# TIME WARPING IN DIGITAL AUDIO EFFECTS

*Gianpaolo Evangelista*

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

## ABSTRACT

Time warping is an important paradigm in sound processing, which consists of composing the signal with another function of time called the warping map. This paradigm leads to different points of view in signal processing, fostering the development of new effects or the conception of new implementations of existing ones. While the introduction of time warping in continuous-time signals is in principle not problematic, time warping of discrete-time signals is not self-evident. On one hand, if the signal samples were obtained by sampling a bandlimited signal, the warped signal is not necessarily bandlimited: it has a sampling theorem of its own, based on irregular sampling, unless the map is linear. On the other hand, most signals are regularly sampled so that the samples at non-integer multiples of the sampling interval are not known. While the use of interpolation can partly solve the problem it usually introduces artifacts. Moreover, in many sound applications, the computation already involves a phase vocoder. In this paper we introduce new methods and algorithms for time-warping based on warped time-frequency representations. These lead to alternative algorithms for warping for use in sound processing tools and digital audio effects and shed new light in the interaction of time warping with phase vocoders. We also outline the applications of time warping in digital audio effects.

## 1. INTRODUCTION

Time warping is, in principle, a simple operation consisting in the composition of the time signal with another function of time called the warping map. As a result, the signal is deformed and its plot vs. time appears as if the original time axis had been warped.

Together with its dual operation defining frequency warping, time warping allows for the introduction of new or known effects, which are often described adopting different points of view, and allows for the mapping of signal representations into other representations. For example, the introduction of vibrato in a signal can be either seen as a result of passing the signal through a time-varying delay line or as a result of time warping the signal. Adding the original signal to a collection of time warped versions of the same signal, one can achieve different realizations of flanging and chorus effects. Even frequency or phase modulation as in FM synthesis could be considered as a version of time warping, where the warping map is one-to-one only for small values of the modulation index.

Used in conjunction with time-frequency or time-scale representations, invertible time and frequency warping help allocating analysis time intervals and / or frequency bands which differ from the ones provided by the original representative elements [1, 2, 3, 4, 5]. This makes it possible to obtain, e.g., non-uniform resolution from the uniform resolution of the original representation, or more sophisticated allocations than the ones sketched by rigid and simplified mathematical rules. In recent times, non-stationary Gabor frames were introduced [6, 7] and linked to time and frequency warping operators subject to redressing methods [8, 9].

In phase vocoder based schemes, time-warping of the windows can be considered as a building block for time stretching sound signals; another building block being the adjustments of the phases to provide alignment of the sinusoidal components across the overlapping stretched windows. In the most conventional implementations, a constant stretching as defined by a linear time-warping map is applied. However, the use of a piecewise linear or curvilinear map generally achieves better results in which, e.g., only the stationary part of the signal is stretched or compressed while the transients, especially at the attack of sounds, are left unaltered.

This paper is organized as follows. In Section 2 we recall the definition of warping as an operator and of its unitary version. We explore useful maps for audio processing and effects in Section 2.1 and evaluate the warped sampling expansion as an algorithm for time warping in Sections 2.2, 2.3 and 2.4. In Section 3 we introduce original methods for time-warping and consider the interaction of time-warping with Gabor frames or phase vocoder in Section 3.1. Two original algorithms for warping are shown in Section 3.2 and 3.3, respectively, where the results of experimentation shown in Section 3.4 provide an assessment of the SNR with test signals together with an analysis of the computational costs found in Section 3.5. In Section 4 we give a brief outline of the use of time warping for time stretching and pitch shifting audio signals. Finally, in Section 5 we draw our conclusions.

Examples and experimental code will be made available at the author's web page:
http://members.chello.at/~evangelista/
under the Sound Examples tab - Time Warping.

## 2. TIME WARPING OPERATORS

Given a function of time $\gamma$, which will play the role of the warping map, a time warping operator $\mathbf{W}_\gamma$ is identical to a composition-by-$\gamma$ operator $\mathbf{C}_\gamma$ acting in the time domain. Thus, we have:

$$s_{tw} = \mathbf{W}_\gamma s = \mathbf{C}_\gamma s = s \circ \gamma, \tag{1}$$

where $s_{tw}$ is the time-warped version of the signal $s$, $\gamma$ is the time warping map and $\circ$ denotes function composition. Thus, for any signal $s(t)$ we have

$$s_{tw}(t) = s(\gamma(t)). \tag{2}$$

Conditions that guarantee the boundedness and invertibility of the warping operators can be found in [8] and references therein.

The conditions for the definition of unitary warping / composition operators are generally less strict than those for the boundedness and invertibility of the non-unitary warping operators (see [8] and references therein). If the warping map $\gamma$ is almost everywhere strictly increasing, one-to-one and differentiable then one can define a unitary time-warping operator $\mathbf{U}_\gamma$ simply by multiplying the non-unitary operator $\mathbf{W}_\gamma$ in (1) by the square root of the magnitude derivative of the map, in which case:

$$s_{tw}(t) = \left[\mathbf{U}_\gamma s\right](t) = \sqrt{\left|\frac{d\gamma}{dt}\right|} s(\gamma(t)). \tag{3}$$

For simplicity, we assume that the warping maps $\gamma$ of interest are almost everywhere increasing, so that they are invertible [8], and that both the first derivatives of $\gamma$ and $\gamma^{-1}$ are essentially bounded from below. Since the maps are increasing, their derivatives are positive so that the magnitude sign under the square root in (3) can be dropped.

### 2.1. Some Maps of Interest for Audio Effects

Time warping with arbitrary maps can be used per se as an audio effect, introducing simultaneous pitch modulation and local or global stretching or compression of the signal. The warped signal can also be mixed with the original signals and / or with other differently warped versions of the signal. As it will be shown in Sections 3 and 4, time warping can also be employed in conjunction with time-frequency representations in order to obtain alternative algorithms for warping and to build modified phase vocoders for time stretching or pitch shifting of audio. In this section we illustrate some time-warping maps that are of interest for the construction of new or known audio effects.

Usually we are only interested in the shape of the map for nonnegative values of time. If needed, in order to define a map over the entire real axis we may extend the generic map by enforcing odd parity: $\gamma(-t) = -\gamma(t)$.

The simplest time-warping map is linear, also known as affine transformation:

$$\gamma_{lin}(t) = \alpha t + c, \tag{4}$$

with inverse

$$\gamma_{lin}^{-1}(t) = \frac{t-c}{\alpha}, \tag{5}$$

where $\alpha \neq 0$ and $c$ are constants.

In the linear map (4), the parameter $\alpha$ is usually chosen as positive in order to maintain the direction of time, while a negative value produces time reversal effects useful, for example, in granular synthesis. With the generic linear map, each sinusoidal component of the signal at frequency $f_0$ is brought to frequency $f_1 = \alpha f_0$. Time-wise, the signal is dilated by a factor $1/\alpha$, which means it is stretched if $\alpha < 1$ and compressed if $\alpha > 1$.

Usually, one selects $c = 0$ in order to map the time origin into itself. However, it is always possible to let $c$ be a negative number in order to introduce a time delay, which might be useful, for example, for online computation of time warping. Indeed, as shown in Fig.1, the identity line $g(t) = t$, which represents the present (the loci of time instants that map into themselves), divides the past (the loci of time instants that map into previous times) from the future (the loci of time instants that map into subsequent times). In the same figure, shown is the map $\gamma(t) = t - d$, where $d$ is a positive number, which completely lies in the past and introduces a uniform delay $d$. For a nonlinear warping map $\gamma(t)$, such that $\gamma(t) - t$ is bounded, it is possible to introduce a delay through composition with a linear delay map so that the whole map does not have points belonging to the future, which makes causal computation possible. That is, given the map $\gamma(t)$, such that $\gamma(t) > t$ in some region, we form the causal map $\tilde{\gamma}(t) = \gamma(t) - d$ where $d \geq \sup_t(\gamma(t) - t)$, which has no points in the future.

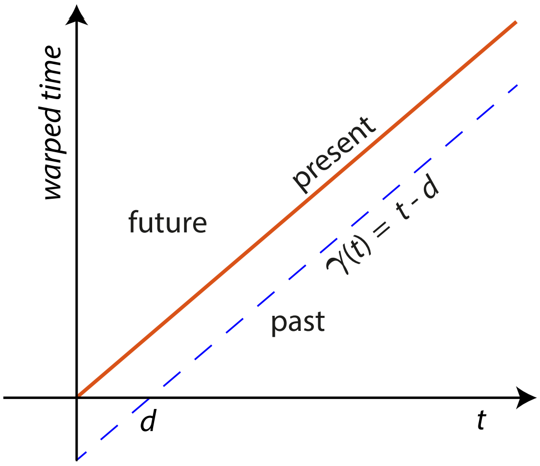

Figure 1: *Subdivision of the time-warped time plane into past, present and future. Also shown is a linear map introducing a delay d.*

Piecewise linear maps can also be exploited in order to produce dynamic warping effects in a simple way, where several versions of (4) and (5) are used on contiguous finite disjoint intervals and their mapped intervals, respectively. In that case, the constant $c$ of each map is chosen to guarantee continuity, where the initial value of the map matches the final value of the previous adjacent map.

Another class of maps of interest is given by the chirps, examples of which are

$$\gamma_l(t) = t + \beta_l t^2, \tag{6}$$

which, when applied to a sinusoidal signal component produces a linear chirp and

$$\gamma_q(t) = t + \beta_q t^3, \tag{7}$$

which gives a quadratic chirp.

When applied to audio signals, the chirp maps produce glissandos. The linear chirp map brings any sinusoidal component of the signal of frequency $f_0$ to a signal having instantaneous frequency $f_0(1 + 2\beta_l t)$. The parameter $\beta_l$ can be selected to achieve a target frequency $f_1$ after a time interval of duration $\tau$, in that case we set $\beta_l = (f_1 - f_0)/2f_0\tau$. It is convenient to express $\beta_l$ in a form that does not depend on $f_0$ as follows:

$$\beta_l = \frac{\rho - 1}{2\tau}, \tag{8}$$

where $\rho$ it the ratio of frequency change in the lapse $\tau$. Only values of $\rho > 1$, corresponding to $\beta_l > 0$, guarantee the invertibility of the map at all times. This is the case of upward chirps where the frequency increases. Downward chirps with $\beta_l < 0$ can still be used on finite length intervals provided that one checks, for invertibility, that the derivative of the map does not change sign.

For $\beta_l > 0$ the inverse of the linear chirp map is

$$\gamma_l^{-1}(t) = \frac{-1 + \sqrt{4\beta_l t + 1}}{2\beta_l}. \tag{9}$$

This map can also be used, by exchanging the roles of the direct and inverse maps, to produce downward chirps, which are, however, not linear.

The quadratic chirp warping map (7) dynamically maps the frequency $f_0$ to the instantaneous frequency $f_0(1 + 3\beta_q t^2)$. Here again, for complete invertibility we require $\beta_q > 0$ and we can express this parameter in terms of the frequency change ratio $\rho$ in the lapse $\tau$ as follows:

$$\beta_q = \frac{\rho - 1}{3\tau^2}, \tag{10}$$

with $\rho > 1$ for an upward chirp. For $\beta_q > 0$, the inverse of the quadratic chirp map is given as follows:

$$\gamma_q^{-1}(t) = \frac{\sqrt[3]{\frac{2}{3\beta_q}}}{Q_{\beta_q}(t)} - \frac{Q_{\beta_q}(t)}{\sqrt[3]{18\beta_q^2}} \tag{11}$$

which gives a downward chirps, where

$$Q_{\beta_q}(t) = \sqrt[3]{\sqrt{81\beta_q^2 t^2 + 12\beta_q} - 9\beta_q t}. \tag{12}$$

To conclude our exploration of relevant warping maps, we consider the phase modulation map

$$\gamma_{pm}(t) = t + I_m sin(2 * \pi f_m t), \tag{13}$$

where $f_m$ is the modulating frequency and $I_m$ is the modulation index expressed in multiples of the carrier frequency, i.e., of the frequency $f_c$ of the sinusoidal component of the signal to which the map is applied. As it is easy to check from its derivative, this warping map is invertible only if $I_m < 2\pi f_m$. It is useful for producing vibratos, for small values of the modulating frequency and for phase modulating audio signals as a special FM effect where the carrier is not a necessarily single sinusoid. The inverse map of (13) cannot be expressed in closed form. When an inverse is desired, one can resort to linear interpolation from the direct map by exchanging the abscissae with the ordinates. Alternatively, one can numerically find for each time point $t$ of interest the zero of the function $\gamma(x) - t$.

Yet another possibility is given by the approximation of the phase modulation map by means of the invertible map

$$\gamma_{apm}(t) = t + \frac{1}{\pi f_m} \tan^{-1}\left(\frac{b_m \sin(2\pi f_m t)}{1 - b_m \cos(2\pi f_m t)}\right), \tag{14}$$

which is inspired from the phase response of a first order real allpass filter. The inverse map of (14) can be readily found by changing the sign of $b_m$. This parameter controls the modulation index and can be optimized for the map to approximate (13). By matching the points of maximum deflection from the linear component $t$ of (13) with the values of (14) at the same points, i.e. at the points $2\pi f_m t = \pi/2 + 2k\pi$, $k \in \mathbb{Z}$, one can see that a good estimate for $b_m$ is $\hat{b}_m = \tan(\pi f_m I_m)$. The detail of the approximation of the phase modulation map with the map is shown in Fig.2. This way we obtain an invertible map in closed form that is more practical for phase modulation effects.

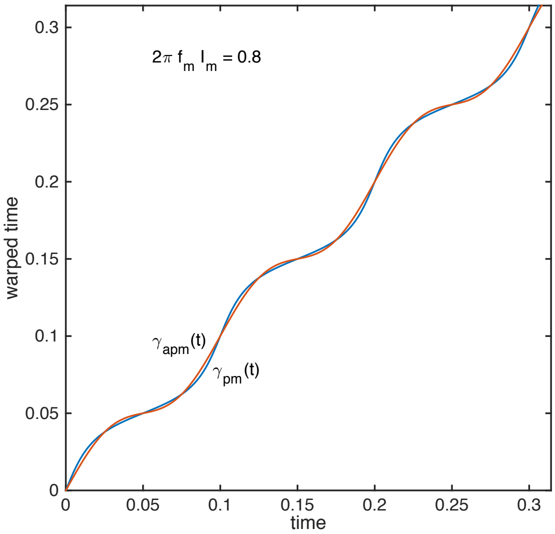

Figure 2: *Detail of the approximation of the phase modulation map with the map $\gamma_{apm}$ in (14).*

### 2.2. Sampling Theorem for Time-Warped Bandlimited Signals

Assume that $s_{tw}(t) = s(\gamma(t))$ as in (2), where the function $\gamma$ is invertible. If the original signal $s(t)$ is bandlimited to $\left]-\frac{f_s}{2}, +\frac{f_s}{2}\right[$, where $f_s/2$ is the Nyquist frequency, then it admits a sampling reconstruction formula

$$s(t) = \sum_n s(nT)\text{sinc}\left(\frac{t}{T} - n\right), \tag{15}$$

where $\text{sinc}\,(t) = \frac{\sin \pi t}{\pi t}$ and $T = 1/f_s$ is the sampling interval. With the simple observation that $s_{tw}(\gamma^{-1}(t)) = s(t)$ one can conclude that $s(nT) = s_{tw}(\gamma^{-1}(nT))$. Thus, by time-warping both sides of (15) one obtains the following sampling reconstruction for the warped signal:

$$s_{tw}(t) = \sum_n s_{tw}(\tau_n)\text{sinc}\left(\frac{\gamma(t)}{T} - n\right), \tag{16}$$

where $\tau_n = \gamma^{-1}(nT)$ are sampling instants that are not regularly spaced unless the map is linear. However, as conjectured in [10], if $\gamma(t)$ is an invertible function, the time-warped signal $s(\gamma(t))$ is not guaranteed to be bandlimited, unless $\gamma(t)$ is an affine map (4). This conjecture is shown to be true for a wide class of maps, including the ones arbitrarily close to a linear map and the piecewise linear ones [11, 10, 12].

The warped sampling expansion (16) constitutes an important algorithm for time-warping discrete-time signals. In fact, knowing that $s(nT) = s_{tw}(\tau_n)$ and disregarding possible aliasing, one can compute the discrete-time time-warped signal by evaluating (16) at uniformly spaced sampling instants $t_r = rT$, for any $r \in \mathbb{Z}$:

$$s_{tw}(rT) = \sum_n s(nT)\text{sinc}\left(\frac{\gamma(rT)}{T} - n\right). \tag{17}$$

Since in computations the sinc function is impractical as it extends over the entire time axis, one can approximate it by a windowed sinc interpolating kernel like the Lanczos kernel

$$\varphi_L(t) = \begin{cases} \operatorname{sinc}\left(\frac{t}{L}\right)\operatorname{sinc}(t) & t \in [-L, +L[ \\ 0 & \text{otherwise} \end{cases} \tag{18}$$

or the von Hann windowed sinc [13]:

$$\varphi_L(t) = \begin{cases} \cos^2\left(\frac{t}{2L}\right)\operatorname{sinc}(t) & t \in [-L, +L[ \\ 0 & \text{otherwise} \end{cases} \tag{19}$$

Here $L$ is an integer parameter which controls the extension of the approximation interval. In both cases, and in many other similar choices of window function, we have $\lim_{L\to\infty} \varphi_L(t) = \operatorname{sinc}(t)$. Thus,

$$s_{tw}(rT) \approx \sum_n s(nT)\varphi_L\left(\frac{\gamma(rT)}{T} - n\right) \tag{20}$$

is a discrete-time approximation of the time warped signal, which is increasingly better as $L$ grows.

The time and frequency domain characteristics of both the Lanczos interpolating kernel and the von Hann windowed sinc are shown in Fig.3. While the two interpolating kernels are very similar in the time domain, the magnitude Fourier transform of Lanczos' kernel shows a slightly steeper frequency roll-off of the main lobe.

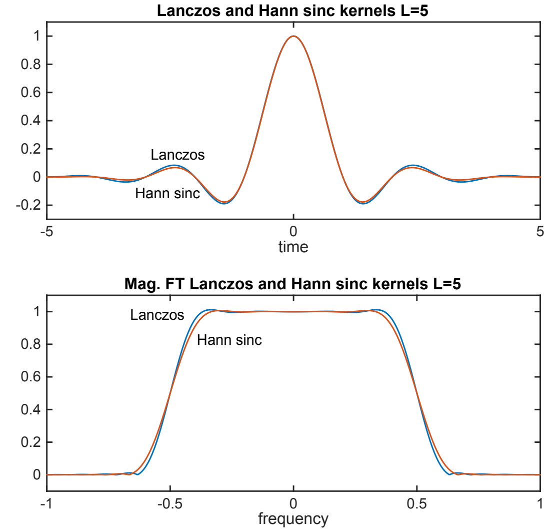

Figure 3: *Lanczos and von Hann windowed sinc interpolating kernels: time domain and magnitude Fourier transforms.*

## 2.3. Experimental Results

In order to assess the quality of the approximation (20) we performed tests with artificial signals which have an analytic closed form and we measured the SNR. The error is estimated as the difference of the signal warped as in (20) and the one obtained by applying the warping map function directly to the artificial signal. Clearly, the results depend on the map and on the signal. In Fig.4 we plot the SNR as a function of the width parameter $L$ for both the Lanczos and the von Hann windowed sinc kernels, choosing as warping map the linear function $\gamma(t) = \alpha t$.

The test signals consisted of 1 KHz sinusoids with smooth envelopes and we chose a sampling rate of 44.1 KHz for the sampling expansion. Varying the parameter $\alpha$ of the warping map, we concluded from the observations that for $\alpha \geq 1$ a 255 dB SNR (not shown in the figure) was achieved in all cases, which means no error above machine precision. The worst cases are for $\alpha \ll 1$; the values of the SNR shown in the figure are computed with $\alpha = 1/16$, which warps the 1 KHz sinusoid down to 67 Hz. The intuition about the lower performance at lower map derivatives lies in the fact that the sample instants $\tau_n$ of $s_{tw}$ in the RHS of (16), of which (20) is an approximation, are linked to the inverse map $\gamma^{-1}(t) = t/\alpha$. Thus, since $\tau_n = nT/\alpha$, their density is lower for smaller values of $\alpha$ and so is the quality of the approximation. However, increasing the sampling rate did not show great benefits. This might be due to the fact that we are dealing with sinusoids still within average Nyquist rate, while increasing the sampling rate proportionally increases the number of the terms in the sampling expansion estimate (20), which brings a proportionally higher approximation error.

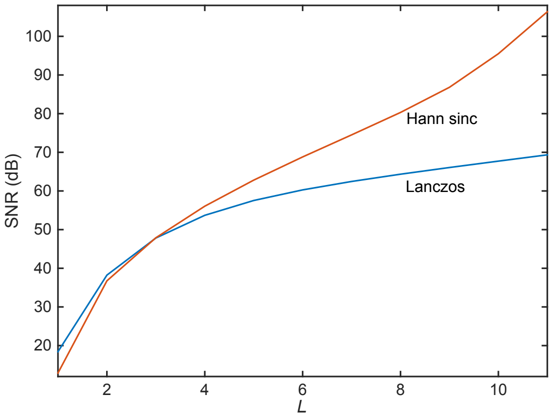

Figure 4: *SNR characteristics of Lanczos and von Hann windowed sinc kernels as a function of the support width parameter L.*

From Fig.4 we remark that for values of $L > 3$ the von Hann windowed sinc outperforms the Lanczos interpolating kernel. A choice of $L = 5$ or larger leads to good approximations where the artifacts are inaudible at 56 dB SNR or better, reaching 106 dB for $L = 11$.

The SNR results in Fig.4 were confirmed up to the first decimal digit in other tests we performed using the linear and quadratic chirp maps described in Section 2.1. We also tried their inverses, which give downwards chirps, achieving similar SNR and acoustic results.

## 2.4. Computational Complexity

The computational complexity of the warped sampling expansion can be easily estimated by observing that in (20), in order to produce the output, any input sample is multiplied by the sinc window kernel and these are added together with the shifted windows of previous and future samples within the window length. The extension of the warped window depends on the warping map. Given an increasing warping map $\gamma$ and the window extension factor $K$, we have that the sampled support of the warped window is the set of $r \in \mathbb{Z}$ such that:

$$\frac{\gamma^{-1}((n-L)T)}{T} \leq r < \frac{\gamma^{-1}((n+L)T)}{T}. \tag{21}$$

For a linear warping map $\gamma(t) = \alpha t$ the width of the support is approximately $2L/\alpha$ samples. For the generic map, in order to compute the average complexity, $\alpha$ can be replaced by the average derivative of the map:

$$\bar{\alpha} = \frac{1}{\Delta} \int_{0}^{\Delta} \frac{d\gamma}{dt} dt = \frac{\gamma(\Delta)}{\Delta}, \tag{22}$$

over a finite interval of duration $\Delta$, where we have assumed $\gamma(0) = 0$, or by its limit as $\Delta \to \infty$, if it exists, for the infinite interval.

In conclusion, the average complexity of the warped sampling theorem based algorithm for time warping is proportional to $2L/\bar{\alpha}$ per sample.

## 3. TIME WARPING AND TIME-FREQUENCY REPRESENTATIONS

The warped sampling expansion illustrated in the previous section is not the only algorithm for time warping signals. In fact, the expansion of the signal in any set of complete orthogonal or biorthogonal functions of time in $L^2(\mathbb{R})$ can be used and so is the expansion into any time domain frame in the same space, provided that either the analysis or the synthesis functions can be expressed in closed analytic form.

In this paper we consider a new approach to the time warping of discrete-time signals based on Gabor frame expansions. While in [14] we considered a similar approach based on filter banks for the computation of frequency warping, time warping has different requirements, which deserve a separate discussion, and the important computational advantage that finite-length windows are still finite-length after warping. Moreover, for audio processing purposes, the filter bank approach to time-warping allows one to control the output bandwidth simply by limiting the number of frequency components.

### 3.1. Gabor Frames

In this paper, we consider the signal expansion over a Gabor frame[1] whose representative elements are obtained by modulating and time-shifting a suitable window function $h_s(t)$:

$$\varphi_{n,m}^{(s)}(t) = \left[\mathbf{T}_{na}\mathbf{M}_{mb}h_s\right](t) = h_s(t-na)e^{j2\pi mb(t-na)}, \tag{23}$$

where $\mathbf{T}_\tau$ is the shift by $\tau$ operator and $\mathbf{M}_\nu$ is the modulation by $\nu$ operator which consists in multiplication by $e^{j2\pi\nu t}$. The constants $a$ and $b$ respectively represent the time-shift sampling interval (hop size) and the frequency sampling interval (distance of the frequency bins), with $ab \leq 1$ a necessary condition for the set $\left\{\varphi_{n,m}^{(s)}(t)\right\}_{n,m\in\mathbb{Z}}$ to form a frame.

We recall that the sequence of functions $\{\varphi_{n,m}\}_{n,m\in\mathbb{Z}}$ in $L^2(\mathbb{R})$ is called a frame if there exist two positive constants $A$ and $B$ such that

$$A\|s\|^2 \leq \sum_{m,n} |\langle s, \varphi_{n,m}\rangle|^2 \leq B\|s\|^2 \quad \forall s \in L^2(\mathbb{R}), \tag{24}$$

where $\|s\|^2 = \langle s, s\rangle$ is the norm square of the signal.

For the Gabor set (23) one can show that perfect reconstruction (PR) is guaranteed if there exists an analysis window $h_a(t)$ such that

$$\sum_n h_a\left(t + \frac{r}{b} - na\right) h_s(t-na) = b\delta_{r,0}, \tag{25}$$

where $\delta_{r,0}$ is the Kronecker delta and the analysis frame is given by

$$\varphi_{n,m}^{(a)}(t) = \left[\mathbf{T}_{na}\mathbf{M}_{mb}h_s\right](t) = h_a(t-na)e^{j2\pi mb(t-na)}. \tag{26}$$

General necessary and sufficient conditions for compact supported and exponentially decaying windows to generate a Gabor frame are given in [15]. For compact supported analysis and synthesis windows with support smaller than $1/b$, the popular overlap-add condition must be satisfied for PR:

$$\frac{1}{b}\sum_n h_a(t-na)h_s(t-na) = 1. \tag{27}$$

It is always possible, in this case, to modify the windows so that the generated frame is tight ($A = B$) so that the analysis and synthesis windows are identical.

If $\left\{\varphi_{n,m}^{(s)}\right\}_{n,m\in\mathbb{Z}}$ forms a Gabor frame with dual frame $\left\{\varphi_{n,m}^{(a)}\right\}_{n,m\in\mathbb{Z}}$, any signal $s(t)$ in $L^2(\mathbb{R})$ can be represented as follows:

$$s(t) = \sum_{m,n} S(na, mb)h_s(t-na)e^{j2\pi mb(t-na)}, \tag{28}$$

where

$$S(\tau,\nu) = \int_{-\infty}^{+\infty} s(t)h_a(t-\tau)e^{-j2\pi\nu(t-\tau)}dt \tag{29}$$

is the Short-Time Fourier Transform of the signal with analysis window $h_a(t)$, so that

$$S(na, mb) = \left\langle s, \varphi_{n,m}^{(a)}\right\rangle \tag{30}$$

is its sampled version on the uniform grid $\{na, mb\}_{n,m\in\mathbb{Z}}$.

A discrete-time version of (28) can be obtained in the same form by uniformly sampling time $t_k = kT$ and by choosing hop-size $a = NT$ and frequency sample interval $b = 1/MT$, where both $N$ and $M$ are integers with $M \geq N$, where, in typical applications, $M = KN$, with the integer $K$ controlling the overlap factor. Furthermore, the summation over the frequency index $m$ is finite in the discrete case.

[1] Throughout this paper we use an equivalent definition of Gabor frames and STFT which includes time shift ($na$ or $\tau$) in the complex sinusoids, so that windows are synchronous with the phase of the exponentials.

### 3.2. Time Warping by Means of Phase Vocoder: Form I

By time-warping both sides of (28) one has the following expansion for the continuous time time-warped signal:

$$s(\gamma(t)) = \sum_{m,n} S(na, mb) h_s\left(\gamma(t) - na\right) e^{j2\pi mb(\gamma(t)-na)}. \quad (31)$$

Due to unitary equivalence [2] through the unitary warping operator $\mathbf{U}_\gamma$, the set

$$\mathbf{U}_\gamma \varphi^{(s)}_{n,m}(t) = \sqrt{\dot{\gamma}(t)} h_s\left(\gamma(t) - na\right) e^{j2\pi mb(\gamma(t)-na)}, \quad (32)$$

where $\dot{\gamma}(t)$ is the time derivative of the warping map, is a frame if and only if the set in (23) is a frame. Thus, an alternate scheme for warping a continuous-time signal consists of projecting the signal over a Gabor analysis frame and compute the expansion (31) over the time-warped frame (32).

An algorithm of interest for time-warping discrete-time signals can be obtained by discretizing (31):

$$\tilde{s}(k) = \sum_{m=-\left\lfloor \frac{M}{2} \right\rfloor}^{+\left\lfloor \frac{M}{2} \right\rfloor} \sum_n S_{n,m} h_s\left(g_k - nNT\right) e^{j\frac{2\pi m}{M}\left(\frac{g_k}{T} - nN\right)}, \quad (33)$$

where $g_k = \gamma(kT)$, while $\tilde{s}(k) = s_{tw}(kT) = s(\gamma(kT))$ is the discrete-time warped signal and

$$S_{n,m} = \sum_k s(kT) h_a\left((k - nN)T\right) e^{-j\frac{2\pi m}{M}(k-nN)}, \quad (34)$$

with $n \in \mathbb{Z}$ and $m = -\left\lfloor \frac{M}{2} \right\rfloor, ..., +\left\lfloor \frac{M}{2} \right\rfloor$, are the Gabor expansion coefficients obtained by projection over the corresponding discrete-time frame. We refer to this algorithm as the Form I computation.

The computation of time warping by means of (34) and (33) only involves the warping of the synthesis window and of the complex exponentials, which are continuous-time functions expressed in closed form so this operation does not pose any problem. Assuming that the warping map is invertible, if the original synthesis window $h_s(t)$ has compact support in $\left[-\frac{KNT}{2}, +\frac{KNT}{2}\right[$, the shifted warped windows $h_s\left(\gamma(t) - nNT\right)$ also have compact support in

$$\left[\gamma^{-1}\left(\left(n - \frac{K}{2}\right)NT\right), \gamma^{-1}\left(\left(n + \frac{K}{2}\right)NT\right)\right[. \quad (35)$$

From this it is easy to find the samples instants $kT$ for which the warped window sequence in (34) is nonzero. Since the numbers $\frac{g_k}{T}$ in (33) are not integer, the computation of the synthesis cannot be performed by means of the IFFT.

### 3.3. Time Warping by Means of Phase Vocoder: Form II

An alternate algorithm for discrete-time time warping by means of Gabor frames, can be obtained by computing the coefficients of the warped signal by projecting it over a frame in the form (23). Thus, we compute the scalar products

$$\tilde{S}(na, mb) = \left\langle \mathbf{U}_\gamma s, \varphi^{(a)}_{n,m} \right\rangle = \left\langle s, \mathbf{U}_{\gamma^{-1}} \varphi^{(a)}_{n,m} \right\rangle, \quad (36)$$

where the last equality is due to the fact that the warping operator is unitary and its adjoint corresponds to unitary warping with the inverse map $\gamma^{-1}$. Thus, (36) is equivalent to projecting the signal over the inversely warped frame, whose elements are:

$$\mathbf{U}_{\gamma^{-1}} \varphi^{(a)}_{n,m}(t) = \sqrt{\dot{\gamma}^{-1}(t)} h_a\left(\gamma^{-1}(t) - na\right) e^{j2\pi mb(\gamma^{-1}(t)-na)}. \quad (37)$$

For the synthesis one uses the dual frame $\left\{\varphi^{(s)}_{n,m}\right\}_{n,m\in\mathbb{Z}}$. Passing to discrete time as in Section 3.2, we have the following algorithm to compute (unitary) warping:

$$\tilde{s}(k) = \sum_{m=-\left\lfloor \frac{M}{2} \right\rfloor}^{+\left\lfloor \frac{M}{2} \right\rfloor} \sum_n \tilde{S}_{n,m} h_s\left((k - nN)T\right) e^{j\frac{2\pi m}{M}(k-nN)}, \quad (38)$$

where $\tilde{s}(k) = \sqrt{\dot{\gamma}(kT)} s(\gamma(kT))$ is the discrete-time unitarily warped signal and

$$\tilde{S}_{n,m} = \sum_k d_k s(kT) h_a\left(g_k - nNT\right) e^{-j\frac{2\pi m}{M}\left(\frac{g_k}{T} - nN\right)} \quad (39)$$

with $n \in \mathbb{Z}$ and $m = -\left\lfloor \frac{M}{2} \right\rfloor, ..., +\left\lfloor \frac{M}{2} \right\rfloor$, are the expansion coefficients obtained by projection over the discrete-time analogue of (37), where $d_k = \sqrt{\dot{\gamma}^{-1}(kT)}$ and $g_k = \gamma^{-1}(kT)$. We refer to this algorithm as the Form II computation of discrete time warping. Since the numbers $\frac{g_k}{T}$ in (39) are not integer, the computation of the analysis coefficients cannot be performed by means of the FFT.

### 3.4. Experimental Results

In this section we provide an assessment of the quality of the algorithms proposed in Sections 3.2 and 3.2 for computing discrete-time time warping by means of generalized Gabor expansions, namely, through Form I in (33) and (34) or through Form II in (38) and (39). For comparison, we used the same sets of closed form signals to evaluate the SNR as we did in the evaluation of the sampling expansion based method in Section 2.3.

The average results for Form I are shown in Fig.5. Here again, the SNR results did not show great variability across the warping maps we tested in our experiments, from linear map to linear and quadratic chirps.

The results obtained by varying the overlap factor $K$ from 2 to 8 and for several values of the hop-size factor $N$ from 64 to 512 are a bit erratic but most of the SNRs are above 100 dB, which are sufficient for most audio applications. The quality of the results generally improves with the length of the window $KN$. We note that at equal window lengths but different overlap factors, e.g., on the $N = 512$ curve with $K = 2$ and the $N = 256$ curve with $K = 4$, we obtain similar SNRs.

In both Form I and Form II algorithm one should choose suitably long windows, as these are time-warped, in the analysis algorithm for Form II and in the synthesis for Form I. Thus, for a given choice of the fixed window length, the warped versions can become too short. The dilation factor of the window locally depends on the time derivative of the map. With this into consideration, with dilation factors smaller than 1 we obtained similar results for Form II to those for Form I at the cost of generally larger window lengths. Typical characteristics of the SNR for the Form II are shown in Fig.6, where we used a linear map (4) with coefficient $\alpha = 0.7$ and $c = 0$. We notice that while the window length is modulated in the analysis, the windows size remains constant in the synthesis, thus involving an extra amount of operations, also depending on the overlap factor.

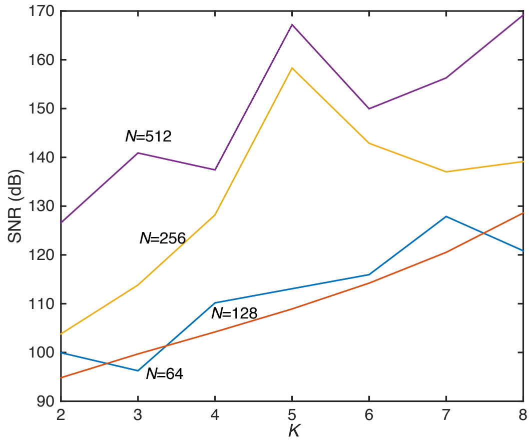

Figure 5: *Average SNR characteristics for the Form I algorithm as a function of the overlap parameter $K$, for several values of the hop size factor $N$.*

### 3.5. Computational Complexity

The computational complexity of the phase vocoder based time-warping algorithms is generally higher than the sampling expansion based algorithm. In fact, in either Form I or Form II algorithms, the analysis and the synthesis cannot be both computed by means of FFT. Thus, a matrix DFT-like form computation is necessary to obtain the warped Fourier transform for each time shift of the window. Due to the variable support of the warped windows, this computation requires on the average an order of $M^2/\bar{\alpha}$ operations in Form I and an order of $\bar{\alpha}M^2$ operations in Form II, where $\bar{\alpha}$ is the average derivative of the map (see discussion in Section 2.4) and $M$ is the number of frequency bins. Each of these computations generates $N$ samples, where $N$ is the hop size.

Thus, the computational complexity of Form I is proportional to $MK/\bar{\alpha}$ operations per sample and that of the Form II to $\bar{\alpha}MK$ operations per samples, where $K$ is the overlap factor such that $M = KN$. Comparing these results with the complexity of the windowed $\mathrm{sinc}$ kernel interpolation in Section 2.4, we see that since to achieve similar SNR values, the width factor $L$ of the kernel can be chosen to be smaller than the length $M$ of the window in Form I and Form II, the latter are computationally less efficient. However, in many audio applications a phase vocoder could already be part of the computational structure of the effect. In that case, time warping can be introduced with little extra effort in order to build dynamic effects.

## 4. TIME WARPING IN TIME STRETCHING AND PITCH SHIFTING

Time warping stretches or compresses signals altering both their pitch and their duration. Often, in sound processing, it is desirable to time stretch the signal without changing the pitch or to pitch shift the signal while preserving its original duration [16]. In the previous parts of this paper we have been considering time warping at the input or output signal level. However, it turns out that for time stretching, time warping in the STFT domain is the best approach.

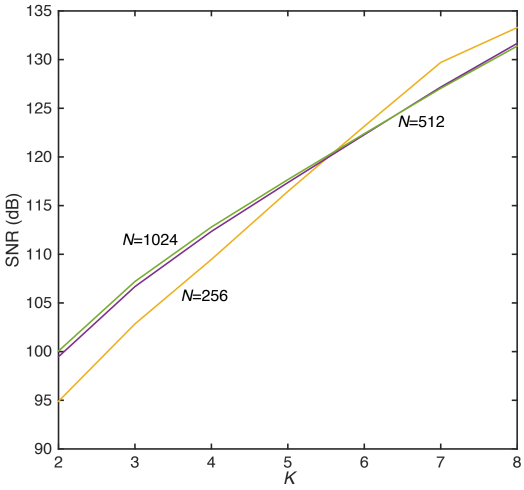

Figure 6: *SNR characteristics for the Form II algorithm for a linear map $\gamma(t) = 0.7 \cdot t$ as a function of the overlap parameter $K$, for several values of the hop size factor $N$.*

In order to time stretch a sound signal without altering its pitch, one wishes to scale or time-warp the envelopes of the partials without altering the oscillation time. We are going to perform a similar derivation of the stretching algorithms to the one for uniform STFT albeit in the warped framework. In order to have an idea of the operations involved, first consider the STFT $S(\tau, \nu)$ (29) of a sinusoidal signal of frequency $f$ with a slowly-varying envelope $a(t)$:

$$s(t) = a(t)e^{j2\pi ft}. \tag{40}$$

If, for any shift $\tau$, the envelope is approximately constant over the support of the shifted analysis window $h_a(t - \tau)$, then we have

$$S(\tau, \nu) \approx a(\tau)H_a(\nu - f)e^{j2\pi f\tau}, \tag{41}$$

where $H_a(\nu)$ is the Fourier transform of the analysis window. Thus, for the simple signal (40), the time characteristics of the magnitude STFT only depends on the amplitude envelope and the time characteristics of the phase only depends on the frequency of the sinusoid. If we time warp only the magnitude STFT, i.e., we let $\tilde{S}(\tau, \nu) = \tilde{a}(\tau)H_a(\nu - f)e^{j2\pi f\tau}$ where $\tilde{a}(\tau) = \sqrt{\dot{\gamma}(t)}a(\gamma(\tau))$, and perform reconstruction via the usual synthesis form:

$$\tilde{s}(t) = \int_{-\infty}^{+\infty} d\nu \int_{-\infty}^{+\infty} d\tau \tilde{S}(\tau, \nu)h_s(t - \tau)e^{j2\pi\nu(t-\tau)}, \tag{42}$$

under the assumption that the warped envelope $\tilde{a}(t)$ is still approximately constant over the support of the shifted windows, we obtain

$$\tilde{s}(t) \approx \tilde{a}(t)e^{j2\pi ft}, \tag{43}$$

which is the dynamically stretched sinusoid with the original frequency $f$.

An alternate equivalent form of (42) consists of pre-unwarping the phase of the STFT with the inverse map $\gamma^{-1}$, then warp the result with the map $\gamma$ and finally perform the synthesis. Clearly, by the unitarity of the warping operator, this is equivalent to performing inverse warping, with respect to time-shift $\tau$, on the synthesis windows:

$$\tilde{s}(t) = \int_{-\infty}^{+\infty} d\nu \int_{-\infty}^{+\infty} d\tau \tilde{\tilde{S}}(\tau,\nu)\tilde{h}_s(t,\tau)e^{j2\pi\nu(t-\gamma^{-1}(\tau))} \quad (44)$$

where

$$\tilde{h}_s(t,\tau) = \sqrt{\dot{\gamma}^{-1}(\tau)}\, h_s(t-\gamma^{-1}(\tau))$$

and

$$\tilde{\tilde{S}}(\tau,\nu) = a(\tau)H_a(\nu-f)\sqrt{\dot{\gamma}^{-1}(\tau)}\, e^{j2\pi f\gamma^{-1}(\tau)}.$$

The latter form (44) can be recognized as a generalization of the most common phase vocoder approach for time stretching. There, for the synthesis we change the hop-size with respect to the original analysis hop-size. This can be seen as a time warping of the hop-size with a linear map.

Once obtained the dynamically time stretched version of the signal from (42) or (44), the dynamically pitch shifted version can be obtained simply by time warping the result with the inverse map $\gamma^{-1}$.

By superposition, if the signal partials fall in sufficiently distant frequency bins, which can be adjusted by properly choosing the frequency resolution, our derivation easily extends to signals made out of several enveloped sinusoids. Of course, in reality, things get a bit more complicated than this outline. In fact, just as in the ordinary phase vocoder, multiple sinusoidal partials can interfere within common analysis bins and make the phase of the STFT have a more complicated dependency on the frequencies of the single partials. Unstable pitch due to vibrato or glissando and transients can alter the simple result. Moreover, in practice, one attempts to time stretch signals directly from a sampled version of the STFT given by the phase vocoder.

Sampled counterparts of (42) and (44) can be readily defined. The measures for robustly adapting these methods for use in time stretching and pitch shifting will be the object of forthcoming work.

## 5. CONCLUSIONS

In this paper we have considered the problem of time-warping discrete-time signals. We compared the algorithm based on the warped sampling expansion with two new methods, Form I and Form II, obtained from the interaction of time warping with a phase vocoder. While the latter require a higher number of operations, all the methods considered achieve high quality in terms of SNR. The phase vocoder based methods have a more flexible design in terms of window length and even transformation scheme. Their use in audio effects could be desirable when a phase vocoder is already present in the computational structure of the effect. Moreover, in conjunction with phase vocoders, time warping is part of the algorithms for time stretching and pitch shifting.

## 6. REFERENCES

[1] R. G. Baraniuk and D. L. Jones, "Warped wavelet bases: unitary equivalence and signal processing," in *1993 IEEE International Conference on Acoustics, Speech, and Signal Processing*, April 1993, vol. 3, pp. 320–323 vol.3.

[2] 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.

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

[4] 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.

[5] 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.

[6] 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.

[7] 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.

[8] G. Evangelista, "Redressing Warped Wavelets and Other Similar Warped Time-Something Representations," in *Proceedings of the Digital Audio Effects Conference (DAFx-17)*, Edinburgh, UK, 2017, pp. 260–267.

[9] G. Evangelista, M. Dörfler, and E. Matusiak, "Arbitrary phase vocoders by means of warping," *Music/Technology*, vol. 7, no. 0, 2013.

[10] D. Cochran and J.J. Clark, "On the sampling and reconstruction of time-warped bandlimited signals," in *International Conference on Acoustics, Speech, and Signal Processing*, Apr 1990, pp. 1539–1541 vol.3.

[11] 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, Oct 1985.

[12] S. Azizi, D. Cochran, and J.N. McDonald, "Reproducing kernel structure and sampling on time-warped spaces with application to warped wavelets," *IEEE Transactions on Information Theory*, vol. 48, pp. 789–790, Mar. 2002.

[13] A.N. Jarrot, C. Ioana, and A. Quinquis, "Toward The Use Of The Time-Warping Principle With Discrete-Time Sequences," *Journal of Computers (JCP)*, vol. 2, no. 6, pp. 49–55, Aug. 2007, NonWOS.

[14] G. Evangelista and S. Cavaliere, "Real-time and efficient algorithms for frequency warping based on local approximations of warping operators," in *Proceedings of the Digital Audio Effects Conference (DAFx-07)*, Bordeaux, France, Sept. 2007, pp. 269–276.

[15] H. Bölcskei and J.E.M. Janssen, "Gabor frames, unimodularity, and window decay," *Journal of Fourier Analysis and Applications*, vol. 6, no. 3, pp. 255–276, May 2000.

[16] J. Driedger and M. Müller, "A review of time-scale modification of music signals," *Applied Sciences*, vol. 6, no. 2, 2016.