Solution of the tutorial on the conversion of sampling frequency and STFT

Roland Badeau

This tutorial aims to carry out filtering in a multirate system (with an application to the conversion of sampling frequency), and to understand a perfect reconstruction filter bank for audio equalization.

1 Conversion of sampling frequency

We want to achieve the conversion of sampling frequency from $F_s = 48kHz$ to $F_s = 32kHz$.

1. Describe and draw the digital processing chain that will permit you to achieve such a conversion (in particular you need to specify the characteristics of the ideal filter $H(e^{2i\pi\nu})$ that you will have to use).

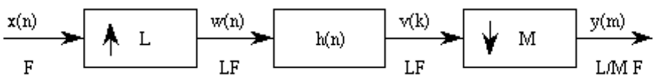

Figure 1: Resampling diagram

The block diagram of resampling is depicted in Figure 1. Here we have $F = F_s$, $L = 2$, and $M = 3$. The frequency response of the anti-aliasing, low-pass filter h is

$$\begin{cases} H(e^{i2\pi\nu}) = L = 2 & \forall |\nu| < \min\left(\frac{1}{2L}, \frac{1}{2M}\right) = \frac{1}{6} \\ H(e^{i2\pi\nu}) = 0 & \forall \min\left(\frac{1}{2L}, \frac{1}{2M}\right) = \frac{1}{6} < |\nu| < \frac{1}{2} \end{cases}.$$

The purpose of the rest of this section is to get an efficient implementation of this conversion of sampling frequency and to compare its performance with that of a direct implementation.
The efficient implementation will be achieved by means of polyphase decompositions (of type I and type II). The use of noble identities and of the equivalence below (Figure 2) will permit you to apply the filtering operations to signals at a lower sampling frequency.

2. Check the equivalence between the two diagrams in Figure 2.

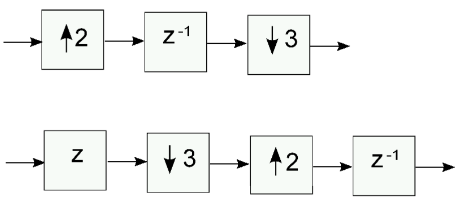

Figure 2: Equivalence

The equivalence in Figure 2 is proved in Figure 3.

3. Find the efficient implementation of this conversion of sampling rate by using two successive polyphase decompositions and the equivalence in Figure 2.

Filter h in Figure 1 has to be decomposed into its type 2 polyphase components at order $L = 2$. Then each of the two polyphase components obtained in this way have to be decomposed into their type 1 polyphase components at order $M = 3$ (actually, the two polyphase decompositions could

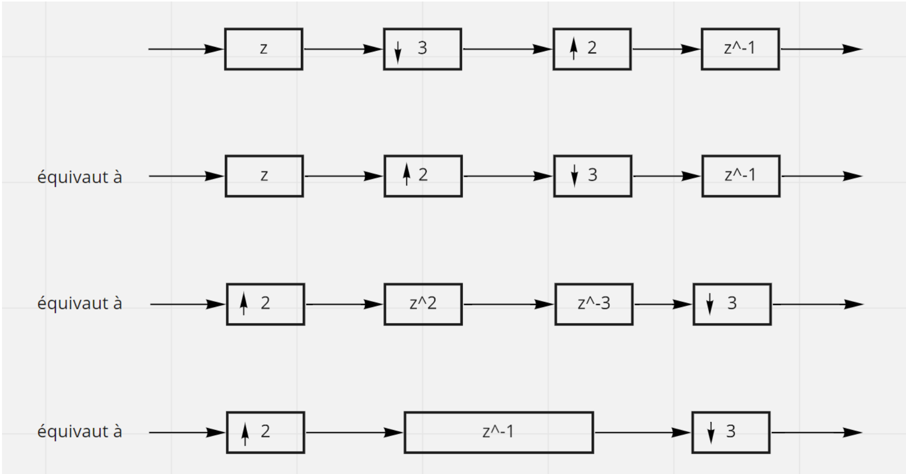

Figure 3: Proof of the equivalence

also be performed in the reverse order, but starting with the type 2 polyphase decomposition makes the derivation more compact). In the end, we get $2 \times 3 = 6$ *polyphase components, which can be switched with the decimators and the insertions of zeros in Figure 1 by using the noble identities and the equivalence of Figure 2, in order to get the efficient implementation of resampling.*

2 STFT audio equalization

2.1 STFT analysis

The definition of the STFT that is referred to as "low-pass convention" is given in discrete time by :

$$W_x(\lambda, b) = \sum_{n \in \mathbb{Z}} x(n) w(n-b) e^{-j2\pi\lambda n}, \tag{1}$$

where $w(n)$ is the analysis window in discrete time, which is supposed summable, real and symmetric.

1. We choose $w(n)$ as a Hann (Hanning) window of length N_w. What is the width of the main lobe as a function of N_w ?

   The width of the main lobe of the Hann window is $\frac{4}{N_w}$ *(in normalized frequency scale).*

2. Note that the expression (1), taken at fixed λ, can be written as a convolution and deduce an interpretation of the STFT in terms of filtering. Explain the role of the corresponding filter (low-pass? band-pass? high-pass?). As a linear phase FIR filter, specify its type (type 1,2,3,4?) according to its length (even or odd).

   Equation (1) *is equivalent to* $W_x(\lambda, b) = (\tilde{h} * \tilde{x})(b)$ *with* $\tilde{x}(n) = x(n)e^{-j2\pi\lambda n}$ *and* $\tilde{h}(n) = w(-n)$. *Filter h is a low-pass filter of cut-off frequency* $\frac{2}{N_w}$. *If the length of h is odd, then h if a type 1 FIR filter, and if the length of h is even, then h if a type 2 FIR filter.*

3. Another definition of the STFT, referred to as "band-pass convention", is given by

$$\tilde{X}(\lambda, b) = \sum_{n \in \mathbb{Z}} x(n+b) w(n) e^{-j2\pi\lambda n}. \tag{2}$$

Explain this latter designation and express $\tilde{X}$ as a function of W_x.

Equation (2) is equivalent to $\tilde{X}(\lambda, b) = (h * x)(b)$ with $h(n) = w(-n)e^{j2\pi\lambda n}$. Filter h is a band-pass filter of center frequency λ and bandwith $\frac{4}{N_w}$, hence the "band-pass" designation. Moreover, with $m = n + b$, we get $\tilde{X}(\lambda, b) = \sum_{m\in\mathbb{Z}} x(m)w(m-b)e^{-j2\pi\lambda(m-b)} = e^{j2\pi\lambda b}W_x(\lambda, b)$.

2.2 Reconstruction

The reconstruction is achieved by an *overlap-add* operation, which will be written as:

$$y(n) = \sum_{u\in\mathbb{Z}} y_s(u, n - uR),$$

where $y_s(u,n) = \mathrm{DFT}^{-1}[\tilde{X}(k,u)](n)\, w_s(n) = \dfrac{1}{M}\sum_{k=0}^{M-1} \tilde{X}(k,u)e^{j2\pi\frac{kn}{M}} \times w_s(n)$

4. Show that a sufficient condition for perfect reconstruction is $f(n) = 1$ $\forall n$ where $f(n) = \sum_{u\in\mathbb{Z}} w(n-uR)w_s(n-uR)$.

   The proof is in the course handout.