The Mellin-Wavelet Transform

Douglas Nelson
Dept. of Defense, 9800 Savage Rd., Ft. Meade, Md.

ABSTRACT

Most of machine speech analysis and processing is based on a warped spectral representation. The intent of this paper is to present a method by which proper warped representations can be computed efficiently. In the case of log-warping functions, the methods of this paper produce a wavelet-like transform as a linear convolution of a single log-warped wavelet basis element and a log-warped representation of the signal. The resulting doubly warped transform is referred to in this paper as a Mellin-Wavelet transform. The majority of the paper is devoted deriving design parameters for implementation of the transform, with speech as the primary application.

INTRODUCTION

The normal process used in computing the warped spectra in speech research, is to first compute a standard short term power spectrum and then construct an approximation of the warped power spectrum by computing weighted averages of the power-spectral coefficients. The errors introduced by this approximation can impact on performance. The warped spectral filters must have non-uniform bandwidth even though the transform on which they are based represent uniform sampling in both time and frequency. In addition, the Schwartz inequality applies to the weighted averages used to approximate the filters.

A theoretically better method for computing the warped spectra is to sample the output of a bank of filters which have the desired spacing and bandwidths. In principal, these filters can be easily constructed, but the computational load is rather heavy. In this paper, a fast method of effectively implementing a constant Q filter bank is developed. In the case of log warped spectra, the method producers a a warped wavelet-like transform based on warpings of both time and frequency by the same logarithmic warping function. Under these warpings, the full wavelet-like transform can be computed with a single short convolution which may be computed using fast convolutional methods. In addition, a windowing function may be applied by modifying the wavelet basis rather than the data. Under this architecture, the windowed log warped transform may be computed with no additional computational load over the un-windowed transform.

CONVENTIONAL WARPED SPECTRUM

Before discussing the M-W transform, we first describe the process normally used to approximate the warped spectrum. As a first step in the process, a short time amplitude spectrum is computed from a 10 to 20 millisecond windowed snapshot of the data. For a real, discrete signal $F(t_k)$ sampled at a uniform sample rate ω_s, the complex spectrum has the representation

$$\hat{F}(\omega_n, N) = \sum_{k=0}^{2(N-1)} F(t_k) W(k) e^{-ik\omega_n}, \tag{1.1}$$

where $\omega_n = \omega_s/n$, $n = 1, \ldots, N$, $i = \sqrt{-1}$ and $W(k)$ is an appropriately chosen windowing function. The corresponding amplitude spectrum has the representation

$$P(\omega_n, N) = \left|\hat{F}(\omega_n, N)\right|. \tag{1.2}$$

The conventional warped amplitude spectrum is functionally equivalent to a matrix product of the form

$$\vec{P}_W = \Psi\vec{P}, \tag{1.3}$$

where Ψ is a positive matrix of filter weights used to combine coefficients of the amplitude spectrum to approximate the desired filter responses and

$$\vec{P} = (P(\omega_1, N), \ldots, P(\omega_N, N))^T. \tag{1.4}$$

For constant Q filterS, the integration times of each filter should be on the order of the reciprocal of the filter bandwidth. Since equation (1.3) is based on uniform integration times, independent of filter bandwidth, there are unnecessary time-frequency ambiguities in the warped spectrum under this representation. If the integration time is adequate to properly contain energy within the narrow filters, the time resolution of the coarser filters suffers. If the integration time is chosen to optimize the time-frequency resolution of the coarser filters, the energy in the narrower filters are corrupted by the out-of-band energy from nearby filters.

A better filter implementation would be to design a bank of properly spaced constant Q filters with the appropriate bandwidths. The integration time of these filters should be dependent on the filter bandwidths and should be on the order of the reciprocal of the filter bandwidths. Since the time-dependent FFT is essentially a bank of uniform bandwidth narrow-band filters with complex output, we can express the constant Q filter outputs as functions of the Fourier coefficients, If $F_\omega(\zeta)$ is the time varying output of the ω^{th} filter and $\hat{F}_j(\zeta)$ is the j^{th} FFT coefficient on the interval I_ζ, then the time-varying power of the ω^{th} filter is of the form

$$P_\omega(\zeta) = \left|\sum a_j \hat{F}_j(\zeta)\right|^2. \tag{1.5}$$

0-7803-2431-5/95 $4.00 © 1995 IEEE

The approximation which is always used in speech processing is

$$\tilde{P}_{\omega}(\zeta) = \sum \left| a_j \hat{F}_j(\zeta) \right|^2 \quad (1.6)$$

Clearly this power estimate is Schwartz inequality approximation. A more subtle point is that it is based on a circular convolution rather than the correct linear convolution. The argument is usually made that the formula (1.6)is correct assuming independence of the coefficients. This is only the case for phasers whose frequencies are in the exact center of the FFT cells.

DESCRIPTION OF TRANSFORM

We now introduce the M-W transform, which is the focus of this paper. The M-W transform is a special case of a larger class of doubly warped transforms which contain both it and the Fourier transform. In general, this class of transforms operate on a warped time-domain signal to produce a spectrum which is warped. In order to uniformly cover the warped spectrum, the time domain warping function must be matched to the frequency domain warping function and the transform basis. In addition, since the filter bandwidths are nonuniform, the intervals of support for the filters are dependent on the warping functions.

In general, a doubly warped transform is of the form

$$\hat{F}_{w_t, w_f, b}(\omega) = \int F(w_t(t)) \overline{b_{w_f(\omega)}(w_t(t))} dt, \quad (2.1)$$

where b is an arbitrary transform basis which covers the warped spectrum and w_t and w_f are time and frequency warping functions. If both warping functions are logarithmic, a wavelet-like transform may be computed as a linear convolution with a single basis element. The resulting transform is a Mellin-Wavelet transform.

In the case of log-warping, the discrete transform represents an exponential sampling of the continuous spectrum. The warping function which drives this sampling is

$$w_{\log}(x \in [x_0, x_1], a, b) = e^{\ln a + \frac{x - x_0}{x_1 - x_0}(\ln b - \ln a)}. \quad (2.2)$$

To complete the transform description, we must specify a basis and a time domain warping function. We consider first a Fourier(Morlet)-like basis consisting of complex phasers of the form

$$\Phi(t) = e^{i\omega t + \phi}, \quad i = \sqrt{-1}. \quad (2.3)$$

To create a 'Morlet' MW basis of the type (2.3), we must warp them in the time domain and define time domain support regions which result in proper filter bandwidths.

For uniform covering of the warped spectrum, the filter support regions must have length proportional to the reciprocal of the filter bandwidth. In the case in which the spectrum is log warped, we see that a time domain warping function of the form (2.2) has the interesting property that translations of intervals under this warping satisfy the desired filter support region property. The problem is that the origin must be avoided since the logarithm is undefined at zero.

If we select a bank of filters which appears uniformly spaced in the log warped frequency domain, then the support intervals L_{ω} for the constant Q filters must satisfy

$$\omega L_{\omega} = const \quad (2.4)$$

Equation (2.4) is equivalent to requiring that basis elements each contain the same number of cycles in their respective support intervals. If we assume that each phaser (2.3) has initial phase $\phi = 0$ at $t = 0$, and we require that the support intervals for each phaser start at the same phase, we have the properties that the basis elements are a time warped wavelet basis, and the support intervals in the warped time domain have the same length. Choosing the basis in this manner guarantees uniqueness of the basis and support intervals, and results in the property that the basis elements and support intervals are translations of each other in the warped time domain and the frequency responses of the resulting filters are translations of each other in the warped frequency domain.

In general, if we choose w_t and w_f to be the logarithmic warping function $w_{\log}$, and b to be any basis with support interval excluding the origin, but satisfying the wavelet basis property

$$b_{\omega_\alpha}\left(k_{\omega_\alpha} t\right) = b_{\omega_\beta}\left(k_{\omega_\beta} t\right), \quad (2.5)$$

we have that the resulting doubly warped transform (2.1) has the property that the resulting warped basis elements

$$b_{\omega, w_L}(t) = b_{\omega}(w_{\log}(t)) \quad (2.6)$$

are translations of each other in the warped time domain, satisfying the property

$$b_{\omega_\alpha}(w_{\log}(t)) = b_{\omega_\beta}(w_{\log}(t) + k(\alpha, \beta)), \quad (2.7)$$

where $k(\alpha, \beta)$ is independent of time

For suitably chosen wavelets, the functions (2.7) will span the spectrum and thus form a basis which we call the Mellin-Wavelet basis for the warped MW transform. This basis has the property that the doubly logarithmically warped transform with this basis may effectively be computed as the linear convolution of any one of the basis elements and the log warped incoming signal. If the original wavelet basis b has finite support, fast convolution methods may be used to compute the linear convolution. Unfortunately, the transform is not necessarily translation invariant since the basis elements are time translations of each other. In order for the transform to be of use, the signal must be preprocessed to properly phase all spectral components.

CALCULATION OF TIME DOMAIN SAMPLING FUNCTION FOR A 'MORLEt' MW TRANSFORM

In designing the sampling function for a constant Q log-warped spectrum, the general form must be an exponential of the form

$$t_n = w(n, N, \alpha, \beta, \zeta) = \zeta + \frac{e^{\alpha n + \beta}}{\omega_s},\ n = 0 \ldots N-1, \quad (3.1)$$

where ζ is the base time for the transform, ω_s is the original uniform sample rate, and N is the sum of the length of the filter in the warped time domain and the number of outputs. The two free parameters, α and β, must be chosen such that the resulting warped filter bank covers the desired frequency range. For 'Morlet' transforms, there is sufficient information to almost completely determine the sampling function. For general MW transforms, the sampling function may be determined from considerations similar to those used in this discussion, but the function parameters are dependent on the choice of the wavelet basis.

We assume that we wish to design a transform with L filters based on a warped convolution of length $N-L$. We further assume that the filters must cover the frequency interval $[\omega_l, \omega_u]$. Under these assumptions, the rate of sampling is related to the derivative of the sampling function, and we have the relationship

$$\left.\frac{\omega_u \partial w}{\omega_l \partial n}\right|_{n = n_0} = \left.\frac{\partial w}{\partial n}\right|_{n = n_0 + L} \quad (3.2)$$

This equation reduces to

$$\alpha = \frac{1}{L}\ln\left(\frac{\omega_u}{\omega_l}\right). \quad (3.3)$$

The choice of β is somewhat arbitrary, but consideration must be given to its choice since it affects the sampling interpolation for small values of n and must not result in a violation of a Nyquist-like condition for large values of n. The sampling interpolation is due to the problem that the exponentially spaced samples are, in general, more closely spaced than the original uniformly spaced samples for small values of n. A Nyquist-like sampling condition is violated if the transform basis wavelet is under sampled on the tail (i.e. for large values of n). To get an upper bound on β, we assume a Nyquist condition which is equivalent to the restriction that the minimum sampling interval for the warped sampling function must be at least one half cycle (for real sampling) of the sinusoid from which the basis wavelet is generated. For complex sampling, the minimum interval must be at least one cycle. The reader is warned that this criterion is estimated and not derived analytically.

Under these conditions, the Nyquist criterion (for real sampling) is equivalent to

$$e^{\alpha(N-1)+\beta} - e^{\alpha(N-2)+\beta} \le \frac{\omega_s}{2\omega_l}. \quad (3.4)$$

(3.5)

This inequality is equivalent to

$$\beta \le \ln\frac{\omega_s}{2\omega_l\left(e^{\alpha(N-1)} - e^{\alpha(N-2)}\right)}. \quad (3.6)$$

$$= \ln\left(\frac{\omega_s}{2\omega_l}\left(\frac{\omega_u}{\omega_l}\right)^{\frac{-(N-2)}{L}}\left(\left(\frac{\omega_u}{\omega_l}\right)^{\frac{1}{L}} - 1\right)^{-1}\right).$$

Performance of the transform may be expected to improve for smaller values of β, but care must be taken to choose β so small that the performance degrades due to interpolation errors caused by over-sampling of the signal for small values of n.

The maximum interpolation required for this choice of β is given by

$$\Delta_{min} = \left(e^{\alpha+\beta} - e^{\beta}\right) \le \frac{\omega_s}{2\omega_l}\left(\left(\frac{\omega_l}{\omega_u}\right)^{\frac{N-2}{L}}\right) \quad (3.7)$$

If Δ_{min} is an integer, there is no interpolation. If $\Delta_{min} < 1$, the samples must be interpolated from the original uniform sampled signal, and if $\Delta_{min} > 1$ the resulting samples are decimations of the original sequence. In any event, for non integer values of Δ_{min}, some filtering algorithm must be used in re-sampling.

As an example, if we assume a typical situation in which ω_s = 8000Hz, ω_l = 300Hz and ω_u = 3000Hz then the following table gives typical design parameters.

(3.8)

N	128	128	256	256	512	512
L	32	64	64	128	128	256
α	0.0720	0.0360	0.0360	0.1800	0.1800	0.0090
β	-3.8807	1.3639	-3.2413	2.0301	-2.5751	2.7097
Δ_{min}	0.0015	0.1433	0.0014	0.1382	0.0014	0.1358

In each of the columns in table (3.7), β was chosen so that the last two samples of a Morlet basis generator are separated by half a period. In the examples in (3.7), choices of L and N in the ratio of approximately 1:2 result in worst case interpolation situations in which the signal would have to be up-sampled for a few samples by a factor of a little less than 7. By contrast, the corresponding situation for L:N ratios of 1:4 results in worst case interpolations requiring up-sampling for the first few samples by a factor of 667 or more. If a windowing function such as a log-warped raised cosine is used, both the region at the beginning of the basis generator where interpolation is worst and the region at the end of the basis generator where the decimation is worst are attenuated by the windowing function. Because of this, a low order, interpolating filter may be used if Δ_{min} is sufficiently large, say 0.1 or greater.

As an example, we can construct a "Morlet" MW basis generator. With the choices of α and β, this basis may be constructed by sampling a complex phaser whose frequency is f_l by the sampling function (3.1). For the phaser

$$\Phi(t) = e^{i\omega_u t},\ i = \sqrt{-1}, \tag{3.9}$$

the sampled basis generator is

$$(\Phi(t_0), \ldots, (\Phi)(t_{N-L-1})). \tag{3.10}$$

All other basis elements are translations of this generator in the warped time domain

For an arbitrary wavelet basis, the MW basis generator is obtained by sampling the mother wavelet with the warped sampling function(3.1).

CALCULATION OF SAMPLE SPACING

Let ω_s, ω_l and ω_u be the original uniform sample rate, the lower cutoff frequency and the upper cutoff frequency and let T_l, T_u and T_m be the times of the first sample, last sample and the m^{th} sample in the transform interval respectively. Since $\log 0 = -\infty$, we must start the correlations at some phase other than zero. We assume that all basis functions start with initial phase $\phi_0 = \pi$. In this case, the time of the first sample is dependent only on the upper cutoff frequency

$$T_l = \frac{\omega_s}{2\omega_u}. \tag{4.1}$$

The sample times are given by

$$T_m = exp\left[\log T_l + \frac{m}{n-1}(\log T_u - \log T_l)\right], m = 0\ldots n-1$$
$$= \left(T_l^{n-m-1} T_u^{m}\right)^{\frac{1}{n-1}}. \tag{4.2}$$

To calculate the upper sample limit, we assume a Nyquist-like criterion that the maximum sample interval between two consecutive samples is such that there are at least two samples for every period of the basis elements (real sampling) or one sample per period (complex sampling). Since the basis elements are warped time translations of each other, for real sampling, this condition is equivalent to

$$2 = \frac{\omega_s}{\omega_l} = T_n - T_{n-2} \tag{4.3}$$
$$= \left(T_l^{-1} T_u^{n}\right)^{\frac{1}{n-1}} - \left(T_l T_u^{n-2}\right)^{\frac{1}{n-1}}$$
$$= \left(\left(\frac{\omega_s}{2\omega_u}\right)^{-1} T_u^{n}\right)^{\frac{1}{n-1}} - \left(\frac{\omega_s}{2\omega_u} T_u^{n-2}\right)^{\frac{1}{n-1}}.$$

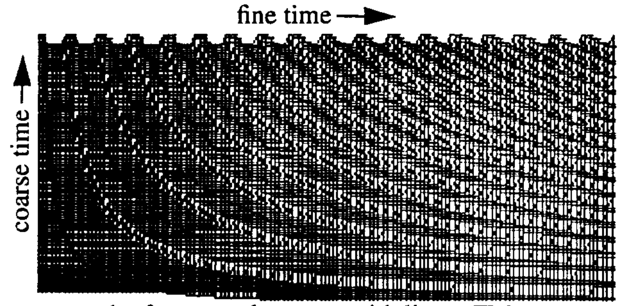
for complex sampling, the equivalent condition is

$$1 = \frac{\omega_s}{\omega_l} = T_n - T_{n-2} \tag{4.4}$$

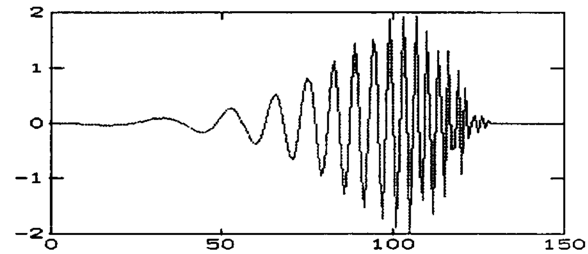
Unfortunately, the closed form solution of equation (2.3) for T_u is not readily apparent, but the equation can easily be solved numerically. For complex sampling, the equivalent formula is

$$\frac{\omega_s}{2\omega_l} = T_n - T_{n-2}. \tag{4.5}$$

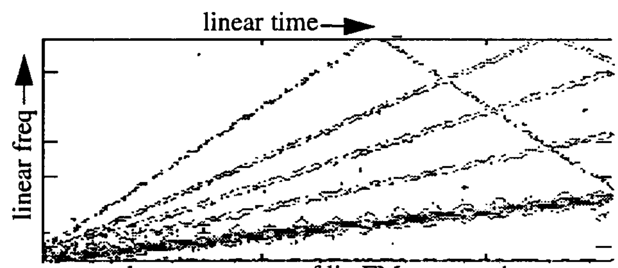
With this final parameter estimate, the "Morlet" design is essentially complete.



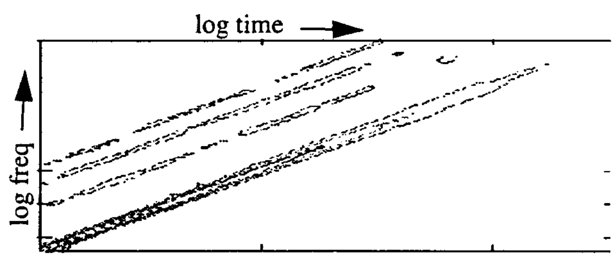
mesh of rectangular wave with linear FM



Real part of windowed 'Morlet' MW basis element



normal power spectra of lin FM rectangular wave



'Morlet' MW transform of linear FM rect. wave

References

[1]Daubechies, I,"Orthonormal Bases of Compactly Supported Wavelets",Comm. Pure & Applied Math.,v.XLI, pp.909-996,1988.

[2]Barniuk,R. and Jones,D.,"Warped Wavelet Basis:Unitary Equivalence and Signal Processing",Proc. IEEE-93 ICASSP, Minneapolis, pp III-320-323,Mar. 1993.

[3]Nelson,D. "Wavelet Based Analysis of Speech Signals", Proc. IEEE ASR-93 Workshop, Snowbird, Ut.,Dec. 12-15.