Music synchronization

Li Su

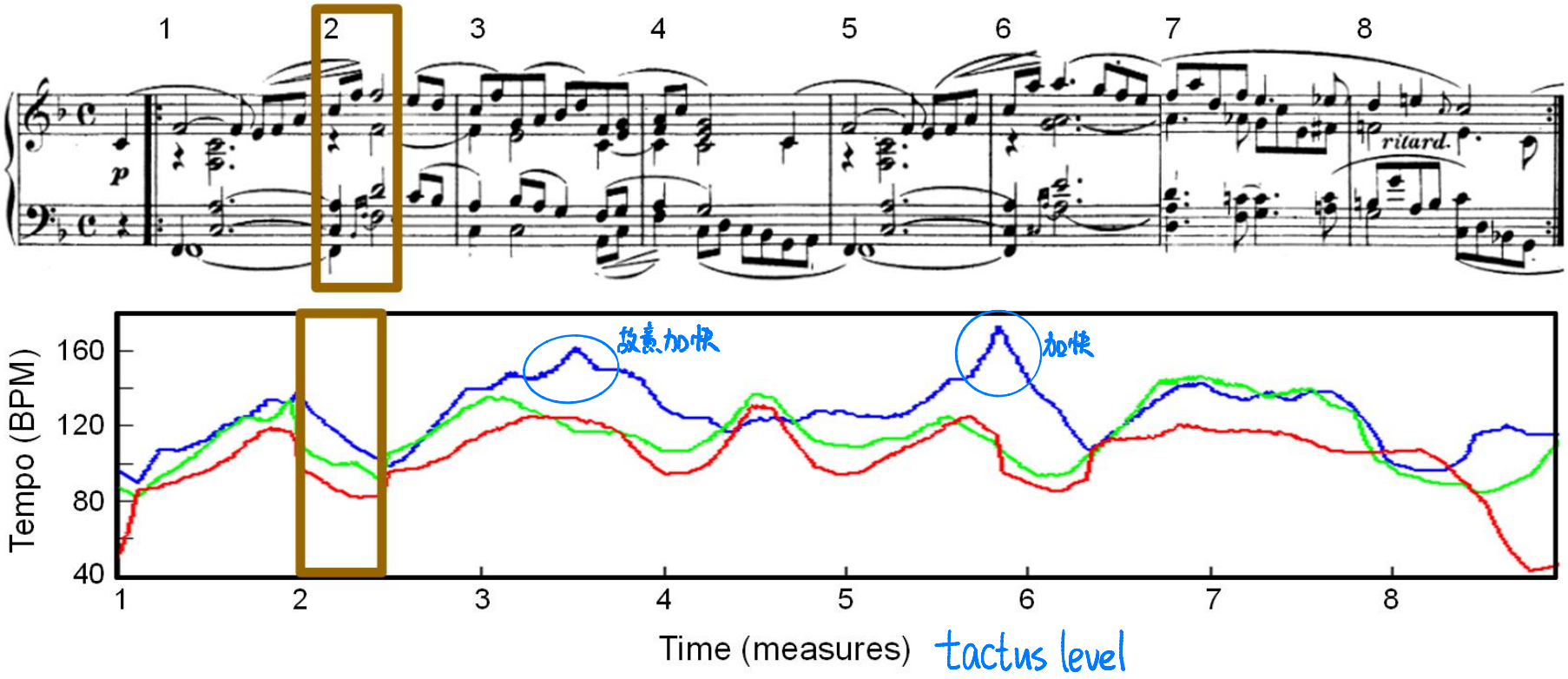
2019/04/30

又稱 alignment
（對齊不同版本的演奏）

Introduction

- Audio synchronization / audio alignment
  - Audio-to-audio alignment (one version of performance to another version, melody to accompaniment, …)
  - Audio-to-score alignment (樂譜) 可看有沒有彈錯，自動翻譜，自動伴奏 (real-time) (已得伴奏譜，想伴奏跟我)
  - Audio-to-lyric alignment 動態歌詞 (KTV)
  - Audio-to-video alignment 影音不同步問題，自動 mashup (拼多人拍同一演唱會，想自動剪接成一部)
- Application
  - Dynamic lyrics
  - Automatic page turner
  - Automatic accompaniment
  - Performance analysis

Application: performance analysis



From: M. Mueller, *Fundamentals of Music Processing*, Chapter 3, Springer 2015

Application: automatic accompaniment

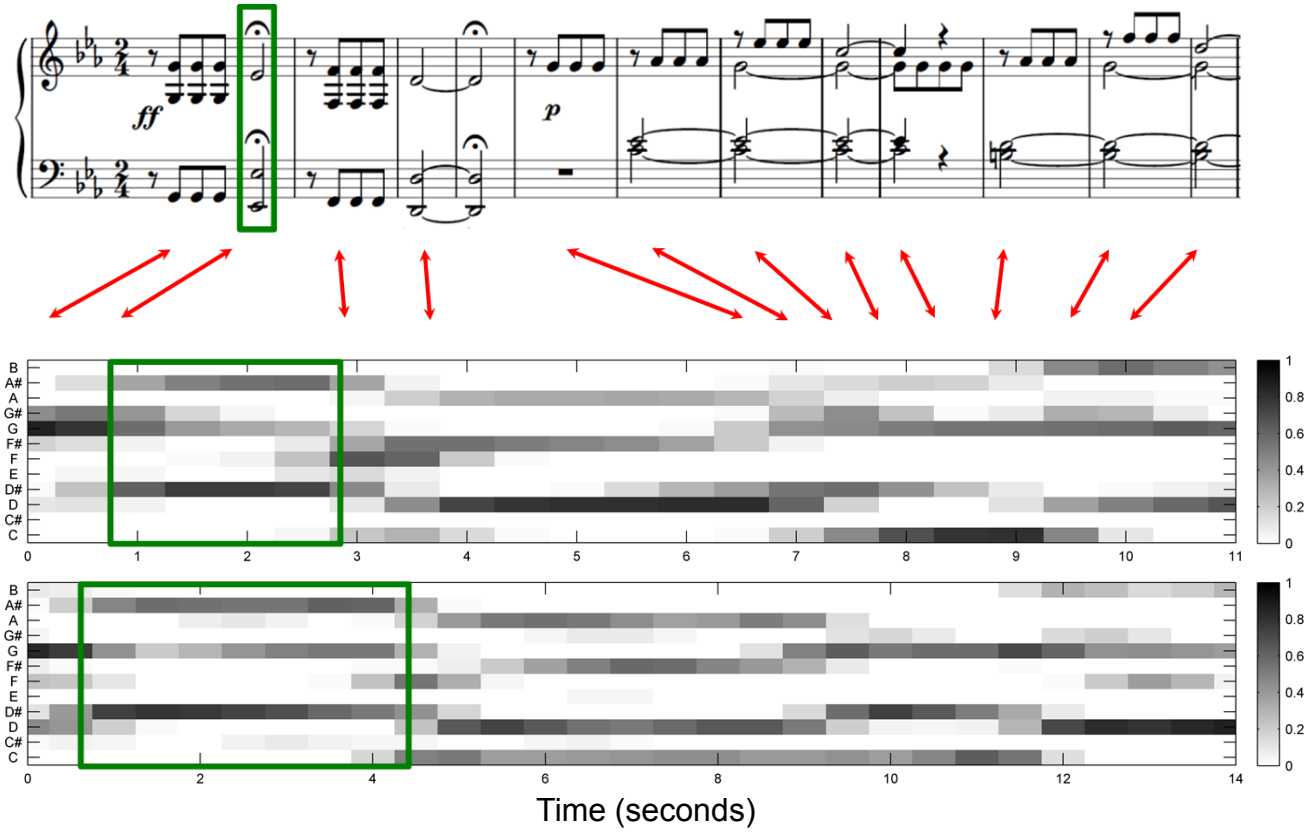
- https://www.youtube.com/watch?v=RnjoxwY3RfA

自動伴奏例子1984，建議去看

Audio-to-score alignment

DTW algorithm?

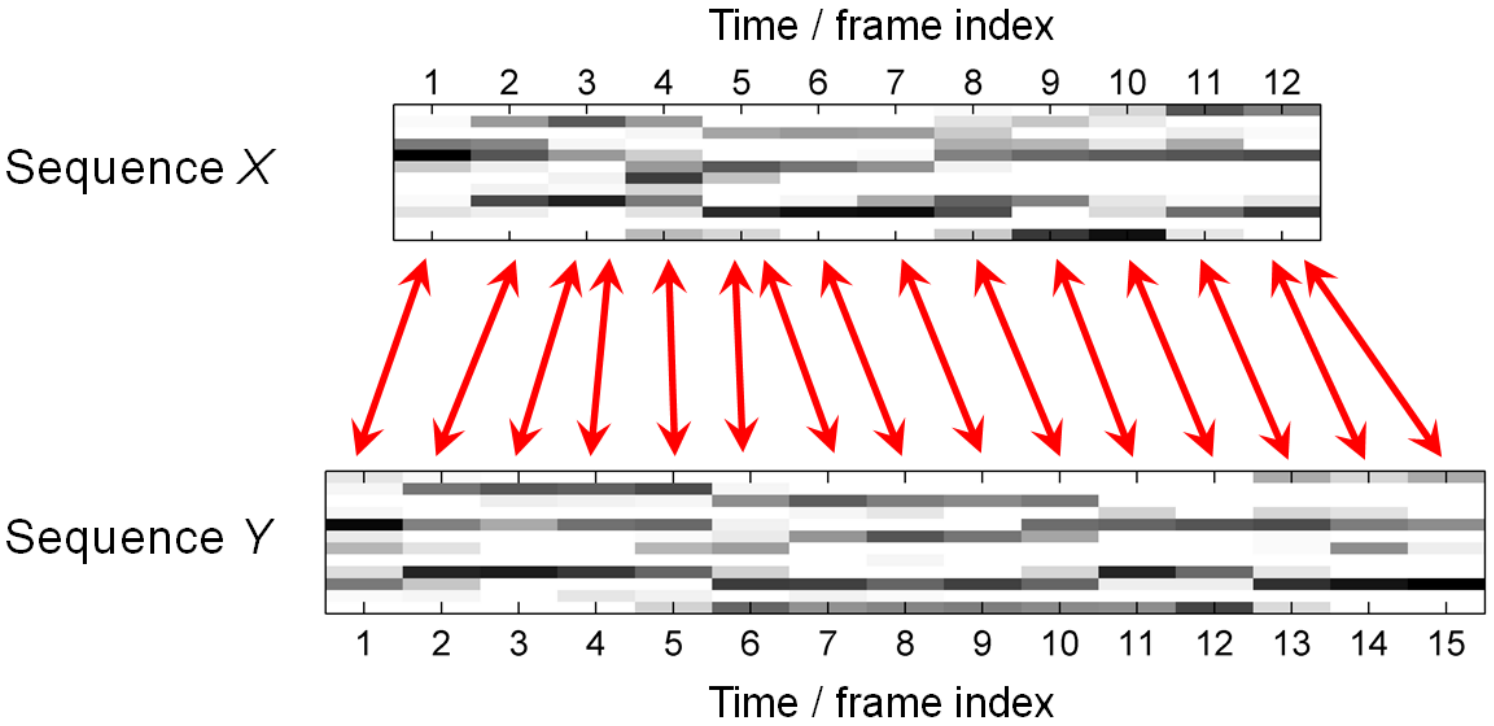
- An example: Beethoven's symphony No. 5



From: M. Mueller, *Fundamentals of Music Processing*, Chapter 3, Springer 2015

Audio-to-audio alignment

- 2 different versions of Beethoven's symphony No. 5



From: M. Mueller, *Fundamentals of Music Processing*, Chapter 3, Springer 2015

The goal of music alignment

- Goal: given the score of a music piece and an audio recording of it, align them in time
- Offline alignment: the whole audio recording is given 已解好不錯 ↳ performance analysis
  - application: performance analysis, feature extraction
- Online alignment: the alignment is done on-the-fly progressively while the recording is being played
  - a.k.a. score following
  - constraint: cannot "look into the future" (i.e. has to be causal)
  - application: automatic accompaniment, page tuner

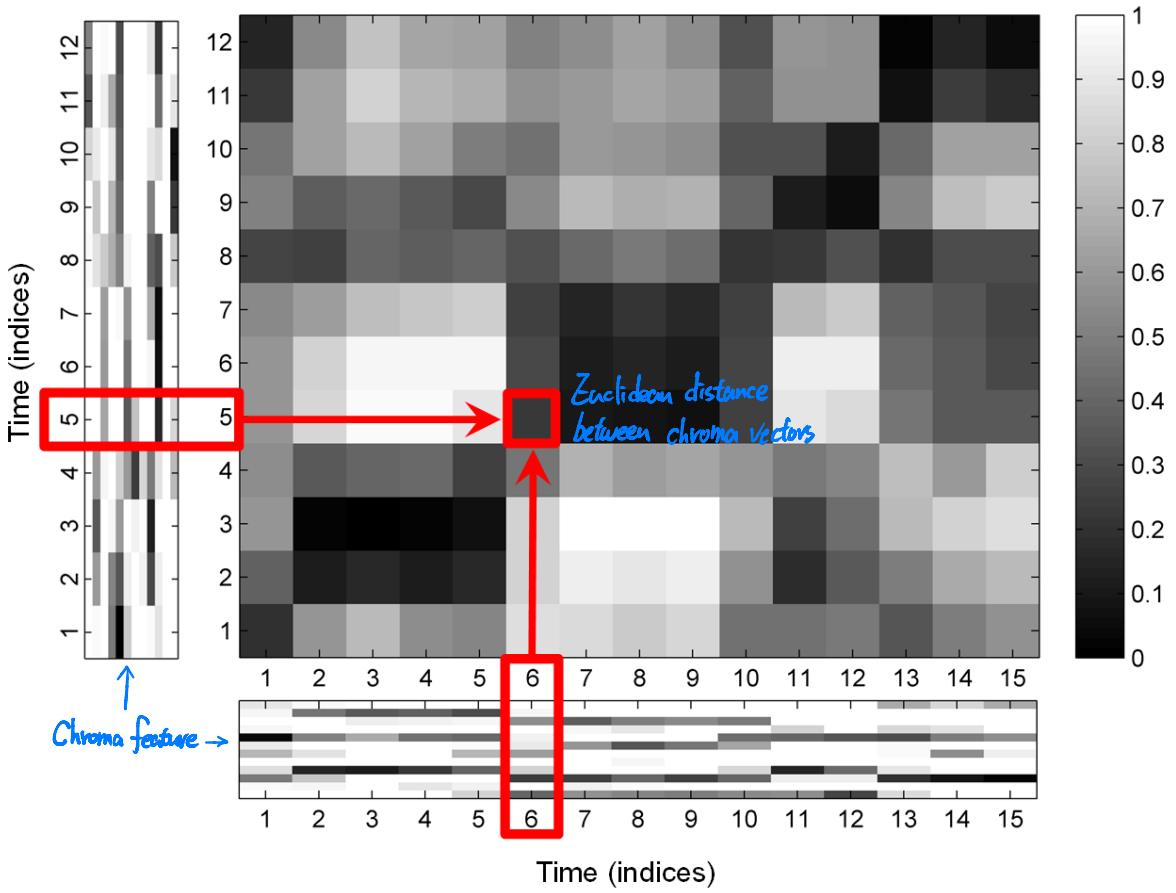
Audio/Midi
Audio/Score
Audio/Audio
) Alignment

Music synchronization problem

- Many music synchronization problems can be viewed as an audio-to-audio synchronization problem
  - For audio-to-MIDI alignment, synthesize the MIDI to audio first
- Question: how to compare two audio sequences and find musically corresponding time positions
  - Feature representation
  - Alignment technique and scenario (offline and online)
- First research dates back to 1984
  - The synthetic performer in the context of live performance, ICMC 1984
  - An on-line algorithm for real-time accompaniment, ICMC 1984

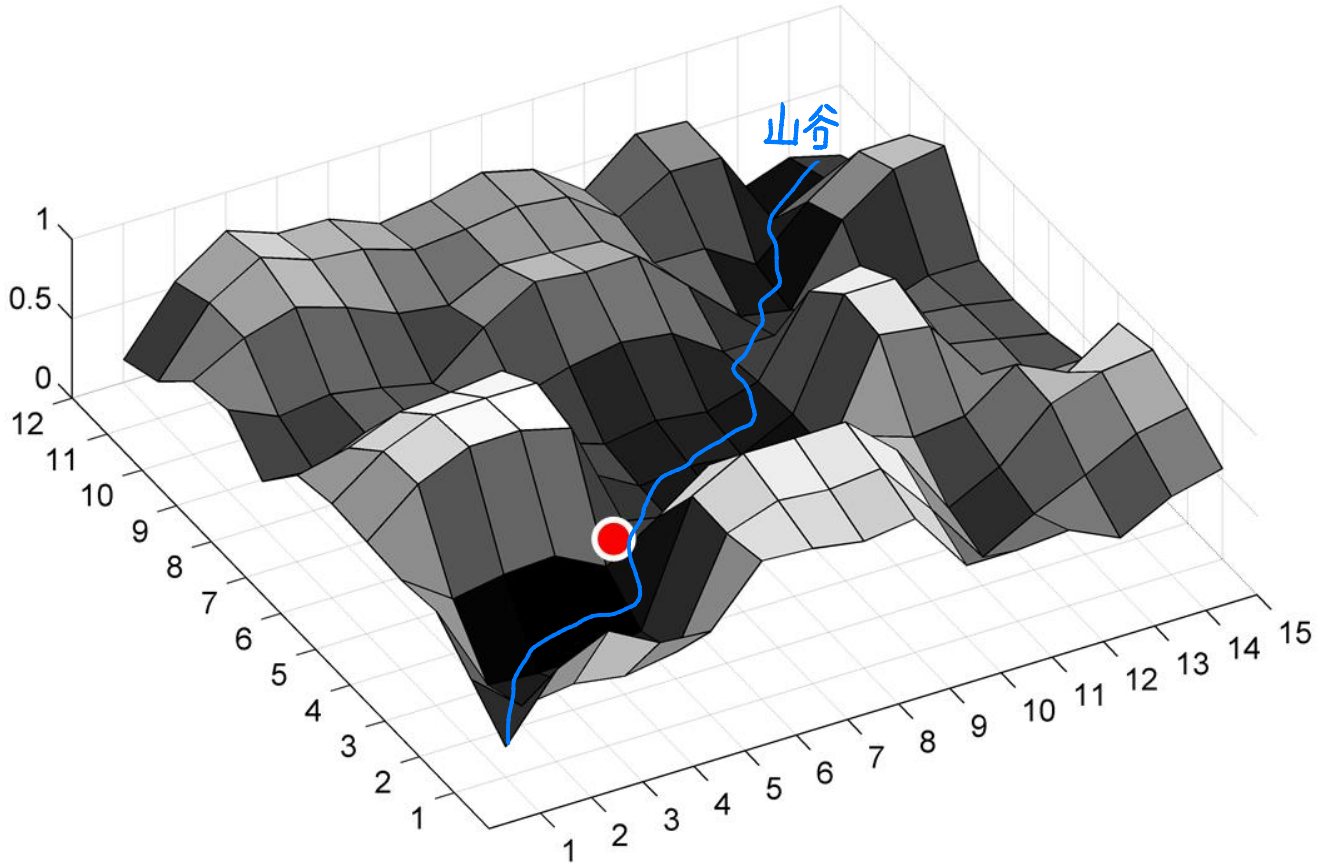
An example (1)

- 2 versions of input audio
- Chroma feature
- Euclidean distance



From: M. Mueller, *Fundamentals of Music Processing*, Chapter 3, Springer 2015

Another view



From: M. Mueller, *Fundamentals of Music Processing*, Chapter 3, Springer 2015

Dynamic time warping (DTW)

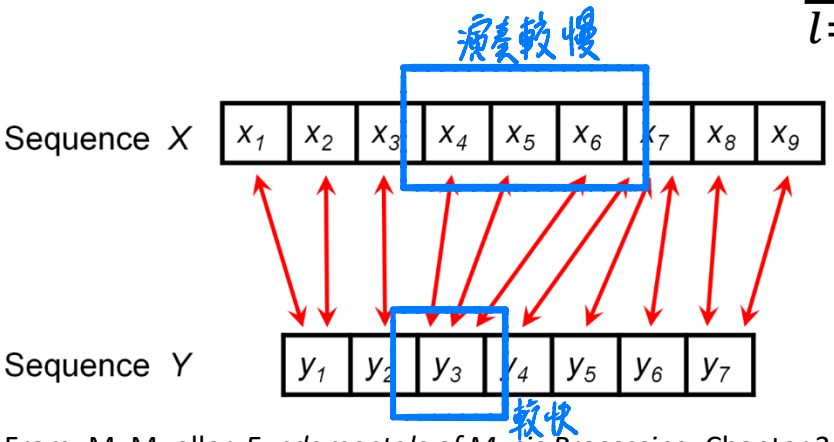
See Thomas Cormen Textbook

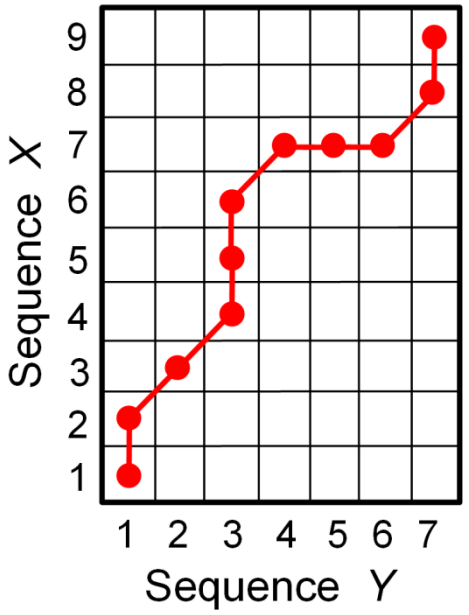
- Find the "minimal distance" between two sequences with different lengths
- Two components of DTW
  - Feature representation for each time instance
  - Distance function for alignment
- The outcome of DTW
  - Alignment path 版本A的第n秒的D。對到的是 版本B的第幾秒的D。?
  - DTW distance distance愈大，兩sequence愈不像
- The DTW algorithm answers that "how the two sequences need to be aligned in order to get minimal distance"

Problem formulation of DTW

- Given two sequences $X = [x_1, x_2, \ldots, x_N]$ and $Y = [y_1, y_2, \ldots, y_M]$, find the warping path $P = [p_1, p_2, \ldots, p_L]$ with $p_l = (n_l, m_l) \in [1:N] \times [1:M]$ in order to minimize the DTW distance between X and Y
- The case of Euclidean distance:

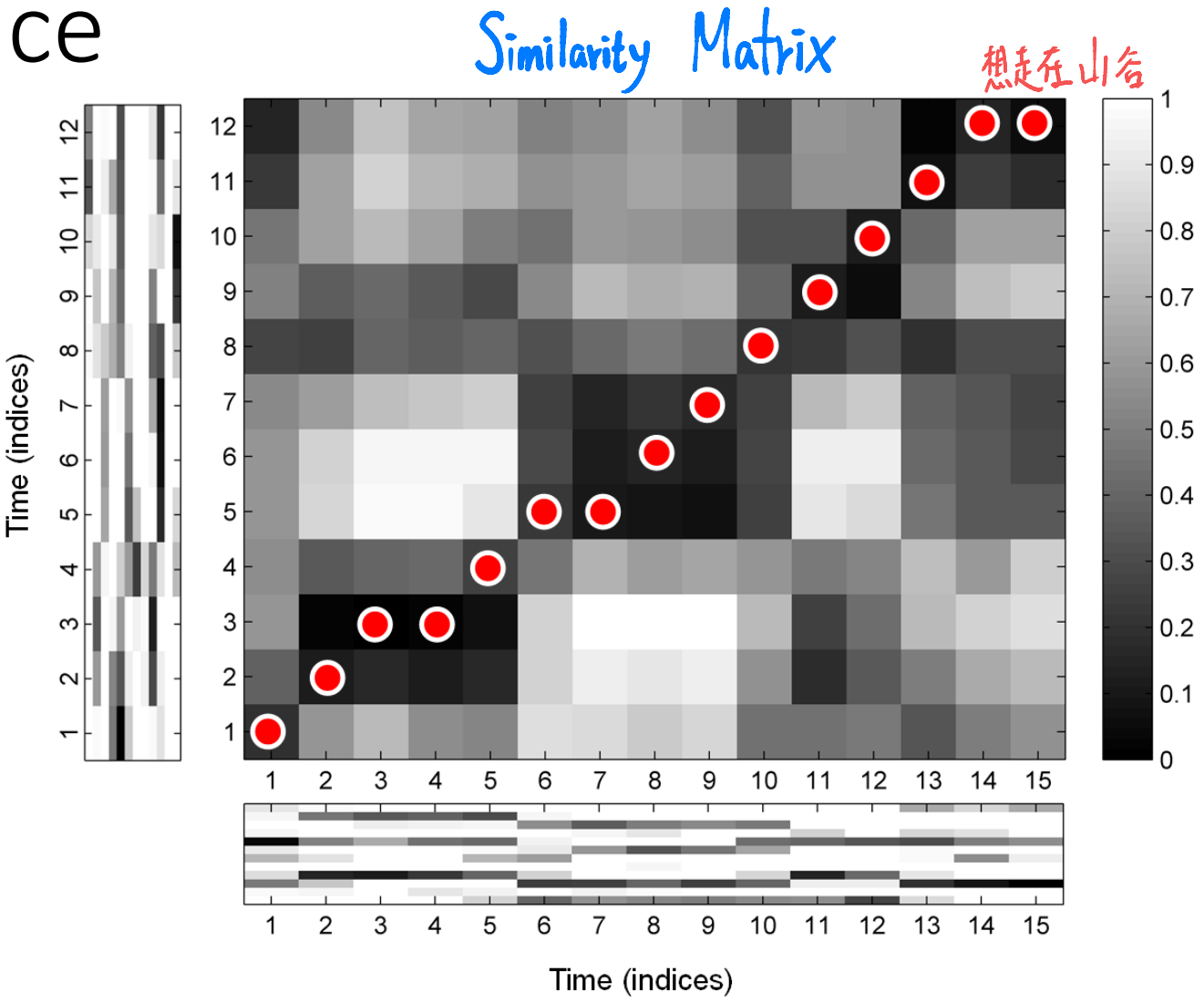
$$DTW(X,Y) = \sum_{l=1}^{L} \left| x_{n_l} - y_{m_l} \right|_2^2$$





From: M. Mueller, *Fundamentals of Music Processing*, Chapter 3, Springer 2015

The alignment path and DTW distance



From: M. Mueller, *Fundamentals of Music Processing*, Chapter 3, Springer 2015

The constraints of DTW

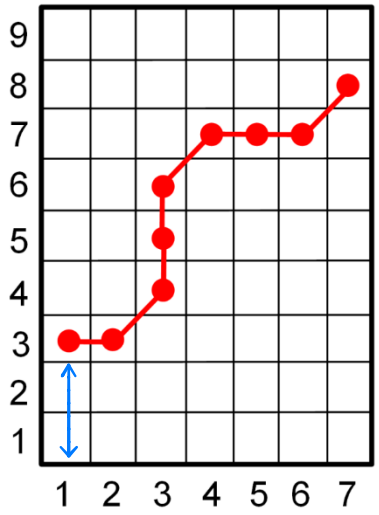
- Constraints
  - boundary condition (anchored beginning, anchored end)
  - monotonically increase
  - step size condition
- Some invalid cases:

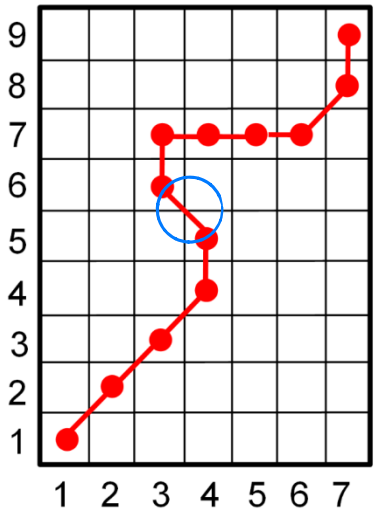
曲子的開始可能在音檔的任意位置之後　指定起點和終點

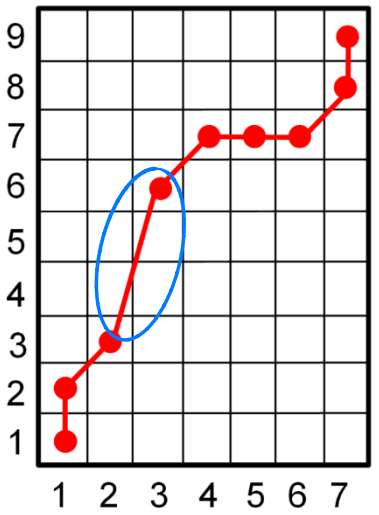
不可走回頭路（只可往右或往上走或右上）

不要跳太遠

From: M. Mueller, *Fundamentals of Music Processing*, Chapter 3, Springer 2015





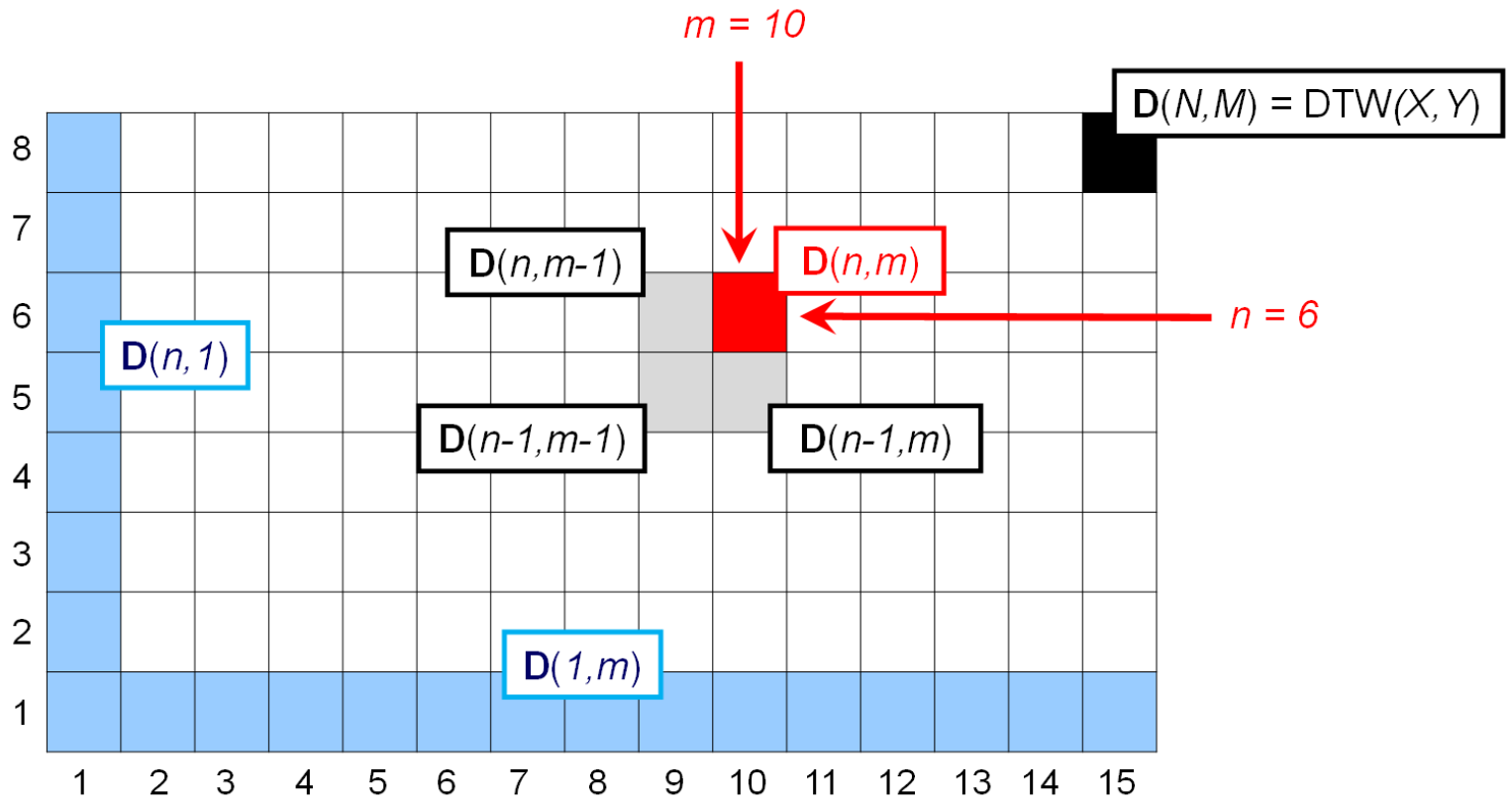


Algorithm

- How to find the warping path?
  - Try all possible paths: $O(2^N 2^M)$
  - Use dynamic programming (DP): $O(NM)$

Algorithm

- Accumulated cost matrix D saves the optimal distance for every step



From: M. Mueller, *Fundamentals of Music Processing*, Chapter 3, Springer 2015

Algorithm

Algorithm: DTW

Input: Cost matrix $\mathbf{C}$ of size $N \times M$
Output: Accumulated cost matrix $\mathbf{D}$
Optimal warping path P^*

Procedure: Initialize $(N \times M)$ matrix $\mathbf{D}$ by $\mathbf{D}(n,1) = \sum_{k=1}^{n} \mathbf{C}(k,1)$ for $n \in [1:N]$ and $\mathbf{D}(1,m) = \sum_{k=1}^{m} \mathbf{C}(1,k)$ for $m \in [1:M]$. Then compute in a nested loop for $n = 2,\ldots,N$ and $m = 2,\ldots,M$:

$$\mathbf{D}(n,m) = \mathbf{C}(n,m) + \min\{\mathbf{D}(n-1,m-1), \mathbf{D}(n-1,m), \mathbf{D}(n,m-1)\}.$$

Set $\ell = 1$ and $q_\ell = (N,M)$. Then repeat the following steps until $q_\ell = (1,1)$:

Increase ℓ by one and let $(n,m) = q_{\ell-1}$.
If $n = 1$, then $q_\ell = (1,m-1)$,
else if $m = 1$, then $q_\ell = (n-1,m)$,
else $q_\ell = \operatorname{argmin}\{\mathbf{D}(n-1,m-1), \mathbf{D}(n-1,m), \mathbf{D}(n,m-1)\}$.
(If 'argmin' is not unique, take lexicographically smallest cell.)

Set $L = \ell$ and return $P^* = (q_L, q_{L-1}, \ldots, q_1)$ as well as $\mathbf{D}$.

From: M. Mueller, *Fundamentals of Music Processing*, Chapter 3, Springer 2015

A simple example of DTW

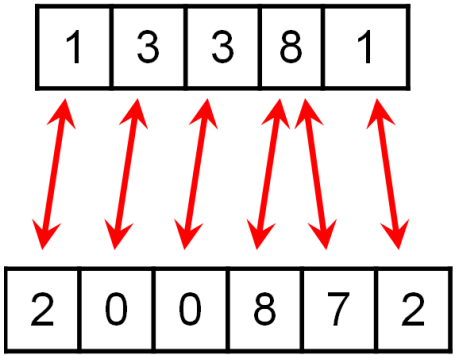
- Two integer-valued sequences

Cost matrix C

1	1	1	1	7	6	1
8	6	8	8	0	1	6
3	1	3	3	5	4	1
3	1	3	3	5	4	1
1	1	1	1	7	6	1
	2	0	0	8	7	2

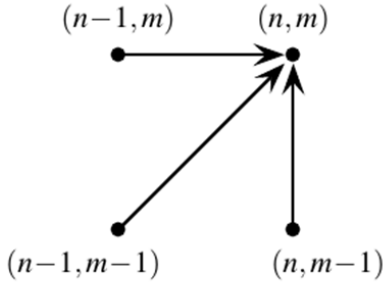
Accumulated cost matrix D

1	10	10	11	14	13	9
8	9	11	13	7	8	14
3	3	5	7	10	12	13
3	2	4	5	8	12	13
1	1	2	3	10	16	17
	2	0	0	8	7	2

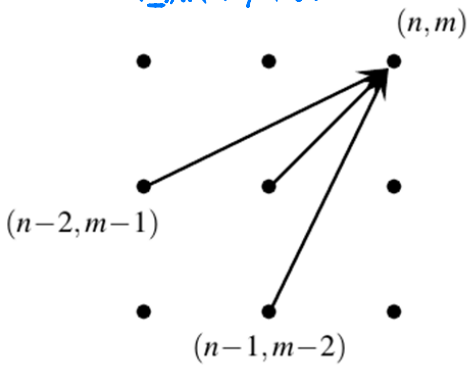


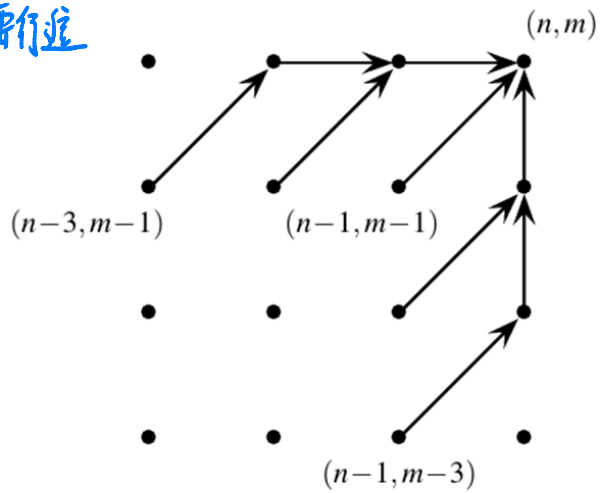
From: M. Mueller, *Fundamentals of Music Processing*, Chapter 3, Springer 2015

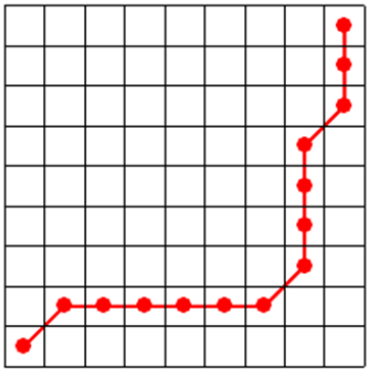
Step Size

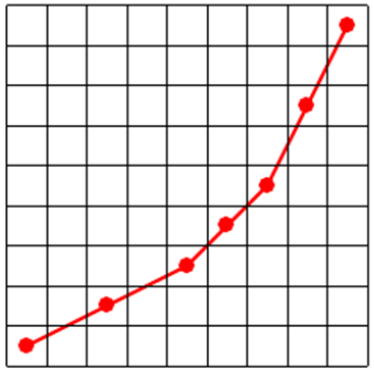


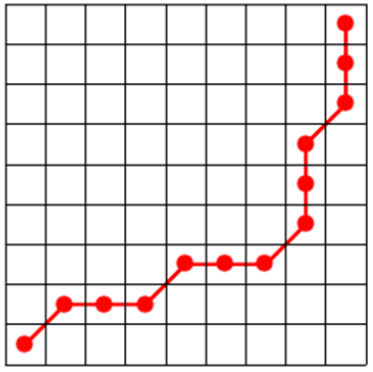
不要直的和横的
因為不太可能有個版本都沒要行近











From: M. Mueller, *Fundamentals of Music Processing*, Chapter 3, Springer 2015

Recurrent formulae

see photo

- $D(n,m) = C(n,m) + \min(D(n-1,m), D(n, m-1), D(n-1,m-1))$

Global constraints

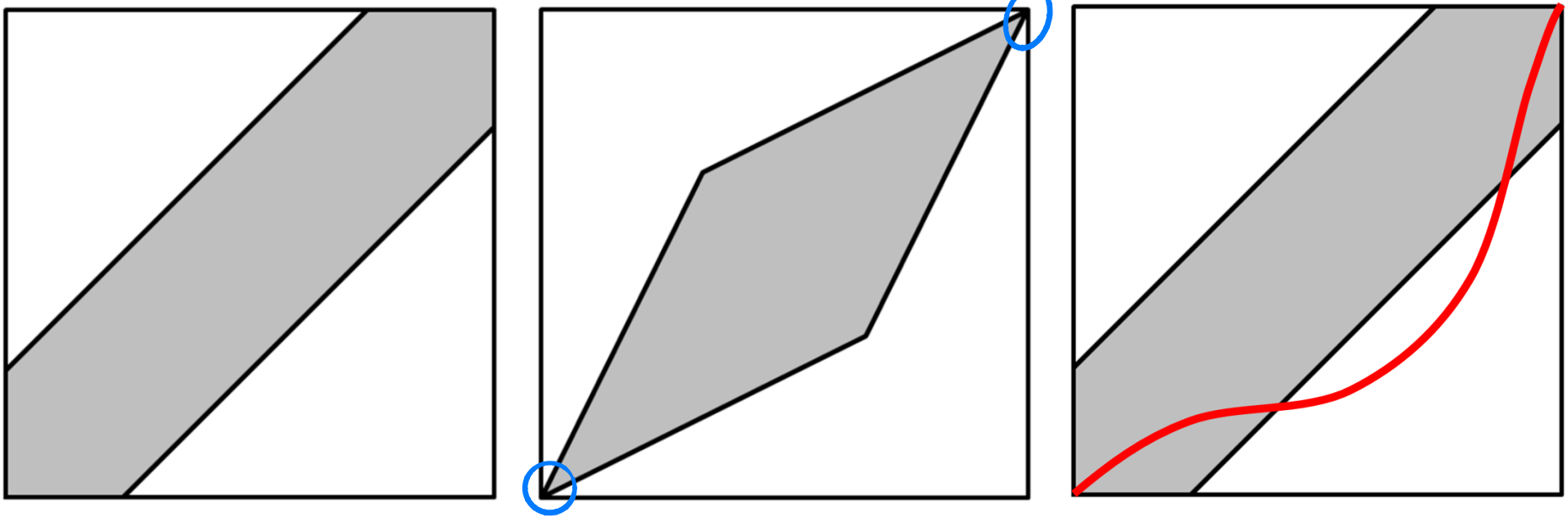
- Useful when the two sequences are similar
- Speed up

大部份的狀況，最好的 alignment path 不會離對角線太遠（只考慮灰色區域）

Sakoe-Chiba band

Itakura parallelogram

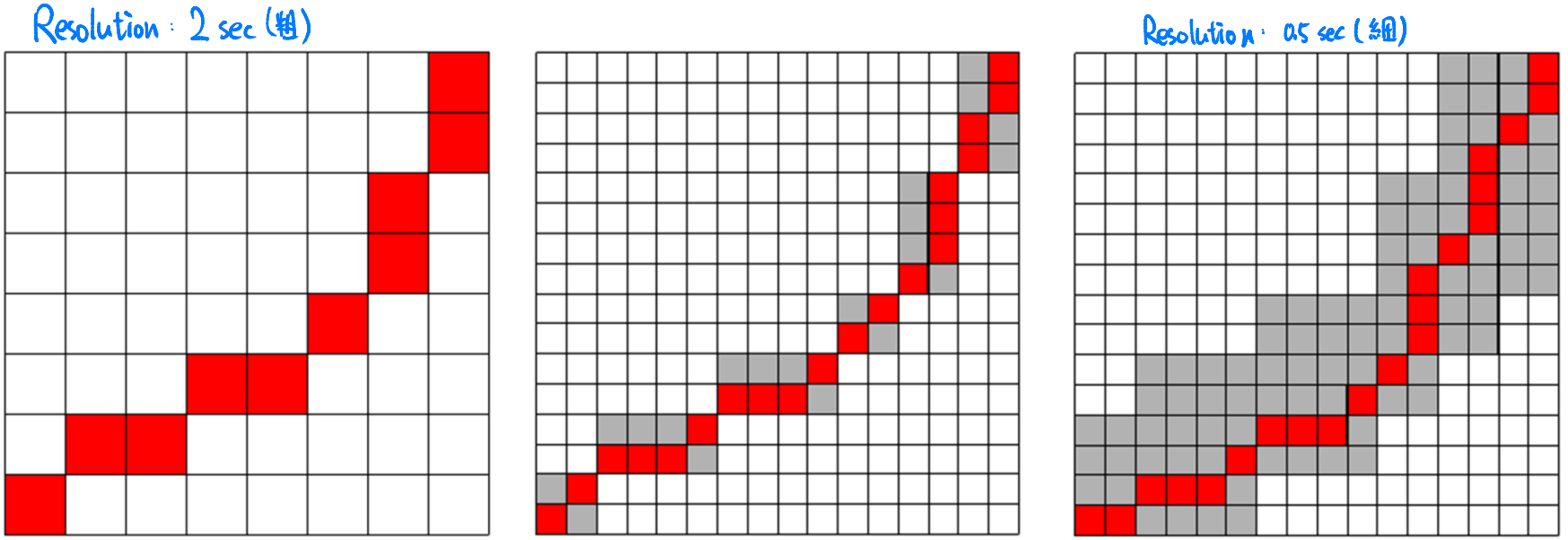
確定頭對頭，尾對尾



From: M. Mueller, *Fundamentals of Music Processing*, Chapter 3, Springer 2015

Multi-scale DTW

- Coarse-to-fine processing



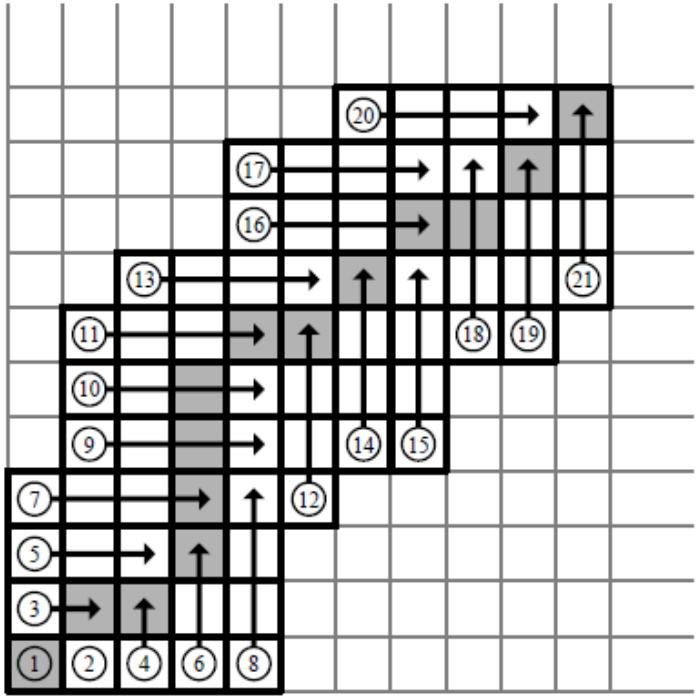
From: M. Mueller, *Fundamentals of Music Processing*, Chapter 3, Springer 2015

沒有backtracking的方法，常用於自動伴奏

Online time warping (OTW)

假設速度不會差太多

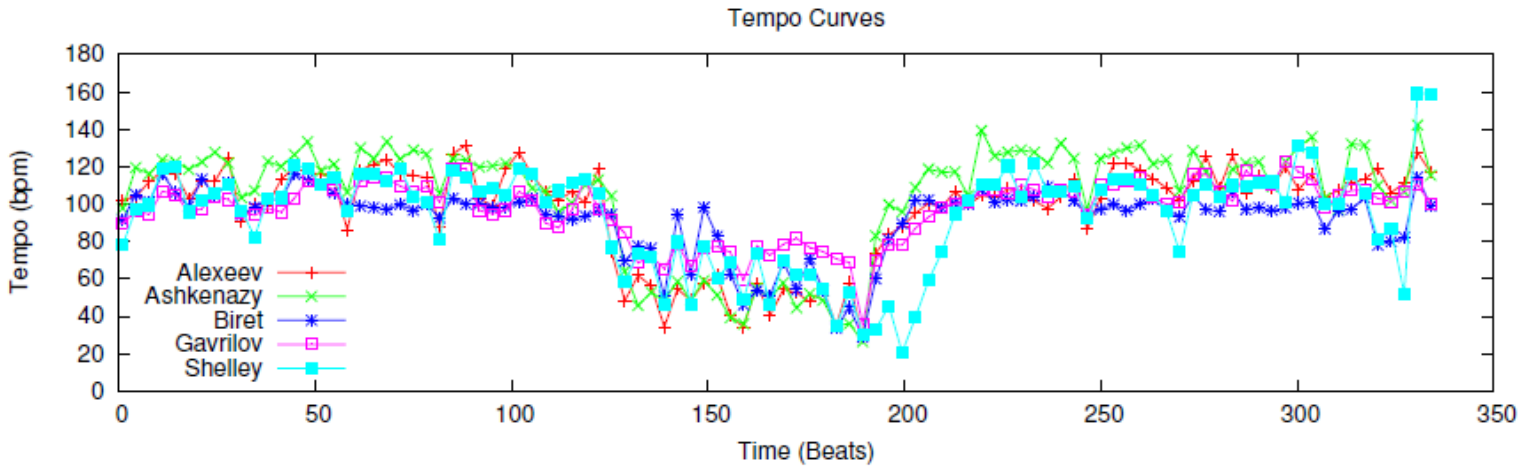
- Linear time without computing the total cost matrix



Dixon, Simon. "Live tracking of musical performances using on-line time warping." *Proceedings of the 8th International Conference on Digital Audio Effects*. 2005.

Online tempo estimation

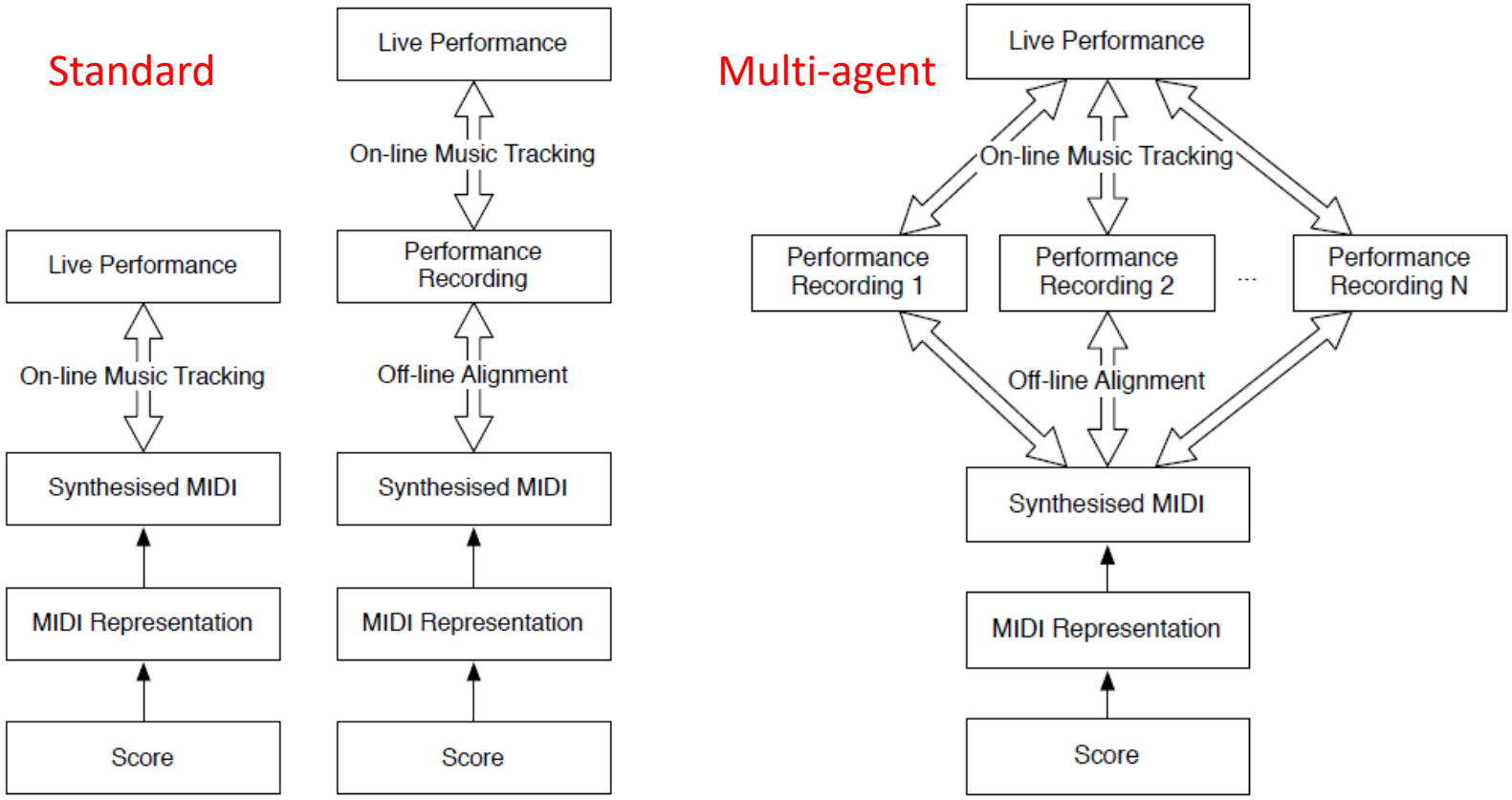
- Estimating the instantaneous tempo (短時節拍推估)
- Prediction with the tempo estimation
- "Starting in the middle" Remaining Challenge：可用於自動翻譜



Arzt, Andreas, and Gerhard Widmer. "Simple tempo models for real-time music tracking." *Proceedings of the Sound and Music Computing Conference (SMC)*. 2010.

Alignment with multiple performance

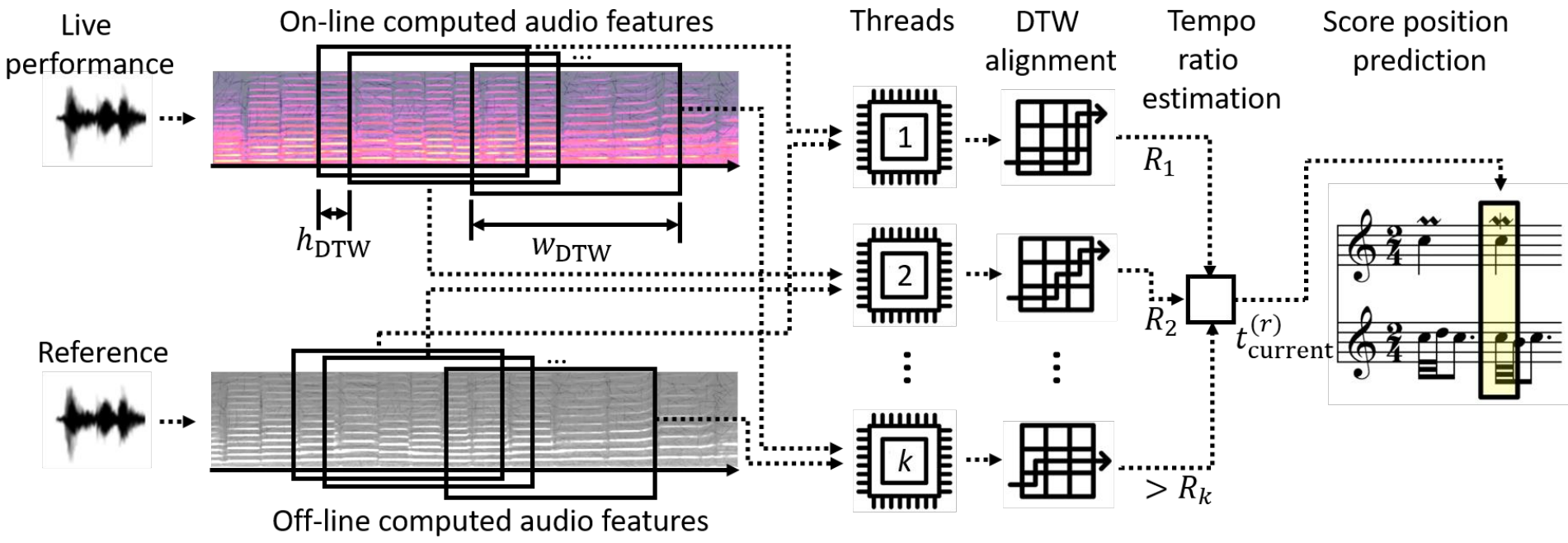
align多個演奏版本



Arzt, Andreas, and Gerhard Widmer. "Real-Time Music Tracking Using Multiple Performances as a Reference." *ISMIR*. 2015.

Alignment with multiple estimation

- Parallel dynamic time warping (PDTW)

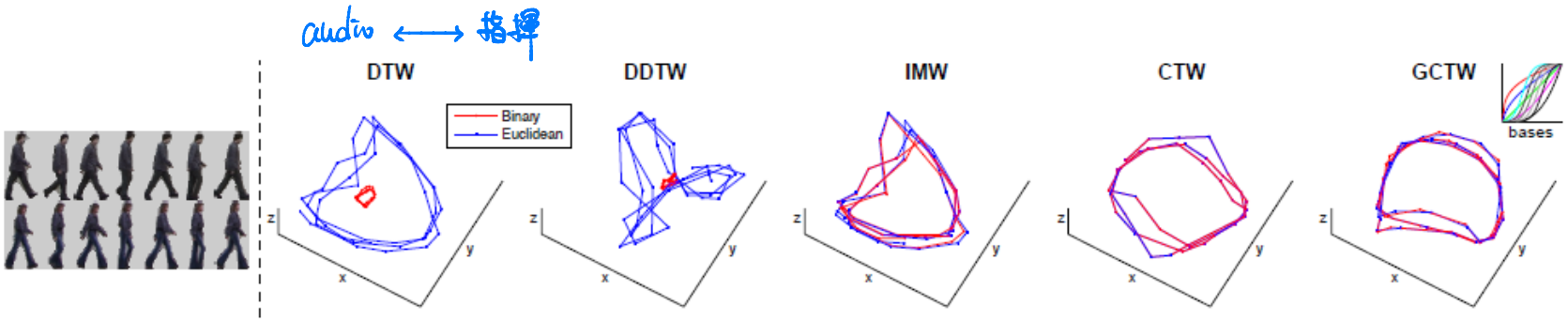


I-Chieh Wei and Li Su, “Online performance tracking with parallel dynamic time warping”

Canonical time warping (CTW)

Audio to other things alignment

- Alignment between different modalities of data



Zhou, Feng, and Fernando De la Torre. "Generalized canonical time warping." *IEEE transactions on pattern analysis and machine intelligence* 38.2 (2016): 279-294.

Canonical correlation analysis (CCA)

(a variant of PCA)

- Given $\mathbf{X} = [\mathbf{x}_1, \ldots, \mathbf{x}_n] \in R^{d_x \times n}$, $\mathbf{Y} = [\mathbf{y}_1, \ldots, \mathbf{y}_n] \in R^{d_y \times n}$, CCA finds the linear combinations of the variables in $\mathbf{X}$ that are most correlated with the linear combinations of the variables in $\mathbf{Y}$
- Find the projection matrices $\mathbf{V_x} \in R^{d_x \times d}$ and $V_y \in R^{d_y \times d}$, $d \leq \min(d_x, d_y)$, to minimize

$$J_{CCA} = \left\| \mathbf{V_x^T X} - \mathbf{V_y^T Y} \right\|_F^2$$

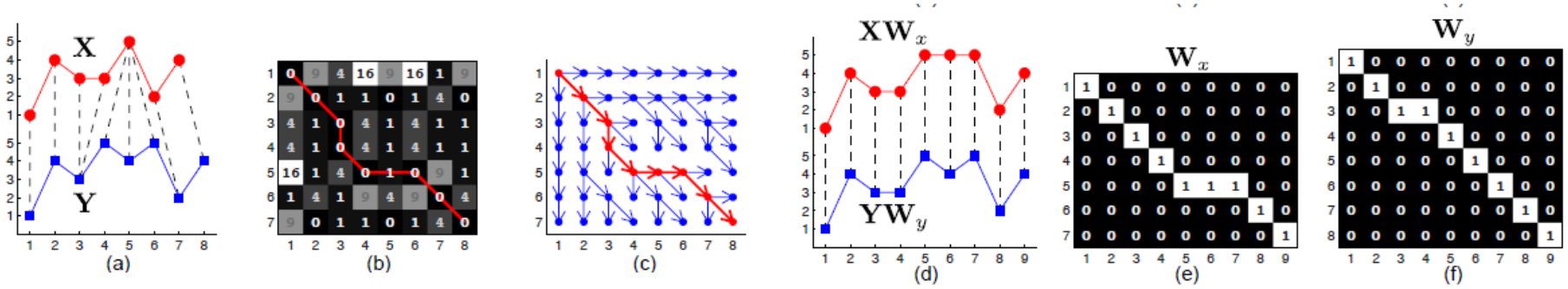
找共同空間

From DTW to CTW

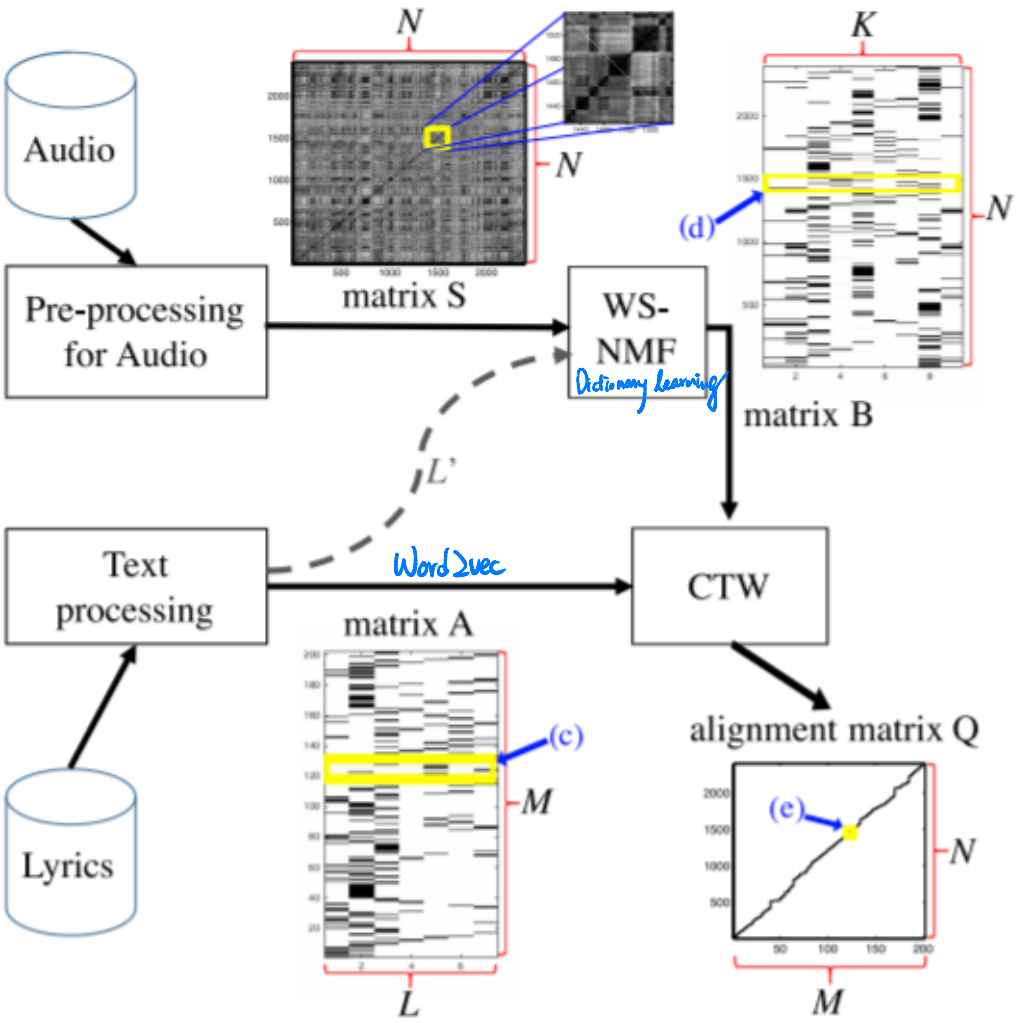
- $\mathbf{X} \in R^{d \times n_x}, \mathbf{Y} \in R^{d \times n_y}$ CTW：在新的空間上做DTW
- Matrix form of DTW: find the warping matrix $\mathbf{W_x} \in \{0,1\}^{n_x \times l}$ and $\mathbf{W_y} \in \{0,1\}^{n_y \times l}$ to minimize

$$J_{DTW} = \left\| \mathbf{XW_x} - \mathbf{YW_y} \right\|_F^2$$

- CTW: minimize $\left\| \mathbf{V_x^T XW_x} - \mathbf{V_y^T YW_y} \right\|_F^2$



Audio-to-lyrics alignment 自動生歌詞



Chang, Sungkyun, and Kyogu Lee. "Lyrics-to-Audio Alignment by Unsupervised Discovery of Repetitive Patterns in Vowel Acoustics." *IEEE Access* 5 (2017)