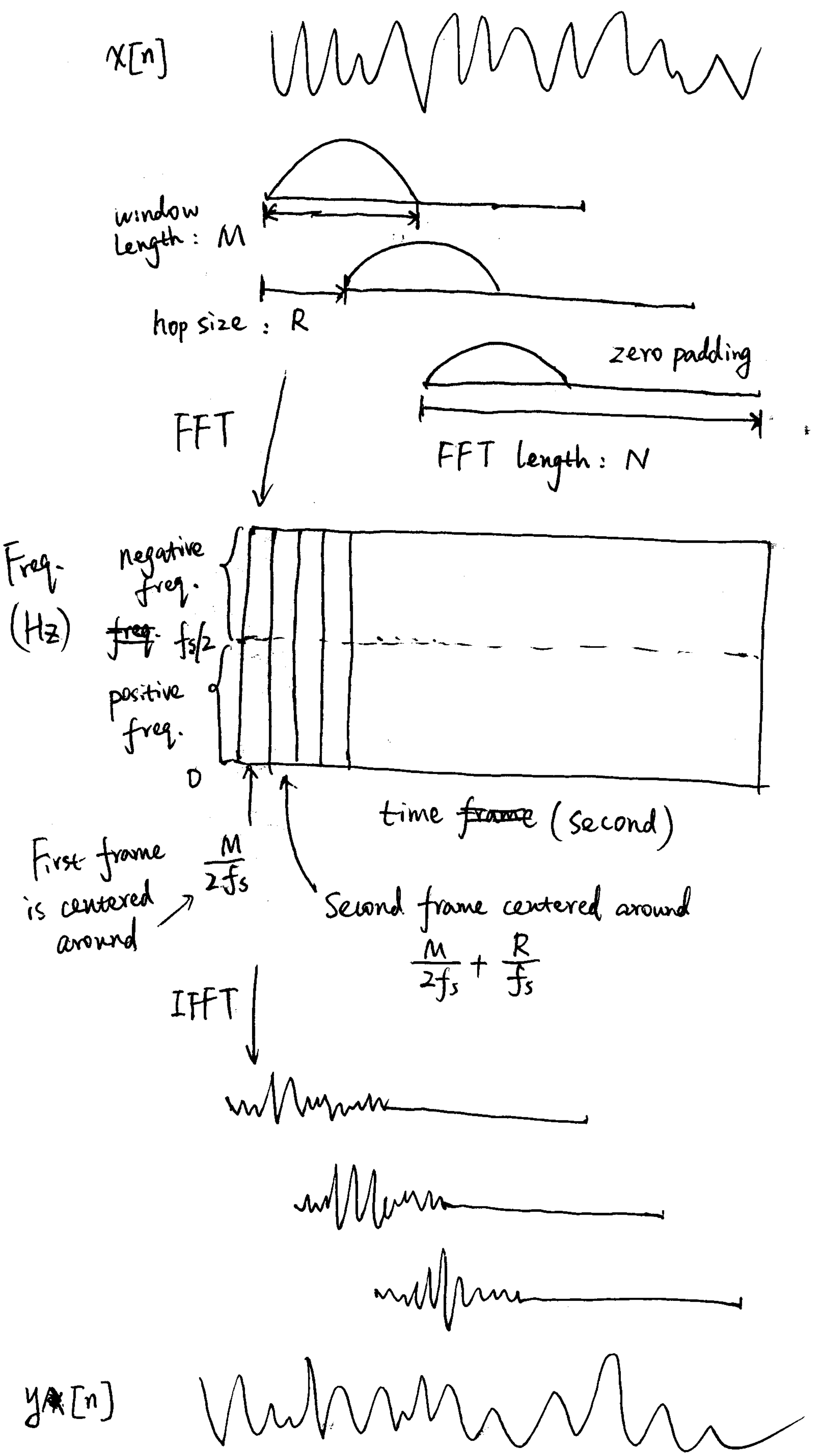
ECE 272/472 Lecture 4: Short Time Fourier Transform

Analyze signal window by window



Spectrogram: $\frac{N+1}{2} \times \#\text{frame}$

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

Processing in frequency domain

Overlap Add (OLA)

reconstruction (resynthesis)

STFT Math

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

$$X_m(\Omega) = \sum_{n=-\infty}^{\infty} x[n]\, \underline{w[n-mR]}\, e^{-j\Omega n}$$

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 OLA

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

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

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

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

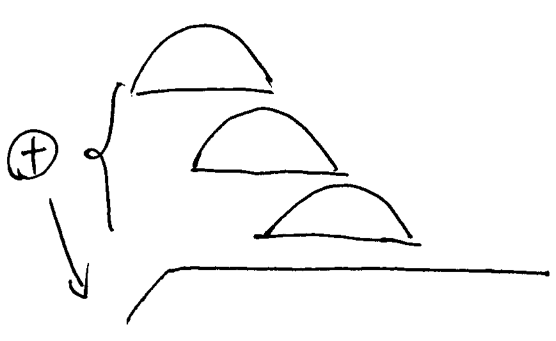
$$y[n] = \frac{1}{2\pi} \int_{-\pi}^{\pi} \underline{\sum_{n'=-\infty}^{\infty} x[n']\, e^{-j\Omega n'}}\, e^{j\Omega n}\, d\Omega$$

$$= \mathrm{IDTFT}\left[\mathrm{DTFT}(x[n'])\right] = x[n]$$

We get perfect reconstruction!

Window and hop size that satisfy $\sum_{m=-\infty} \underset{\text{window}}{w}[n-m\underset{\text{Hop size}}{R}] = 1$ is called

Constant Overlap Add (COLA) window.



Rect window : ~~COLA.t~~ $R \le M$ $\quad R \in \mathbb{Z}$

Triangular window: $\quad R = \frac{M}{2}, \frac{M}{3}, \cdots, \frac{M}{k}$ $\quad R \in \mathbb{Z}$

Hamming window: $\quad R = \frac{M}{2}, \frac{M}{4}, \frac{M}{6}, \frac{M}{8}, \cdots \frac{M}{2k}$, $R \in \mathbb{Z}$

Hanning window $\quad R =$ - - - - - - - - - - - - - - , $R \in \mathbb{Z}$

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

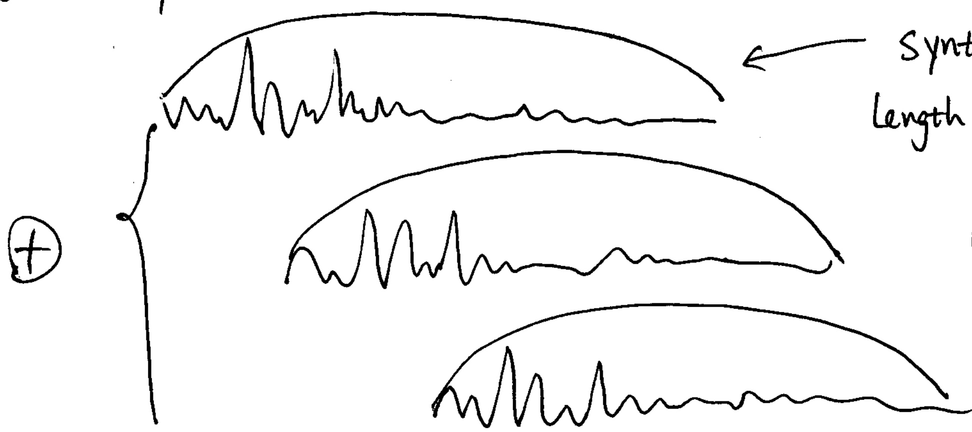
② Any window satisfies COLA(1).

Note: for Hamming and Hann window, it should be modified a little bit to satisfy 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 <u>and</u> there is often "blocking effects" in resynthesized signal, i.e., audible discontinuities at frame boundaries.

Use "synthesis window" to smooth out these discontinuities.



Synthesis window $v[n]$

Length : N = FFT size.

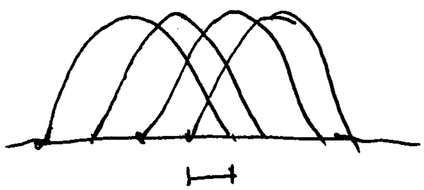
Length could also be the same as analysis window length M.

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 Hann window

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

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



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

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

$\therefore$ If we apply hann window with 25% hop size at both analysis and synthesis stages, we get COLA with 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, $\therefore$ 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 $\frac{g_s \cdot T}{f_s}$ secon

to play one cycle of the sine wave (which should ~~be~~ finish in T seconds).

Therefore, the speed is changed by a factor $\cancel{\frac{g_s}{f_s}}$. $\frac{f_s}{g_s}$

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 g_s. If the original period was T, then the new period is $\frac{f_s \cdot T}{g_s}$

Therefore, speed is changed by a factor of $\frac{g_s}{f_s}$

And pitch - - - - - - - - - - - - - - - - as well!

Question: How to achieve independent speed and pitch change?

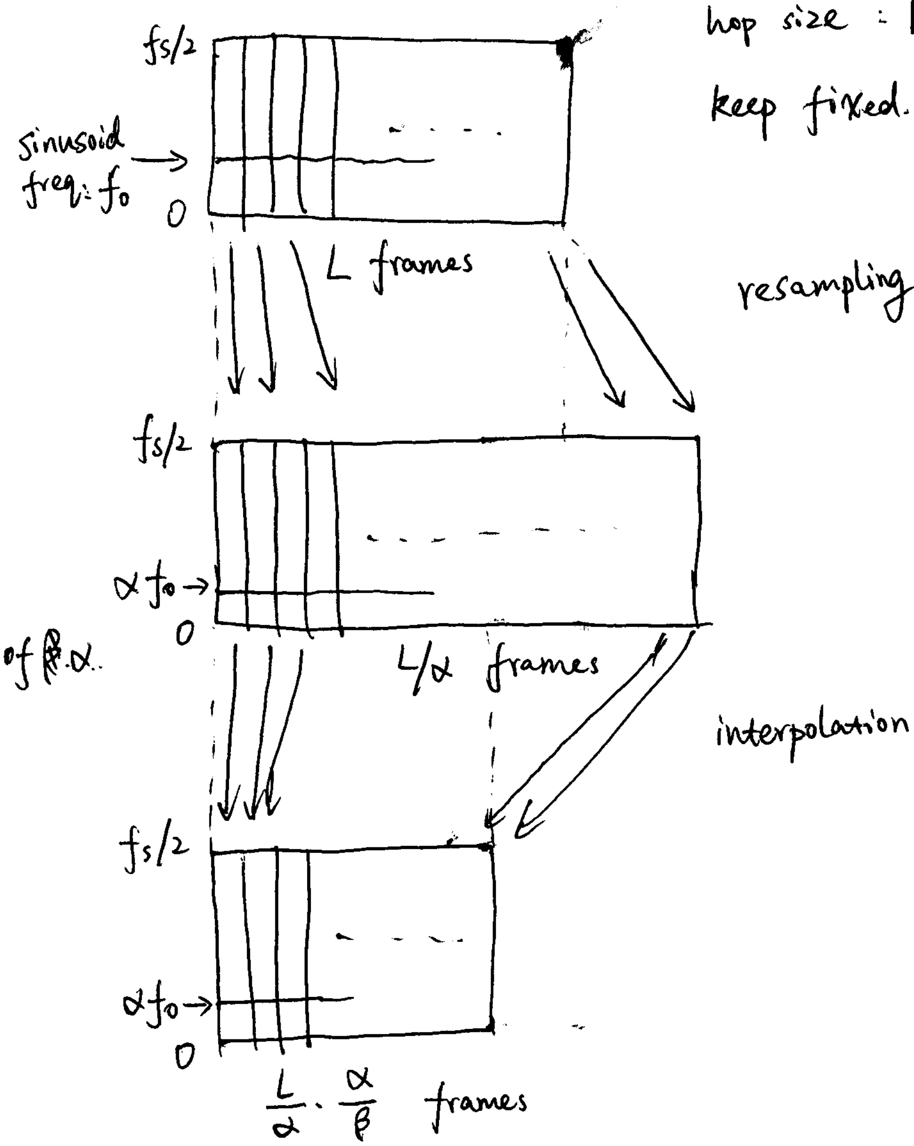
Phase Vocoder (pitch change factor: α / speed - - - - - - β). spectrogram

frame size: M
hop size: R
keep fixed.

① original signal $x[n]$

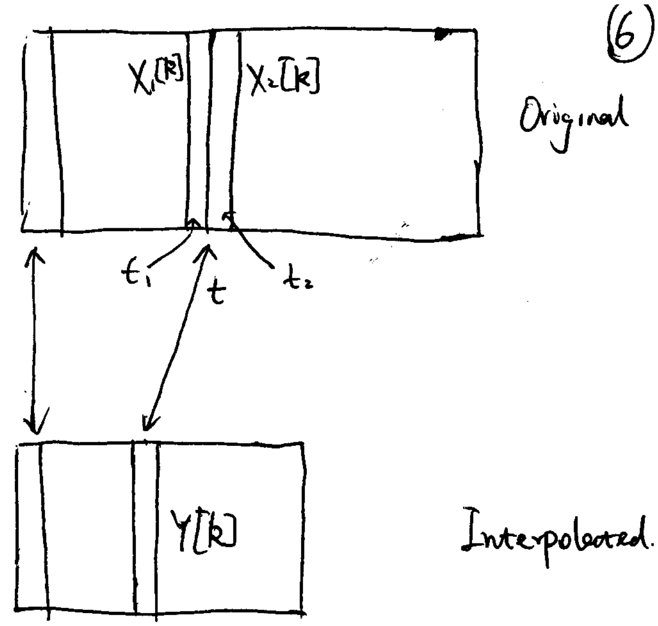
② resample $x[n]$ with $\frac{f_s}{\alpha}$ to achieve pitch change factor of α.
and speed change factor of $\beta \cdot \alpha$.

③ interpolate spectorgram to achieve speed factor of β.



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 <u>phase advance</u> 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.