Digital Speech Processing—
Lecture 10

Short-Time Fourier Analysis Methods - Filter Bank Design
Review of STFT

1. \[ X_{\hat{n}}(e^{j\hat{\omega}}) = \sum_{m=-\infty}^{\infty} x[m]w[\hat{n} - m]e^{-j\hat{\omega}m} \]

- function of \( \hat{n} \) (looks like a time sequence)

- function of \( \hat{\omega} \) (looks like a transform)

\[ X_{\hat{n}}(e^{j\hat{\omega}}) \] defined for \( \hat{n} = 1, 2, 3, \ldots \); \( 0 \leq \hat{\omega} \leq \pi \)

2. Interpretations of \( X_{\hat{n}}(e^{j\hat{\omega}}) \)

1. \( \hat{n} \) fixed, \( \omega = \hat{\omega} \) variable; \( X_{\hat{n}}(e^{i\omega}) = \text{DTFT} \left[ x[m]w[\hat{n} - m] \right] \)

\( \Rightarrow \) DFT View \( \Rightarrow \) OLA implementation

2. \( n = \hat{n} \) variable, \( \hat{\omega} \) fixed; \( X_n(e^{j\hat{\omega}}) = x[n]e^{-j\hat{\omega}n} * w[n] \)

\( \Rightarrow \) Linear Filtering view \( \Rightarrow \) filter bank implementation

\( \Rightarrow \) FBS implementation
Review of STFT

3. Sampling Rates in Time and Frequency ⇒
   recover $x[\hat{n}]$ exactly from $X_{\hat{n}}(e^{j\omega})$
   
   1. time: $W(e^{j\omega})$ has bandwidth of $B$ Hertz ⇒
      $2B$ samples/sec rate

      Hamming Window: $B = \frac{2F_S}{L}$ (Hz) ⇒ sample at

      $\frac{4F_S}{L}$ (Hz) or every $L/4$ samples

   2. frequency: $w[n]$ is time limited to $L$ samples ⇒
      need at least $L$ frequency samples to avoid time aliasing
Review of STFT

- with OLA method can recover $x(n)$ \textit{exactly} using lower sampling rates in either time or frequency, e.g., can sample every $L$ samples (and divide by window), or can use fewer than $L$ frequency samples (filter bank channels); but these methods are highly subject to aliasing errors with any modifications to STFT

- can use windows (LPF) that are longer than $L$ samples and still recover with $N < L$ frequency channels; $\Rightarrow$ e.g., ideal LPF is infinite in time duration, but with zeros spaced $N$ samples apart where $F_s/N$ is the BW of the ideal LPF
Review of STFT

Fig. 6.11 Equivalent linear system relating $y_k(n)$ and $y(n)$ to $x(n)$.

$$h_k[n] = w[n]e^{jn\omega}$$

$$\tilde{H}(e^{j\omega}) = \sum_{k=0}^{N-1} H_k(e^{j\omega}) = 1$$

$$\tilde{h}[n] = \sum_{k=0}^{N-1} h_k[n] = \delta[n] = w[n]p[n]$$

$$p[n] = N \sum_{r=-\infty}^{\infty} \delta[n - rN]$$

$$\tilde{h}[n] = N \sum_{r=-\infty}^{\infty} w[rN] \delta[n - rN] = w[0] + w[N] + ...$$

⇒ need to design digital filters that match criteria for exact reconstruction of $x[n]$ and which still work with modifications to STFT
Tree-Decimated Filter Banks
Tree-Decimated Filter Banks

• can sample STFT in time and frequency using lowpass filter (window) which is moved in jumps of $R<L$ samples if $L \leq N$ where:
  – $L$ is the window length,
  – $R$ is the window shift, and
  – $N$ is the number of frequency channels

• for a given channel at $\omega = \omega_k$, the sampling rate of the STFT need only be twice the bandwidth of the window Fourier transform
  – down-sample STFT estimates by a factor of $R$ at the transmitter
  – up-sample back to original sampling rate at the receiver
  – final output formed by convolution of the up-sampled STFT with an appropriate lowpass filter, $f[n]$
Filter Bank Channels

Fully decimated and interpolated filter bank channels; (a) analysis with bandpass filter, down-shifting frequency and down-sampling; (b) synthesis with up-sampler followed by lowpass interpolation filter and frequency up-shift; (c) analysis with frequency down-shift followed by lowpass filter followed by down-sampling; (d) synthesis with up-sampling followed by frequency up-shift followed by bandpass filter.
Full implementation of STFT analysis/synthesis system with channel decimation by a factor of $R$, and channel interpolation by the same factor $R$ (including a box for short-time modifications.
Review of Down and Up-Sampling

\[ x[n] \xrightarrow{\downarrow R} x_d[n] = x[nR] \]

\[ X_d(e^{j\omega}) = \frac{1}{R} \sum_{r=0}^{R-1} X(e^{j(\omega-2\pi r)/R}) \]

- aliasing addition

\[ x[n] \xrightarrow{\uparrow R} x_u[n] = \begin{cases} x[n/R] & n = 0, \pm1, \pm2, \ldots \\ 0 & \text{otherwise} \end{cases} \]

\[ X_u(e^{j\omega}) = X(e^{j\omega R}) \]

- imaging removal

10
Filter Bank Reconstruction

- assuming no short time modifications

\[ X_{rR}(k) = X_{rR}(e^{j\omega_k}) = \sum_{m \in W_{rR}} w(rR - m)x(m)e^{-j\omega_km} \]

\[ y(n) = \frac{1}{N} \sum_{k=0}^{N-1} P(k)Y_{rR}(e^{j\omega_k})e^{j\omega_kn} \]

\[ = \frac{1}{N} \sum_{k=0}^{N-1} P(k)[X_{rR}[k]f[n - rR]] e^{j\omega_kn} \]

\[ = \sum_{m \in W_{rR}} x(m) \left[ \sum_{r=-\infty}^{\infty} w(rR - m)f(n - rR) \right] \frac{1}{N} \sum_{k=0}^{N-1} P(k)e^{j\omega_k(n-m)} \]

- recognizing that when \( P(k) = 1, \forall k \)

\[ \frac{1}{N} \sum_{k=0}^{N-1} e^{j\omega_k(n-m)} = \sum_{q=-\infty}^{\infty} \delta(n - m - qN) \]

- can show that the condition for \( y(n) = x(n), \forall n \) is

\[ \sum_{r=-\infty}^{\infty} w(rR - n + qN)f(n - rR) = \delta(q) \Rightarrow m = n \]
Filter Bank Reconstruction

- frequency domain equations, assuming:
  \[ h_k[n] = w[n]e^{j\omega_k n}; \quad g_k[n] = f[n]e^{j\omega_k n}; \]
  \[ H_k(e^{j\omega_k}) = W(e^{j(\omega - \omega_k)}); \quad G_k(e^{j\omega_k}) = F(e^{j(\omega - \omega_k)}); \]
  \[ Y(e^{j\omega}) = \frac{1}{N} \sum_{k=0}^{N-1} P(k) G_k(e^{j\omega_k}) \cdot \left[ \frac{1}{R} \sum_{\ell=0}^{R-1} H_k(e^{j(\omega - \frac{2\pi \ell}{R})}) X(e^{j(\omega - \frac{2\pi \ell}{R})}) \right] \]
  \[ = X(e^{j\omega}) \cdot \frac{1}{RN} \sum_{k=0}^{N-1} P(k) G_k(e^{j\omega_k}) H_k(e^{j\omega}) + \sum_{\ell=1}^{R-1} X(e^{j(\omega - \frac{2\pi \ell}{R})}) \cdot \frac{1}{RN} \sum_{k=0}^{N-1} P(k) G_k(e^{j\omega_k}) H_k(e^{j(\omega - \frac{2\pi \ell}{R})}) \]
  \[ = \tilde{H}(e^{j\omega}) \cdot X(e^{j\omega}) + \text{aliasing terms} \]
Filter Bank Reconstruction

- conditions for perfect reconstruction ($\hat{Y}(e^{j\omega}) = X(e^{j\omega})$)

$$\tilde{H}(e^{j\omega}) = \frac{1}{RN} \sum_{k=0}^{N-1} P(k)G_k(e^{j\omega_k})H_k(e^{j\omega}) = 1$$

Flat Gain

and

$$\frac{1}{RN} \sum_{k=0}^{N-1} P(k)G_k(e^{j\omega_k})H_k(e^{j(\omega - \frac{2\pi\ell}{R})}) = 0 \text{ for } \ell = 1, 2, ..., R$$

Alias Cancellation
• practical solution for the case $w(n) = f(n)$, $N = R = 2$ (2-band solution) where the conditions for exact reconstruction become

$$\frac{1}{4} \left[ P(0)F(e^{j\omega})W(e^{j\omega}) + P(1)F(e^{j(\omega-\pi)})W(e^{j(\omega-\pi)}) \right] = 1$$

and

$$\frac{1}{4} \left[ P(0)F(e^{j\omega})W(e^{j(\omega-\pi)}) + P(1)F(e^{j(\omega-\pi)})W(e^{j(\omega-2\pi)}) \right] = 0$$

• aliasing cancels exactly if $P(0) = -P(1) = 1$ (with $F(e^{j\omega}) = W(e^{j\omega})$) giving

$$\frac{1}{4} \left[ (W(e^{j\omega}))^2 - (W(e^{j(\omega-\pi)}))^2 \right] = e^{-j\omega M}$$

$$\hat{y}(n) = x(n - M)$$
Decimated Analysis-Synthesis Filter Bank

When aliasing is completely eliminated (by suitable choice of filters), the overall analysis-synthesis system has frequency response:

\[
\tilde{H}(e^{j\omega}) = \frac{1}{RN} \sum_{k=0}^{N-1} P[k] W_e(e^{j(\omega-\omega_k)})
\]

where

\[
W_e(e^{j\omega}) = W(e^{j\omega}) F(e^{j\omega})
\]

\[
\tilde{h}[n] = w_e[n] p[n]
\]

where

\[
w_e[n] = w[n] * f[n]
\]

\[
p[n] = \frac{1}{RN} \sum_{k=0}^{N-1} P[k] e^{j\omega_k n}
\]
Decimated Analysis-Synthesis Filter Bank

To design a decimated analysis/synthesis filter bank system, need to determine the following:

1. the number of channels, $N$, and the decimation/interpolation ratio $R$. (Usually determined by the desired frequency resolution)

2. the window, $w[n]$, and the interpolation filter impulse response, $f[n]$. (These are lowpass filters that should have good frequency-selective properties such that the bandpass channel responses do not overlap into more than one band on either side of the channel).

3. the complex gain factors, $P[k]$, $k=0,1,\ldots,N-1$. (These constants are important for achieving flat overall response and alias cancellation.)
Maximally Decimated Filter Banks

- we show here that it is possible to obtain exact reconstruction with $R=N$ and $N<L$.
  - $R=N$ is termed maximal decimation because this is the largest decimation factor that can be used with an $N$-channel filter bank and still achieve exact reconstruction.
  - nominal bandwidth of the bandpass channel signal is increased from $\pm \pi/N$ to $\pm \pi$
  - down-sampling by $N$ accomplishes frequency down-shifting normally accomplished by modulation
Maximally Decimated Filter Banks

Analysis

Synthesis
Analysis-Synthesis Filter Bank
(all modulators removed)
Two Channel Filter Banks
Two Channel Filter Banks

For this two-band system, we have:

\[ R = N = 2; \quad \omega_k = \pi k, \quad k = 0, 1 \]

Applying the frequency domain analysis we get:

\[
Y(e^{j\omega}) = \frac{1}{2} \left[ G_0(e^{j\omega})H_0(e^{j\omega}) + G_1(e^{j\omega})H_1(e^{j\omega}) \right] X(e^{j\omega})
\]

\[
+ \frac{1}{2} \left[ G_0(e^{j\omega})H_0(e^{j(\omega-\pi)}) + G_1(e^{j\omega})H_1(e^{j(\omega-\pi)}) \right] X(e^{j(\omega-\pi)})
\]

not including multiplier \(1/N\) or the gain factors \(P[0], P[1]\)

Aliasing can be eliminated completely by choosing the filters such that:

\[
\left[ G_0(e^{j\omega})H_0(e^{j(\omega-\pi)}) + G_1(e^{j\omega})H_1(e^{j(\omega-\pi)}) \right] = 0 \text{ (aliasing cancelation)}
\]

Perfect reconstruction requires:

\[
\frac{1}{2} \left[ G_0(e^{j\omega})H_0(e^{j\omega}) + G_1(e^{j\omega})H_1(e^{j\omega}) \right] = 1 \text{ (flat gain condition)}
\]
Quadrature Mirror Filter (QMF) Banks

Alias cancellation achieved if the filters are chosen to satisfy the conditions:

\[ h_1[n] = e^{j\pi n} h_0[n] \iff H_1(e^{j\omega}) = H_0(e^{j(\omega - \pi)}) \]
\[ g_0[n] = 2h_0[n] \iff G_0(e^{j\omega}) = 2H_0(e^{j\omega}) \]
\[ g_1[n] = -2h_1[n] \iff G_1(e^{j\omega}) = -2H_0(e^{j(\omega - \pi)}) \]

Note that there is only a single lowpass filter, with impulse response \( h_0[n] \) on which all other filters are based; filter has nominal cutoff of \( \pi/2 \) rad with a narrow transition to a stopband with adequate attenuation to isolate the two bands.

The filter \( h_1[n] = e^{j\pi n} h_0[n] \) is the complementary highpass filter.

The filters \( h_0[n] \) and \( h_1[n] \) are called Quadrature Mirror Filters (QMF).
Quadrature Mirror Filter (QMF) Banks

Substituting for the filter relationships we get:

\[ 2H_0(e^{j\omega})H_0(e^{j(\omega-\pi)}) - 2H_0(e^{j(\omega-\pi)})H_0(e^{j(\omega-2\pi)}) = 0 \]

i.e., perfect alias cancellation for any choice of lowpass filter

\[ Y(e^{j\omega}) = \left( H_0(e^{j\omega}) \right)^2 - \left( H_0(e^{j(\omega-\pi)}) \right)^2 \] \[ X(e^{j\omega}) = \tilde{H}(e^{j\omega})X(e^{j\omega}) \]

For perfect reconstruction we require that:

\[ \left[ \left( H_0(e^{j\omega}) \right)^2 - \left( H_0(e^{j(\omega-\pi)}) \right)^2 \right] = e^{-j\omega M} \]

i.e., flat magnitude with delay of \( M \) samples, due to the delay of the causal filters of the filter bank
Quadrature Mirror Filters

Assume FIR lowpass filter of length $L$ samples such that $h_0[n] = h_0[(L - 1) - n], 0 \leq n \leq L - 1$; this filter has linear phase with a delay of $(L - 1)/2$ samples and a frequency response of the form:

$$H_0(e^{j\omega}) = A_0(e^{j\omega})e^{-j\omega(L-1)/2} \quad (L - 1) \text{ must be odd}$$

where $A_0(e^{j\omega})$ is real. Using this lowpass filter in the filter bank, the condition for perfect reconstruction becomes:

$$\tilde{H}(e^{j\omega}) = \left[\left(A_0(e^{j\omega})\right)^2 - e^{j\pi(L-1)}\left(A_0(e^{j(\omega-\pi)})\right)^2\right]e^{-j\omega(L-1)}$$

If $L - 1$ is odd, we get:

$$\tilde{H}(e^{j\omega}) = \left[\left(A_0(e^{j\omega})\right)^2 + \left(A_0(e^{j(\omega-\pi)})\right)^2\right]e^{-j\omega(L-1)}$$

linear phase is guaranteed, but not perfect reconstruction
Quadrature Mirror Filters

For flat gain we need to satisfy the condition:

\[
\left[ (A_0(e^{j\omega}))^2 + (A_0(e^{j(\omega-\pi)})^2 \right] = 1
\]

Solution is to use high order linear phase QMF filters

Johnston proposed an algorithm for design of the basic lowpass filter that minimized the deviation from unity while providing a desired level of stopband attenuation
Quadrature Mirror Filters

\[ h_0[n] = w[n] = g_0[n] \quad h_1[n] = w[n]e^{j\pi n} = g_1[n] \]
Quadrature Mirror Filters

- practical realization of QMF filters
- note that aliasing is cancelled, but the overall frequency response is not perfectly flat
Quadrature Mirror Filters

(a) Impulse Response of Base Lowpass Filter

(b) Frequency Response of Lowpass and Highpass Filters

(c) Magnitude of Composite Frequency Response
The motivation for the subband approach is to process different speech properties independently in each band and thus being able to localize the band (compact bandwidth) is important.

Also, based on the original motivation of short-time analysis, temporal focus (within a limited time duration) is also important.

Filter bank design is thus result of a trade-off between perfect reconstruction and bandwidth concentration in a joint criterion:

\[
J = \alpha \int_{\omega_0}^{\pi} |W(e^{j\omega})|^2 d\omega + (1 - \alpha) \int_{0}^{\pi} (1 - \left| T(e^{j\omega}) \right|^2) d\omega
\]

- Stop-band leakage
- Deviation from PR

When processing (e.g., quantization, a non-linear operation) is involved, harmonic distortions are inevitable and perfect reconstruction cannot be achieved, as seen in the simple example.

There exists a design procedure (Smith-Barnwell) for QMF filters that attempts to meet the design constraints as best as possible
Smith-Barnwell QMF Design

1. Start with a sequence of length, say, $2L-1$ ($L$ even) that satisfies $g(n) + (-1)^n g(n) = 2C_0 \delta(n)$, that is, even samples = 0, except at $n=0$. $C_0$, the value of the sequence at $n=0$, needs to be large enough to satisfy the positivity condition (see below).

2. Factor $G(z) = \mathcal{F}\{g(n)\} = W_0(z)W_0(z^{-1})$
   such that all roots of $W(z)$ are inside the unit circle.

3. Positivity condition: $G(e^{j\omega}) = W_0(e^{j\omega})W_0(e^{-j\omega}) = \left|W_0(e^{j\omega})\right|^2 > 0$

4. The mirror channel: $W_1(e^{j\omega}) = -W_0(-e^{-j\omega})e^{-j\omega N}$, $N = L-1$
   $W_1(z) = -W_0(-z^{-1})z^{-N}$ or equivalently $w_1(n) = (-1)^n w_0(N-n)$

5. To eliminate aliasing: $F_0(z) = W_1(-z)$ and $F_1(z) = -W_0(-z)$
   $F_0(z) = W_1(-z)$ or equivalently $f_0(n) = (-1)^n w_1(n)$
   $F_1(z) = -W_0(-z)$ or equivalently $f_1(n) = -(1)^n w_0(n)$
Smith-Barnwell QMF Design - Example

$L = 8, N = L - 1 = 7$  

$$G(z) = \mathcal{Q}\{g(n)\} = W_0(z)W_0(z^{-1})$$

Length of $w_0(n) = L$, then $g(n)$ has length $2L - 1$  
choose $g(n) = \frac{\sin(\pi n/2)}{\pi n}$, $n = 0, \pm 1, \pm 2, \cdots, \pm 7$  
g(n) = [-0.0455 0.0000 0.0637 -0.0000 -0.1061 0.0000 0.3183 0.6 0.3183 0.0000 -0.1061 -0.0000 0.0637 0.0000 -0.0455]

at $n = 0$, 0.1 is added to satisfy the positivity condition

Factorize $G(z) = W_0(z)W_0(z^{-1})$  
$W_0(z)$: length $L$ with all roots inside the unit circle

$$w_0(n) = [1 \ 0.8091 \ 0.3426 \ -0.2332 \ -0.2032 \ 0.1590 \ 0.1182 \ -0.1461]$$

$$w_1(n) = (-1)^n w_0(N-n) = [-0.1461 \ -0.1182 \ 0.1590 \ 0.2032 \ -0.2332 \ -0.3426 \ 0.8091 \ -1]$$

$$f_0(n) = (-1)^n w_1(n) = [-0.1461 \ 0.1182 \ 0.1590 \ -0.2032 \ -0.2332 \ 0.3426 \ 0.8091 \ 1]$$

$$f_1(n) = -(-1)^n w_0(n) = [-1 \ 0.8091 \ -0.3426 \ -0.2332 \ 0.2032 \ 0.1590 \ -0.1182 \ -0.1461]$$

all arrays start with $n=0$
QMF Frequency Responses (Example)
Tree-Structured Filter Banks
Tree-Structured Filter Banks

- Stage 1: \( F_s / 2 \)
- Stage 2: \( F_s / 4 \)
- Stage 3: \( F_s / 8 \)
Sampling Pattern for Tree-Structured Filter Banks
Tree-Structured Filter Banks

Stage 3
\( (F_s / 8) \)

Stage 2
\( (F_s / 4) \)

Stage 1
\( (F_s / 2) \)
Tree-Structured QMF Filterbank
Quantization noise is concentrated in the band that it is generated in.

\[
P_{\hat{y}}(e^{j\omega}) = \sigma_{e_0}^2 |G_0(e^{j\omega})| + \sigma_{e_1}^2 |G_1(e^{j\omega})|
\]
Subband Coding

- advantages of subband coding
  - each subband can be encoded according to the perceptual criterion appropriate to that band
  - quantization noise can be confined to the band that produces it
  - low energy bands can be encoded so as to produce less noise
- applicable to coding in the 9.6-16 Kbps range
Practical Example of Subband Coder

Sub-band coder designs for 16, 24 or 32 kb/s

<table>
<thead>
<tr>
<th>Band</th>
<th>Decimation Factor From 8 kHz</th>
<th>Band Edges (Hz)</th>
<th>Sub-band Sampling Rates (Hz)</th>
<th>(bits/sample) for I (kb/s) =</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>8</td>
<td>0-500</td>
<td>1000</td>
<td>4</td>
</tr>
<tr>
<td>2</td>
<td>8</td>
<td>500-1000</td>
<td>1000</td>
<td>4</td>
</tr>
<tr>
<td>3</td>
<td>4</td>
<td>1000-2000</td>
<td>2000</td>
<td>2</td>
</tr>
<tr>
<td>4</td>
<td>4</td>
<td>2000-3000</td>
<td>2000</td>
<td>2</td>
</tr>
<tr>
<td>5</td>
<td>4</td>
<td>3000-4000</td>
<td>2000</td>
<td>0</td>
</tr>
</tbody>
</table>

- subband coder using tree-structured filter bank (Crochiere, 1981)
Subband Coder Subjective Quality

![Graph showing relative preference for different bit rates of ADPCM and single sub-band coders.](image-url)
Subband Coder Waveforms

\[ x(n) \]

\[ x_1(n) \]

\[ 2.5 \cdot x_2(n) \]

\[ 2.5 \cdot x_3(n) \]

\[ 2.5 \cdot x_4(n) \]

TIME (ms)
Concept of Time-Frequency Resolution

• When interested in the changing characteristics of the signal, short time Fourier transform or short time spectral analysis is used. 

\[ F(\omega, t) = \int_{-\infty}^{\infty} f(\tau) g(t - \tau) e^{-j\omega \tau} d\tau \]

\( g(t) \) is the window function providing focus in time.

Temporal spread : 
\[ \sigma_t^2 = \int_{-\infty}^{\infty} (t - t_c)^2 |g(t)|^2 dt \cdot \left[ \int_{-\infty}^{\infty} |g(t)|^2 dt \right]^{-1} \]

Frequency spread : 
\[ \sigma_\omega^2 = \int_{-\infty}^{\infty} (\omega - \omega_c)^2 |G(\omega)|^2 d\omega \cdot \left[ \int_{-\infty}^{\infty} |G(\omega)|^2 d\omega \right]^{-1} \]

If \( g(t) \) is Gaussian, it achieves the minimum time-bandwidth product \( \sigma_t \sigma_\omega = 0.5 \).

Continuous prolate spheroidal wave function: a bandlimited signal that have the highest energy concentration in a specified time interval.

• To maintain a similar level of uncertainty, the window function should be short to provide time resolution for high frequency components and long to provide frequency resolution for low frequency components.
Wavelet-Based Methods

What is in the Wavelet Transform block? What is the difference?
Wavelet Transform

- Built on set of expansion (or basis) functions that are derived from a mother wavelet through scaling and translation:

Let \( \psi(t) \) be a mother wavelet: 
\[
\psi_{a,\tau}(t) = \frac{1}{\sqrt{a}} \psi \left( \frac{t - \tau}{a} \right)
\]

Continuous Wavelet Transform:
\[
F_w(a, \tau) = \int_{-\infty}^{\infty} f(t) \psi_{a,\tau}^*(t) dt = \frac{1}{\sqrt{a}} \int_{-\infty}^{\infty} f(t) \psi^* \left( \frac{t - \tau}{a} \right) dt = \frac{1}{\sqrt{a}} f(\tau) \psi^* \left( -\frac{\tau}{a} \right)
\]

\[
F_w(a, \tau) = \frac{1}{\sqrt{a}} \int_{-\infty}^{\infty} f(at) \psi^* \left( t - \frac{\tau}{a} \right) dt
\]

stretch or compress the signal for analysis
\[
|F_w(a, \tau)|^2 = \text{scalogram}; \quad |F(\omega, \tau)|^2 = \text{spectrogram}
\]

Let \( \Psi(\omega) = F\{\psi(t)\} \) and \( \Psi_{a,\tau}(\omega) = F\{\psi_{a,\tau}(t)\} \)

Inverse transform:
\[
f(t) = \frac{1}{C_\psi} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} F_w(a, \tau) \psi_{a,\tau}(t) \frac{da \, d\tau}{a^2}
\]

Admissibility condition: 
\[
C_\psi = \int_0^{\infty} \frac{|\Psi_0(\omega)|^2}{\omega} d\omega < \infty \quad \Rightarrow \quad \Psi(0) = 0
\]
Wavelet and Fourier Transform

\[ \Psi(\omega) = F\{\psi(t)\} \quad \text{and} \quad \Psi_{a,\tau}(\omega) = F\{\psi_{a,\tau}(t)\} \quad F(\omega) = F\{f(t)\} \]

\[ \psi_{a,\tau}(t) = \frac{1}{\sqrt{a}} \psi\left(\frac{t-\tau}{a}\right) \quad \Leftrightarrow \quad \Psi_{a,\tau}(\omega) = \sqrt{a} \Psi(a\omega)e^{-j\tau\omega} \]

\[ F_w(a, \tau) = \frac{1}{\sqrt{a}} \int_{-\infty}^{\infty} f(t) \psi^*\left(\frac{t-\tau}{a}\right) dt = \frac{1}{\sqrt{a}} f(\tau) \psi^*\left(-\frac{\tau}{a}\right) \]

\[ W(a, \omega) = F(F_w(a, \tau)) = F\left\{ \frac{1}{\sqrt{a}} f(\tau) \psi^*\left(-\frac{\tau}{a}\right) \right\} = \sqrt{a} F(\omega) \Psi(a\omega) \]

\( W(a, \omega) \) is the frequency domain representation of the wavelet transform.
Wavelet and Its Spectrum

\[ \Psi(t) \]

\[ \sigma \]

\[ \Psi(t/a) \]

\[ a\sigma, 0 < a < 1 \]

\[ a\sigma, 1 < a \]

\[ \omega_0 / a \]

\[ \sigma_\omega / a \]
Scaling and Translation

- For discrete wavelet transform,
  \[ a = a_{0}^{-m} \text{ and } \tau = n \tau_{0} a_{0}^{-m} \quad m, n \text{ are integers} \]
  \[ \psi_{a,\tau}(t) = \psi_{m,n}(t) = a_{0}^{m/2} \psi(a_{0}^{m} t - n \tau_{0}) \quad m, n \in \mathbb{Z} \]

  If \( a_{0} = 2 \) and \( \tau_{0} = 1 \), dyadic or octave sampling
  \[ \psi_{m,n}(t) = 2^{m/2} \psi(2^{m} t - n) \quad m, n \in \mathbb{Z} \]

  If the set \( \{\psi_{m,n}(t)\} \) is complete, it is called affine wavelets.

  Orthogonality condition:
  \[ \int_{-\infty}^{\infty} \psi_{m,n}(t) \psi_{p,q}^{*}(t) dt = \delta(m - p)\delta(n - q) \]

  \[ F_{w}(a, \tau) = w_{m,n} = a_{0}^{m/2} \int_{-\infty}^{\infty} f(t) \psi(a_{0}^{m} t - n \tau_{0}) dt \]

  and
  \[ f(t) = \sum_{m} \sum_{n} w_{m,n} \psi_{m,n}(t) \]

  The orthogonality condition may not be easy to satisfy for a general set of “wavelets;” but it is still possible to invert discrete wavelet transform, sometimes.
STFT and CWT

Constant bandwidth

Constant-Q bandwidth
Discrete Wavelet Transforms
Other Simpler Transforms or Filter Banks

- Short-time analysis can also be viewed as data transformation, executed in a block-by-block manner, using for example:
  - Discrete Cosine Transform
  - Discrete Sine Transform
  - Filter bank based on DFT/FFT

- General remarks:
  - Discussion so far aims at a filter bank framework that allows perfect reconstruction (i.e., no loss of information in the original signal if desired);
  - But many speech processing systems do not require perfect reconstruction; rather, one may want to focus on “extraction” of key parameters or features from the speech signal;
  - Therefore, other analysis methods (e.g., transformations that do not have inverse) may still apply.
DFT and DCT

N-point DFT (Discrete Fourier Transform)

\[ X(k) = \sum_{n=0}^{N-1} x(n)e^{-j(2\pi/N)kn} \]

\[ \tilde{x}(n) = \frac{1}{N} \sum_{k=0}^{N-1} X(e^{j(2\pi/N)k})e^{j(2\pi/N)kn} = \sum_{r=-\infty}^{\infty} x(n + rN) \]

Possible excessive discontinuity at edges

Relates to real part of 2N-point DFT when the signal is symmetric

\[ X(k) = \sum_{n=0}^{2N-1} x(n)e^{-j(2\pi/N)kn} \quad \text{and} \quad x(n) = x(2N - n) \rightarrow X(k) = 2\sum_{n=0}^{N-1} x(n)\cos(2\pi kn / N) \]

But, be careful about the point of symmetry!!
Discrete Cosine Transforms

Extensively used in audio, image and video

**DCT-I**

\[
c_k = \frac{1}{2} \left( x(0) + (-1)^k x(N-1) \right) + \sum_{n=1}^{N-2} x(n) \cos \left( \frac{kn\pi}{N-1} \right)
\]

**DCT-II**

\[
c_k = \sum_{n=0}^{N-1} x(n) \cos \left[ \frac{k\pi}{N} \left( n + \frac{1}{2} \right) \right]
\]

Most common DCT; equivalent (up to a scale factor of 2) to a DFT of 4N real inputs of even symmetry with the even-indexed elements set to zero; i.e., an artificial shift of half sample in the original sampling setup.

**DCT-III** (IDCT)

\[
c_k = \frac{1}{2} x(0) + \sum_{n=1}^{N-1} x(n) \cos \left[ \frac{n\pi}{N} \left( k + \frac{1}{2} \right) \right]
\]

even around \( x_0 \) and odd around \( x_N \); \( c_k \) is even around \( k = -\frac{1}{2} \) and odd around \( k = N - \frac{1}{2} \).

Still more variants.
Discrete Sine Transform

\[ S = \left[ s_{ij} \right]_{i,j=0}^{N-1} \]

2-D case; eliminate one dimension for 1-D

\[ s_{ij} = \sqrt{\frac{2}{N+1}} \sin \left( \frac{(j+1)(i+1)\pi}{N+1} \right), \quad i, j = 0, 1, 2, \ldots, N - 1 \]

Relates to imaginary part of \( \sim 2N \)-point DFT

3-point sequence

8-point anti-symmetric sequence

Sequence corresponds To 8-point DFT
Discrete (Time) Fourier Transform - Revisited

• The DTFT (discrete time Fourier transform) of an $N$-point sequence is

$$X(e^{j\omega}) = \sum_{n=0}^{N-1} x(n)e^{-j\omega n}$$

Frequency $\omega$ is continuous.

• Sample the DTFT at

$$\omega_k = (2\pi/N)k, \quad k = 0, 1, \ldots, N - 1.$$  

• The result is the DFT (Discrete Fourier Transform)

$$X\left(e^{j\omega}\right)|_{\omega = 2\pi k / N} = \sum_{n=0}^{N-1} x(n)e^{-j(2\pi/N)kn} = X(k)$$

• If we compute the inverse DFT, we obtain

$$\tilde{x}(n) = \frac{1}{N} \sum_{k=0}^{N-1} X(e^{j(2\pi/N)k})e^{j(2\pi/N)kn} = \sum_{r=-\infty}^{\infty} x(n + rN)$$
DFT for Power Spectral Density Estimation

\[ X(e^{j\omega}) = \sum_{n=0}^{N-1} x(n)e^{-j\omega n} \]

\[ S(\omega) = |X(e^{j\omega})|^2 \]

: k\(^{th}\) channel processing (weighting) function

\[ S_k(\omega) = S(\omega)P_k(\omega), \quad \forall k \]

Examples of non-uniform filter banks

Perfect reconstruction may not be a requirement.