show how further warping in the time-frequency plane can redress the warped representations.

2.3. Redressing the Warped STFT

To address the problem of realigning the frequency warped STFT $[\mathcal{S}_{\tilde{\theta}}s](\tau,\nu)$, consider its Fourier transform w.r.t. the time variable τ. This can be written in the form (4) by replacing the Fourier transform of the signal with that of the frequency warped signal:

$$\left[\widehat{\mathcal{S}\mathbf{W}_{\tilde{\theta}}s}\right](f,\nu) = \overline{\hat{h}_{0,0}(f-\nu)}\sqrt{\tfrac{d\theta}{df}}\,\hat{s}(\theta(f)). \tag{11}$$

Recall that f is the frequency variable conjugate to time τ in the time-frequency plane. Performing unitary frequency warping on this variable by means of the inverse frequency map θ^{-1} one obtains:

$$\left[\mathbf{W}_{\tilde{\theta}^{-1}}\widehat{\mathcal{S}\mathbf{W}_{\tilde{\theta}}s}\right](f,\nu) = \overline{\hat{h}_{0,0}(\theta^{-1}(f)-\nu)}\hat{s}(f), \tag{12}$$

where we have used the fact that

$$1 = \frac{d\left[\theta(\theta^{-1}(f))\right]}{df} = \frac{d\theta}{d\alpha}\Big|_{\alpha=\theta^{-1}(f)}\frac{d\theta^{-1}}{df}. \tag{13}$$

The redressed frequency warped STFT (12) is again in the form of a time-invariant filtering operation (convolution in time domain) where the filters are frequency warped versions of the modulated windows in (4). As a result, the dispersive delays in the analysis elements (9) are brought back to non-dispersive delays, the Fourier transform of the redressed analysis elements being

$$\hat{\tilde{\tilde{h}}}_{\tau,\nu}(f) = \left[\mathbf{T}_\tau\widehat{\mathbf{W}_{\tilde{\theta}}h_{0,\nu}}\right](f) = \hat{h}_{0,0}(\theta^{-1}(f)-\nu)e^{-j2\pi f\tau}. \tag{14}$$

It is possible to interpret (12) as the similarity transformation $\mathbf{W}^\dagger_{\tilde{\theta}}\mathcal{S}\mathbf{W}_{\tilde{\theta}}$ on the STFT operator, which is time-shift covariant.

3. REDRESSED WARPED GABOR FRAMES

In this section we review the definition of Gabor and warped Gabor frames. We would like to apply the same redressing method used in the previous section to counteract dispersion and realign time. However, the Gabor expansion coefficients are time-frequency samples of the STFT so that only a discrete version of time-frequency unwarping can be set forth.

3.1. Gabor frames

Given a window function h and two sampling parameters $a, b > 0$, the set of functions

$$\mathcal{G}(h,a,b) = \{\mathbf{T}_{na}\mathbf{M}_{mb}h : q,n \in \mathbb{Z}\} \tag{15}$$

is called a *Gabor system*. A signal s can be projected over a Gabor system by taking the scalar products $\langle s, \mathbf{T}_{na}\mathbf{M}_{mb}h\rangle$. These are exactly evaluations of the STFT of a signal with window h at the time-frequency grid of points (na, qb). Here we have defined the Gabor system using the same convention as in the definition (1) of the STFT. Usually, Gabor systems are defined with a reverse order of time-shift and frequency modulation operators, i.e. $\{\mathbf{M}_{mb}\mathbf{T}_{na}h : q,n \in \mathbb{Z}\}$. However, the extra phase factors that are introduced to convert from one definition to the other are perfectly irrelevant when establishing properties of the system. Even in the computation the extra phase factors cancel out in the analysis-synthesis algorithm, so they can be ignored.

A sequence of functions $\{\psi_l\}_{l\in I}$ in the Hilbert space $\mathcal{H}$ is called a frame if there exist both positive constant lower and upper bounds A and B, respectively, such that

$$A\|s\|^2 \le \sum_{l\in I}|\langle s,\psi_l\rangle|^2 \le B\|s\|^2 \quad \forall s \in \mathcal{H}, \tag{16}$$

where $\|s\|^2 = \langle s, s\rangle$ is the norm square or total energy of the signal. Frames generate signal expansions, i.e., the signal can be perfectly reconstructed from its projections over the frame.

A Gabor system that is a frame is called a *Gabor frame*. In this case, the signal can be reconstructed from the corresponding samples of the STFT. While not unique, reconstruction can be achieved with the help of a dual frame, which in turn is a Gabor frame generated by a dual window $\tilde{h}$. Perfect reconstruction depends on the choice of the window and the sampling grid. One can show that there exist no Gabor frames when $ab > 1$.

3.2. Warping Gabor frames

From (16) it is easy to see that any unitary operation on a frame results in a new frame with the same frame bounds A and B [5]. In particular, unitary operators can be applied to Gabor frames to obtain new frames. Depending on the operator, the resulting frames are not necessarily of the Gabor type, as the atoms are not generated by shifting and modulating a single window function.

Conceptually, starting from a Gabor frame (analysis) $\{\varphi_{n,q}\}_{q,n\in\mathbb{Z}}$ and dual frame (synthesis) $\{\gamma_{n,q}\}_{n,q\in\mathbb{Z}}$:

$$\begin{aligned}\varphi_{n,q} &= \mathbf{T}_{na}\mathbf{M}_{qb}h \\ \gamma_{n,q} &= \mathbf{T}_{na}\mathbf{M}_{qb}g,\end{aligned} \tag{17}$$

where h and g are dual windows, warped frames can be generated by unitarily warping the signal s prior to analysis and unitarily unwarping it after the synthesis:

$$\begin{aligned}s = \mathbf{U}_{\tilde{\theta}}^{\dagger} \sum_{n,q\in Z} \langle \mathbf{U}_{\tilde{\theta}} s, \varphi_{n,q}\rangle \gamma_{n,q} = \\ \sum_{n,q\in Z} \left\langle s, \mathbf{U}_{\tilde{\theta}}^{\dagger}\varphi_{n,q}\right\rangle \mathbf{U}_{\tilde{\theta}}^{\dagger}\gamma_{n,q},\end{aligned} \tag{18}$$

where $\mathbf{U}_{\tilde{\theta}}$ is a unitary frequency warping operator. Defining the frequency warped frame (analysis) $\{\tilde{\varphi}_{n,q}\}_{q,n\in\mathbb{Z}}$ and dual frame (synthesis) $\{\tilde{\gamma}_{n,q}\}_{n,q\in\mathbb{Z}}$ as follows:

$$\begin{aligned}\tilde{\varphi}_{n,q} &= \mathbf{U}_{\tilde{\theta}}^{\dagger}\varphi_{n,q} = \mathbf{U}_{\tilde{\theta}^{-1}}\mathbf{T}_{na}\mathbf{M}_{qb}h \\ \tilde{\gamma}_{n,q} &= \mathbf{U}_{\tilde{\theta}}^{\dagger}\gamma_{n,q} = \mathbf{U}_{\tilde{\theta}^{-1}}\mathbf{T}_{na}\mathbf{M}_{qb}g,\end{aligned} \tag{19}$$

one obtains the signal expansion

$$s = \sum_{n,q\in Z} \langle s, \tilde{\varphi}_{n,q}\rangle \tilde{\gamma}_{n,q}. \tag{20}$$

Just as Gabor frames can be obtained by uniformly sampling the integral STFT, the warped frames can be obtained as a result of nonuniform sampling in time-frequency. Nonuniform sampling theorems based on a time warping map were introduced in [6] and their adaptation to frequency sampling is immediate. Applications of frequency warping to time-frequency analysis date back to [7]. However, warped Gabor frames suffer from the same problem as the warped STFT: as a result of frequency warping, the time organization of the analysis and synthesis systems is disrupted; the windows are time-shifted with frequency dependent shifts. Indeed the Fourier transforms of the warped Gabor frame elements are

$$\hat{\tilde{\varphi}}_{n,q}(f) = \sqrt{\tfrac{d\theta^{-1}}{df}}\,\hat{h}(\theta^{-1}(f) - qb)e^{-j2\pi\theta^{-1}(f)na}, \tag{21}$$

which bear frequency dispersive delays. In other words dispersive time samples are produced by the direct application of the warped frame analysis. Similar problems are encountered when time-warping Gabor frames.

The magnitude Fourier transforms $\hat{h}(\theta^{-1}(f) - qb)$ of a set of frequency warped modulated windows corresponding to $1/3$ octave frequency resolution is shown in Fig. 1, together with a scaled version $\frac{1}{b}\theta^{-1}$ of the warping map, which maps warped frequency to fractional band number, i.e., the integer values of $\frac{1}{b}\theta^{-1}$ correspond to the center frequencies of the bands.

3.3. Redressing Warped Gabor Frames

The evaluation of the warped Gabor expansion coefficients

$$\tilde{c}_{n,q} = \langle s, \tilde{\varphi}_{n,q}\rangle \tag{22}$$

is identical to that of a time-frequency sampled warped STFT. In order to redress the frequency warped STFT into a time covariant representation we have introduced additional inverse frequency warping with respect to the time variable τ in the time-frequency plane. However, in the warped Gabor frames (19) this variable is sampled at instants na. Therefore, in order to parallel our warped STFT redressing procedure in the warped Gabor frames case, one can only apply a discrete-time form of frequency warping to the time index n.

It is possible to show [8, 9] that if the discrete-time frequency warping map ϑ is one-to-one and onto $[-\frac{1}{2}, +\frac{1}{2}[$, and almost everywhere differentiable there, then the set of sequences

$$\eta_m(n) = \int_{-\frac{1}{2}}^{+\frac{1}{2}} \sqrt{\tfrac{d\vartheta}{d\nu}}\, e^{j2\pi(n\nu - m\vartheta(\nu))} d\nu, \tag{23}$$

where $n, m \in \mathbb{Z}$, forms an orthonormal basis of $\ell^2(\mathbb{Z})$. These are recognized as generalized Laguerre sequences [10, 11, 12], which are the inverse discrete-time Fourier transforms of warped harmonic complex sinusoids in the frequency domain interval $[-\frac{1}{2}, +\frac{1}{2}[$. The map ϑ can be extended over the entire real axis as congruent modulo 1 to a 1-periodic function.

Given a sequence $\{x(n)\}$ in $\ell^2(\mathbb{Z})$, the scalar products

$$\tilde{x}(m) = \langle x, \eta_m\rangle_{\ell^2(\mathbb{Z})} \tag{24}$$

generate another sequence $\{\tilde{x}(m)\}$ in $\ell^2(\mathbb{Z})$, which satisfies

$$\hat{\tilde{x}}(\nu) = \sqrt{\tfrac{d\vartheta^{-1}}{d\nu}}\,\hat{x}(\vartheta^{-1}(\nu)), \tag{25}$$

where the ˆ symbol, when applied to sequences, denotes discrete-time Fourier transform. Thus, $\overline{\eta_m(n)}$ defines the nucleus of an inverse unitary frequency warping $\ell^2(\mathbb{Z})$ operator $\mathbf{D}_{\tilde{\vartheta}^{-1}} = \mathbf{D}_{\tilde{\vartheta}}^{\dagger}$. Clearly, the transposed conjugate sequences $\mu_m(n) = \overline{\eta_n(m)}$ form the nucleus of a unitary frequency warping $\ell^2(\mathbb{Z})$ operator $\mathbf{D}_{\tilde{\vartheta}}$.

In order to limit or eliminate time dispersion in the frequency warped Gabor expansion, one can apply the discrete-time frequency warping operator $\mathbf{D}_{\tilde{\vartheta}^{-1}}$ to the time sequence of expansion coefficients over the warped Gabor frame (22), i.e., with respect to index n. Since the operator is applied only on the time index, for generality, one can include dependency of the map and of the sequences η_n on the frequency index q, which will be useful in the sequel. The new coefficients are obtained as follows:

$$\begin{aligned}\tilde{\tilde{c}}_{n,q} = \left[\mathbf{D}_{\tilde{\vartheta}_q^{-1}}\tilde{c}_{\bullet,q}\right](n) = \sum_{m\in\mathbb{Z}} \overline{\eta_{n,q}(m)}\,\langle s, \tilde{\varphi}_{m,q}\rangle = \\ \left\langle s, \sum_{m\in\mathbb{Z}} \eta_{n,q}(m)\tilde{\varphi}_{m,q}\right\rangle.\end{aligned} \tag{26}$$

In order to reconstruct the signal from the coefficients $\tilde{\tilde{c}}_{n,q}$ one can first recover the coefficients $\tilde{c}_{n,q}$, which stems from the completeness and orthogonality of the set $\{\eta_{n,q}\}_{n\in\mathbb{Z}}$, and then combine them with the dual warped frame elements:

$$s = \sum_{n,q\in\mathbb{Z}} \tilde{c}_{n,q}\tilde{\gamma}_{n,q} = \sum_{n,q\in\mathbb{Z}}\sum_{m\in\mathbb{Z}} \tilde{\tilde{c}}_{m,q}\eta_{m,q}(n)\tilde{\gamma}_{n,q}. \tag{27}$$

Hence, defining the redressed frequency warped Gabor analysis and synthesis frames as follows:

$$\begin{aligned}\tilde{\tilde{\varphi}}_{n,q} &= \mathbf{D}_{\tilde{\vartheta}_q^{-1}}\tilde{\varphi}_{\bullet,q} = \sum_m \eta_{n,q}(m)\tilde{\varphi}_{m,q} \\ \tilde{\tilde{\gamma}}_{n,q} &= \mathbf{D}_{\tilde{\vartheta}_q^{-1}}\tilde{\gamma}_{\bullet,q} = \sum_m \eta_{n,q}(m)\tilde{\gamma}_{m,q},\end{aligned} \tag{28}$$

from (26) and (27) we have:

$$s = \sum_{n,q \in Z} \tilde{\tilde{c}}_{n,q} \tilde{\tilde{\gamma}}_{n,q} = \sum_{n,q \in (Z)} \left\langle s, \tilde{\tilde{\varphi}}_{n,q} \right\rangle \tilde{\tilde{\gamma}}_{n,q}. \tag{29}$$

Indeed, the redressing discrete-time warping transformation is based on an orthonormal and complete expansion in $\ell^2(\mathbb{Z})$, which leads to the unitary equivalence of the redressed warped frames with the warped frames.

Exploiting the periodicity of the discrete-time redressing frequency warping map one can show that the Fourier transforms of the redressed frame is

$$\hat{\tilde{\tilde{\varphi}}}_{n,q}(f) = A(f)\hat{h}(\theta^{-1}(f) - qb)e^{-j2\pi n \vartheta_q(a\theta^{-1}(f))}, \tag{30}$$

where

$$A(f) = \sqrt{\frac{d\theta^{-1}}{df}} \sqrt{\frac{d\vartheta_q}{d\nu}}\Bigg|_{\nu = a\theta^{-1}(f)}. \tag{31}$$

Hence, the effect of the dispersive delays would be counteracted if

$$\vartheta_q(a\theta^{-1}(f)) = d_q f \tag{32}$$

for any $f \in \mathbb{R}$, where d_q are positive constants controlling the time scale in each frequency band. In this case, the Fourier transforms of the redressed frame elements simply become:

$$\hat{\tilde{\tilde{\varphi}}}_{n,q}(f) = \sqrt{\tfrac{d_q}{a}}\hat{h}(\theta^{-1}(f) - qb)e^{-j2\pi n d_q f}. \tag{33}$$

Furthermore, if all d_q are identical, all the time samples would be aligned to a uniform time scale throughout frequencies.

However, each map ϑ_q is constrained to be congruent modulo 1 to a 1-periodic function, while the global warping map θ can be arbitrarily selected. Furthermore, having to be one-to-one in each unit interval, the functions ϑ_q can at most experience an increment of 1 there.

The problem of linearizing the phase is illustrated in Fig. 2, where the black curve is the amplitude scaled warping map $d_q\theta(\nu)$ and the gray curve represents the map $\vartheta_q(a\nu)$, which is $1/a$-periodic, both plotted in the abscissa $\nu = \theta^{-1}(f)$. Amplitude scaling the warping map θ allows the values of the map to lie in the range of the discrete-time warping map ϑ_q. The amplitude scaling factors are the new time sampling intervals d_q of the redressed warped Gabor expansion.

If the window h is chosen to have compact support in the frequency domain, which is the so called "painless" case, one can exactly eliminate the dispersive delays with the help of (28). In fact, suppose for simplicity that the bandwidth of the window h is Kb, with K a positive integer, i.e., $\hat{h}(f) = 0$ for $|f| \geq Kb/2$. The choice of the initial sampling interval a allows all the maps $\{\vartheta_q\}_{q\in\mathbb{Z}}$ to be arbitrarily specifiable to match $d_q\theta(\nu)$ in the intervals where the Fourier transforms of the warped modulated windows (warped frame elements) are nonzero. Hence, condition (32) only needs to be satisfied by the map ϑ_q in this interval. Equivalently, we require

$$\vartheta_q(a\nu) = d_q\theta(\nu), \quad (q - \tfrac{K}{2})b < \nu < (q + \tfrac{K}{2})b, \tag{34}$$

which is possible if on one hand the variation of the argument of the map ϑ_q in (34) satisfies

$$a[(q + \tfrac{K}{2})b - (q - \tfrac{K}{2})b] = Kab \leq 1 \tag{35}$$

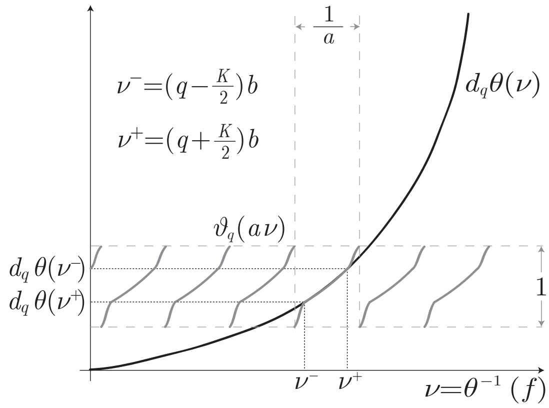

Figure 2: Locally eliminating dispersion by means of discrete-time frequency warping. Black line: curve derived from the original map θ by amplitude scaling. Gray line: discrete-time frequency warping characteristics for local delay linearization.

and, on the other hand, if also the variation of the map ϑ_q over the warped modulated window bandwidth satisfies

$$d_q[\theta((q + \tfrac{K}{2})b) - \theta((q - \tfrac{K}{2})b)] = d_q B_q \leq 1, \tag{36}$$

where $B_q = \theta((q + \frac{K}{2})b) - \theta((q - \frac{K}{2})b)$ is the full bandwidth of the warped modulated window. The first of these conditions only requires $ab \leq 1/K$, which does not depend on q and can be satisfied assigning sufficient redundancy (oversampling) of the initial Gabor frame. Incidentally, this is the same condition for the original Gabor system to form a frame. A valid choice is $K = 2$, which requires $ab \leq 1/2$. For the second condition, one needs to select $d_q \leq 1/B_q$, as intuitively clear from the sampling theorem. If there is an upper bound B to the bandwidths B_q then one can choose identical $d_q = 1/B$, $q \in \mathbb{Z}$, to satisfy the sampling condition with uniform rates.

In the general case, a perfect time realignment of the components is not guaranteed. By construction, the redressed warped Gabor systems are guaranteed to be frames for any choice of the maps ϑ_q satisfying the stated periodicity conditions, even when the phase is not completely linearized. Locally, within the essential bandwidths of the warped modulated windows it is possible to linearize the phase of the complex exponentials in (30).

4. ONLINE COMPUTATION AND APPROXIMATIONS

The warping map design method to eliminate dispersive sampling in the frequency warped Gabor elements is exact when the elements are compactly supported in the frequency domain. This type of frames are definitely suitable for offline computation using simple and efficient frequency domain techniques [1].

Since the computation of Gabor expansion coefficients is not causal, in online computation one requires the frame elements to have compact support in the time domain. Starting with a finite duration window, one can linearize the phase and choosing suitable sampling parameters, one can eliminate dispersion within the essential bandwidths of the warped modulated windows [4]. However, this is still not sufficient for online computation purposes.

In fact, generally, the modulated frequency warped windows will not have compact support in the time domain even if the original window had this property.

In order to provide an approximation suitable for online computation, one can observe that the window h is narrow band low pass and the warping map is differentiable. Therefore, in the argument of $\hat{h}$ in (30) one can expand $\theta^{-1}(f)$ in Taylor series around the point $\theta(qb)$. Truncating to first order, which corresponds to a local linearization of the warping map within the bandwidth of the window, one obtains:

$$\theta^{-1}(f) \approx qb + \frac{1}{\tau_q}(f - \theta(qb)), \tag{37}$$

where

$$\tau_q = \left.\frac{d\theta}{df}\right|_{f=qb} \tag{38}$$

is the group delay associated to the warping map $\theta(f)$ at frequency $f = qb$. Thus, we have the following approximation:

$$\hat{g}_q(f) = \hat{h}(\theta^{-1}(f) - qb) \approx \hat{h}\left(\frac{f-\theta(qb)}{\tau_q}\right) \tag{39}$$

Thus, in this approximation, the window $g_q(t)$ is simply obtained by dilating and modulating the prototype window h, in which the local group delay acts as scaling factor:

$$g_q(t) = \tau_q h(\tau_q t) e^{j2\pi\theta(qb)t}. \tag{40}$$

Hence, if the prototype window has compact support in the time domain, all its approximate warped modulated versions will have compact support.

In order to perform online computations of the redressed frequency warped Gabor expansion, one can start from a prototype window h that has compact support in the time domain, where aliasing is canceled in the time-domain through overlap-add, such as the time-domain cosine window, given by

$$h(t) = \begin{cases} \sqrt{\frac{2b}{R}} \cos \frac{\pi t}{T} & \text{if } -\frac{T}{2} \leqslant t < +\frac{T}{2} \\ 0 & \text{otherwise} \end{cases} \tag{41}$$

where T is the total duration of the window, $R > 1$ is an integer, b is the frequency sampling interval and we let the time shift parameter $a = T/R$.

In the redressed frame (30) one replaces the warped modulated windows by the scaled windows in (39). Furthermore, one performs redressing in the essential bandwidth and considers uniform time sampling within each analysis band. This requires suitable setting of the time-frequency sampling rates, which we are goind to illustrate for the cosine window example.

The Fourier transform of the cosine window, given by

$$\hat{h}(\nu) = \sqrt{\frac{b}{2R}}(\mathrm{sinc}(\nu T - \tfrac{1}{2}) + \mathrm{sinc}(\nu T + \tfrac{1}{2})), \tag{42}$$

is plotted in Fig. 3, from which one can see that the main lobe has bandwidth $3/T = 3/Ra$. Assuming this as the essential bandwidth in which to linearize the phase, in order to satisfy (32) here, one needs to select $R \geq 3$, which is the analogon of (35), and $d_q B_q \leq 1$, which is the analogon of (36), where now $B_q = \theta(qb + \frac{3}{2T}) - \theta(qb - \frac{3}{2T})$.

Concurrently, the parameter T can be selected according to the smallest required essential bandwidth. For example, in the case of a tempered scale warping map, in order to have sufficient

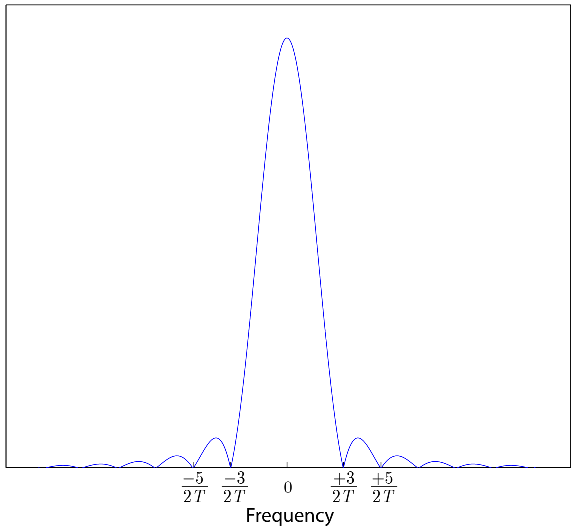

Figure 3: Magnitude Fourier transform of the cosine window.

frequency resolution one can select $\frac{3}{2T} = f_0$, where f_0 is the frequency of the smallest tone to be represented, so that adjacent tones fall away from the main frequency lobe of the window, which gives $a = \frac{3}{2Rf_0}$.

The frequency shift parameter b must be chosen so that $ab \leq 1/R$ for the original Gabor system to be a frame. For $R = 3$ and the chosen value of a, this gives $b \leq 2f_0/3$. However, in practice one would like the tones of the scale to be adequately represented by the warped bands; moreover, narrower bands improve the approximation of the warped modulated windows with the scaled modulated cosine windows. In our examples we chose $b = f_0/3$. The quality of the approximation can be evaluated by comparing the magnitude Fourier transform of the widows, shown in Fig. 4 for the case of 1/3 octave warping map for the centerband frequency of 356.02 Hz. The two modulated windows are shown in the time domain in Fig. 5. One can see that the scaled cosine modulated window closely approximates the warped window on a finite interval, truncating its tails.

As a refinement of the finite length window approximation, one can consider the truncation of the modulated warped windows on a larger interval than that offered by the approximating scaled cosine. The length of the interval can be estimated at 1.5 the support of the approximating cosine window. In this case one can obtain a reconstruction error norm in the order of 10^{-5} of the norm of the signal for the 1/3-octave warping map. Informal perceptual tests show no audible artifacts attached to the approximate analysis and synthesis procedure. A deeper analysis of the approximation error will be the object of a forthcoming paper.

Since the center frequencies of the warped Gabor frames are not equally spaced, the computation of the transform cannot be directly performed by means of the Fast Fourier Transform. Real multirate filterbanks can be designed by combining the complex conjugate channels. Therefore the complexity is linear in the number of samples, where the number of channels is a proportional factor.

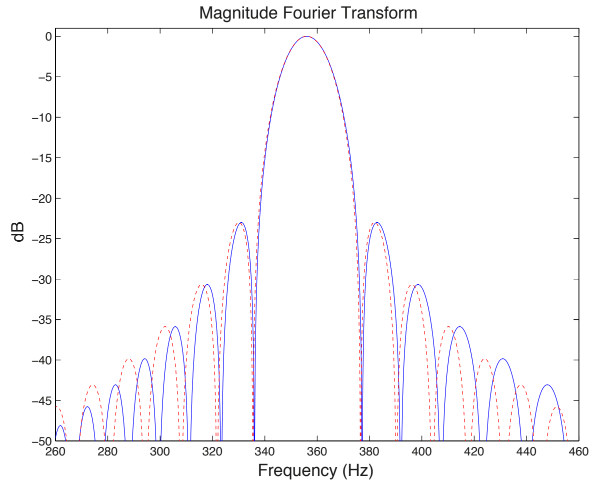

Figure 4: Magnitude Fourier transforms of the warped modulated window (dotted line) and of the approximating modulated scaled cosine window (solid line), calculated for a 1/3 octave time-frequency representation.

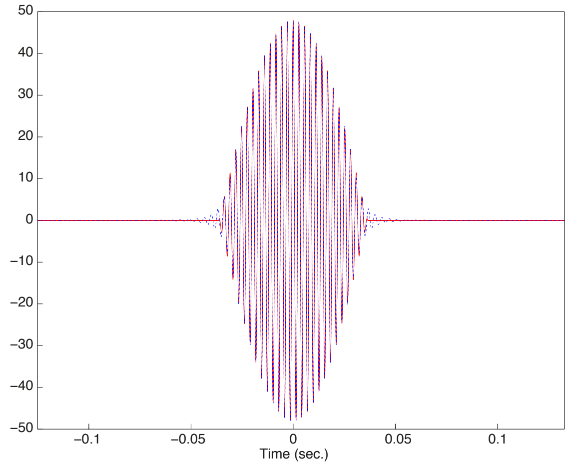

Figure 5: Warped modulated window (dotted line) and of the approximating modulated scaled cosine window (solid line), calculated for a 1/3 octave time-frequency representation.

5. CONCLUSIONS

In this paper, we have introduced approximation methods suitable for the online computation of the analysis and synthesis of time-frequency representations with arbitrary allocation of the frequency bands based on frequency warping. The problems arising from the dispersive sampling introduced by warping are solved by introducing a further warping operation in time-frequency.

The approximation of the frequency warped modulated windows consists in a local linearization of the warping map, which corresponds to time scaling and modulating a prototype window. The effect of dispersion is minimized within the essential bandwidths of the frame elements when these are selected, in order to fulfill causal computational needs, to have compact support in the time domain. A further refinement is obtained by directly truncating the modulated windows on larger intervals than the approximating cosine windows, obtaining higher accuracy in the reconstruction at the cost of larger storage, as the windows can be precomputed offline.

6. REFERENCES

[1] G. A. Velasco, N. Holighaus, M. Dörfler, and T. Grill, “Constructing an invertible constant-Q transform with nonstationary Gabor frames,” in *Proceedings of the Digital Audio Effects Conference (DAFx-11)*, Paris, France, 2011, pp. 93–99.

[2] P. Balazs, M. Dörfler, F. Jaillet, N. Holighaus, and G. A. Velasco, “Theory, implementation and applications of nonstationary Gabor Frames,” *Journal of Computational and Applied Mathematics*, vol. 236, no. 6, pp. 1481–1496, 2011.

[3] G. Evangelista, M. Dörfler, and E. Matusiak, “Phase vocoders with arbitrary frequency band selection,” in *Proceedings of the 9th Sound and Music Computing Conference*, Copenhagen, Denmark, 2012, pp. 442–449.

[4] G. Evangelista, “Warped Frames: dispersive vs. nondispersive sampling,” in *Proceedings of the Sound and Music Computing Conference (SMC-SMAC-2013)*, Stockholm, Sweden, 2013, pp. 553–560.

[5] R. G. Baraniuk and D. L. Jones, “Unitary equivalence : A new twist on signal processing,” *IEEE Transactions on Signal Processing*, vol. 43, no. 10, pp. 2269–2282, Oct. 1995.

[6] J. J. Clark, M. Palmer, and P. Lawrence, “A transformation method for the reconstruction of functions from nonuniformly spaced samples,” *IEEE Transactions on Acoustics, Speech and Signal Processing*, vol. 33, no. 5, pp. 1151–1165, 1985.

[7] C. Braccini and A. Oppenheim, “Unequal bandwidth spectral analysis using digital frequency warping,” *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 22, pp. 236–244, 1974.

[8] P. W. Broome, “Discrete orthonormal sequences,” *Journal of the ACM*, vol. 12, no. 2, pp. 151–168, Apr. 1965.

[9] L. Knockaert, “On Orthonormal Muntz-Laguerre Filters,” *IEEE Transactions on Signal Processing*, vol. 49, no. 4, pp. 790 –793, apr 2001.

[10] G. Evangelista, “Dyadic Warped Wavelets,” *Advances in Imaging and Electron Physics*, vol. 117, pp. 73–171, Apr. 2001.

[11] G. Evangelista and S. Cavaliere, “Frequency Warped Filter Banks and Wavelet Transform: A Discrete-Time Approach Via Laguerre Expansions,” *IEEE Transactions on Signal Processing*, vol. 46, no. 10, pp. 2638–2650, Oct. 1998.

[12] G. Evangelista and S. Cavaliere, “Discrete Frequency Warped Wavelets: Theory and Applications,” *IEEE Transactions on Signal Processing*, vol. 46, no. 4, pp. 874–885, Apr. 1998, special issue on Theory and Applications of Filter Banks and Wavelets.