THE SHORT-TIME LAGUERRE TRANSFORM: A NEW METHOD FOR REAL-TIME FREQUENCY WARPING OF SOUNDS

Gianpaolo Evangelista

Audio-Visual Communications Laboratory (LCAV),
Federal Institute of Technology, Lausanne, Switzerland,
evangelista@epfl.ch

ABSTRACT

The Laguerre transform and its time-varying version are general DSP techniques for frequency-warping audio signals. By means of this exact and computable transform, the frequency content of the original signal is altered via a mapping of the frequency axis. The interesting effects that are obtained include transformation of harmonic into inharmonic sounds, time-varying pitch shifting for vibrato and modulation, dispersive model-based pitch shifting of piano tones and sound morphing.

However, the computation of the Laguerre transform is non-causal and not suitable for real-time applications. Due to the frequency dependent group delay of the basis elements of the transform, the time-localization of the signal is altered, e.g., the amplitude envelopes of the partials are non-uniformly deformed according to the frequency range. As a result, the Laguerre transform of long audio signals leads to the time fusion of certain frequency components that were originally related to distinct events.

In this paper we introduce the Short-Time Laguerre Transform (STLT) as a new method for frequency warping, which circumvents most of the drawbacks of the Laguerre Transform. The sound stream is subdivided into overlapping frames and the windowed Laguerre transform is computed on each frame. Due to the time spread of the components, this process generates frames whose size is theoretically infinite while in practice not conformal to the size of the original frames. We address the problem of forming the warped signal by means of overlap add techniques.

Keywords: signal transformations, frequency warping, Laguerre transform, dispersive delay lines

1 INTRODUCTION

Frequency warping consists of modifying the frequency spectrum of a signal $s(k)$ by displacing its frequency content to other frequencies [2]. This is achieved by means of a map $\Omega = \theta(\omega)$ transforming each frequency into a new one according to the following law:

$$\hat{S}(\theta(\omega)) = S(\omega) . \tag{1}$$

Frequency warping may be performed with the help of arbitrary maps and the quality of the effect is intimately related to the mathematical properties of the function $\theta(\omega)$. If the map is invertible then the warped frequency spectrum of a discrete-time signal $s(k)$ is obtained as follows:

$$\hat{S}(\omega) = S(\theta^{-1}(\omega)) = \sum_k s(k) e^{-jk\theta^{-1}(\omega)} . \tag{2}$$

The time domain counterpart of (2) gives the warped signal $\hat{s}(n)$ in terms of the original signal $s(k)$:

$$\hat{s}(n) = \frac{1}{2\pi} \int_{-\pi}^{\pi} \hat{S}(\omega) e^{jn\omega} d\omega = \sum_k s(k) h_n(k) , \tag{3}$$

where

$$h_n(k) = \frac{1}{2\pi} \int_{-\pi}^{\pi} e^{j\left[n\omega - k\theta^{-1}(\omega)\right]} d\omega .$$

This shows that frequency warping a signal is equivalent to orthogonally project it onto the set $h_n(k)$. It can be shown that if $\theta(\omega)$ is continuously differentiable and maps the interval $[-\pi, +\pi]$ one-to-one onto itself, then the sequences

$$h_n(k) = IDTFT\left[e^{-jn\theta(\omega)} \frac{d\theta}{d\omega} \right] \tag{4}$$

form a complete set -- useful for the expansion of any finite energy signal -- biorthogonal to the set

$$g_n(k) = IDTFT\left[e^{-jn\theta(\omega)} \right] .$$

In this case

$$s(k) = \sum_n \hat{s}(n) g_n(k) .$$

If $\theta(\omega)$ is monotonically increasing one can factor the (positive) derivative

$$\frac{d\theta}{d\omega} = \left| F_0(\omega) \right|^2$$

and obtain the following orthogonal and complete set

$$f_n(k) = IDTFT\left[e^{-jn\theta(\omega)} F_0(\omega) \right] . \tag{5}$$

Projection on this set obtains frequency warping combined with spectrum scaling:

$$F_0(\omega) \hat{S}(\theta(\omega)) = S(\omega) .$$

An arbitrary frequency band $[B_0 , B_1]$ is mapped into a warped band [$\theta(B_0)$, $\theta(B_1)$]. Orthogonal warping has the following property: in the warped frequency band the warped spectrum has the same energy as the original spectrum in the original band [B_0, B_1]:

$$\int_{B_0}^{B_1} |S(\omega)|^2 d\omega = \int_{\theta(B_0)}^{\theta(B_1)} |\hat{S}(\omega)|^2 d\omega \ .$$

The sequences (5) can be generated by iteratively filtering the sequence $f_0(k)$ by the all-pass

$$A(\omega) = e^{-j\theta(\omega)} .$$

If the input signal is causal then only the causal part of $f_n(k)$ gives non-zero contribution to the expansion and one can make these filters into causal filters. However, the transfer functions are not necessarily rational and their implementation can only be approximate. One can show that the unique continuously differentiable, one-to-one and onto map $\theta(\omega)$ fixing the points 0 and π that can be implemented by means of a rational transfer function is obtained by the sign-reversed phase of a first order real all-pass filter [1]:

$$A(z) = \frac{z^{-1} - b}{1 - bz^{-1}} \quad \text{with } -1 < b < 1 \ .$$

In this case

$$\theta(\omega) = -\ \arg\ A(e^{j\omega}) = \omega + 2\tan^{-1}\left(\frac{b\sin\omega}{1 - b\cos\omega}\right) \quad (6)$$

belongs to a one-parameter family of warping maps. The corresponding unitary signal transform is called the Laguerre transform. The basis elements $\lambda_n(k)$ are discrete-time counterparts of the Laguerre functions and their z-transforms are given by the following iteration:

$$\Lambda_0(z) = \frac{\sqrt{1-b^2}}{1-bz^{-1}}$$

$$\Lambda_r(z) = \Lambda_0(z) A(z)^{r+1}, \qquad r=0,1,\ldots \ . \quad (7)$$

A structure for computing frequency warping via the Laguerre transform is shown in Fig. 1. It consists of a dispersive delay line implemented in a chain of all-pass filters. Each filter introduces a frequency dependent group delay

$$\tau_g(\omega) = \frac{d\theta}{d\omega} = \frac{1-b^2}{1-2b\cos\omega + b^2} = |\Lambda_0(\omega)|^2 .$$

The signal is time-reversed, filtered by $\Lambda_0(z)$ and fed to the delay line. The output of the filter $\Lambda_0(z)$ and of each delay element is read at time k=0 to form the frequency warped signal $\hat{s}(n)$. Theoretically, an infinite-length delay line is required in order to compute the transform. However, one can show [3] that for a finite-length N full-band input signal the line can be truncated to a number M of sections equal to the maximum group delay experienced by the N samples:

$$M \approx \frac{1+|b|}{1-|b|} N .$$

The computational complexity of the warping structure is of the order of N^2. The inverse transform can be computed by means of the same structure of Fig. 1 provided that one reverses the sign of the Laguerre parameter. This can be shown to be equivalent to compute the transform over the transposed sequences $\lambda_n^T(k) = \lambda_k(n)$, which have the same form as (7), with a reversed sign of b.

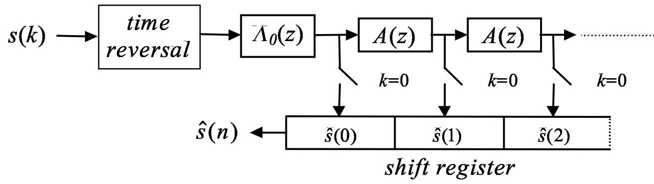


Fig. 1 Dispersive delay-line structure for frequency warping via the Laguerre transform.

2 The Short-Time Laguerre Transform

The Laguerre transform is not suitable for real-time frequency warping. In order to compute the transform by means of the structure in Fig. 1 all the signal samples must be available and, because of the time-reversal operation, samples are processed in reverse order. Another intrinsic drawback of the transform is that long dispersive delays considerably alter the time structure of the signal, e.g., according to the sign of the parameter b, higher frequencies are perceived long after or long before lower frequencies. In order to implement real-time frequency warping an approximate and modified scheme must be employed. As a first step, consider a finite-length N window sequence $w(n)$ satisfying the following requirement:

$$\sum_r w(n - rL) = 1 \quad \text{for some integer } L > 0 . \qquad (8)$$

A trivial example is given by the rectangular window with L=N:

$$rect_L(n) = \begin{cases} 1 & if\ n = 0,\ldots,L\text{-}1 \\ 0 & otherwise \end{cases} .$$

The Hanning window gives a better-behaved example:

$$h_N(n) = \tfrac{1}{2} rect_N(n)\left(1 - \cos\left(\tfrac{2n\pi}{N}\right)\right)$$

with $N = RL$, R integer. Given a window $w(n)$ satisfying (8) one can put the Laguerre transform in the following framewise form:

$$\hat{s}(n) = \sum_r \sum_{k=0}^{N-1} w(k) s(k + rL) \lambda_n(k + rL) . \qquad (9)$$

By taking the DTFT of both sides of (9) one obtains

$$\hat{S}(\omega) = \sum_r e^{-jrL\theta^{-1}(\omega)} Q_r(\omega) , \qquad (10)$$

where

$$Q_r(\omega) = \sum_{k=0}^{N-1} w(k) s(k + rL) \Lambda_k^T(\omega) \qquad (11)$$

is the DTFT of the Laguerre transform of the windowed signal $w(k)s(k+rL)$, i.e.,

$$q_r(n)=\sum_{k=0}^{N-1} w(k)s(k+rL)\lambda_n(k)\,. \qquad (12)$$

Equation (10) denotes an alternate way of computing the Laguerre transform, which has no practical advantage other than being in block form. However, equation (12) can be interpreted as a Short-Time Laguerre Transform (STLT). Due to their frequency dependent delay the terms $e^{-jrL\theta^{-1}(\omega)}$ in (10) introduce a strong time spreading of the signal components in different areas of the frequency spectrum. One way to resynchronize these components is to replace $L\theta^{-1}(\omega)$ by a linear phase term. In this way one can frequency warp the short-time frequency spectrum while preserving the long-term evolution of the signal. In other words we replace (10) with a new definition of frequency warping obtained by overlap adding the short time warped spectra:

$$\tilde{S}(\omega)=\sum_r e^{-jr\beta\omega}Q_r(\omega)\,. \qquad (13)$$

The problem is: how do we choose the linear delay coefficient β so that the transitions between warped segments are smooth enough to avoid amplitude modulation or envelope distortion effects? The answer requires a deeper investigation of the terms $Q_r(\omega)$. Since

$$DTFT[w(k)s_r(k)](\omega)=\tfrac{1}{2\pi}\int_{-\pi}^{+\pi} W(\omega-\Omega)S_r(\Omega)d\Omega$$

where $s_r(k)=s(k+rL)$, k=0,1…,L-1 is the r-th signal frame, then (11) can be written as follows:

$$Q_r(\omega)=\Lambda_0^T(\omega)\tfrac{1}{2\pi}\int_{-\pi}^{+\pi} W(\theta^{-1}(\omega)-\Omega)S_r(\Omega)d\Omega\,. \qquad (14)$$

By performing the change of variable

$$\alpha=\theta(\Omega-\theta^{-1}(\omega))+\omega$$

in the integral, (11) becomes

$$Q_r(\omega)=\Lambda_0^T(\omega)\tfrac{1}{2\pi}\int_{-\pi}^{+\pi} \widehat{W}(\omega-\alpha)S_r(\theta^{-1}(\omega)+\theta^{-1}(\alpha-\omega))d\alpha\,,$$

where

$$\widehat{W}(\omega)=\left|\Lambda_0^T(\omega)\right|^2 W(\theta^{-1}(\omega))$$

is a warped version of the window. Since $W(\omega)$ is narrow band then also $\widehat{W}(\omega)$ is narrow band and the largest contributions to the integral are for $\alpha\approx\omega$. Hence

$$Q_r(\omega)\approx\tfrac{1}{2\pi}\int_{-\pi}^{+\pi} \widehat{W}(\omega-\alpha)\Lambda_0^T(\alpha)S_r(\theta^{-1}(\alpha))d\alpha\,.$$

In this approximation

$$q_r(n)\approx\hat{w}(n)\hat{s}_r(n)\,,$$

where

$$\hat{s}_r(n)=IDTFT[\Lambda_0^T(\omega)S_r(\theta^{-1}(\omega))](n)$$

is a warped version of the signal frame, while

$$\hat{w}(n)=IDTFT[\widehat{W}(\omega)](n)$$

and

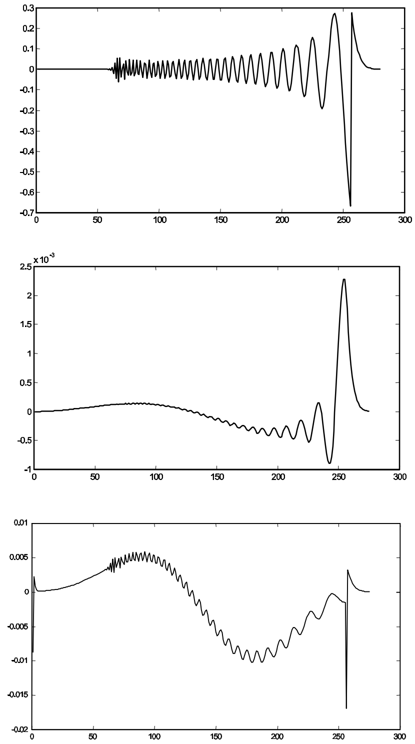
$$\widehat{W}(\omega)\approx\left|\Lambda_0^T(0)\right|^2 W(\theta^{-1}(0)*\omega)=\frac{1-b}{1+b}W\left(\tfrac{1-b}{1+b}\omega\right).$$

Therefore, each output frame $q_r(n)$ is approximately equal to a warped version of the input frame modulated by a scaled version of the window. As a result, if the window $w(n)$ satisfies (8) then the warped window $\hat{w}(n)$ will approximately satisfy

$$\sum_r \hat{w}(n-rK)\approx 1 \quad \text{for } K=round\left(\tfrac{1-b}{1+b}\right)L\,.$$

Hence, we have found that the factor β required in (13) for performing a synchronized overlap add operation on the warped frames is

$$\beta=\tfrac{1-b}{1+b}\,.$$



(a) (b) (c)

Fig. 2 Difference error of warped vs. scaled window: (a) rectangular, (b) Hanning and (c) Chebychev.

The quality of the approximation depends on the choice of the window. This is illustrated in Fig. 2 where warped windows are compared with scaled windows and the difference error is reported. For a warping parameter $b=-\frac{1}{3}$ one has $\beta=2$, hence $K=2L$. This allows for a comparison free of rounding effects. The Hanning window compares favorably with both the rectangular and Chebychev window, with a maximum error of the order of 10^{-3}. Notice that, with proper normalization, the Hanning window satisfies property (8) for any integer L dividing the length of the window N, i.e., for N=QL, Q integer. For this reason the Hanning window is a good candidate for the computation of the short-time Laguerre transform.

3. Experiments and Results

We tested the STLT algorithm on several isolated sounds and musical phrases of a single instrument. Compared to the strict frequency-warping algorithm via Laguerre transform, the STLT offers the advantage of preserving the time structure of the sounds. This is especially important in sequences, which would otherwise be scrambled or their time organization destroyed by pure warping. However, the STLT presents some drawbacks. The choice of the window length influences the quality of warping. A short window generally introduces a low-pass filtering effect of the tones, while a long window alters the envelopes of the tones, especially on sharp attacks. In other words, long windows provide a more exact frequency-warping algorithm while shorter windows tend to preserve the boundaries between sounds and more faithfully reproduce their transitions. This effect is somewhat reduced by choosing a large overlap factor.

The approximation described in the previous section is illustrated in Fig. 3, where to a warped frame of a trumpet sound we superimposed the properly β-stretched Hanning window. The fit is very good and provides experimental support of our hypotheses. In order to use overlap add techniques, the factor βL must be rounded to an integer. In this way (13) becomes a sum of time domain delayed warped frames. We experimented with several factors and warping parameters. The error due to rounding does not seem to affect the overall quality of the signal. The expected effect would be an amplitude modulation of the warped signal.

We conclude that the warping effect adds richness and texture to natural or synthetic sounds. The strict harmonic relationships of voiced sounds can be broken at will. Warping is present in natural instruments such as piano in the low register [7, 8]. The STLT provides a real-time solution for warping the short-time spectrum while preserving the macrostructure of sound. The algorithm can be generalized to time-varying warping [9,10], by changing the warping factor at each frame. For slow variations of the parameter this creates minor distortion of the overlap add method. This technique allows us to insert interesting pitch and frequency-content modulations of the signal, including vibrato and detuning effects.

Other applications of warping include the realization of perfect reconstruction filter banks based on perceptual scales [4, 6] and the decomposition of inharmonic sounds in transients and noise plus resonant component via the Frequency-Warped Wavelet Transform [3,5].

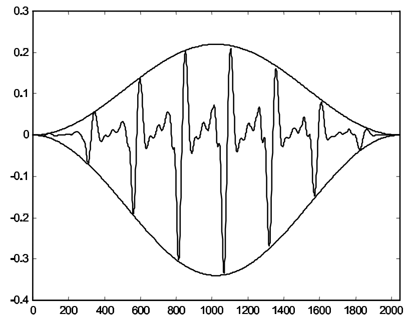


Fig. 3 Warped frame of trumpet sound with superimposed stretched Hanning window.

References

[1] P. W. Broome, "Discrete Orthonormal Sequences," *J. Assoc. Comput. Machinery,* vol. 12, no. 2, pp. 151-168, 1965.

[2] A.V. Oppenheim and D. H. Johnson, "Discrete Representation of Signals," *Proc. IEEE,* vol. 60, pp. 681-691, June 1972.

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

[4] G. Evangelista, S. Cavaliere, "Discrete Frequency Warped Wavelets: Theory and Applications," *IEEE Trans. on Signal Processing,* special issue on Theory and Applications of Filter Banks and Wavelets, vol. 46, no. 4, pp. 874-885, April 1998.

[5] G. Evangelista, S. Cavaliere, "Analysis and Regularization of Inharmonic Sounds via Pitch-Synchronous Frequency Warped Wavelets," *Proc. of ICMC '97,* pp. 51-54, Thessaloniki, Greece, Sept. 1997.

[6] G. Evangelista, S. Cavaliere, "Auditory Modeling via Frequency Warped Wavelet Transform," Proc. EUSIPCO '98, pp. 117-120.

[7] I. Testa, G. Evangelista and S. Cavaliere, "A Physical Model of Stiff Strings," *Proc. of the Institute of Acoustics (Internat. Symp. on Music and Acoustics, ISMA'97) Vol.19:Part 5 (1997) Book 1*, pp. 219-224, Edinburgh, U.K., Aug. 1997.

[8] S. A. Van Duyne and J. O. Smith, "A Simplified Approach to Modeling Dispersion Caused by Stiffness in Strings and Plates," *Proc. of ICMC'94.*

[9] G. Evangelista, S. Cavaliere, "Audio Effects Based on Biorthogonal Time-Varying Frequency Warping," to appear on *Applied Signal Processing*, Springer-Verlag, Dec. 2000.

[10] G. Evangelista, S.Cavaliere, "Time-Varying Frequency Warping: Results and Experiments," *Proc. of the 99 Digital Audio Effects Workshop (DAFx99)*, pp. 13-16, Trondheim, Norway, Dec. 1999.