REAL-TIME AND EFFICIENT ALGORITHMS FOR FREQUENCY WARPING BASED ON LOCAL APPROXIMATIONS OF WARPING OPERATORS

Gianpaolo Evangelista

Sound Technology, Media Division
Inst. of Technology, Linköping University
Norrköping, Sweden
giaev@itn.liu.se

Sergio Cavaliere

ACEL
Dept. of Physical Sciences
Federico II University of Naples, Italy
sergio.cavaliere@na.infn.it

ABSTRACT

Frequency warping is a modifier that acts on sound signals by remapping the frequency axis. Thus, the spectral content of the original sound is displaced to other frequencies. At the same time, the phase relationship among the signal components is altered, nonlinearly with respect to frequency. While this effect is interesting and has several applications, including in the synthesis by physical models, its use has been so far limited by the lack of an accurate and flexible real-time algorithm. In this paper we present methods for frequency warping that are based on local approximations of the warping operators and allow for real-time implementation. Filter bank structures are derived that allow for efficient realization of the approximate technique. An analysis of the error is also presented, which shows that both numerical and perceptual errors are within acceptable limits. Furthermore, the approximate implementation allows for a larger variety of warping maps than that achieved by the classical (non-causal) first-order allpass cascade implementation.

1. INTRODUCTION

This paper is concerned with the computation of the frequency warping operation on signals, such as a musical tone of duration of about one second or longer. In other audio applications, frequency warping is often applied to filters and to filter design, in which case the computation presents little or no problem, the results are classical and well studied [1] and the technique is implementable in real-time. On the contrary, warping long-length signals presents large computational problems, both from the point of view of complexity and of causality, which hamper the possibility of using this technique in real-time applications, at least in exact form. A few plugins partially exploiting the musical capabilities offered by frequency warping are available on the market [2, 3].

In this paper, we present a simple approximation method that yields efficient and causal structures, in the form of unconventional multirate filter banks, which allow for real-time computation of frequency warping, for a large class of warping maps. At similar computational cost, the algorithm largely improves the quality of warping with respect to the one presented in [4, 5]. The new approximation is based on the intuitive idea that frequency warping a narrowband signal, such as a sinusoidal wave packet, simply amounts to properly time scale the amplitude envelope and to remodulate the signal to warped frequency. The scaling factor of the envelope approximately depends on the derivative of the warping map at warped frequency.

If a wideband signal is decomposed into sinusoidal wave packets, e.g., by means of a Gabor expansion or short-time Fourier transform [6, 7], then the warped signal can be generated by scaling, remodulating and properly delaying each packet. Due to the dispersive character of frequency warping, the original time shift of the wave packets transforms into a frequency dependent delay, which can be approximated to a constant delay for wave packets at each given frequency. Fortunately, scaling and delay transform in a covariant way by warping. As it can be expected, this results into generalized phase vocoder filter bank structures [8] where different resampling factors are applied to the various channels. Moreover, scaling is achieved by enforcing unequal downsampling and upsampling rates in each channel.

The efficient filter bank structures for the computation are discussed in this paper together with an analysis of the complexity and of the error. Our results show that both the numerical and the perceptual errors are largely tolerable.

The applications of the approximate warping filter bank structure range from signal representations [9] to synthesis and digital audio effects [10, 5]. In particular, in the physical models of piano strings and rods, the frequency warping section can be cascaded to a digital waveguide in order to simulate dispersion [11]. At comparable computational cost and similar warping map flexibility, this is a viable alternative to the use of dispersive waveguides that does not require the design of very high order allpass filters [12].

2. FREQUENCY WARPING OPERATORS

Frequency warping is completely characterized by a frequency transformation, i.e., by a map $\theta(\omega)$ of the frequency axis into or onto itself. Given a signal $s(t)$ with Fourier transform $S(\omega)$, the Fourier transform of the warped signal is the function

$$S(\theta(\omega)) = \tfrac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} s(t)e^{-j\theta(\omega)t}dt. \qquad (1)$$

Thus, frequency warping can be defined as action of the operator

$$\mathfrak{W}_\theta = \mathfrak{F}^\dagger \mathfrak{C}_\theta \mathfrak{F}, \qquad (2)$$

where $\mathfrak{F}$ is the Fourier transform operator and $\mathfrak{F}^\dagger$ its adjoint. The symbol $\mathfrak{C}_\theta$ denotes the composition by θ operator, i.e.,

$$[\mathfrak{C}_\theta S](\omega) = [S \circ \theta](\omega) = S(\theta(\omega)). \qquad (3)$$

The arbitrary shape of the map θ defines the character of the frequency warping operations. In most applications, $\theta(\omega)$ is a continuously increasing and differentiable antisymmetric function mapping zero frequency to zero frequency, as shown in Figure 1. However, more general forms in which the warping map $\theta(\omega)$ is not
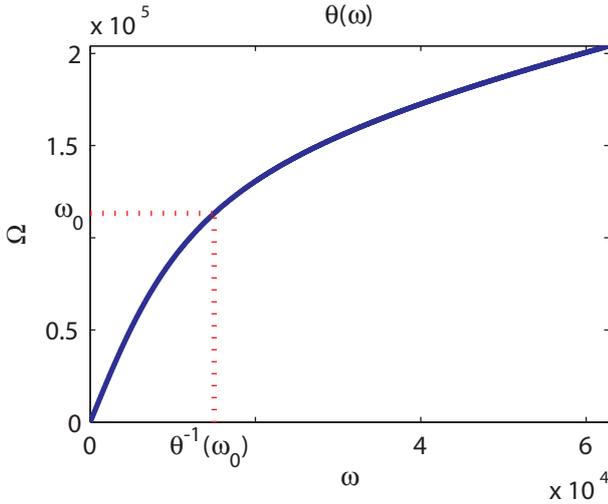
Figure 1: *A typical warping map.*

one-to-one can be considered provided that local invertibility is guaranteed.

If the map is one-to-one and almost everywhere differentiable then a unitary form of the warping operator can be defined by the action

$$\tilde{S}(\omega) = [\mathfrak{U}_\theta S](\omega) = \sqrt{\left|\frac{d\theta}{d\omega}\right|} S(\theta(\omega)). \tag{4}$$

We will assume henceforth that the map is increasing so that its first derivative is non-negative and the absolute value can be dropped in (4). The main property of the unitary warping operator is to preserve in-band energy. This is an important property even from the perceptual point of view: since warping a given frequency band results in a dilated or compressed band, without normalization the warped band may be perceived as louder or fainter, respectively. If the map θ is strictly increasing, then the warping operator is invertible and the action of the inverse is given by

$$\begin{aligned}
\left[\mathfrak{U}_\theta^\dagger S\right](\omega) &= \left[\mathfrak{U}_\theta^{-1} S\right](\omega) = [\mathfrak{U}_{\theta^{-1}} S](\omega) \\
&= \sqrt{\frac{d\theta^{-1}}{d\omega}} S(\theta^{-1}(\omega)),
\end{aligned} \tag{5}$$

where the symbol $\dagger$ denotes the adjoint operator, which is identical to the inverse in view of unitarity.

A spectral peak of the signal at $\omega = \omega_0$ may result into one or more peaks of the warped signal located at the roots Ω of the equation $\theta(\Omega) = \omega_0$, if any. If the map θ is one-to-one and onto then the peak at $\omega = \omega_0$ transforms into a peak at $\theta^{-1}(\omega_0)$. In this sense, frequency warping is a frequency dependent modulation technique, where each component sinusoid is modulated to a different frequency.

2.1. Warping Quasi-Sinusoidal Signals

It is instructive to investigate on how an amplitude modulated sinusoidal signal of the form

$$s(t) = g(t)e^{j\omega_0 t}, \tag{6}$$

where $g(t)$ is a real smooth envelope, is transformed by frequency warping. In this case we have

$$S(\omega) = G(\omega - \omega_0), \tag{7}$$

hence the Fourier transform of the warped signal is

$$\tilde{S}(\omega) = \sqrt{\frac{d\theta}{d\omega}} G(\theta(\omega) - \omega_0). \tag{8}$$

If the amplitude envelope $g(t)$ is narrowband then $G(\theta(\omega) - \omega_0)$ is nonzero only in a small neighborhood of $\theta^{-1}(\omega_0)$. Therefore, if the map is continuous and differentiable in this neighborhood, we can expand $\theta(\omega)$ in a Taylor series about $\omega = \omega_0$. Truncation to first order yields:

$$\begin{aligned}
\theta(\omega) &\approx \theta(\theta^{-1}(\omega_0)) + \beta(\omega - \theta^{-1}(\omega_0)) \\
&= \omega_0 + \beta(\omega - \theta^{-1}(\omega_0)),
\end{aligned} \tag{9}$$

where

$$\beta = \left.\frac{d\theta}{d\omega}\right|_{\omega=\theta^{-1}(\omega_0)} = \left[\left.\frac{d\theta^{-1}}{d\omega}\right|_{\omega=\omega_0}\right]^{-1}. \tag{10}$$

Substituting (9) into (8) and approximating, within the warped band, the first derivative of θ with the constant β obtains

$$\tilde{S}(\omega) \approx \sqrt{\beta} G(\beta(\omega - \theta^{-1}(\omega_0))). \tag{11}$$

Equation (11) shows that warping a narrowband signal is approximately equivalent to scaling the signal envelope by β and to remodulating to the warped frequency $\theta^{-1}(\omega_0)$. Indeed, from (11) and the Fourier scale theorem, we obtain:

$$\tilde{s}(t) \approx \frac{1}{\sqrt{\beta}} g\left(\frac{t}{\beta}\right) e^{j\theta^{-1}(\omega_0)t}. \tag{12}$$

Consequently, the frequency warped version of a wideband signal represented in terms of narrowband Gabor grains can be obtained approximately by individually scaling and remodulating the grains.

Another essential ingredient for the approximation of warping is dispersion. It is easy to see that the warped version of the shifted signal $s(t - \tau)$ has Fourier transform

$$[\mathfrak{F}\mathfrak{U}_\theta s(t - \tau)](\omega) = \sqrt{\left|\frac{d\theta}{d\omega}\right|} e^{-j\theta(\omega)\tau} S(\theta(\omega)). \tag{13}$$

Accordingly, each frequency component of the signal is time shifted by a different amount controlled by the warping map θ, acting as a multiplier of the time shift τ. If $s(t)$ is the narrowband signal in (6) then, within the same approximation as in (11), we have

$$[\mathfrak{F}\mathfrak{U}_\theta s(t - \tau)](\omega) \approx e^{-j[\beta\omega + \omega_0 - \beta\theta^{-1}(\omega_0)]\tau} \tilde{S}(\omega). \tag{14}$$

As the result of warping, the delayed enveloped sinusoid is shifted by $\beta\tau$. Hence, the warped group delay β acts as a multiplier of time shift. The phase correction term $\left[\beta\theta^{-1}(\omega_0) - \omega_0\right]\tau$ adjusts the phase delay of the unwarped sinusoid to that of the warped one. Indeed, it is easy to show that the approximately warped version of a running sinusoid amplitude modulated by a delayed envelope, i.e., of a signal of the form:

$$v(t) = g(t - \tau)e^{j\omega_0 t}, \tag{15}$$

is the following:

$$\tilde{v}(t) = \mathfrak{U}_\theta v(t) \approx \frac{1}{\sqrt{\beta}} g\left(\frac{t}{\beta} - \tau\right) e^{j\theta^{-1}(\omega_0)t}. \tag{16}$$

The last approximation can be written in the compact form:

$$\mathfrak{U}_\theta \mathfrak{M}_{\omega_0} \mathfrak{T}_\tau g(t) \approx \mathfrak{M}_{\theta^{-1}(\omega_0)} \mathfrak{T}_{\beta\tau} \mathfrak{D}_\beta g(t), \tag{17}$$

where $\mathfrak{M}_{\omega_0}$ is the modulation operator

$$\mathfrak{M}_{\omega_0} g(t) = e^{j\omega_0 t} g(t), \tag{18}$$

$\mathfrak{T}_\tau$ is the time-shift operator

$$\mathfrak{T}_\tau g(t) = g(t - \tau), \tag{19}$$

and $\mathfrak{D}_\beta$ is the dilation operator

$$\mathfrak{D}_\beta g(t) = \frac{1}{\sqrt{\beta}} g\left(\frac{t}{\beta}\right). \tag{20}$$

In other words, since (17) can be rewritten as follows:

$$\mathfrak{D}_\beta^{-1} \mathfrak{T}_{\beta\tau}^{-1} \mathfrak{M}_{\theta^{-1}(\omega_0)}^{-1} \mathfrak{U}_\theta \mathfrak{M}_{\omega_0} \mathfrak{T}_\tau g(t) \approx g(t), \tag{21}$$

the low-pass window $g(t)$ must be selected as an approximate eigenfunction of the unitary operator on the left hand side, with eigenvalue 1.

It is interesting to note that both the modulation and the dilation operator are particular cases of unitary warping operators, where the map θ is chosen as $\omega - \omega_0$ and $\beta\omega$, respectively. Indeed, using this fact and the commutation rule $\mathfrak{T}_{\beta\tau} \mathfrak{D}_\beta = \mathfrak{D}_\beta \mathfrak{T}_\tau$, one can show that (21) is equivalent to

$$\mathfrak{U}_\psi \mathfrak{T}_\tau g(t) \approx \mathfrak{T}_\tau g(t) \tag{22}$$

with

$$\psi(\omega) = \theta\left(\frac{\omega}{\beta} + \theta^{-1}(\omega_0)\right) - \omega_0. \tag{23}$$

In other words, the time-shifted window must be close to an eigenfunction of the incremental warping operator $\mathfrak{U}_\psi$.

3. APPROXIMATE FREQUENCY WARPING THROUGH GABOR FRAMES

By means of Gabor expansions [7, 6], signals are represented in terms of a collection of windowed sinusoids:

$$s(t) = \sum_{q,n \in \mathbf{Z}} S_{q,n} g_{q,n}(t), \tag{24}$$

where

$$g_{q,n}(t) = \mathfrak{M}_{2\pi q a} \mathfrak{T}_{n\tau} g(t) = e^{j2\pi q a t} g(t - n\tau); \ q, n \in \mathbf{Z}, \tag{25}$$

in which q is the frequency index and n the time index. The representative elements are obtained by time-shifting and modulating a unique window function $g(t)$. The representation is complete provided that the frame operator

$$\mathfrak{P} s(t) = \sum_{q,n \in \mathbf{Z}} \langle s, g_{q,n} \rangle \, g_{q,n}(t) \tag{26}$$

is invertible, where the symbol $\langle,\rangle$ denotes scalar product in $\mathbf{L}^2(\mathbf{R})$. This is true if there exist two finite non-zero constants A and B such that

$$A \leqslant \sum_{q,n \in \mathbf{Z}} |\langle s, g_{q,n} \rangle|^2 \leqslant B, \tag{27}$$

in which case (25) is said to be a frame and the expansion coefficients $S_{q,n}$ in (24) can be computed – not uniquely in the general case – as the scalar products

$$S_{q,n} = \langle s, \hat{g}_{q,n} \rangle, \tag{28}$$

where $\hat{g}_{q,n} = \mathfrak{P}^{-1} g_{q,n}(t)$ is the dual frame. For the Gabor frame (27) can only be satisfied if the time-frequency sampling grid is sufficiently fine, i.e., if $a\tau \leqslant 1$.

The definition of frame (27) and of frame operator (26) is independent on the way the frame elements $g_{q,n}(t)$ are generated, i.e., it extends to non-Gabor frames in which the frame elements are not generated by time-shift and modulation. For the Gabor frame one can show that

$$\hat{g}_{q,n}(t) = \mathfrak{M}_{2\pi q a} \mathfrak{T}_{n\tau} \hat{g}(t) = e^{j2\pi q a t} \hat{g}(t - n\tau); \ q, n \in \mathbf{Z}, \tag{29}$$

i.e., the dual Gabor frame is also obtained by time-shifting and modulating a unique dual window function $\hat{g}(t)$, so that (28) takes on the form of a short-time Fourier transform with inverse (24).

Any unitary operation on the frame results in a new frame with the same frame bounds A and B [13]. In particular, when applied to a Gabor frame, the unitary warping operator $\mathfrak{U}_\theta$ generates the frequency warped frame and dual frame

$$\begin{aligned} \tilde{g}_{q,n}(t) &= \mathfrak{U}_\theta \mathfrak{M}_{2\pi q a} \mathfrak{T}_{n\tau} g(t), \\ \tilde{\hat{g}}_{q,n}(t) &= \mathfrak{U}_\theta \mathfrak{M}_{2\pi q a} \mathfrak{T}_{n\tau} \hat{g}(t). \end{aligned} \tag{30}$$

A discrete-time version of warped frame was introduced in [14] for the non-uniform time-frequency representation of signals. As we showed in the previous Section, the warped Gabor frame and its dual are not Gabor frames. Unless the warping map is linear, there is no exact commutation rule between warping and time-shift operators. However, reasoning as in (17), one can show that

$$\tilde{g}_{q,n}(t) \approx \mathfrak{M}_{\theta^{-1}(2\pi q a)} \mathfrak{T}_{n\beta_q \tau} \mathfrak{D}_{\beta_q} g(t), \tag{31}$$

where

$$\beta_q = \left.\frac{d\theta}{d\omega}\right|_{\omega = \theta^{-1}(2\pi q a)} = \left[\left.\frac{d\theta^{-1}}{d\omega}\right|_{\omega = 2\pi q a}\right]^{-1}, \tag{32}$$

provided that $g(t)$ is sufficiently smooth and narrowband, and similarly for the dual frame.

In this paper we will not focus on the unitarily equivalent warped frame representation. Rather, we seek approximations of the warping operator through Gabor frame representation. The main idea is that when (31) holds and the signal is represented as in (24) then

$$\begin{aligned} \mathfrak{U}_\theta s(t) &= \sum_{q,n \in \mathbf{Z}} S_{q,n} \mathfrak{U}_\theta g_{q,n}(t) \\ &\approx \sum_{q,n \in \mathbf{Z}} S_{q,n} \mathfrak{M}_{\theta^{-1}(2\pi q a)} \mathfrak{T}_{\beta_q \tau} \mathfrak{D}_{\beta_q} g(t). \end{aligned} \tag{33}$$

Notice that (33) is a peculiar "irregular" Gabor-like expansion in which the windows are differently scaled by the dilation operators $\mathfrak{D}_{\beta_q}$ and the modulation frequencies $\theta^{-1}(2\pi q a)$ are not harmonically related. Moreover, the expansion coefficients $S_{q,n}$ are obtained as in (28) by computing the scalar product of the signal with the unwarped dual Gabor frame $\hat{g}_{q,n}(t)$.

An alternate (dual) scheme consists, in principle, in warping the signal before computing the Gabor coefficients and using these coefficients for the expansion on a conventional Gabor frame. Thus, one can compute the coefficients

$$\tilde{S}_{q,n} = \langle \mathfrak{U}_\theta s, \hat{g}_{q,n} \rangle = \left\langle s, \mathfrak{U}_\theta^\dagger \hat{g}_{q,n} \right\rangle. \tag{34}$$

As shown in the rightmost term of (34), this process is unitarily equivalent to computing the scalar product of the signal over the inversely warped dual frame:

$$\tilde{\hat{g}}_{q,n}(t) = \mathfrak{U}_\theta^\dagger \hat{g}_{q,n}(t) = \mathfrak{U}_{\theta^{-1}} \mathfrak{M}_{2\pi qa} \mathfrak{T}_{n\tau} \hat{g}(t). \qquad (35)$$

Summarizing the results of this Section, approximate frequency warped can be computed by means of Gabor-like expansions in which either the frame elements are approximately warped by scaling and frequency dependent modulation or the dual frame elements are approximately inversely warped by scaling and frequency dependent modulation.

The resulting piecewise warping map approximation is shown in Figure 2 for a very reduced number of bands. There, a warping curve is approximated by tangent linear segments in partly overlapping bands. The edges and centerbands of the ideal uniform bandwidth frame elements can be read on the ordinates axis. These are transformed according to the inverse map θ^{-1} into non-uniformly spaced frequencies as read on the abscissae axis (dotted lines). As a result of the approximation, the band edges are transformed according to the tangent lines and can be traced in the figure as dashed lines.

We remark that, according to the local convexity of the warping map, the bands resulting from the local scaling approximation of warping may either overlap in the frequency domain or there can be gaps. However, as the number of bands increases, the overlapping portions or the gaps become less and less pronounced. Other piecewise linear approximations are possible, e.g., by tracing chord segments joining the band edges. The one we selected exactly warps the centerband frequencies.
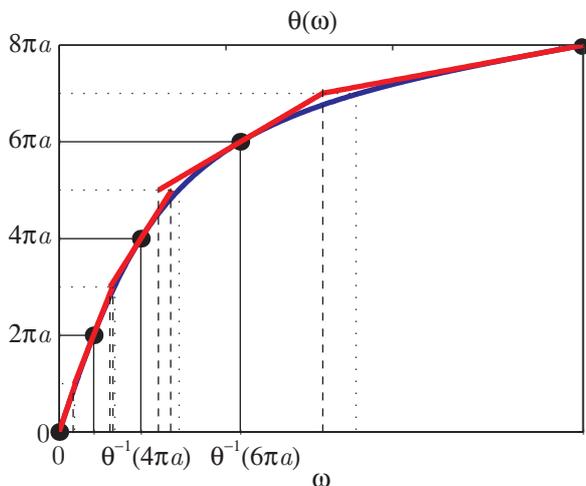
Figure 2: *Piecewise linear approximation of warping map with a very small set of points: thin solid lines denote center bands and dotted lines denote band edges(initial and warped).*

4. DISCRETE-TIME FREQUENCY WARPING

The discrete-time counterpart of the frequency warping approximation easily follows from the continuous time version discussed in the previous sections. The discrete-time Gabor set requires only a finite number of bands M to cover the normalized frequency range $\omega \in [-\pi, +\pi)$, with frequency resolution $a = 2\pi/M$. The discrete-time Gabor frame elements $g_{q,n}(r)$ are obtained by shifting and modulating a unique window sequence $g(n)$, where only integer shifts multiple of an integer $N \leq M$ are allowed, thus

$$g_{q,n}(r) = \mathfrak{M}_{\frac{2\pi q}{M}} \mathfrak{T}_{nN} g(r) = e^{j\frac{2\pi q}{M}r} g(r - nN), \qquad (36)$$

for $q = 0, 1, ..., M - 1$ and $n \in \mathbf{Z}$, and similarly for the dual frame $g_{q,n}(r)$.

A few remarks on the warping maps are however necessary. In fact, since frequency warping transforms bandwidths, the restriction of an invertible analog frequency map to the interval $[-\pi, +\pi)$ is not necessarily one-to-one and onto over this interval. For a strictly increasing map mapping zero frequency to zero frequency one needs to require that $\pm\pi$ is mapped to $\pm\pi$. If the map maps $[-\pi, +\pi]$ one to one onto a smaller interval then aliasing-like phenomena occur, similar to downsampling, unless the original signal sequence is suitably bandlimited to the smaller interval $\theta([-\pi, +\pi))$. Vice versa, if the map maps $[-\pi, +\pi]$ one to one onto a larger interval then spectral replication occurs, similar to upsampling, unless the domain of the map is restricted to the smaller interval $\theta^{-1}([-\pi, +\pi))$. In other words, in order to avoid both aliasing and imaging one needs to restrict both domain and co-domain of the warping map to the interval $[-\pi, +\pi]$, keeping in mind that there can be frequencies that are not mapped by any frequency or frequencies that do not map to any frequency. In the latter case, only the scalar products of the signal with a subset of the Gabor frame elements, with respect to the frequency index q, are required in order to compute warping.

The choice of the analysis and synthesis windows, respectively $\hat{g}(n)$ and $g(n)$, is arbitrary, provided that the discrete-time counterpart of (27) is satisfied. Ideally, the synthesis window should match eigenfunctions of the incremental warping operator $\mathfrak{U}_\psi$ as in (22). However, the Fourier transform of an eigenfunction of the unitary warping operator corresponds to the square root of the derivative of an eigenfunction of the composition operator $\mathfrak{C}_\psi$, up to an additive constant. Using the theory developed in [15], we were able to prove that the only eigenfunctions of $\mathfrak{U}_\psi$ with eigenvalue 1 are constant, corresponding to a Dirac pulse at $\omega = 0$ in the frequency domain. In order to approach this shape, we can select the window to be narrowband lowpass, which leads to conventional window design in short-time spectral analysis. We let M be an integer multiple of N, i.e., $M = KN$, with K a small integer (usually $K = 2$ or 3), and determine a length M lowpass symmetric window $g(n)$ such that

$$\sum_{n \in \mathbf{Z}} g^2(r - nN) = \frac{1}{KM}. \qquad (37)$$

In this case, by choosing identical analysis and synthesis windows: $g(n) = \hat{g}(n)$, one obtains a tight frame, i.e., a frame with bounds $A = B = 1$ in (27). This is similar to orthogonal bases, however, the frame is complete but not a basis. A popular choice is

$$g^2(r) = \frac{1 - \cos\left(\frac{2\pi r}{M}\right)}{KM}; \; r = 0, 1, ..., M - 1, \qquad (38)$$

i.e., $g^2(r)$ is the Von Hann window and

$$g(r) = \sqrt{\frac{2}{KM}} \sin\left(\frac{\pi r}{M}\right); \; r = 0, 1, ..., M - 1 \qquad (39)$$

is the sine window, which we will enforce in this paper. One of the advantages of this choice is that the sinusoidal form of the window

leads to a single frequency term, which warps to another single frequency term. Therefore, except for terms deriving from the finite length of the window, the shape of the warped window is very close to that of a dilated window.

Computation of the expansion coefficients

$$S_{q,n} = \langle s, g_{q,n} \rangle = \sum_{r \in \mathbf{Z}} s(r) e^{-j \frac{2\pi}{M} qr} g(r - nN) \quad (40)$$

can be performed using the analysis section of the filter bank structure in Figure 3, which corresponds to the analysis section of a phase vocoder [16, 17, 8], with impulse responses

$$g_q(r) = e^{j \frac{2\pi}{M} qr} g(M - r) = e^{j \frac{2\pi}{M} qr} g(r). \quad (41)$$

A delay of $K - 1$ samples is introduced in the coefficient sequence $S_{q,r}$ in order to make computation causal.

The approximate synthesis of the warped signal requires a discrete counterpart of (33). The discrete-time window can be considered as the samples of a continuous-time function $g(t)$ taken at unit sampling rate. In order to form the discrete-time warping synthesis windows one can apply the dilation operation $\mathfrak{D}_{\beta_q} g(t)$ to the continuous-time window and then sample at unit sampling rate. This operation yields the window

$$h_q(r) = \sqrt{\frac{1}{\beta_q}} g\left(\frac{r}{\beta_q}\right). \quad (42)$$

However, the result is sharper if only the window length M is modulated to some other integer M_q. To the purpose we let M_q be the closest integer to $\beta_q M$ and we let

$$h_q(r) = \sqrt{\frac{M}{M_q}} g\left(\frac{rM}{M_q}\right); \ r = 0, 1, ..., M - 1. \quad (43)$$

Similarly, for the discrete-time counterpart of the shift operator $\mathfrak{T}_{\beta_q \tau}$, we enforce the integer shifts $nN_q = nM_q/K$. Both dilation and shifting operations are quantized in the discrete-time warping algorithm, which is a source of error that can be controlled by choosing a sufficiently large window length M and a sufficiently small overlap factor K. This is preferable with respect to the non-integer choice, which causes amplitude modulation since the window would not satisfy the overlap add identity (37). In this form, the approximate warped signal can be obtained as output of the synthesis filter bank in Figure 3. There, the analysis coefficients are phase corrected, upsampled with a different upsampling rate in each channel and interpolated by unequal length modulated windows

$$\tilde{g}_q(r) = \mathfrak{M}_{\theta^{-1}\left(\frac{2\pi q}{M}\right)} h_q(r) = e^{-j\theta^{-1}\left(\frac{2\pi q}{M}\right) r} h_q(r). \quad (44)$$

The phase correction terms

$$\gamma_q = N_q \theta^{-1} \left(\frac{2\pi q}{M}\right) - \frac{2\pi q N}{M}, \quad (45)$$

which are the discrete analogues of (14), originate from the filter bank implementation of the phase vocoder, where the modulating terms in the frame elements need to be time shifted for both the analysis and synthesis procedures to be put in the form of (resampled) convolution. To the purpose, we use the following commutation rule:

$$\mathfrak{M}_{\omega_0} \mathfrak{T}_\tau = e^{j\omega_0 \tau} \mathfrak{T}_\tau \mathfrak{M}_{\omega_0} \quad (46)$$

in both the analysis and synthesis. While in conventional phase vocoders the phase correction factors cancel out, in reason of the heterogeneous analysis and synthesis sections these factors are indeed necessary for correct computation in the approximate warping filter bank scheme.

An alternate structure, with similar properties, to that shown in Figure 3 can be worked out from (34) and (35). It consists of an approximate inverse warping analysis section, followed by a conventional phase vocoder synthesis section. Furthermore, an extension to rational numbers N_q and M_q is possible, which achieves a smaller error but requires a more costly implementation in terms of upsamplers and downsamplers by possibly high factors.

5. PERFORMANCE

In this Section we give a brief account of the performance of the approximate warping algorithm, in terms of operation count and numerical and perceptual error.

5.1. Computational Cost

The analysis section of the approximate warping filter bank in Figure 3 can be efficiently implemented with a length M FFT, with a cost of $O(M \log M)$ operations, which are repeated every N input samples, for a total rate of $O(K \log M)$ operations per sample. The synthesis filter bank section cannot be efficiently computed by means of FFT since the modulating frequencies of the warped frame elements are not harmonically related. For real signals, only half of the filters in the bank must be computed since the other half are complex conjugated versions applied to complex conjugated coefficient sequences. Therefore, a set of $M/2$ real impulse responses can be derived by combining each pair of complex conjugates sections of index q and $M - q$. For each expansion coefficient, the generic section q filter generates N_q output samples; the other $M_q - N_q = (K - 1)N_q$ samples need to be prepared for the overlap with subsequent outputs in the overlap-add scheme. This operation requires M_q multiplies of the coefficient $S_{q,r}$ times the corresponding precomputed synthesis window. Thus, each filter section requires $M_q/N_q = K$ operations per output sample. For M even there are $M/2 + 1$ sections and for M odd there are $\lfloor M/2 \rfloor + 1$ sections. In each length N_q output segment there are K slices of shifted windows to be added together. Therefore, the total cost of the complete synthesis section is $(\lfloor M/2 \rfloor + 1) K$ multiplies and $\lfloor M/2 \rfloor (K - 1)$ adds per sample. Notice that in this computational scheme the phase correction factors are not needed as they are embedded in the modulated windows both in the analysis and in the synthesis.

For a finite-length input signal, the overall cost of the approximate frequency warping structure grows linearly. This should be compared with the non-causal allpass cascade structures in classical computation of the Laguerre transform for frequency warping [18, 14], whose complexity grows quadratically with the number of samples.

5.2. Numerical and Perceptual Error

A direct estimate of the warping approximation error bounds provides the following characteristics: both the ℓ^2 error norm and the absolute error decrease as $M^{-3/2}$ as M grows. Therefore, the error can be reduced by choosing a sufficiently wide window. Moreover, due to the fact that the approximation of the delays tends to
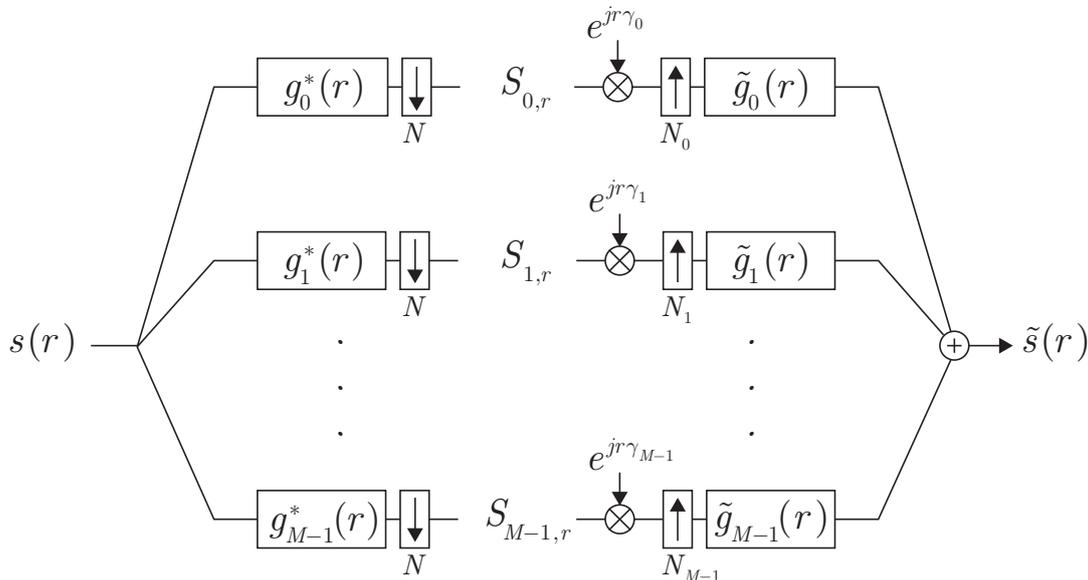
Figure 3: *Efficient warping structure based on Gabor filter bank for the analysis and approximated warped Gabor filter bank for the synthesis.*

cumulate in time, the error grows with the length of the input signal. With this respect, for fixed window length M, our estimates provide an error trend of $O(L^{5/4}/M^2)$ as the length L of the signal grows. For example, for a window length of 2400 samples, the error remains confined within 10% for about $1\ s$ at 44.1 kHz sampling rate. The error is also proportional to the sum of the coefficients β_q, showing that the error depends on the steepness of the warping curve. Moreover, the influence of the overlap factor over the error decays modestly as $K^{-1/2}$. However, since N decreases as K grows, the heavier quantization of the delays and of the window shifts tends to increase the error. Therefore, from the error point of view increasing K is not beneficial.

We extensively tested the theoretical error estimates against the numerical error. In this task we employed a number of sources like noise and tones of musical instruments with sharp or slow attack and decay. In the comparison we used the one-parameter family of Laguerre maps, as only in this case an exact algorithm to compute warping is available. The experimental results confirm the theoretical estimates and show proportionality with the given trends by small constants, confirming that the theoretical estimates provide worst case bounds.

Additionally, we evaluated the perceptual error introduced by warping. For the purpose, for a set of instrument sounds, we evaluated the error of the approximation with respect to Laguerre maps. Using a procedure similar to that found in MPEG coders, we computed the masking curves of the signal over the error. As the error is coherent with the signal, masking is likely to occur. This is due to the fact that the most relevant part of the error originates from the approximation of frequency dependent delays as constant

within narrow frequency bands and from their quantization. In our results, the approximation error rarely exceeds the masking threshold. In sounds of approximately $1\ s$ duration and using a window length of $M = 2400$ samples, the masking threshold overshooting occurred in less than 1% of the frames, mostly located at the attack and at the decay segments of the tones.

6. APPLICATIONS

The frequency warping effect introduces a coloration in the signal in which harmonically related partial can become more or less inharmonic, with consequent beating and floating. The effect is per se interesting and can be applied to the sounds of several instruments and, especially, strings. Examples employing the proposed algorithm on a variety of sounds, together with a comparison with the exact algorithm in the available cases (Laguerre maps) can be found at the following URL: http://www.itn.liu.se/~giaev/soundexamples.html.

With properly selected maps, frequency warping can be used in physical models as an alternative to dispersive delay lines. In fact, a dispersive delay line can be considered as a warped delay line in which the elementary delay (or a group of them) is replaced by an allpass filter. The structure is therefore equivalent to a delay line cascaded with a warping section, provided that all the inputs are inversely warped. A dispersive waveguide may contain several dispersive digital lines, each going from the excitation input to the boundaries or vice versa and to and from the output pick-up point. Often the inputs are control signals that can be either prewarped

off-line or, in the case of noise, are unaffected by warping. Each dispersive chunk is formed by a chain of allpass filters, or, what is equivalent, by a high order all-pass filter. The design of heterogeneous length allpass filters simulating the proper dispersion is very hard and still an open problem [12]. Rather, it may be convenient to simulate a non-dispersive waveguide, moving dispersion outside the closed loop, which is accomplished by warping the output signal.

In contrast to the one parameter family of Laguerre maps in allpass cascade implementations, the proposed warping algorithm leaves ample freedom in the choice of warping map. In fact, fixed the window length M, both the discrete set of warped frequencies $\omega_q = \theta^{-1}(2\pi q/M)$ and the slopes β_q can be arbitrarily selected. The latter control the scale of the warped signal about the corresponding frequency ω_q, where large values of β_q generate longer signals. This property can be used in digital audio warping effects in order to change the decay rates of the output signal in a frequency dependent fashion.

While the cascade allpass structure for Laguerre warping are strongly non-causal and impractical for real-time computation, the proposed warping algorithm lends itself to real-time computation. However, the limitation of our algorithm toward real-time is that the upsampling rates N_q in each channel of the synthesis structure should never be smaller than the corresponding downsampling rate N of the analysis section, otherwise data would be missing for the computation of the current sample. Since the ratio $N_q/N \approx \beta_q$, then (see (32)) a condition for real-time operation is that the first derivative of the warping map should not be smaller than 1. Warping maps having this characteristic bring frequencies in $[-\pi, +\pi]$ to frequencies in a smaller interval around zero frequency.

In the synthesis by physical models, the previous limitation can be circumvented by calibrating the non-dispersive waveguide to a much higher pitch than the target one, so that the tone is brought back to the desired height by warping. If the warping map has a specific form $\hat{\theta}(\omega)$, as dictated by physical laws or by experimental measures, one can easily transform this into a map having derivative greater than one simply by multiplying it by a constant $\alpha > 1$. With the new map $\theta(\omega) = \alpha\hat{\theta}(\omega)$, a frequency ω_0 warps to frequency $\tilde{\omega}_0 = \theta^{-1}(\omega_0/\alpha)$. Therefore, in order to achieve the desired frequency $\omega_d = \hat{\theta}^{-1}(\omega_0)$ one simply needs to start with frequency $\alpha\omega_0$. The linearity of this relationship ensures that the relationships among the desired frequencies, e.g., the harmonics to be warped to inharmonic partials with a given law, is not altered. The only limitation is therefore the reduced frequency range due to the $[-\pi, +\pi]$ clipping of the scaled warping map. This can be handled by oversampling.

In digital audio effects, the above limitation toward real-time may be too severe. However, a workaround is possible by embedding pitch-shifting in the modified phase vocoder scheme. This is indeed directly possible by frequency bin reassignment and by phase correction in the analysis section [19, 20].

The latency of the approximate warping scheme is proportional, via the sampling interval, to the window shift integer parameter N. The latter is the minimum number of samples in order to produce a set of coefficients, assuming that the previous $(K-1)N$ samples are known or zero. For a fixed window length M, one can reduce N by increasing the overlap factor K. However, as we have seen in Section 5, reducing N increases the computational cost and the quantization error of the delays and window shifts in the synthesis section. This results in lower adherence to the established warping rule but in no other annoying effect.

Therefore, a trade-off between latency and precision needs to be achieved. For $K = 2$ and $M = 2200$, latency is estimated at about $25\ ms$ at a sampling rate of 44.1 kHz.

7. CONCLUSION

In this paper we introduced an efficient scheme for frequency warping audio signals, which overcomes several limitations of allpass chain based systems, notably those of the restricted family of achievable maps and the non-causality of the computational structure. The approximate warping algorithm was shown to introduce a negligible error, both numerical and perceptual, which can be controlled by proper choice of the length of the window. With some limitations and workarounds, the algorithm has a real-time implementation and it can be suitable as a digital audio effect or in the synthesis by physical models.

8. REFERENCES

[1] A. Constantinides, "Spectral transformations for digital filters," *Proc. Inst. Elec. Eng.*, vol. 117, pp. 1585–1590, Aug. 1970.

[2] GRM-INA, "FreqWarp plugin," http://www.grmtools.org/quicktour/vstqtst/warp.html.

[3] Prosoniq, "π-Warp plugin," http://products.prosoniq.com/cgi-bin/register?service=showdetail\&refno=36.

[4] G. Evangelista, "The short-time Laguerre transform: a new method for real-time frequency warping of sounds," in *Proc. of the Int. Computer Music Conf. (ICMC 2000)*, Berlin, Germany, Aug. 2000, pp. 380–383.

[5] G. Evangelista, "Time and frequency warping of musical signals," in *Digital Audio Effects*. Wiley, Chichester, UK, 2002.

[6] D. Gabor, "Theory of communication," *J. IEE*, vol. 93, no. 26, pp. 429–457, 1946.

[7] J. J. Benedetto and D. F. Walnut, "Gabor frames for L^2 and related spaces," in *Wavelets: Mathematics and Applications*, J.J. Benedetto and M.W. Frazier, Eds., pp. 97–162. CRC Press, 1994.

[8] M. E. Portnoff, "Time-frequency representations of digital signals and systems based on short time Fourier analysis," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 28, pp. 55–69, 1980.

[9] G. Evangelista and S. Cavaliere, "Discrete frequency warped wavelets: Theory and applications," *IEEE Trans. on Signal Processing*, vol. 46, no. 4, pp. 874–885, April 1998, Special issue on Theory and Applications of Filter Banks and Wavelets.

[10] G. Evangelista and S. Cavaliere, "Audio effects based on biorthogonal time-varying frequency warping," *EURASIP Journ. on Applied Signal Processing*, vol. 2001, no. 1, pp. 27–35, March 2001.

[11] I. Testa, G. Evangelista, and S. Cavaliere, "Physically inspired models for the synthesis of stiff strings with dispersive

waveguides," *EURASIP Journ. on Applied Signal Processing*, vol. 2004, no. 7, pp. 964–977, July 2004, special issue on Model-Based Sound Synthesis.

[12] J. S. Abel and J. O. Smith, "Robust design of very high-order allpass dispersion filters," in *Proc. of the Int. Conf. on Digital Audio Effects (DAFx-06)*, Montreal, Quebec, Canada, Sept. 18–20, 2006, pp. 13–18.

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

[14] A. V. Oppenheim, D. H. Johnson, and K. Steiglitz, "Computation of spectra with unequal resolution using the Fast Fourier Transform," *Proc. IEEE*, vol. 59, pp. 299–301, Feb. 1971.

[15] P. S. Bourdon and J. H. Shapiro, "Mean growth of Koenigs eigenfunctions," *J. American Math. Soc.*, vol. 10, no. 2, pp. 299+325, April 1997.

[16] M. Dolson, "The phase vocoder: A tutorial," *Comp. Music J.*, vol. 10, no. 4, pp. 14–27, 1986.

[17] J. L. Flanagan and R. M. Golden, "Phase vocoder," *Bell System Tech. J.*, pp. 1493–1509, November 1966.

[18] P. W. Broome, "Discrete orthonormal sequences," *J. Assoc. Comput. Machinery*, vol. 12, no. 2, pp. 151–168, 1965.

[19] J. Laroche and M. Dolson, "New phase vocoder technique for pitch-shifting, harmonizing and other exotic effects," in *Proc. IEEE Workshop on Appl. Signal Proc. Audio and Acoustics*, New Paltz, NY, 1999.

[20] M. Puckette, "Phase-locked vocoder," in *Proc. IEEE ASSP Conf. Appl. Signal Proc. Audio and Acoustics*, New Paltz, NY 1995.