Infinite length windows for short-time Fourier transform

S. Tassart
Analysis-synthesis team, Ircam, PARIS, FRANCE
Stephan.Tassart@ircam.fr, http://www.ircam.fr/equipes/analyse-synthese/tassart

Abstract

This paper presents an extension of the Short-time Fourier transform (STFT) to the case of infinite rational windows. The choice of a suitable window for the STFT is a major issue in signal analysis. The ability to use an infinite impulse response filter as an analysis or synthesis window opens new perspectives.

1 Introduction

The short-time Fourier transform (STFT) is a widely-used signal processing tool for sound analysis and synthesis. It is commonly used, for example, in time-frequency analysis as well as in the phase vocoder. Since only a finite number of operations is possible in a computer implementation of the STFT, in most cases the analysis and the synthesis window are of finite length. Using such analysis or synthesis windows, usually leads to a tradeoff between time resolution, frequency resolution and amplitude of side lobes. Different optimal windows have been developed for several problems: Hann, Hamming, Kaiser windows...

In this paper, we propose an algorithm for computing, in a finite number of operations, the STFT of a signal windowed by a rational infinite length window, i.e. by the impulse response of an ARMA filter. This work extends the STFT theory to infinite length windows.

We present one particular result of this work concerning the problem of designing new optimized windows for specific analysis or synthesis problems. Infinite length windows do not follow the same constraints as finite length windows do. For instance, when designing optimal low pass filters, the criteria comprise minimizing ripples, maximizing the slopes, and the flatness of the transfer function. We present new window families, and some tradeoff criteria adapted to the problem of the tracking of partials for additive sound synthesis.

2 Theory

Given a sequence of complex numbers $(x_n)_{n\in\mathbb{Z}}$ and an analysis window $(w_n)_{n\in\mathbb{Z}}$, we form the STFT $X_n(e^{j\omega})$ (see [1, 9]) by:

$$\forall n \in \mathbb{Z},\ X^n(e^{j\omega}) = \sum_{m\in\mathbb{Z}} x_{n-m} w_m e^{-j(n-m)\omega} \qquad (1)$$

2.1 Notation

We define here a compact set of notation in order to highlight the key points of the demonstrations to follow. We set respectively:

- $\vec{w}_p$, a complex vector of size N defined as $\vec{w}_p = (w_{pN}, w_{pN+1}, \dots, w_{pN+N-1})$,
- $\vec{x}_p$, a complex vector of size N defined as $\vec{x}_p = (x_{pN}e^{-jpN\omega}, \dots, x_{pN+N-1}e^{-j(pN+N-1)\omega})$,
- $\langle\vec{y}|\vec{z}\rangle$, a symmetrical bilinear function operating on two vectors of size N, $\vec{y}$ and $\vec{z}$:

$$\langle\vec{y}|\vec{z}\rangle = \langle\vec{z}|\vec{y}\rangle = \sum_{m=1}^{N} y_m z_{N-m+1}$$

Given this bloc notation, the STFT from Eq. (1) is entirely recasted as an infinite bloc summation:

$$\forall k \in \mathbb{Z},\ X^{kN}(e^{j\omega}) = \sum_{m\in\mathbb{Z}} \langle\vec{w}_m|\vec{x}_{k-m}\rangle \qquad (2)$$

Following (2), $\langle\vec{v}|\vec{x}_k\rangle$ is to be interpreted as the result of the short-time Fourier transform $Y_{\vec{v}}^{kN}(e^{j\omega})$ of $(x_n)_{n\in\mathbb{Z}}$ using the N-point analysis window $\vec{v}$.

2.2 Bloc recursion

We suppose here that the causal analysis window $(w_n)_{n\in\mathbb{N}}$ is the infinite impulse response of an ARMA filter, whose transfer function is a rational function $W(z) = \tilde{B}(z)/\tilde{A}(z)$ of order P. In this case, there exists a set of vectors $(\vec{b}_q)_{q\in[0,P-1]}$ and coefficients $(a_p)_{p\in[1,P]}$, $a_P \neq 0$, describing a bloc recursion for the analysis window, δ_q being the Kronecker symbol:

$$\forall k \in \mathbb{Z},\ \vec{w}_k = \sum_{p=1}^{P} a_p \vec{w}_{k-p} + \sum_{q=0}^{P-1} \vec{b}_q \delta_{k-q} \qquad (3)$$

Note that the complex roots $(\gamma_p)_{p\in[1,P]}$ of the polynomial $z^P - \sum_{p=1}^{P} a_p z^{P-p}$ are deduced from those of $\tilde{A}(z)$, $(\tilde{\gamma}_p)_{p\in[0,P]}$, by the relation: $\gamma_p = \tilde{\gamma}_p^N$. The set of vectors $(\vec{b}_q)_{q\in[0,P-1]}$ is determined by evaluating the $N \cdot P$ first values of the analysis window $(w_n)_{n\in\mathbb{N}}$:

$$\forall q < P,\ \vec{b}_q = \vec{w_q} - \sum_{p=1}^{q} a_p \vec{w}_{q-p} \quad (4)$$

2.3 Bloc STFT

In the first step we compute a first order linear combination $c_0 X^{kN} + c_1 X^{(k-1)N}$ of two successive STFT, assuming only the causality of the window $(w_n)_{n\in\mathbb{N}}$:

$$\begin{aligned}
&c_0 X^{kN}(e^{j\omega}) + c_1 X^{(k-1)N}(e^{j\omega}) = \\
&c_0\langle \vec{w}_0|\vec{x}_k\rangle + c_0\sum_{m=1}^{+\infty}\langle \vec{w}_m|\vec{x}_{k-m}\rangle + c_1\sum_{m=0}^{+\infty}\langle \vec{w}_m|\vec{x}_{k-1-m}\rangle \\
&= c_0\langle \vec{w}_0|\vec{x}_k\rangle + c_0\sum_{m=1}^{+\infty}\langle \vec{w}_m|\vec{x}_{k-m}\rangle + c_1\sum_{m=1}^{+\infty}\langle \vec{w}_{m-1}|\vec{x}_{k-m}\rangle \\
&= c_0\langle \vec{w}_0|\vec{x}_k\rangle + \sum_{m=1}^{+\infty}\langle c_0\vec{w}_m + c_1\vec{w}_{m-1}|\vec{x}_{k-m}\rangle
\end{aligned}$$

The linear combination of STFT has been split into one term depending only on the first values of the analysis window and another term corresponding to the infinite sum exhibiting the same linear combination transposed to bloc-windows.

In a very similar way, the P-order linear combination $\sum_{p=0}^{P} c_p X^{(k-p)N}$ can also be split into two parts. The first part originates from the P first initial values, $(\vec{w}_p)_{p\in[0,P-1]}$. The second part corresponds to an infinite sum where the linear combination of STFT is transposed into a linear combination of bloc-windows:

$$\begin{aligned}
\sum_{p=0}^{P} c_p X^{(k-p)N}(e^{j\omega}) = \sum_{q=0}^{P-1}\left\langle \sum_{p=0}^{q} c_p\vec{w}_{q-p}\,\middle|\, \vec{x}_{k-q}\right\rangle + \\
\sum_{m=P}^{+\infty}\left\langle \sum_{p=0}^{P} c_p\vec{w}_{m-p}\,\middle|\, \vec{x}_{k-m-P}\right\rangle \quad (5)
\end{aligned}$$

The $(c_p)_{p\in[0,P]}$ coefficients have to be chosen so that the infinite sum disapears.

If the analysis window $(w_n)_{n\in\mathbb{N}}$ is infinite, but rational, then, we know from section 2.2 that it admits a set of coefficients $(c_p)_{p\in[0,P]}$ simplifying each term $\sum_{p=0}^{P} c_p\vec{w}_{m-p}$ to 0 for $m \geq P$ (see (3)). For this set of coefficients and with the help of (4), the first part of (5) can therefore be recasted as a linear combination on $(\vec{b}_q)_{q\in[0,P-1]}$. In other words the analysis window recursion (3) is transposed in terms of the following STFT recursion:

$$X^{kN}(e^{j\omega}) = \sum_{p=1}^{P} a_p X^{(k-p)N}(e^{j\omega}) + \sum_{q=0}^{P-1}\langle \vec{b}_q|\vec{x}_{k-q}\rangle \quad (6)$$

This last expression shows that whenever the analysis window is infinite, the STFT is computed in a finite number of steps. The benefit of the autoregressive structure of the analysis window is transposed in a vectorial autoregressive structure in the STFT.

The term $\sum_{q=0}^{P-1}\langle \vec{b}_q|\vec{x}_{k-q}\rangle$ is to be interpreted, following (2), as a finite length window STFT. Its analysis window is actually a truncated version of $(w_n)_{n\in\mathbb{N}}$, reduced to its first PN points. Thus, (6) shall be considered as a recursive STFT, based on a PN-point analysis window with an overlap of $(P-1)N$ points.

Note that the STFT analysis and synthesis stages are always dual from each other. For instance, the overlap-add (OLA) reconstruction method results by duality from the Fourier transform interpretation of the STFT [1]. Thus, (6) shall also be interpreted as a dual expression from an infinite response *synthesis* filter reconstruction method. Such an infinite response synthesis filter may help to design an oversampling filter or a long term correlation ... Unfortunatly, it does not seem possible to verify the perfect reconstruction condition when both analysis and synthesis windows are infinite (the least square method proposed in [4] for the reconstruction of the STFT leads for instance to an anticausal filter).

2.4 Time-frequency tradeoff

In continuous time, the time-frequency resolution of a window happens to be a tradeoff between a measure of its bandwidth D_ω and a measure of its duration D_t. When these measures are chosen to be the standard deviation of respectively the time density for D_t and the spectral density for D_ω, then the gaussian window is known to minimize the product $D_{t\omega} = D_\omega D_t$ which is here considered as a time-frequency criterium (see [2]). Some other time-frequency criteria (equivalent noise bandwidth, the -3dB bandwidth ...) leading to different properties and tradeoff have already been proposed in [5].

For finite length discrete windows, the chosen criterium is rather a CPU-frequency than a time-frequency tradeoff. Both are usually linked since the length of a window is proportionnal to its computational cost. For instance, Harris in [5] compares the -3dB bandwidth of analysis windows of same *length*.

For infinite length discrete windows, we have to take into account a time-frequency criterium but also the computational cost. As a matter of fact, the time-frequency resolution of infinite length discrete sequences has not been widely studied from a theoritical point of view [8]. While we recognize that a similar Heisenberg *uncertainty principle* exists in discrete time, usual acceptations of bandwidth and duration do not meet this principle: the duration D_n can be made as small as one desires whereas the bandwidth D_ω remains finite. Here we propose to estimate the time-frequency resolution of discrete windows by adapting continuous-time relations. For a fast decreasing sequence, $(w_n)_{n\mathbb{Z}}$, $W(e^{j\omega})$ being its Fourier transform, the k^{th} moment of the time and frequency density exists and is defined as:

$$\left\langle \omega^k \right\rangle_\omega = \int_{-\pi}^{\pi} \omega^k |W(e^{j\omega})|^2 d\omega \quad (7)$$

$$\left\langle n^k \right\rangle_n = \sum_{n\in\mathbb{Z}} n^k |w_n|^2 \quad (8)$$

The discrete *duration* D_n and *bandwidth* D_ω are then defined as standard deviations:

$$D_\omega^2 = \frac{\left\langle \omega^2 \right\rangle_\omega}{\langle 1 \rangle_\omega} - \left(\frac{\langle \omega \rangle_\omega}{\langle 1 \rangle_\omega}\right)^2 = \frac{\left\langle \omega^2 \right\rangle_\omega}{\langle 1 \rangle_\omega} \quad (9)$$

$$D_n^2 = \frac{\left\langle n^2 \right\rangle_n}{\langle 1 \rangle_n} - \left(\frac{\langle n \rangle_n}{\langle 1 \rangle_n}\right)^2 \quad (10)$$

We want now to replace, for instance, D_ω by $f(D_\omega)$ in such a way that the product $D_{n\omega} = D_n f(D_\omega)$ could be considered as a time-frequency estimation and meet an uncertainty principle. It would seem necessary that $f(D_\omega)$ diverges when D_n tends to zero. When D_n is large enough, $f(D_\omega) \underset{0}{\sim} D_\omega$ seems enough to fullfill the constrain. Eq. (11) gives a reasonable estimation of the time-frequency resolution of discrete windows. We conjecture this quantity to be greater than 1/2 for any discrete sequence:

$$D_{n\omega} = \frac{2}{\sqrt{3}} D_n \tan\left(\frac{\sqrt{3} D_\omega}{2}\right) \quad (11)$$

It should be noted that the well-known bilinear transform maps a discrete-time frequency into a continuous-time frequency by means of the trigonometric tangent function; that may justify the choice for the function f. The normalization factor $\sqrt{3}/2$ is chosen in order to fit the bandwidth of the unit impulse.

2.5 Exponential window

The exponential causal (single-sided) window is defined as: $\forall n \geq 0,\ w_n = a^n$ and 0 elsewhere. This exponential window verifies a first order recursion, i.e $P = 1$ and $\forall n \geq 0,\ w_n = a w_{n-1} + \delta_n$

This equation is easily recasted as a vector recursion, the vector $\vec{w}_0$ being made from the N first values of exponential window. Applying relation (6) to the last recursion leads to the following relationship, already shown in [10]:

$$X^{kN}(e^{j\omega}) = a^N X^{(k-1)N}(e^{j\omega}) + Y_{\vec{w}_0}^{pN}(e^{j\omega})$$

The time-frequency resolution of the analysis window evaluated from (11) depends on a. It diverges for a in the neighbourhood of 0 and 1, and admits a minimum $D_{n\omega} = 1.24$ at $a = 0.42$. The coefficient a^N may also be viewed as a forgetting factor in an adaptative scheme of the STFT algorithm. This algorithm is also known as an *exponential average* [7].

2.6 Discussion

This section points out details which slighty differ from the common STFT. Usually, the length N of the analysis window is directly related to the time-frequency resolution since the window shape is stretched to the correct size and the N-point FFT gives N *bins* uniformly spaced in frequency. With the recursive STFT, N is no longer linked to time resolution.

The overlap factor is commonly understood as the rate of advancement of the analysis window relative to the length of the window. In theory it corresponds to a decimation coefficient, and therefore is related to the bandwidth of the analysis window. However, for applications where this analysis stage precedes a transformation and a synthesis stage, the overlap factor must be chosen with regard to the type of transformation planned. A 75%-overlap rate STFT is quite usual for common transformations such as time-stretching [6]. This overlap factor is also a means to estimate the computational cost per unit of time of processed signal. The 75%-overlap rate STFT corresponds therefore to a 4^{th} order filter.

The effective overlap factor should rather be understood as the rate of advancement of the analysis window compared to its duration. The duration of common analysis windows (Hann, Blackman, Hamming...) is far less than their length. One should expect similar effective overlap factors for finite or infinite length windows. This factor is related on one hand to the duration of the analysis window and on the other hand to the number of frequency bins used for evaluating the Fourier transforms (i.e. N, the size of the FFT). The overall steps needed for designing a infinite length analysis window are summed up here:

- choose the computational cost (overlap factor),

	N	D_n	D_ω	$D_{n\omega}$
Rectangular	256	74	0.10	7.7
Rectangular	2048	590	0.036	22
Blackman -92dB	256	26	0.020	0.52
Blackman -92dB	2048	210	0.0030	0.62
Hann	2048	290	0.0018	0.51
Hann-Poisson(0.5)	2048	260	0.0019	0.51
$A(10^{-4}, 21.6)$	2048	290	0.0018	0.51
$A(1.8, 0.92)$	2048	200	0.0027	0.55
Butterworth order 3	2048	570	0.0011	0.61
Butterworth order 4	2048	690	0.00098	0.67

- deduce the order of the filter,
- choose N the numbers of bins of FFT,
- deduce the bandwidth of the filter,
- design the ARMA filter,
- compute the duration of its impulse response,
- evaluate the effective overlap factor.

3 Extraction of spectral peaks

In the context of sound analysis, Depalle and Hélie in [3] proposed an efficient method to improve the estimation of frequency, amplitude and phase of partials of a signal based on a parametric modeling of the short-time Fourier transform. Frequency estimation is highly sensitive to the analysis window shape, and nosidelobe windows were necessary to prevent false detections due to local minima. A small bandwidth improves the conditionning for the algorithm whereas a small effective duration minimizes the smoothing effect of the time variation of parameters. Unfortunately the estimation of the time-frequency resolution of the windows presented in [3] is not adapted to infinite length windows. In the following results, N gives the number of bins in the FFT and an estimation of the computational cost (also related to the filter's order), D_n, D_ω and $D_{n\omega}$ are evaluated following (10), (9) and (11) as an estimation of the time-frequency characteristics of the window.

Butterworth filters have been chosen in the process of filter design since they are characterized by a magnitude response that is maximally flat in the passband and monotic overall. These filters sacrifice the rolloff steepness for monotonicity, and are therefore well suited for the forementionned algorithm.

The 3 firsts windows presented in the table (rectangular, Blackman, Hann) have sidelobes on the contrary of the remaining other windows. The table shows that the Butterworth filters achieve approximatively the same resolution as a large -92dB Blackman window, but it seems also that optimal nosidelobe A-windows designed in [3], or nosidelobe Hann-Poisson windows, have still better time-frequency characteristics. However, in every case, the bandwidth of the Butterworth filter is sharper, causing the effective overlap factor to be smaller, allowing one to decrease N and to increase the order of the filter in applications where the quality of the transformation depends on this factor (phase vocoder).

4 Conclusion

This paper has demonstrated how to extend the STFT to the case of infinite analysis or synthesis windows. We have also proposed different ways to compare the characteristics of these windows to those of finite length windows. Future work comprises design of new windows and application of this extension to the STFT to other analysis or synthesis algorithms.

References

[1] J. B. Allen and L. R. Rabiner. A unified approach to short-time Fourier analysis and synthesis. *Proc. IEEE*, 65(11):1558–1564, November 1977.

[2] L. Cohen. *Time-frequency analysis.* Signal Processing series. Prentice Hall, 1995.

[3] P. Depalle and T. Hélie. Extraction of spectral peak parameters using a short-time Fourier transform modeling and no sidelibe windows. In *Proc. of IEEE ASSP Workshop on App. of Sig. Proc. to Audio and Acoust.*, Mohonk, New Paltz, USA, October 1997.

[4] D. W. Griffin and J. S. Lim. Signal estimation from modified short-time Fourier transform. *IEEE Trans. Acoust., Speech, and Signal Process.*, ASSP-32(2):236–242, April 1984.

[5] F. J. Harris. On the use of windows for harmonic analysis with the discrete Fourier transform. *Proc. IEEE*, 66(1):51–83, January 1978.

[6] J. Laroche and M. Dolson. About this phasiness business. In *Proc. Int. Computer Music Conf. (ICMC'97)*, pages 55–58, Thessaloniki, September 1997.

[7] R. Miquel. *Le Filtrage Numérique par microprocesseurs.* Éditests, 1985.

[8] J. Pearl. Time, frequency, sequency and their uncertainty relations. *IEEE Trans. on Info. Theory*, 19:225–229, March 1973.

[9] M. R. Portnoff. Time-frequency representation of digital signals and systems based on short-time Fourier analysis. *IEEE Trans. Acoust., Speech, and Signal Process.*, ASSP-28(1):55–69, February 1980.

[10] T. Saso. On short-time Fourier transform with single-sided exponential window. *Signal Processing*, 55:141–148, 1996.