A FAST MELLIN TRANSFORM WITH APPLICATIONS IN DAFX

Antonio De Sena and Davide Rocchesso

Dipartimento di Informatica, Università di Verona

ABSTRACT

Many digital audio effects rely on transformations performed in the Fourier-transformed (frequency) domain. However, other transforms and domains exist and could be exploited. We propose to use the Mellin transform for a class of sound transformations. We present a fast implementation of the Mellin transform (more precisely a Fast Scale Transform), and we provide some examples on how it could be used in digital audio effects.

1. INTRODUCTION

Fast realizations of the Discrete Fourier Transform are widely used in order to produce audio signal transformations by operating in the transformed (frequency) domain [1]. These realizations are used as if they were approximations of an underlying continuous-time Fourier Transform, and the transformations rely on properties such as the magnitude invariance to time shifts, the relative auditory unimportance of phase for stationary signals, and the interpretation of spikes in the transformed domain as periodic components in the time domain [2, 3].

Many other transforms with different properties have been devised in order to make certain operations easier or certain features more easily visible. Among these, the Mellin Transform, and its restricted version called the Scale Transform, can represent a signal in terms of *scale*. The scale can be interpreted, similarly to frequency, as a physical attribute of signals [4]. Thus, we can conceive digital audio effects that work by handling the signal in the scale domain, with transformation of the magnitude and/or phase of the Mellin image. This is technically feasible as long as fast and accurate realizations of these transforms are available.

Other useful applications can be done using the Mellin transform. For example Patterson and Irino [5] have proposed to use a particular bidimensional version of this transform for vowel normalization.

Digital audio effects, such as time/pitch scaling, using the Mellin transform and the phase vocoder were previously proposed [6], but the realization relied on non-uniform sampling or re-sampling, and no considerations on the speed, accuracy, and feasibility of these operations were given. In image processing, effects such as localized denoising have been proposed as based on the scale (Mellin) transform [7]. Since the Mellin transform can be interpreted as a Fourier transform working on logarithmic time, it relies on warping the time axis. Effects based on time and frequency warping, using the Fast Fourier Transform (FFT) or dispersive delay lines, were presented in [8].

In Section 2 we briefly introduce the Mellin and scale transforms, and we provide an interpretation of the transform and its relation with the Fourier transform. Section 3 shows how a fast discrete version of the scale transform is implemented, using exponential resampling and the FFT algorithm. Section 4 presents some digital audio effects obtained by transformations in the Mellin domain.

2. THE SCALE AND MELLIN TRANSFORMS

The Mellin transform of a function f is defined as:

$$M_f(p) = \int_0^\infty f(t)\, t^{p-1}\, \mathrm{d}t\ , \tag{1}$$

where $p \in \mathbb{C}$ is the Mellin parameter. The scale transform is a particular restriction of the Mellin transform on the vertical line $p = -jc + \frac{1}{2}$, with $c \in \mathbb{R}$. Thus, the scale transform is defined as:

$$D_f(c) = \frac{1}{\sqrt{2\pi}} \int_0^\infty f(t)\, \mathrm{e}^{(-jc-\frac{1}{2})\ln t}\, \mathrm{d}t. \tag{2}$$

The scale inverse transform is given by

$$f(t) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^\infty D_f(c)\, \mathrm{e}^{(jc-\frac{1}{2})\ln t}\, \mathrm{d}c. \tag{3}$$

The key property of the scale transform is the scale invariance. This means that if f is a function and g is a scaled version of f, the transform magnitude of both functions is the same. A scale modification is a compression or expansion of the time axis of the original function that preserves the signal energy. Thus, a function $g(t)$ can be obtained with a scale modification from a function $f(t)$, if $g(t) = \sqrt{\alpha} f(\alpha t)$, with $\alpha \in \mathbb{R}^+$. When $\alpha < 1$ we get a scale expansion, when $\alpha > 1$ we get a scale compression. Given a scale modification with parameter α, the scale transforms of the original and scaled signals are related by

$$D_g(c) = \alpha^{jc} D_f(c). \tag{4}$$

This property derives from a similar property of the Mellin transform. In fact, if $h(t) = f(\alpha t)$, then

$$M_h(p) = \alpha^{-p} M_f(p). \tag{5}$$

In both (4) and (5), scaling is reflected by a multiplicative factor for the transforms, and for (4) such factor reduces to a pure phase shift.

2.1. The scale transform interpretation

A parallel can be drawn between the properties of the Fourier and scale transforms. In particular, we can define a *scale periodicity* as follows: a function $f(t)$ is said to be scale periodic with period $\mathcal{T}$ if it satisfies $f(t) = \sqrt{\mathcal{T}} f(t\mathcal{T})$, where $\mathcal{T} = b/a$, with a and b period starting point and period ending point respectively (see Figure 1). $C_0 = 2\pi/\ln \mathcal{T}$ is the “fundamental scale” associated with the periodic function. By analogy with the Fourier theory, we can define a "scale series" and Parseval theorem. Very important is the "exponential sampling theorem" [9] that, like the Shannon theorem, allows a perfect reconstruction of a scale-band limited signal from its samples. These samples must be distributed exponentially in time according to $p_k = \mathcal{T}_s^k$, with $k \in \mathbb{Z}$, $\mathcal{T}_s = \mathrm{e}^{\pi/C_m}$, and C_m the signal maximum scale.

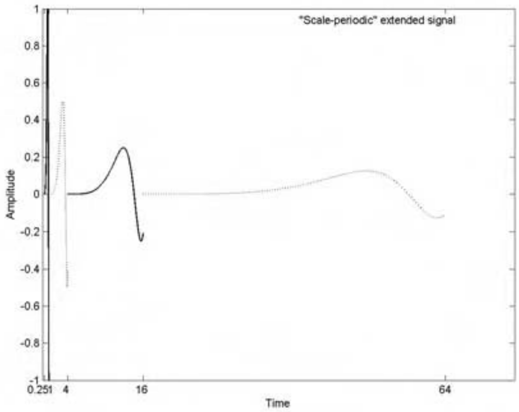

Figure 1: *Scale-periodic extension of a base signal defined between $a = 1s$ and $b = 4s$*

2.2. Relation with the Fourier transform

From its definition and interpretation, the Mellin transform provides a tight correspondence with the Fourier transform. More precisely, the Mellin transform with the parameter $p = -jc$ can be interpreted as a logarithmic-time Fourier transform. Similarly, we can define the scale transform of a function $f(t)$ using the Fourier transform of a function $g(t)$, with $g(t)$ obtained from $f(t)$ by time-warping f and multiplying the result by an exponential function. This result can be generalized for any p defined as $p = -jc + \beta$, with $\beta \in \mathbb{R}$.

3. THE FAST MELLIN TRANSFORM

Practical modifications of signals in the Mellin domain can be achieved only if an accurate and fast discrete realization of the Mellin transform is available. In [10], an algorithm based on the extraction of the analytic signal was proposed, and it is now available for both `matlab` and `scilab`[1]. However, such realization requires the specification of a lower and upper frequency bounds, its complexity appears to be quadratic in the number of samples, and it displays strong side lobes in the scale domain (see Figure 4).

We realized a Fast Mellin Transform (FMT) by exploiting the analogy between the Mellin and Fourier transforms, as a sequence of exponential time-warping, multiplication by an exponential, and Fast Fourier Transform, as represented in Figure 2.

The Mellin transform with parameter $p = -jc + \beta$ (with $\beta \in \mathbb{R}$) of $f(t)$ is identical to the Fourier transform of $\mathrm{e}^{t\beta} f(\mathrm{e}^t)$:

$$D[f(t)] = F[\mathrm{e}^{t\beta} f(\mathrm{e}^t)], \tag{6}$$

[1] `http://www.inrialpes.fr/is2/people/pgoncalv/`

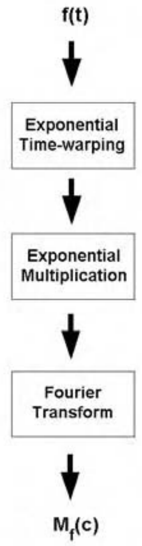

Figure 2: *Implementation of the Fast Mellin Transform*

where $F[\cdot]$ and $D[\cdot]$ refer to the Fourier transform and scale transform, respectively. The Fourier transform is commonly computed in time $\mathcal{O}(N \log N)$ on N samples, by means of the FFT. While the multiplication by an exponential is trivially done in $\mathcal{O}(N)$, we must find an algorithm for performing the exponential time-warping. This last problem can be seen as an exponential sampling of the continuous time signal. Generally we have only a uniformly (Shannon) sampled signal, thus the problem can be seen as re-sampling a discrete-time sequence, and this can be solved using interpolation. In theory, we should use a sinc interpolator (based on Shannon sampling theory), but the overall complexity turns out to be too high. However, we can approximate this interpolator by means of a natural cubic spline, and have the linear complexity associated with resolution of a tridiagonal matrix. Using this interpolator we can resample the original function obtaining an exponentially-sampled version.

The parameters needed by the resampling process are the exponential sampling step $\mathcal{T}_s$ and the number of exponential samples N_{exp}. If the original signal has been uniformly sampled by taking n samples with time step T_s starting[2] at time T_s, we can show that $\mathcal{T}_s = 1 + 1/n$ and $N_{exp} \simeq n \ln n$. Figure 3 shows an example of distribution of exponential samples derived from a sequence of uniform samples.

The algorithm has an asymptotic complexity that depends only on the FFT, as this is the most (computationally) complex part of the entire process (the spline interpolation block is linear in N_{exp} and the exponential multiplication block is also linear). The asymptotic complexity of the entire process is $\mathcal{O}(N_{exp} \ln N_{exp})$ or, in terms of the number n of uniform samples, $\mathcal{O}(n \ln^2 n)$.

The accuracy of the Fast Mellin Transform in providing an approximation to the continuous-time Mellin transform is good.

[2] If the signal starts at an arbitrary point $a > 0$ we can "scale-shift" the signal using the equation $\overline{f}(t) = \sigma^{\beta} f(\sigma t)$, where a is the original starting point, $\sigma = a/T_s$ and $\overline{f}(t)$ is the scale-shifted signal.

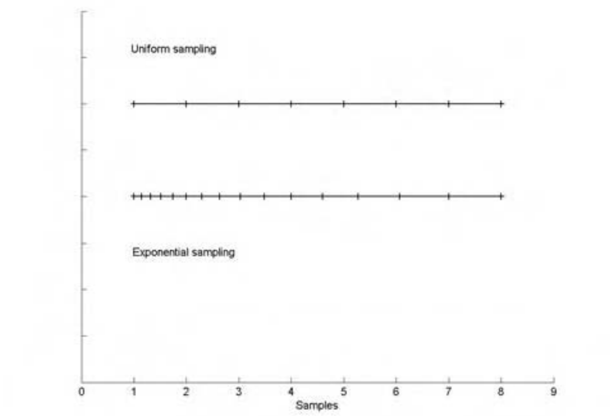

Figure 3: *Uniform sampling and (critical) exponential resampling*

For example, Figure 4 provides a comparison of the magnitude of the FMT with the theoretical continuous-time Mellin transform and with the realization proposed in [10]. In this example we've worked with a step function signal created using 128 samples, a sampling frequency of $8000Hz$ and setting the firsts 50 samples to 1 and the others to 0.

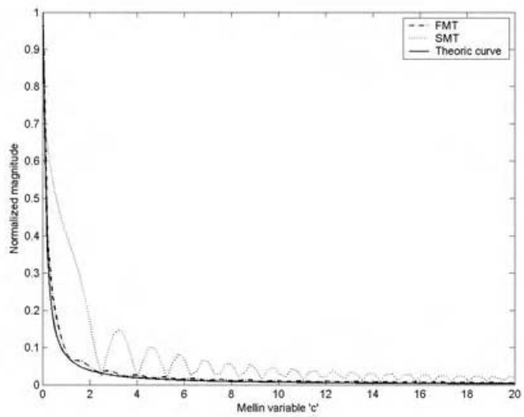

Figure 4: *Scale transform (magnitude) of a step function: continuous-time transform (solid) and its approximations with the realizations by [10] (dashed) and by the authors (dashdotted)*

4. DIGITAL AUDIO EFFECTS IN MELLIN (SCALE) DOMAIN

In this Section we show how to realize some digital audio effects using the scale domain.

4.1. Time scaling in Mellin domain

A straightforward yet useful effect is time compression or expansion with signal energy preservation. For two signals that are one the scaled copy (with factor α) of the other, the scale transforms have the same magnitude, and a difference in the phase. Let $f(t)$ and $g(t)$ be those two signals, with $g(t) = \sqrt{\alpha} f(\alpha t)$ and $\alpha \in \mathbb{R}^{+}$. Using the fact that $D_g(c) = \alpha^{jc} D_f(c)$ we can obtain $g(t)$ from $f(t)$ by applying the scale transform to $f(t)$, adding the linear contribution $c \ln \alpha$ to the the phase, and anti-transforming the result. Some care has to be taken in the choice of α: if it is too high the signal that we get from scale-compression will have frequency components that can cause aliasing. Conversely, if α is too low, we may end up cropping the signal in time.

Figure 5 shows a signal and its time-compressed version, obtained by adding a linear offset to the phase of the scale transform (6). What is depicted in Figure 5 is essentially a resampling.

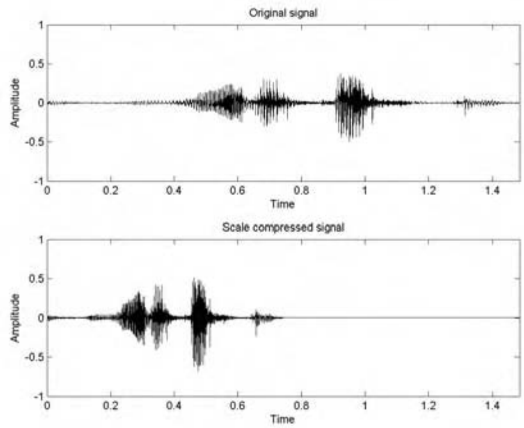

Figure 5: *Original audio signal and a scaled version (Scaled factor $\alpha = 2$) obtained using the* FMT.

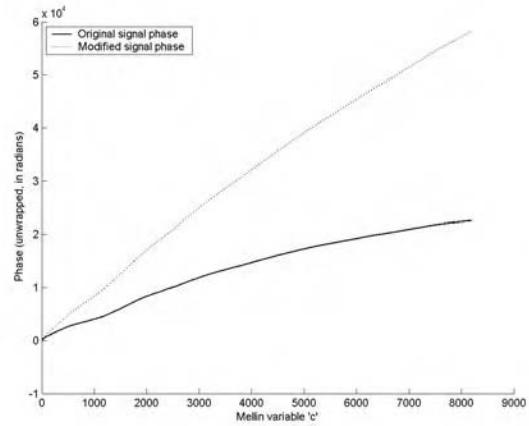

Figure 6: *Signal phase and modified signal phase.*

If the added phase contribution is not linear, then we can achieve simultaneous resampling and time warping. As compared to other resampling methods, such as the windowed-sinc interpolation [11] implemented in `octave` or `matlab`, using the scale transform does not introduce any benefit neither in accuracy nor in efficiency. However, the possibility to work directly in the phase domain adding contributions to the nominal phase gives the possibility to "sculpt" the temporal behavior of the signal in just the same way as audio practitioners sculpt the frequency behavior with softwares such as `audiosculpt`[3].

4.2. Signal reconstruction using only magnitude or phase

In order to better understand the respective roles of magnitude and phase in the Mellin transform, we can transform a signal and reconstruct it using only the phase or only the magnitude. Figure 7 shows the sonogram of a test signal. Figure 8 reports the sonogram of the signal reconstructed by replacing the magnitude response with a constant (top) or by replacing the phase response. We notice that the phase-only reconstruction preserves the temporal location of the main events that remain well visible (and audible). This is similar to the highlighting of edges in images reconstructed from their (phase-only) Mellin transform [7].

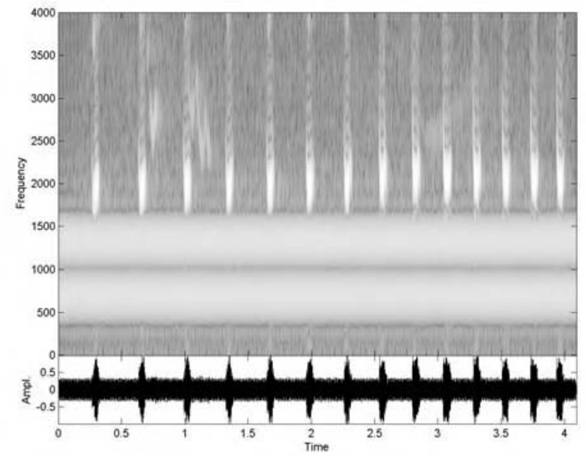

Figure 7: *Original signal sonogram.*

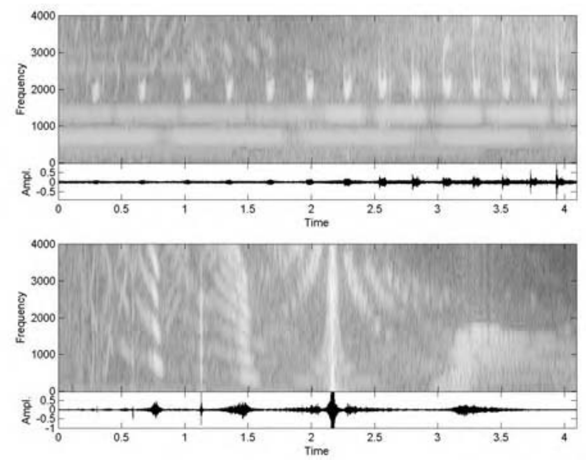

Figure 8: *Signal reconstructed with phase only (top) and with magnitude only (bottom).*

4.3. Low-pass and high-pass filtering

In this Section we show what a low-pass filter or a high-pass filter do (in the Mellin domain). To low-pass filter, we simply set to zero all magnitude components that are found between a cutoff scale and the the signal maximum scale. Observing the results from a classical Fourier-based viewpoint (see Figure 9, top), we can interpret this filter like a time-varying low-pass filter. The filter cutoff frequency exponentially approaches zero in time. The speed of convergence depends on the cutoff scale. The high-pass filter behaves symmetrically, gradually moving the cutoff frequency toward zero (see Figure 9, bottom).

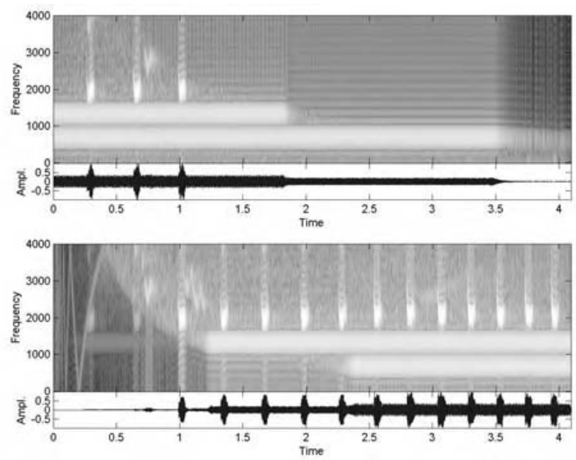

Figure 9: *Low pass (top) and high pass (bottom) filtered signal sonogram.*

4.4. Phase with random deviations

In this experiment we introduce a random deviation to the phase. This deviation grows linear by the scale and adds up to the unwrapped phase. In this example the deviation doesn't exceed the 0.04% of the phase, and this is enough to destroy the fine temporal structure without loosing the most important events (see Figure 10).

5. CONCLUSION

We presented a new implementation and some audio applications of the discrete Mellin transform. The `octave/matlab` code will be made publicly available. In the future, more sophisticated interpolation schemes will be tried in order to improve the accuracy, and the extension to a sliced-time framework (a kind of Short-Time Mellin Transform) will be attempted.

6. REFERENCES

[1] D. Arfib, "Time-frequency processing," in *Digital Audio Effects*, U. Zölzer, Ed., pp. 237–292. John Wiley and Sons, Ltd., Chichester, Sussex, UK, 2002.

[3] `http://www.ircam.fr`

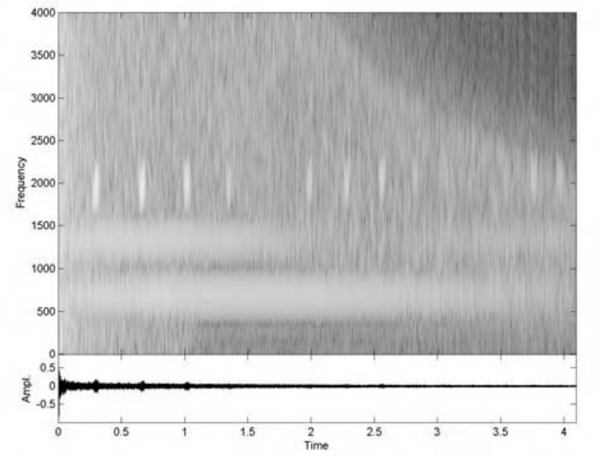

Figure 10: *Phase with growing random deviation.*

[2] J. O. Smith, *Mathematics of the Discrete Fourier Transform (DFT)*, W3K Publishing, `http://www.w3k.org/books/`.

[3] D. Rocchesso, *Introduction to Sound Processing*, Edizioni di Mondo Estremo, Firenze, 2003, `http://www.mondo-estremo.com/`.

[4] L. Cohen, "The scale representation," *IEEE Trans. on Signal Processing*, vol. 41, no. 12, pp. 3275–3291, December 1993.

[5] T. Irino and R. D. Patterson, "Segregating information about the size and the shape of the vocal tract using a time-domain auditory model: The stabilised wavelet-Mellin transform," *Speech Communication*, vol. 36, no. 3, pp. 181–203, March 2002.

[6] J. Garas and P. C. W. Sommen, "Warped linear time invariant systems and their application in audio signal processing," in *Proc. IEEE Int. Conf. on Acoustics, Speech, and Sig. Proc.*, Phoenix, AZ, March 1999.

[7] G. Cristóbal and L. Cohen, "Scale in images," in *Proc. SPIE Int. Symposium on Optical Sc., Eng. and Instr.*, Denver, CO, 1996.

[8] G. Evangelista, "Time and frequency warping musical signals," in *Digital Audio Effects*, U. Zölzer, Ed., pp. 439–447. John Wiley and Sons, Ltd., Chichester, Sussex, UK, 2002.

[9] H. Sundaram, S. D. Joshi, and R. K. P. Bhatt, "Scale periodicity and its sampling theorem," *IEEE Trans. on Signal Processing*, vol. 45, no. 7, pp. 1862–1864, July 1997.

[10] J. Bertrand, P. Bertrand, and J. P. Ovarlez, "Computation of affine time-frequency distributions using the fast mellin transform," in *Proc. IEEE Int. Conf. on Acoustics, Speech, and Sig. Proc.*, 1992, pp. 117–120.

[11] J. O. Smith and P. Gossett, "A flexible sampling-rate conversion method," in *Proc. Int. Conf. Acoustics, Speech and Signal Processing*, 1984, pp. 19.4.1–19.4.2.