SIGNAL PROCESSING

Meaning of Hop-Size in Filter Bank Interpretation of Short Term Fourier Transform

Asked 4 years, 8 months ago Active 1 year, 1 month ago Viewed 3k times

3

The Short Term Fourier Transform (STFT) is used to obtain the time-varying spectrum from a signal. There are two ways to understand it - the overlap add interpretation, and the filter bank interpretation. The overlap-add method is where the signal is cut up into windowed segments in time and the DFT of each segment is taken. The window is advanced forward by R samples in time, where R is called the hop-size.

The filter bank interpretation is where the window is seen as a low-pass filter, and the segment is shifted in frequency by an amount equal to the time segment. The hop-size here is interpreted as a down-sampling in frequency.

I am unable to understand this interpretation and would like to know how R is interpreted as down-sampling of the signal in frequency. This would also clarify the constraints on choosing a hop-size R, to prevent aliasing in frequency (as having a hop-size R means down-sampling in frequency).

fft frequency-spectrum short-time-ft

Share Improve this question Follow

edited May 6 '20 at 17:13
Gideon Genadi Kogan
416 2 11

asked Oct 4 '16 at 9:25
Samyukta Ramnath
299 2 11

2 Answers

Active Oldest Votes

2

Your message was apparently unfinished ('When reading about the DCT...'). I thus will be starting from it.

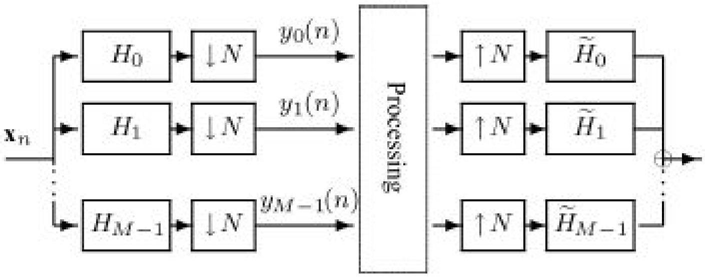
The JPEG compression format is an example of hopping. A DCT-filter bank (Discrete Cosine Transform) is applied on the whole image, and subsampled by 8×8 (in both directions). Another interpretation is via a rectangular window, implemented in a fake overlap-add version, as the overlap is zero (8×8 blocks are almost disjoint). The DCT has a very simple interpretation, as it consists of an infinite continuous cosine basis, discretized and windowed with a rectangular window.

The traditional Short Term Fourier Transform (STFT) is an improvement, with the use of an overlapping window. In the continuous setting, an STFT with a window h can be inverted with a lot of dual windows g (under very mild assumptions on h and g).

Going to discrete implementations, certain choices of overlap, hop and window can lead to easy inversion. Inversion, or synthesis is the turning point: you can "analyse" any signal with any window, or hop. But not all designs allow synthesis or inversion, although it should be permitted more easily than in the critical and orthogonal framework, because of the potential redundancy.

There are instance of windows allowing discrete, non-redundant inversion, like the Modified DCT, the LOT (lapped orthogonal transform), the MLT (modulated lapped transform), and a handful of others (GenLOT, GULLOT). They are used for instance in the mp3 (mpeg-1 layer 3) style of audio compression.

In the above case, the hop is equal to the number of channels, similarly to the case of the JPEG DCT: 8-channel DCT, and a jump of 8 pixels in each direction. Allowing only a window for all the channels is somewhat restrictive. The most efficient framework is that of oversampled filter banks:



There is no direct concept of a global window for all M-band filters, $M \geq N$. Instead, some windowing is embedded in each one. In the above picture, the N corresponds to your R. My understanding is that the downsampling takes place in the time-domain (hopping), but it results in a shift in frequency (and aliasing), as filter H_i outputs are sent to the base-band.

Then, given a set of analysis filters H_i (including some overlapping), and if M is greater than N, you can very often cancel all the aliasing induced by the N downsampling with an infinity of inverse filter banks. One issue is to find ONE suitable inverse.

This is discussed for instance in Optimization of Synthesis Oversampled Complex Filter Banks, J. Gauthier et al., IEEE Transactions on Signal Processing, 2009:

> An important issue with oversampled FIR analysis filter banks (FBs) is to determine inverse synthesis FBs, when they exist. Given any complex oversampled FIR analysis FB, we first provide an algorithm to determine whether there exists an inverse FIR synthesis system. We also provide a method to ensure the Hermitian symmetry property on the synthesis side, which is serviceable to processing real-valued signals. As an invertible analysis scheme corresponds to a redundant decomposition, there is no unique inverse FB. Given a particular solution, we parameterize the whole family of inverses through a null space projection. The resulting reduced parameter set simplifies design procedures, since the perfect reconstruction constrained optimization problem is recast as an unconstrained optimization problem. The design of optimized synthesis FBs based on time or frequency localization criteria is then investigated, using a simple yet efficient gradient algorithm.

Share Improve this answer Follow

answered Oct 4 '16 at 19:26
Laurent Duval
27k 3 25 87

1

I have not seen it in a book but it seems to me that if the STFT definition is:

$$X(\tau,\omega)=\int_{-\infty}^{\infty}x(t)w(t-\tau)e^{-i\omega t}dt$$

Then we can define the filter as

$$h(t)=w(t)e^{-i\omega t}$$

where $w(t)$ can be interpreted as a low-pass filter and the exponent as frequency shifting method (see Shift Theorem). Next we define

$$g(t)=h(-t)$$

where both have the same frequency content with a phase shift between them (see the conjugation property of the Fourier Transform). Thus

$$X(\tau,\omega)=e^{-i\omega\tau}\int_{-\infty}^{\infty}x(t)w(t-\tau)e^{-i\omega(t-\tau)}dt=e^{-i\omega\tau}\int_{-\infty}^{\infty}x(t)g(\tau-t)dt$$

and here we got the convolution of the signal with a bandpass function $g(t)$, with some phase shift (which we might correct or neglect)

$$X(\tau,\omega)=e^{-i\omega\tau}(x*g)(\tau)$$

Since we are implementing the Discrete STFT, τ resolution is given by the hop, R.