E9:261

26-02-2016

Recap …

- GMMs and EM algorithm
  - Relationship between neighboring samples is not considered.
- Time alignment and normalization
  - Dynamic Programming algorithms
    - recursive solution

Dynamic Programming

1. Initialization

$$\varphi_1(i,n) = \zeta(i,n)$$
$$\xi_1(n) = i$$
$$\text{for } n = 1, 2, \ldots, N.$$

2. Recursion

$$\varphi_{m+1}(i,n) = \min_{1 \le \ell \le N} [\varphi_m(i,\ell) + \zeta(\ell,n)]$$
$$\xi_{m+1}(n) = \arg\min_{1 \le \ell \le N} [\varphi_m(i,\ell) + \zeta(\ell,n)]$$
$$\text{for } n = 1, 2, \ldots, N \text{ and } m = 1, 2, \ldots, M-2$$

3. Termination

$$\varphi_M(i,j) = \min_{1 \le \ell \le N} [\varphi_{M-1}(i,\ell) + \zeta(\ell,j)]$$
$$\xi_M(j) = \arg\min_{1 \le \ell \le N} [\varphi_{M-1}(i,\ell) + \zeta(\ell,j)]$$

4. Path Backtracking

$$\text{optimal path} = (i, i_1, i_2, \ldots, i_{M-1}, j),$$

where

$$i_m = \xi_{m+1}(i_{m+1}), \qquad m = M-1, M-2, \ldots, 1$$

with $i_M = j$.

"Fundamentals of Speech Recognition", Rabiner and Juang

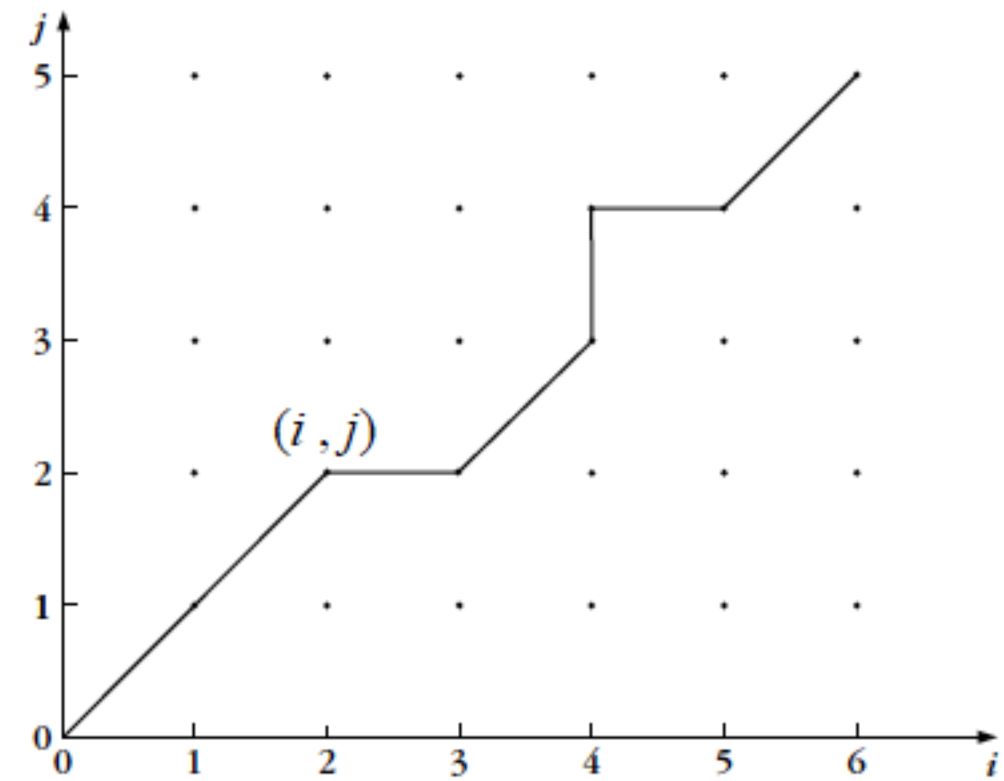
Dynamic Programming

- For an optimal path passing through $(i\ ,\ j)$:

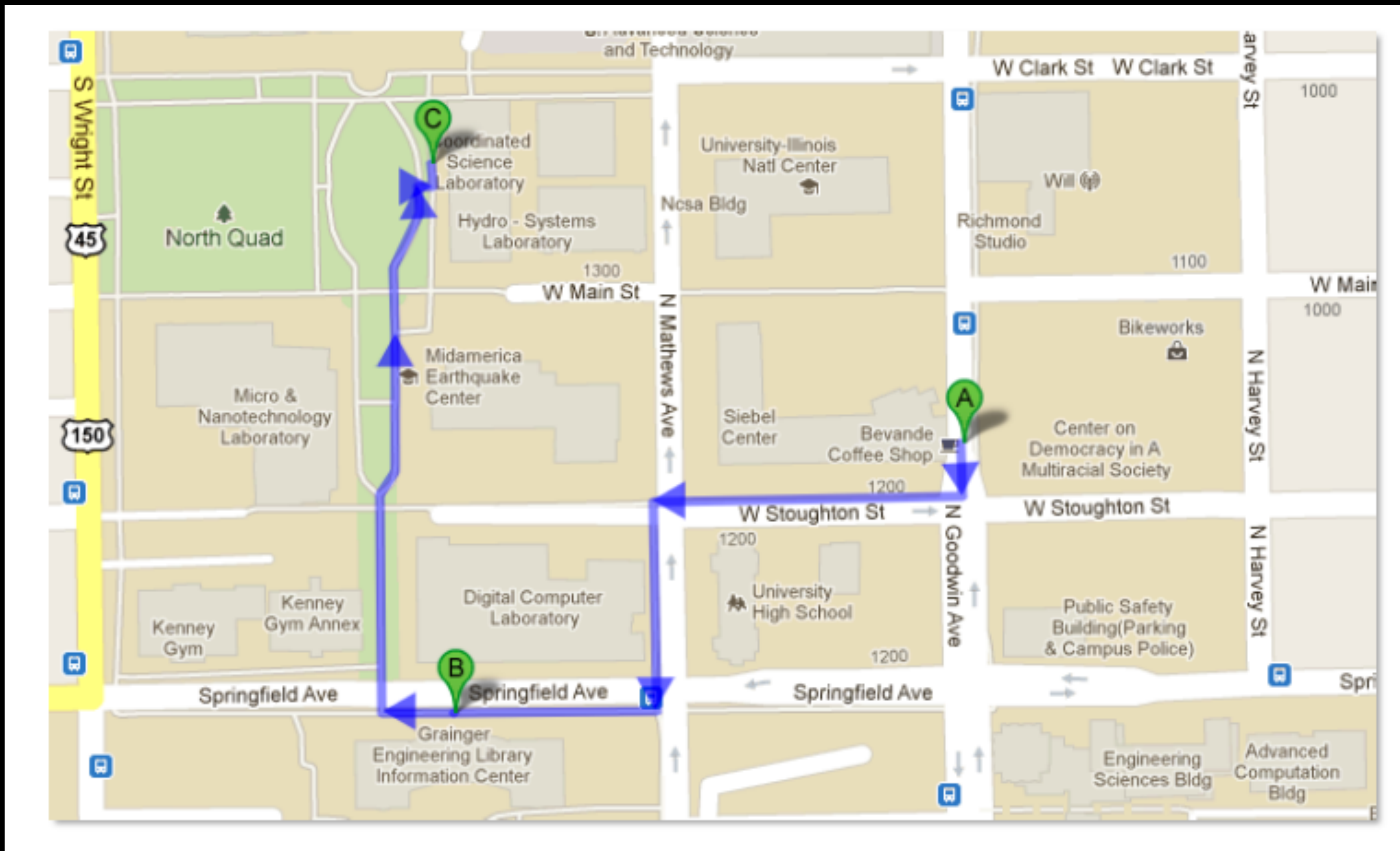
$$(i_0, j_0) \overset{opt}{\longrightarrow} (i_f, j_f)$$

- Then:

$$(i_0, j_0) \overset{opt}{\longrightarrow} (i_f, j_f) = \left\{ (i_0, j_0) \overset{opt}{\longrightarrow} (i, j), (i, j) \overset{opt}{\longrightarrow} (i_f, j_f) \right\}$$



Dynamic Programming in Real Life



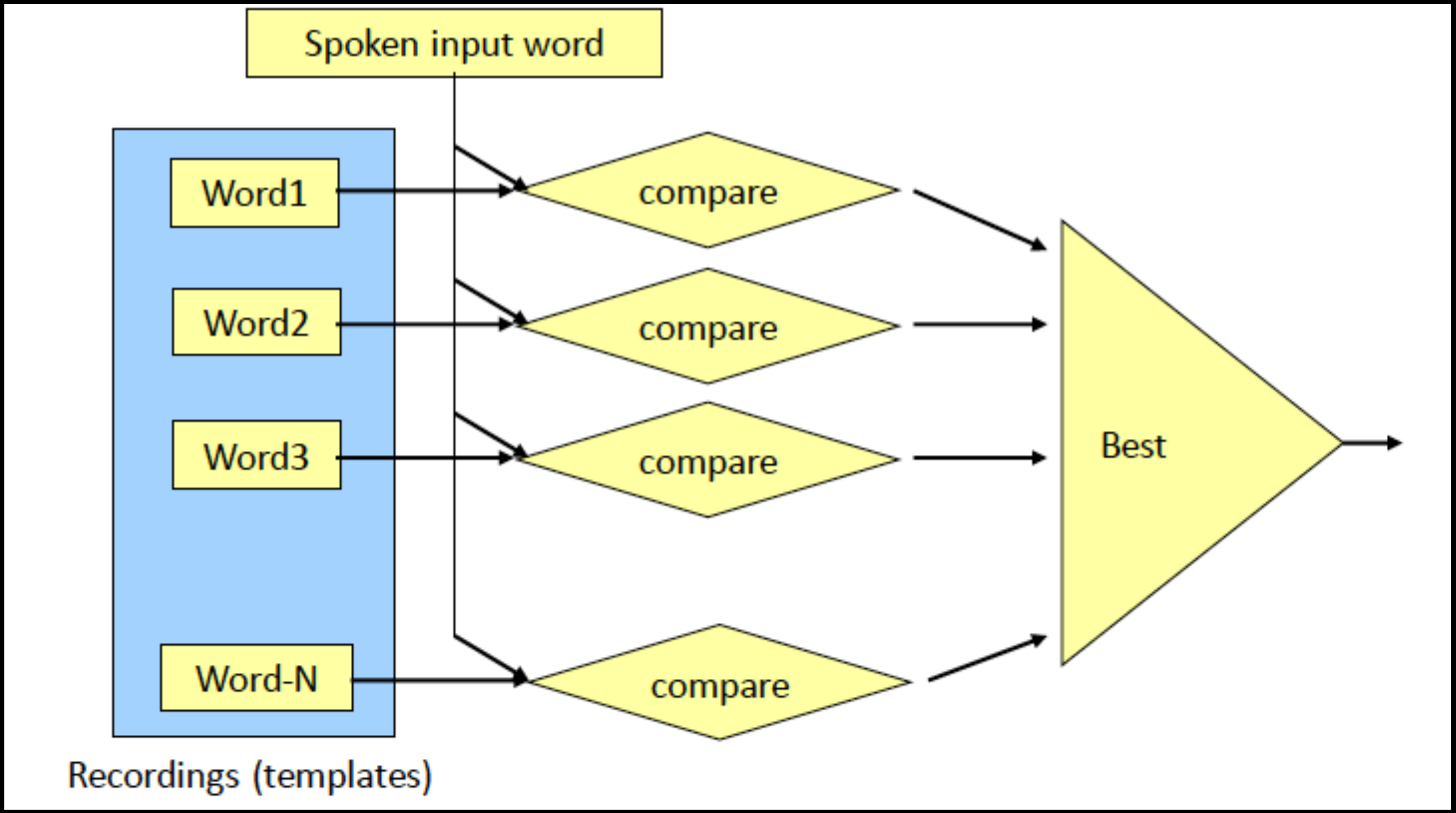
"Learning Time Series", UIUC Lecture Series

Template Matching

- Store "templates" for all words to be recognized
  - Template = example recording
    - Actually feature sequence from example recording
- Compute distance of input test data to all templates, select the closest

"Design and Implementation of Speech Recognition", CMU, Spring 2013

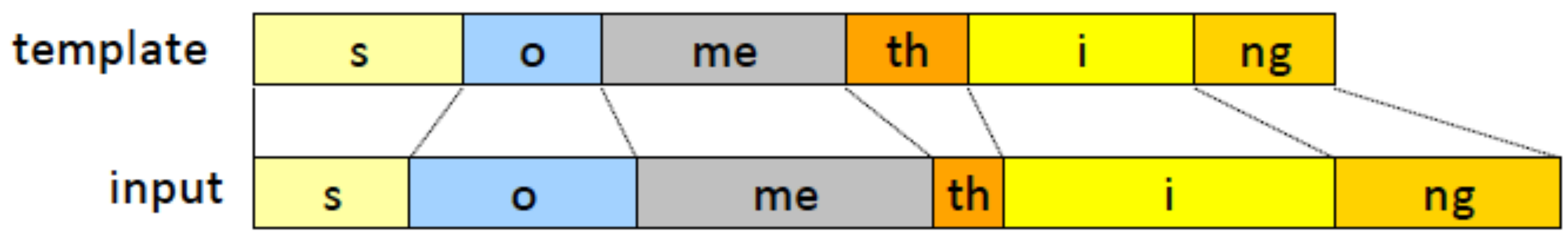
Template Matching



"Design and Implementation of Speech Recognition", CMU, Spring 2013

Template Matching

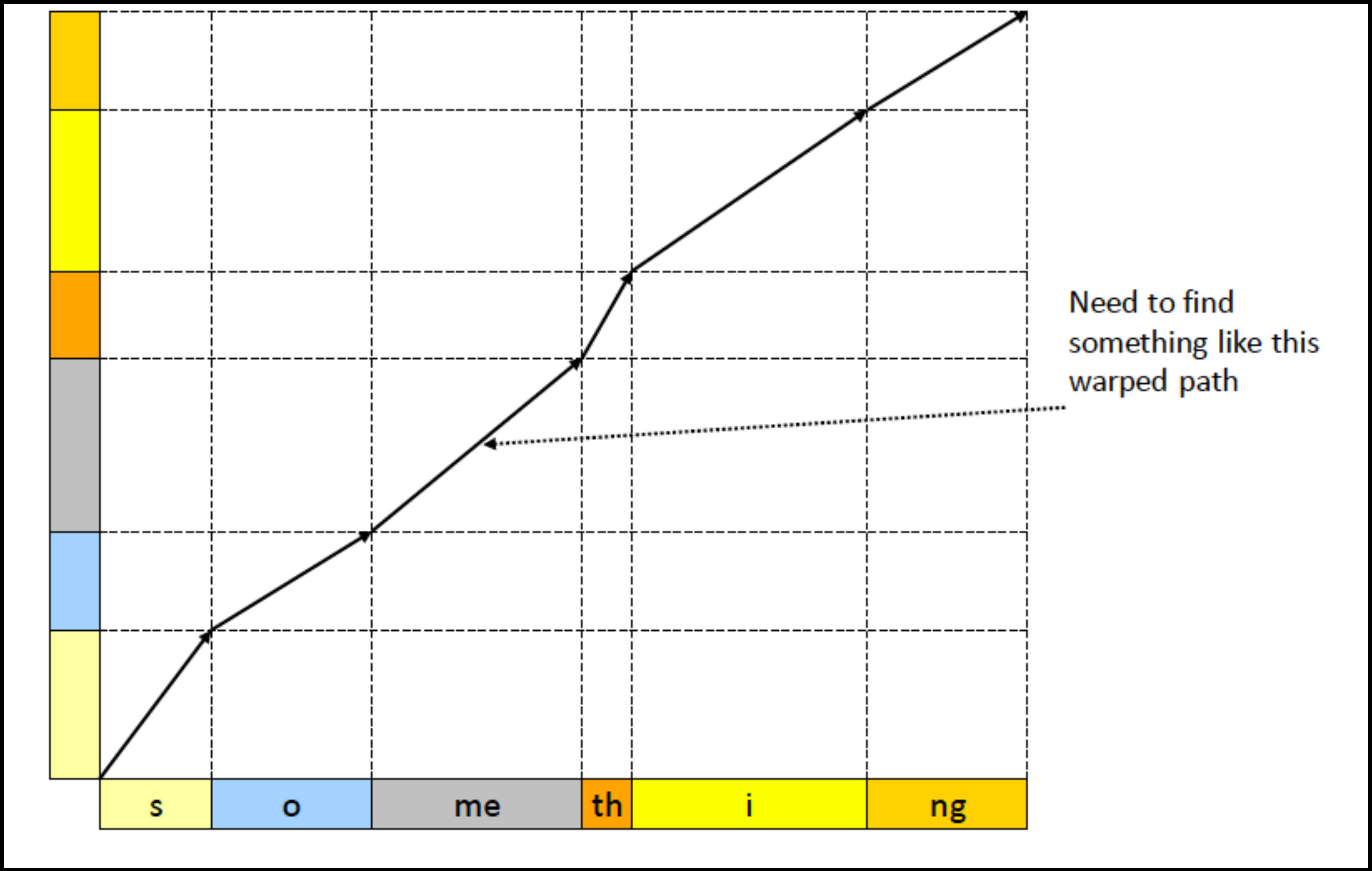
- Problem: Input and template may be different lengths



- Worse – the change in length may not be uniform

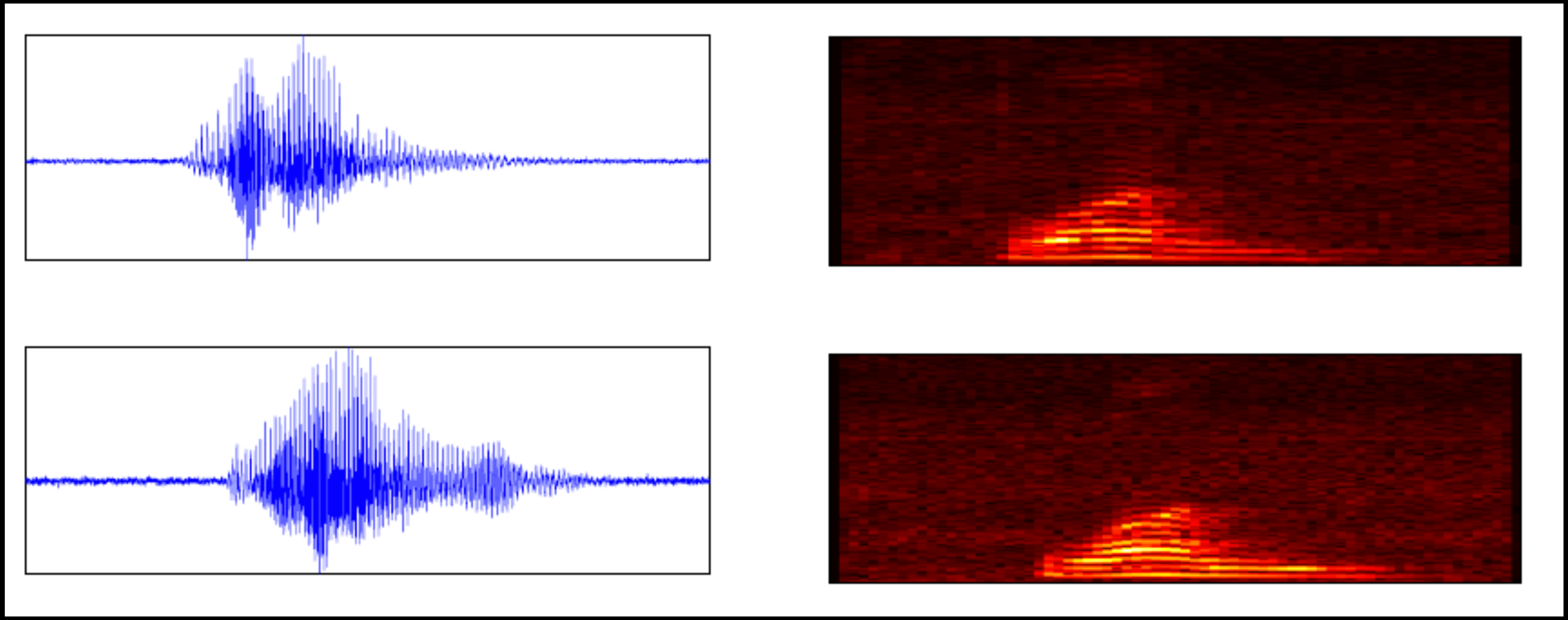
"Design and Implementation of Speech Recognition", CMU, Spring 2013

Template Matching



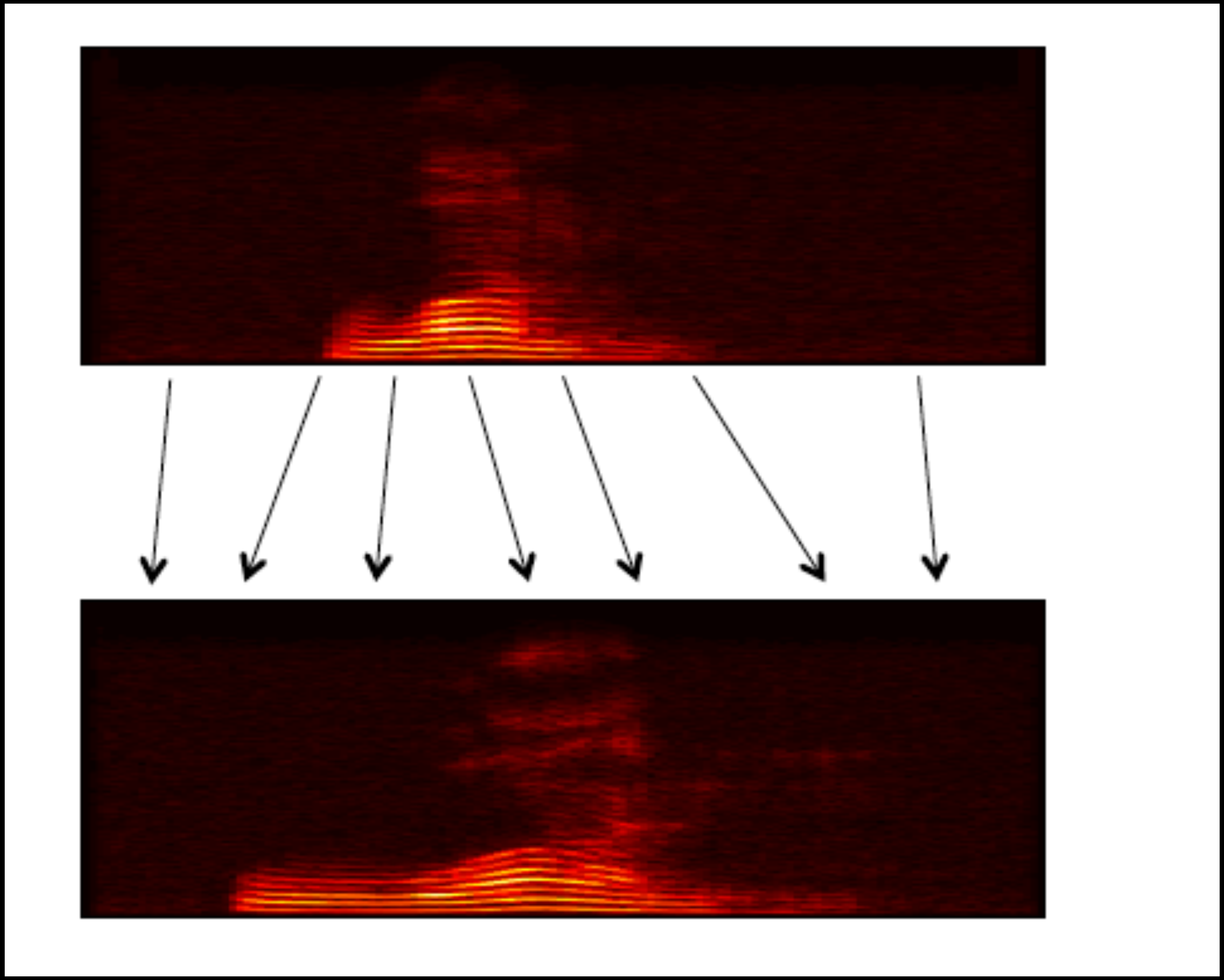
"Design and Implementation of Speech Recognition", CMU, Spring 2013

Template Matching



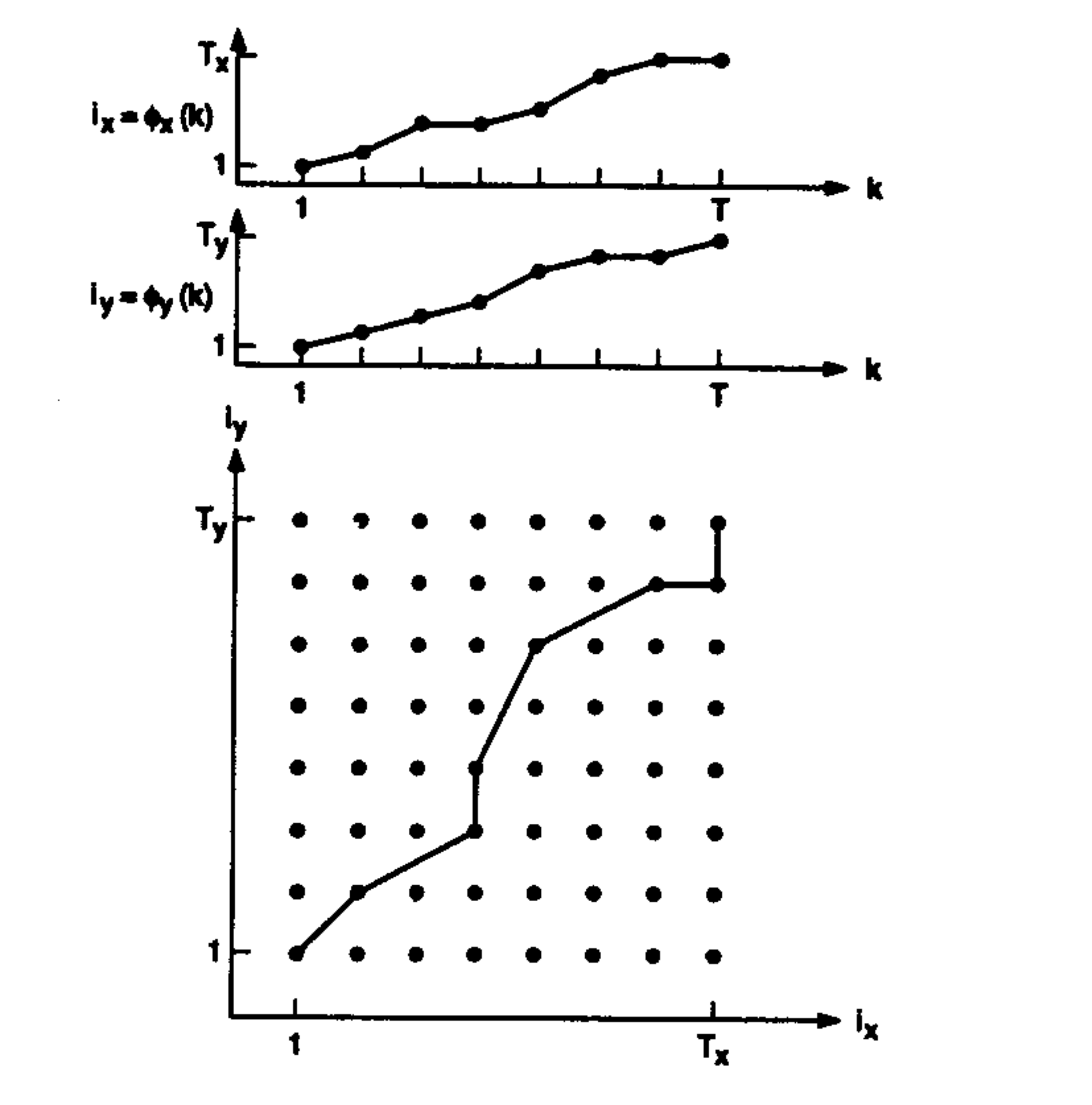
"Learning Time Series", UIUC Lecture Series

Template Matching



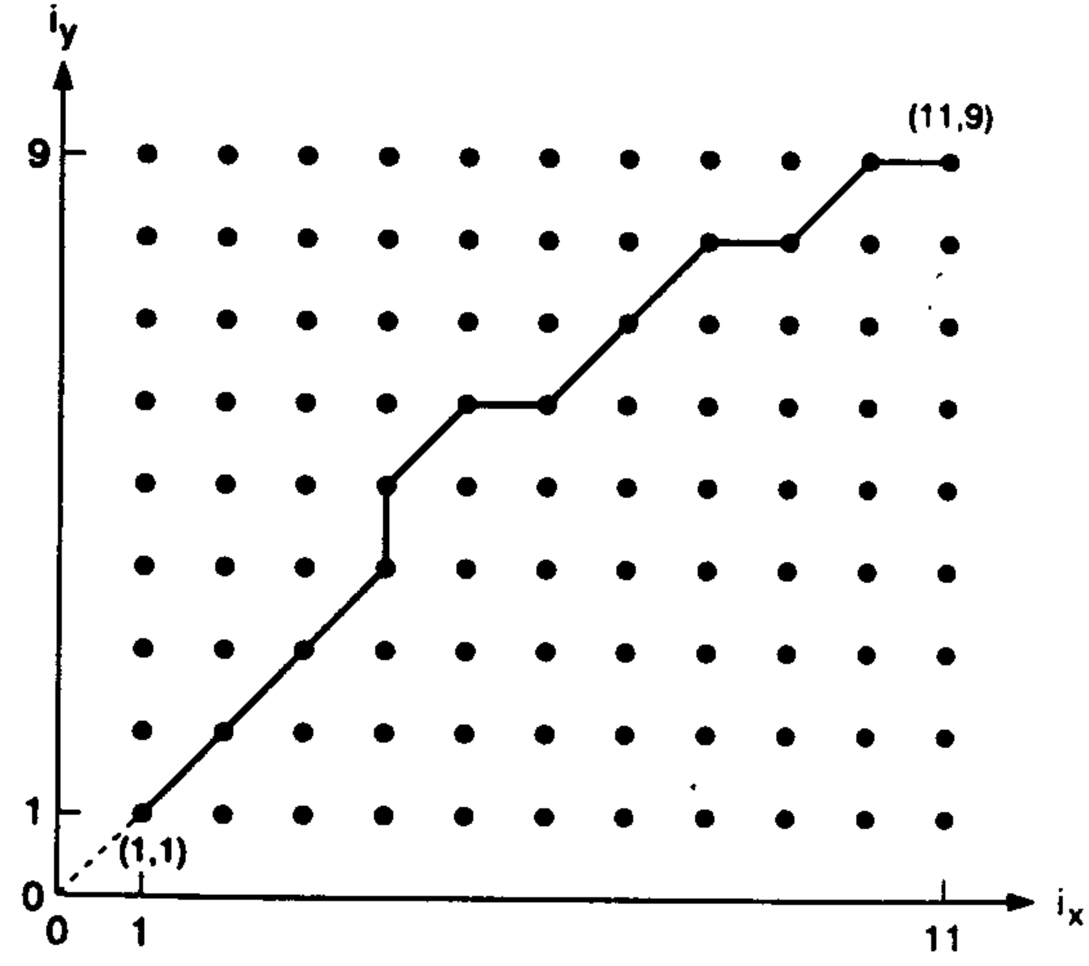
“Learning Time Series”, UIUC Lecture Series

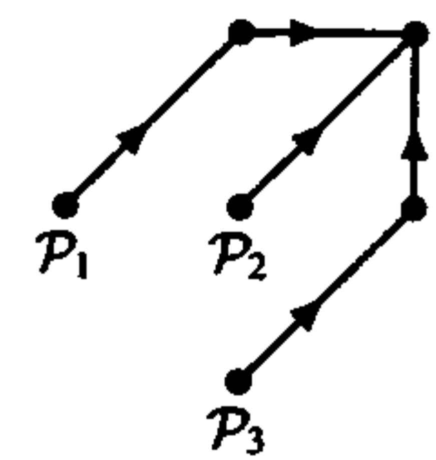
DTW



"Fundamentals of Speech Recognition", Rabiner and Juang

DTW





$$\mathcal{P}_1 \rightarrow (1,1)(1,0)$$
$$\mathcal{P}_2 \rightarrow (1,1)$$
$$\mathcal{P}_3 \rightarrow (1,1)(0,1)$$

$$\mathcal{P} \rightarrow \mathcal{P}_2\ \mathcal{P}_2\ \mathcal{P}_2\ \mathcal{P}_3\ \mathcal{P}_1\ \mathcal{P}_2\ \mathcal{P}_1\ \mathcal{P}_1.$$

“Fundamentals of Speech Recognition”, Rabiner and Juang

Disadvantages of DTW

- **Based on heuristics**
- **Creating templates from large number of examples can be hard.**
- **With large vocabulary sizes, computationally intractable.**
- **Not a statistical method.**

EARLY VOICE RECOGNITION