Analyze signal window by window

\[ X[n] \]

- Window length: \( M \)
- Hop size: \( R \)
- Zero padding

\[ \text{FFT} \]
- FFT length: \( N \)

\[ \text{Freq. (Hz)} \]
- Negative freq.: \( f_s/2 \)
- Positive freq.: \( 0 \)

First frame is centered around \( \frac{M}{2f_s} \)

Second frame centered around \( \frac{M}{2f_s} + \frac{R}{f_s} \)

\[ \text{Spectrogram: } \frac{N+1}{2} \times \text{frames} \]

Usually only show positive freq. as negative freq. spectrum is conjugate symmetric.

Processing in frequency domain

\[ \text{IFFT} \]

Overlap Add (OLA)

Reconstruction (resynthesis)
STFT Math

Let $X_m(\Omega)$ be the DTFT spectrum of the $m$-th frame windowed signal.

$$X_m(\Omega) = \sum_{n=-\infty}^{\infty} x[n] W[n-mR] e^{-j2\pi n}$$

Window centered around $mR$, i.e., shifted version of $W[n]

Suppose we don't modify $X_m(\Omega)$, but resynthesize $x[n]$ from it using DLA

$$y[n] = \sum_{m=-\infty}^{\infty} \frac{1}{2\pi} \int_{-\pi}^{\pi} X_m(\Omega) e^{j2\pi n} d\Omega$$

$$= \sum_{m=-\infty}^{\infty} \frac{1}{2\pi} \int_{-\pi}^{\pi} \sum_{n'=-\infty}^{\infty} x[n'] W[n'-mR] e^{-j2\pi n'} e^{j2\pi n} d\Omega$$

$$= \frac{1}{2\pi} \int_{-\pi}^{\pi} \sum_{n'=\infty}^{\infty} x[n'] \sum_{m=-\infty}^{\infty} W[n'-mR] e^{-j2\pi n'} e^{j2\pi n} d\Omega$$

If $\sum_{m=-\infty}^{\infty} W[n'-mR] = 1$, then

$$y[n] = \frac{1}{2\pi} \int_{-\pi}^{\pi} \sum_{n'=\infty}^{\infty} x[n'] e^{-j2\pi n'} e^{j2\pi n} d\Omega$$

$$= \text{IDTFT}[\text{DTFT}(x[n'])] = x[n]$$

We get perfect reconstruction!

Window and hop size that satisfy $\sum_{m=-\infty}^{\infty} W[n-mR] = 1$ is called

Constant Overlap Add (COLA) window.
Rect Window: \( \text{COLA}(R \leq M, R \in \mathbb{Z}) \)

\[ R= \frac{M}{2}, \frac{M}{3}, \ldots, \frac{M}{k}, R \in \mathbb{Z} \]

\n
Triangular Window:

\[ R= \frac{M}{2}, \frac{M}{3}, \ldots, \frac{M}{k}, R \in \mathbb{Z} \]

Hamming Window:

\[ R= \ldots, R \in \mathbb{Z} \]

1. If a window \( w \) satisfies \( \text{COLA}(R) \), it also satisfies \( \text{COLA}(\frac{R}{k}) \) for \( k \in \mathbb{N} \) and \( \frac{R}{k} \in \mathbb{Z} \).

2. Any window satisfies \( \text{COLA}(1) \).

Note: for Hamming and Hann Window, it should be modified a little bit to satisfy \( \text{COLA} \). Use `hamming(M, 'periodic')`, `hann(M, 'periodic')` in Matlab.

**Weighted Overlap Add (WOLA)**

If we have modified the spectrogram before transforming it back to time domain, we will not get perfect reconstruction and there is often "blocking effects" in resynthesized signal, i.e., audible discontinuities at frame boundaries.

Use "synthesis window" to smooth out these discontinuities.

\[
\text{Synthesis window } v[n] \quad \text{Length: } N = \text{FFT size.}
\]

Length could also be the same as analysis window length \( M \).

Similarly, we want \( \sum_{m} w[n-MR] \cdot v[n-mR] = 1 \) for perfect reconstruction.
If we let \( W[n] \) and \( V[n] \) be the same Hamming window

\[
W[n] = V[n] = \cos^2 \left( \frac{\pi}{M} n \right), \quad \text{i.e., } W[n] \cdot V[n] = \cos^4 \left( \frac{\pi}{M} n \right)
\]

And let \( R = \frac{M}{4} \) \((\text{i.e., } 25\%) \) over hop size)

\[
\cos^4 \left( \frac{\pi}{M} n \right) + \cos^4 \left( \frac{\pi}{M} (n - \frac{M}{4}) \right) + \cos^4 \left( \frac{\pi}{M} (n - \frac{M}{2}) \right) + \cos^4 \left( \frac{\pi}{M} (n - \frac{3M}{4}) \right)
\]

sum of 4 windows. \( = \frac{3}{2} \)

If we apply Hamming window with 25% hop size at both analysis and synthesis stages, we get COLA with \( R \) a gain of \( \frac{3}{2} \).

**Speed Change & Pitch Shift.**

1) Suppose \( X[n] \) has sampling rate of \( f_s \), sampled from a sine wave with period \( T \). Each period is represented by \( f_s \cdot T \) samples.

If we resample \( X[n] \) with sampling rate \( g_s \), i.e., we interpolate from \( X[n] \) (with appropriate LP filtering) to prevent aliasing) into \( Y[n] \), where \( y[n] \) represents the same sine wave, but uses \( g_s \cdot T \) samples for one period.

If we play back \( Y[n] \) with sampling rate \( f_s \), then it takes \( \frac{g_s \cdot T}{f_s} \) second to play one cycle of the sine wave (which should be \( T \) seconds).

Therefore, the speed is changed by a factor \( g_s \cdot \frac{f_s}{g_s} \cdot \frac{f_s}{g_s} \)

And the pitch is changed by a factor \( \frac{f_s}{g_s} \) as well!
2) Suppose \( X[n] \) has sampling rate \( f_s \), but is now played back with sampling rate of \( 9f_s \). If the original period was \( T \), then the new period is \( \frac{f_s T}{9f_s} \).

Therefore, speed is changed by a factor of \( \frac{9f_s}{f_s} \).

And pitch \( \frac{1}{\alpha} \) as well!

**Question:** How to achieve independent speed and pitch change?

**Phase Vocoder** (pitch change factor \( \alpha \), speed \( \beta \) )

1. **original signal** \( X[n] \)
2. resample \( x[n] \) with \( \frac{f_s}{\alpha} \) to achieve pitch change factor of \( \alpha \) and speed change factor of \( \frac{\beta}{\alpha} \).
3. interpolate spectrogram to achieve speed factor of \( \beta \).

 dục frame size: \( M \)

hop size: \( R \)

keep fixed.

resampling

\[
\text{frames: } L / \alpha \text{ frames, } L / \beta \text{ frames}
\]
Spectrogram Interpolation

1) Magnitude spectrogram interpolation

① Figure out corresponding time $t$ in original spectrogram.

② Find the left and right frames.
   $t_1, t_2$

③ Linear interpolation
   $\lambda = \frac{t - t_1}{t_2 - t_1}$
   
   $|Y[k]| = (1 - \lambda) |X_1[k]| + \lambda |X_2[k]|$

2) Phase reconstruction.

① Let the phase of first interpolated spectrum the same as the first spectrum in the original.

② Let phase advance from $\angle Y[k]$ to its next frame be the same as the phase advance from $\angle X_1[k]$ to $\angle X_2[k]$

This will make sure the phase change coherent.