

Sound wave synthesis

Li Su

Introduction

- In the previous lectures, we focused on sound analysis.
- Now let's turn to sound **synthesis** tasks related to the creation and modification of sounds.
- How to create/modify sounds?
- Related tasks:
 - Synthesis 合成
 - Morphing 變快變慢變高變低
 - Separation source separation 有些你錯不錯

Sound synthesis methods

- Filtering (linear or nonlinear):
 - Equalizers, distortion
- Physical modeling:
 - Dynamical systems, digital waveguides, source-filter models
- Generative models:
 - Gaussian mixture models (GMM), neural networks
- Signal decomposition and reconstruction:
 - Sinusoidal plus noise modeling
 - Short-time Fourier transform
 - Gabor frames
 - Wavelet transform
 - Non-negative matrix factorization (NMF) and sparse coding

Analysis and synthesis

- Synthesis can be regarded as the inverse problem of analysis
 - Determine the characteristics of the desired signal first, then synthesize the signal according to the characteristics
- Similar ideas:
 - Encoder for analysis, decoder for synthesis
 - Discriminator for analysis, generator for synthesis
- Reconstruction
 - If the analysis-synthetic framework can recover the original signal without loss
- Generation
 - If the analysis-synthesis can generate data not seen (but similar) in the training set

Sinusoidal modeling

我們感興趣的大多是週期性訊號

- Most music signals of interests are periodic
- Sinusoidal plus noise model

$$s(t) = \sum_{k=1}^{K \text{ harmonics}} A_k(t) \cos(\omega_k(t) + \phi_k) + n(t)$$

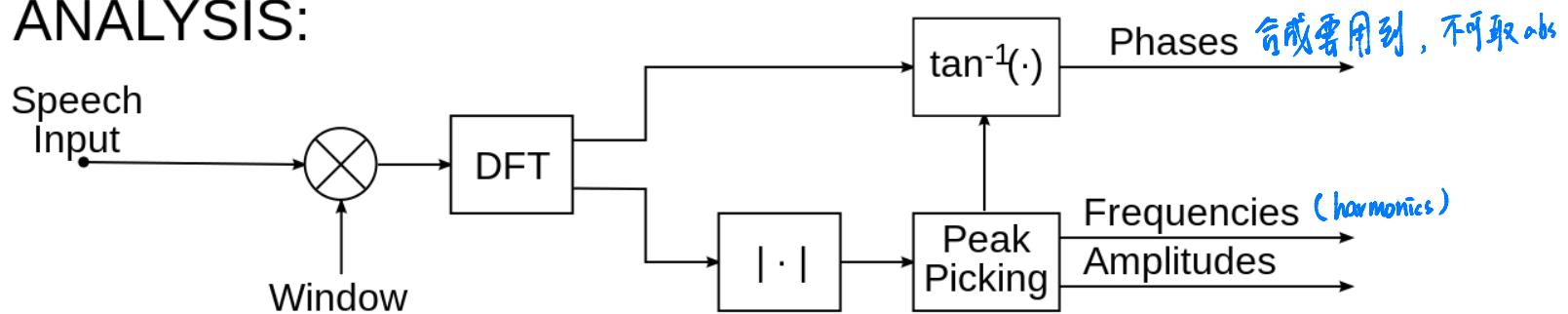
- Both $A_k(t)$ and $\omega_k(t)$ are slow-varying

對 piano, guitar 還不錯
音樂人蔣不OK

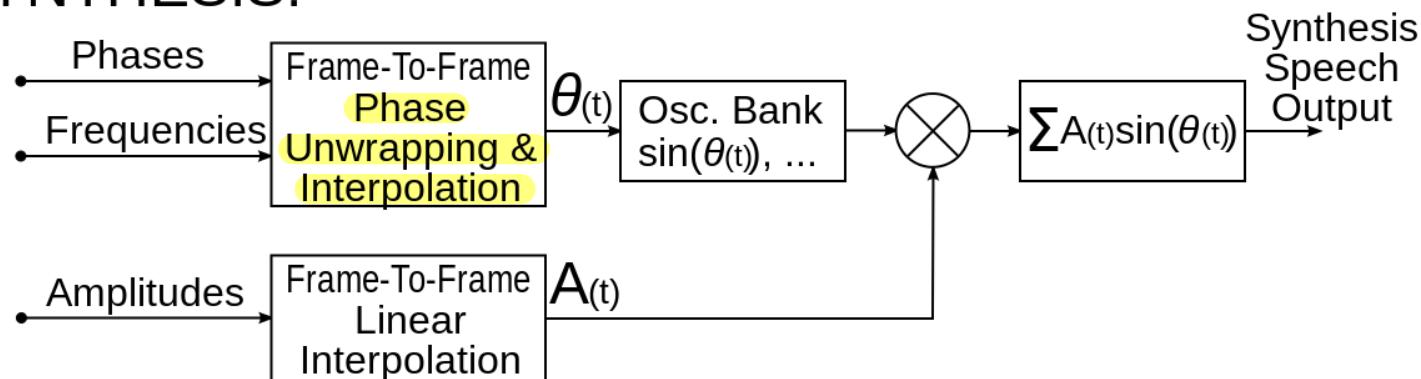
Sinusoidal modeling

- **Analysis-synthesis framework**

ANALYSIS:



SYNTHESIS:



From: https://en.wikipedia.org/wiki/Audio_time_stretching_and_pitch_scaling

Signal reconstruction from STFT

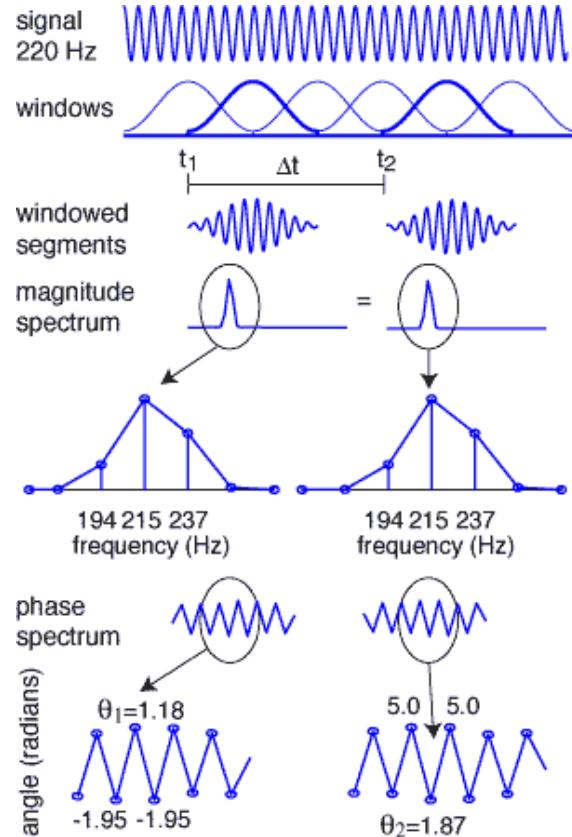
- DFT in the analysis-synthesis framework: fast Fourier transform (FFT) for analysis, then inverse fast Fourier transform (IFFT) for synthesis
- STFT in the analysis-synthesis framework: what is the inverse short-time Fourier transform (ISTFT)?
- Not only IDFT, but also needs to convert the 2-D image to a 1-D signal
- **Overlap-and-add (OLA) method:** add back the frames after IFFT according to the timestamp
- **Perfect reconstruction:** 75% for Hamming window, for example

An example

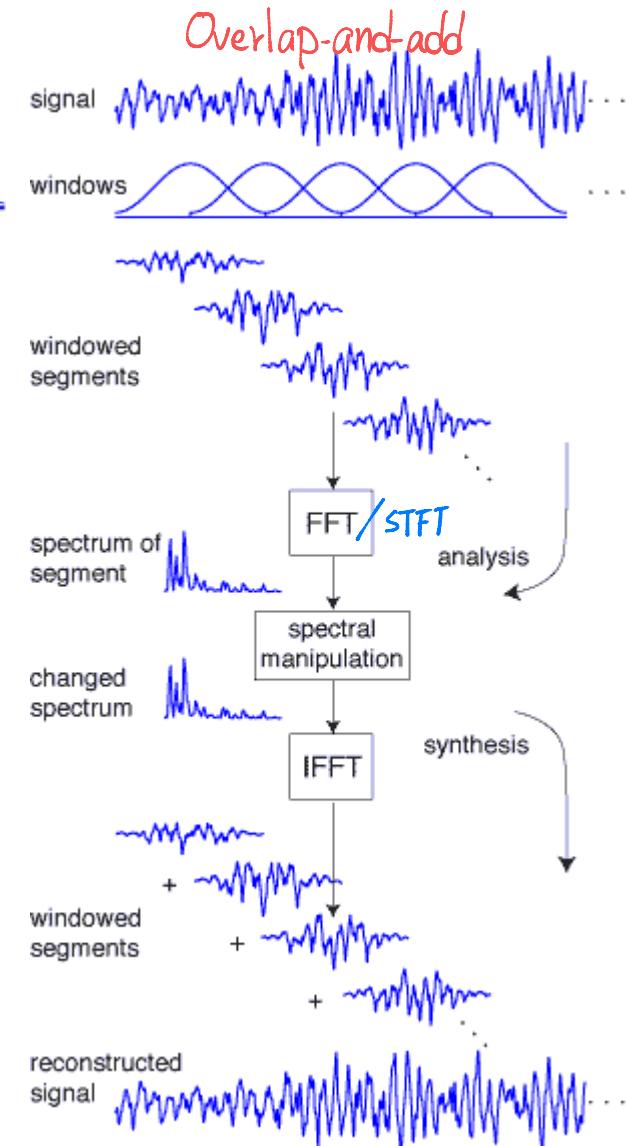
- Manipulation of
 - Amplitude spectrum
 - Phase spectrum
- Phase vocoder

time-stretching & pitch-shifting

可以自己錄音
(試把 phase 擄掉)



(Has been proved to be able to perfectly reconstruct signal)
overlap Δt 重疊 Δt → perfect reconstruction



From: William A. Sethares, "A Phase Vocoder in Matlab"
<http://sethares.engr.wisc.edu/vocoders/phasevocoder.html>

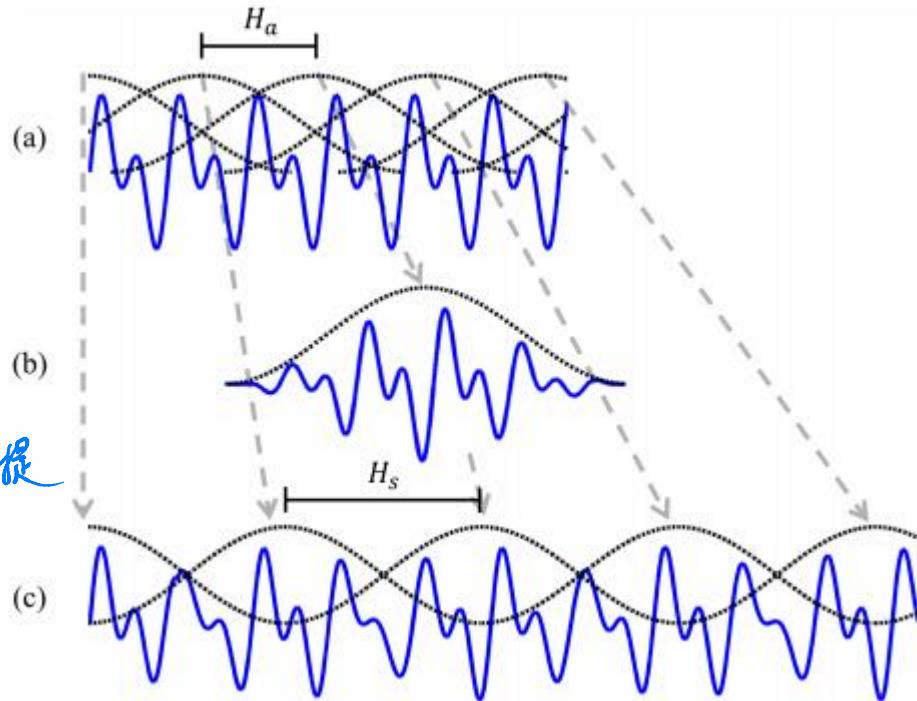
Time-scale modification (TSM)

訊號拉長或縮短

Downsample 無法音高、音高會變高

- How to stretch a sound?
- OLA method for TSM
 - Overlapped windows
 - Analysis hop size H_a
 - Synthesis hop size H_s

用不同 hop length 加回去，要注意 perfect reconstruction 的前提



Driedger, Jonathan, and Meinard Müller. "TSM Toolbox: MATLAB Implementations of Time-Scale Modification Algorithms." *DAFx*. 2014.

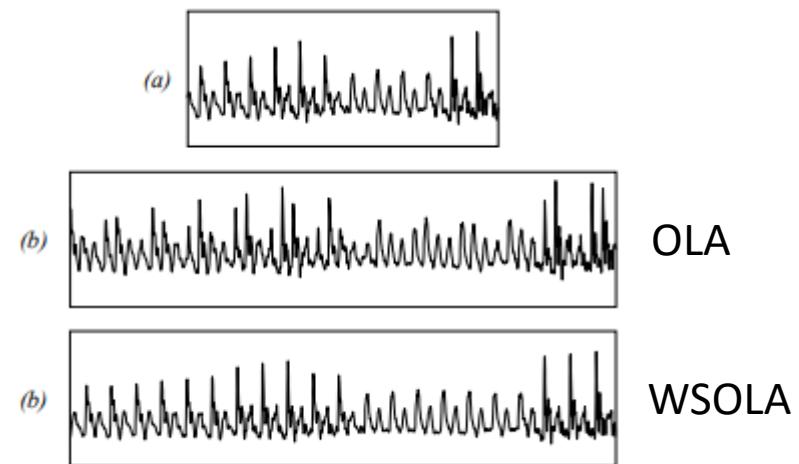
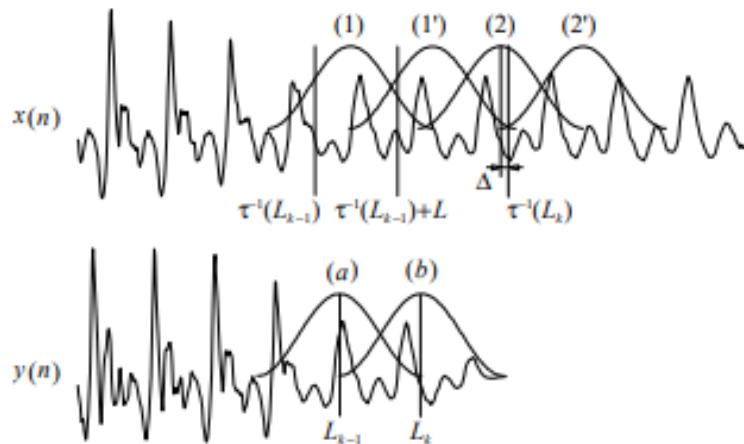
Time-scale modification (TSM)

- The problems of OLA:
 - Phase discontinuities
 - Harmonic structures not preserved
- Solutions: maintain the waveform as much as possible
 - Waveform similarity overlap-Add (WSOLA)
 - Phase vocoder
 - And others

Waveform similarity overlap-Add (WSOLA)

希望相鄰的window框到的waveform能很像

- Analysis frame position tolerance Δ_{max}
 - The position of each analysis frame in the input signal may be shifted on the time-axis by $\Delta \in [-\Delta_{max}, \Delta_{max}]$ such that the waveforms of two overlapping synthesis frames are as similar as possible in the overlapping regions



Verhelst, Werner, and Marc Roelands. "An overlap-add technique based on waveform similarity (WSOLA) for high quality time-scale modification of speech." in ICASSP 1993

Phase vocoder

對 spectrogram 內插再對插完的 spectrogram 作 IFFT
(phase 也要內插)

- WSOLA approaches the problem of phase discontinuities in the time domain
- Phase vocoders approach this in the frequency domain
- Adapting the phases of these sinusoids such that no phase discontinuities introduced
 - Preserve the rate of phase change between slices =
instantaneous frequency $f_i = \frac{1}{2\pi} \frac{d}{dt} \Phi(f, t)$
 - Horizontal phase coherence
 - Vertical phase coherence
- Good for STFT operation

Magnitude-only reconstruction

不要 phase 的 reconstruction , 跟據複變函數的性質, 若 函數可微, 實部和虛部為封偱關係
此方法有此假設

- Just manipulate the spectrogram and resynthesize the signal by it: is it possible?
- Griffin & Lim algorithm: let $|Y(f, t)|$ be the **modified** magnitude STFT of $x(t)$, find $STFT^{-1}\{|Y(f, t)|\}$
- By iteration

$$y^{(n)}(t) = STFT^{-1}\{|Y(f, t)|\Phi^{(n)}(f, t)\}$$
$$\Phi^{(n+1)}(f, t) = \angle \left\{ STFT\{y^{(n)}(t)\} \right\}$$

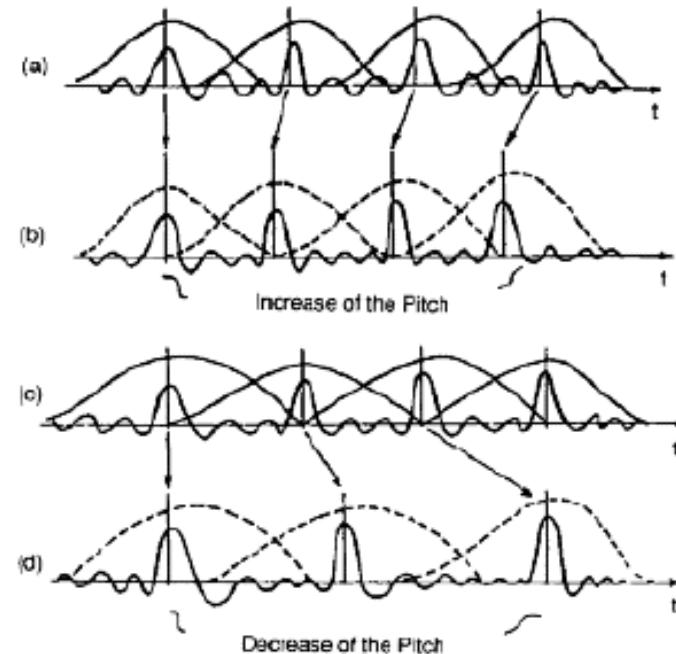
- Converge to a solution, but slowly
- Phase retrieval from magnitude only: a developing field

Pitch scaling

常用: resample + low-pass filter

- One simple method: **time stretching + resampling**
- Time-domain pitch-synchronous overlap-add (TD-PSOLA) for pitch scaling
 - Pitch marking
 - Waveform extraction

若已知音高，可将window size和hop length设置为语音周期



Chalamandaris, Aimilios, et al. "An efficient and robust pitch marking algorithm on the speech waveform for TD-PSOLA." ICSIPA 2009

Tools

- TSM toolbox (in MATLAB)
 - <https://www.audiolabs-erlangen.de/resources/MIR/TSMtoolbox/>
- Phase vocoder (*also available in Python*)
 - <http://www.ee.columbia.edu/ln/rosa/matlab/pvoc/>