東京大学 工学部 計数工学科/物理工学科

応用音響学：音声認識 (E2)

連続単語 DP matching

嵯峨山 茂樹 ＜sagayama@hil.t.u-tokyo.ac.jp＞

東京大学 工学部 計数工学科 http://hil.t.u-tokyo.ac.jp/

謝辞： システム情報第一研究室勉強会資料を利用

参考文献

- 北 研二・中村 哲・永田 昌明「音声言語処理」森北出版
- 中川 聖一「パターン情報処理」丸善
- 古井 貞熙「音声情報処理」森北出版
- 谷萩 隆嗣「音声と画像のディジタル信号処理」コロナ社
- 中川 聖一「確率モデルによる音声認識」コロナ社

DPマッチングによる連続単語音声認識

連続単語音声認識において用いる際には、膨大な計算を回避するために以下のような効率良く処理を行う方法が提案されている。

- **2段DP法 ..... NEC 中央研究所: 迫江、千葉**
- **Level-Building法 ..... Bell Labs: Myers & Rabiner**
- **One-Pass DP法 ..... RSRE: J. Bridle / Philips: H. Ney / NEC 中研: 迫江 (“Clockwise DP”)**

基本概念: 連続単語音声認識の原理

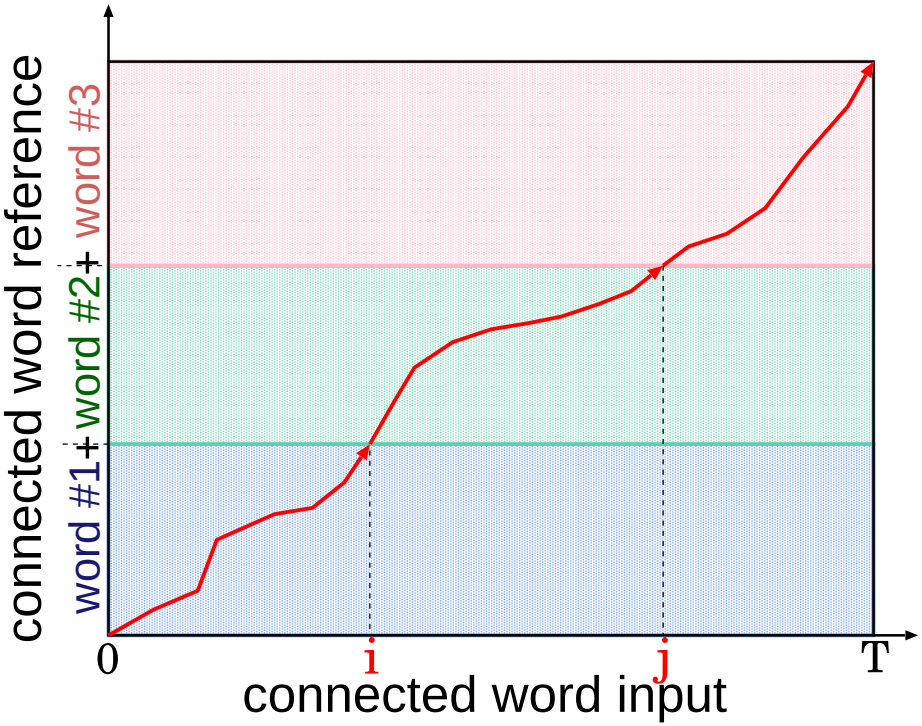


図1. ・第1単語から第n単語までそれぞれ複数候補を差し替えて距離計算したい。
・ n単語を連結した標準パターンとDPを行えば良いが、組合せは膨大。
・ 各単語のDP計算面が再利用できれば、効率は高い。
・ 単語境界では、各時刻の累積距離最小値だけ残せば良い。(他は影響しない)

端点フリーDP法の概念

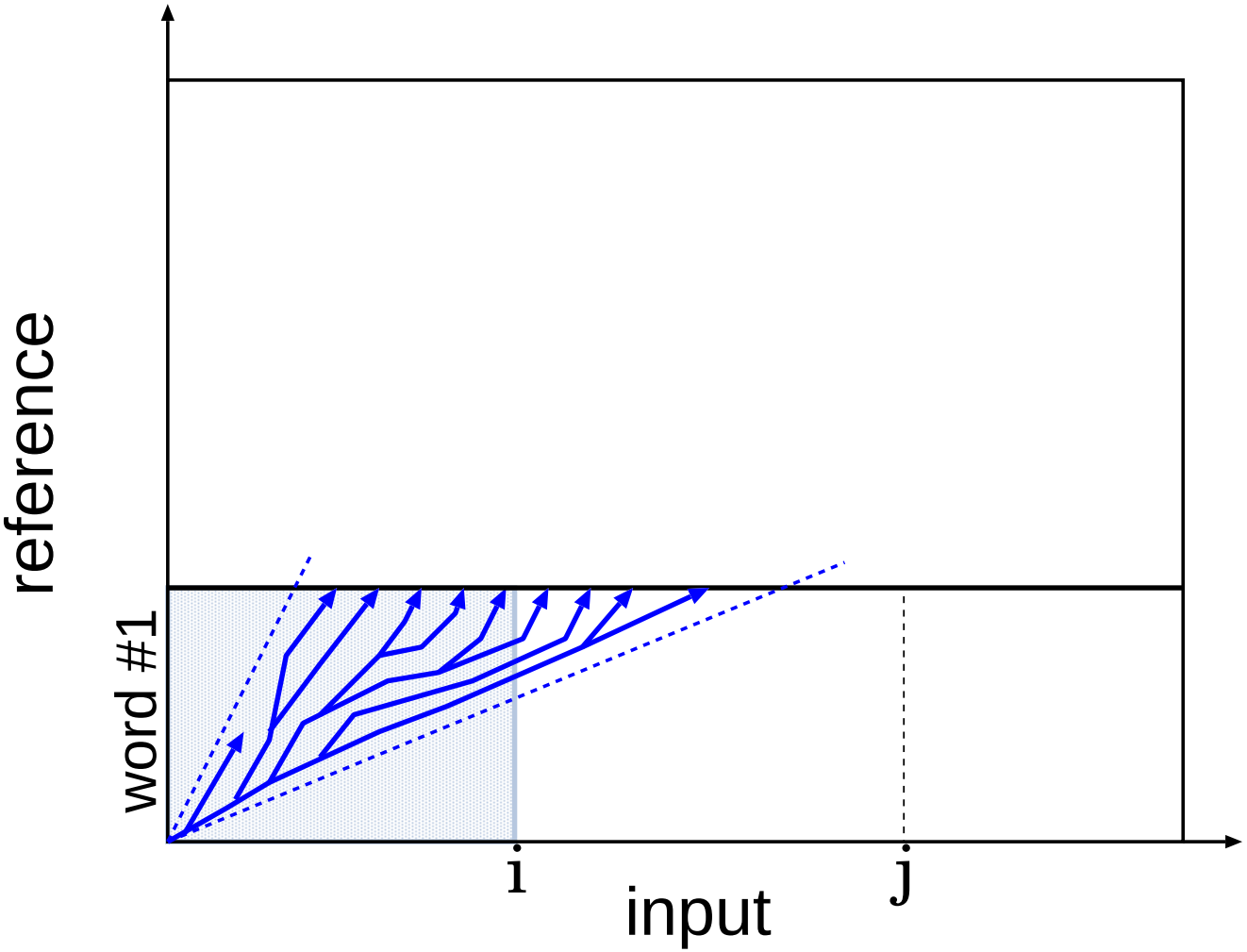


図2. 端点フリーDP – 終端を固定せず、ある区間について最適経路が求まる。

二段DPの計算の第１段：端点フリーDP法

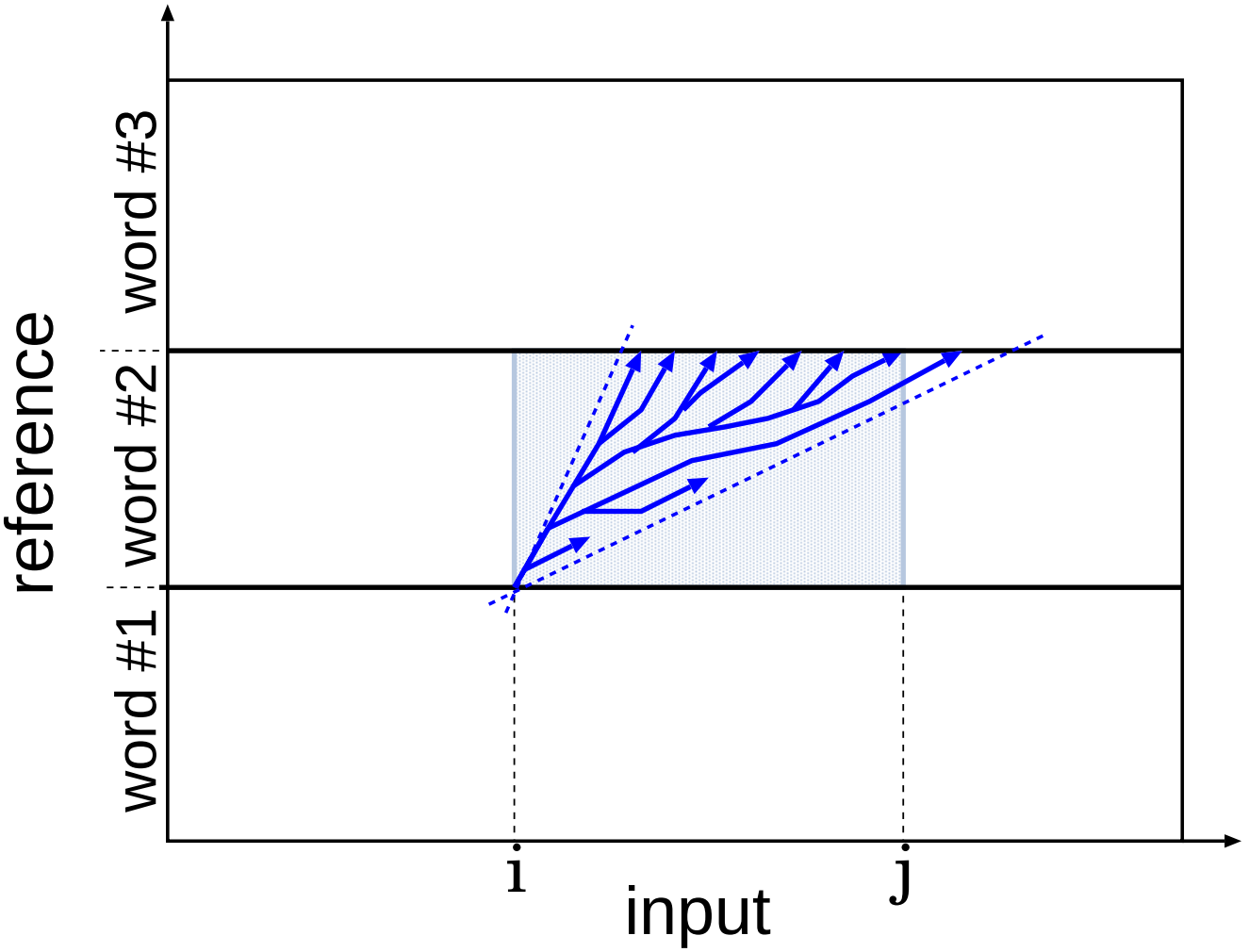


図3. DPにより始端iごとに終端jまでと単語kとの距離$d(i,j,k)$を幅のあるjに対して効率良く計算し、同一区間(i,j)ごとに最小にする単語kとともに記憶する。

二段DPの計算の第２段

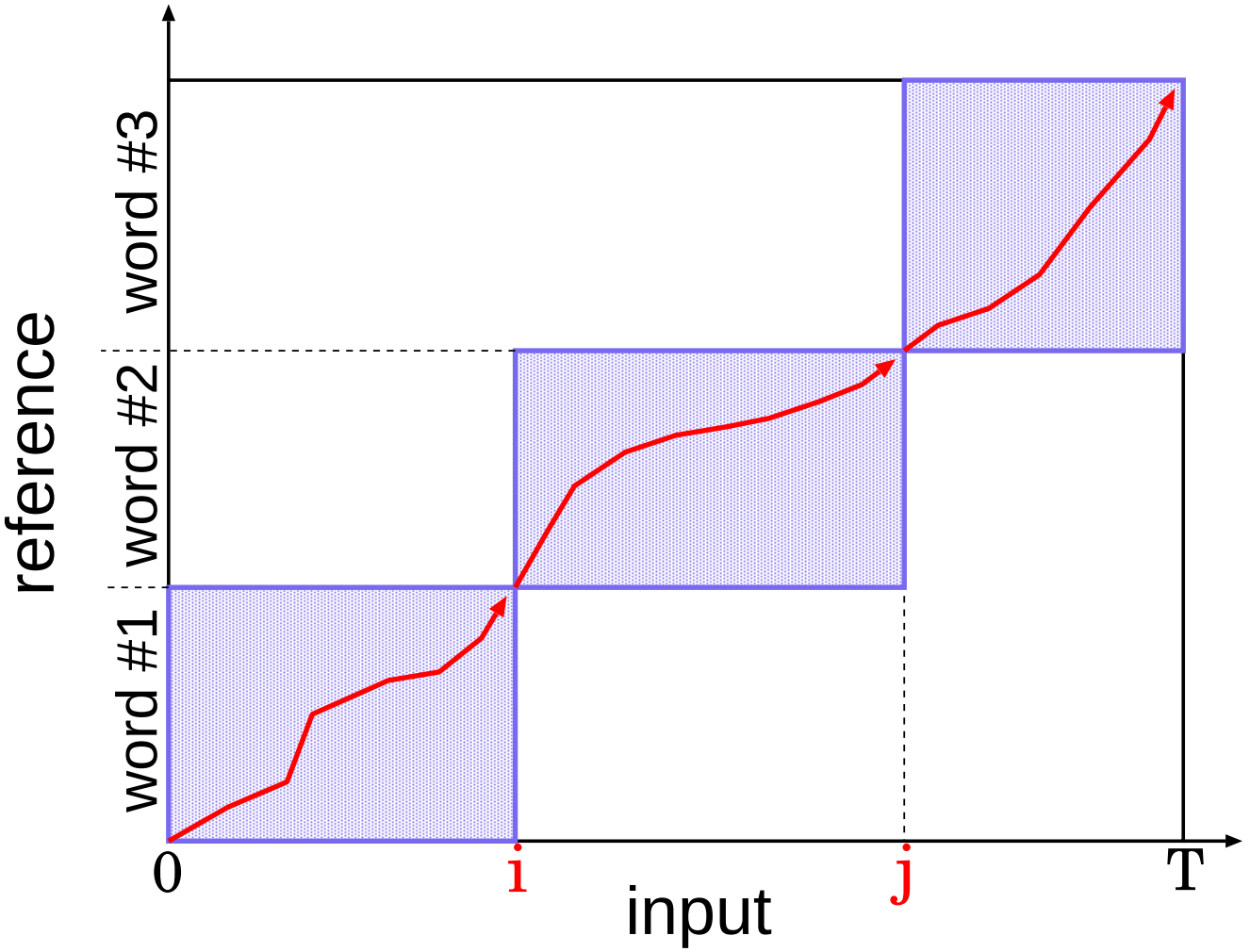


図4. 2段DP法の概念 – 矩形のDP計算領域をつなげる。$d(0,i) + d(i,j) + d(j,T)$ を最小にする最適な接続点 i, j をDPにより求める。経路を traceback して連続単語認識。

2段DP法

1. まず入力音声の任意の始点以後と各単語標準パターンとの終端自由の片端点フリーDPマッチングを行う。
2. 入力音声の時点sから時点tまでとの距離が最小となる単語を$\omega(s,t)$，その距離を$D(s,t)$とし、記憶する。
   ($1 \leq s < t \leq T$, Tは入力音声の長さ)
3. 入力音声全体における累積距離が最小となる単語系列を求める．

$$D_{Td} = \min_{\{m_j\},k} \{D(1,m_1) + D(m_1+1,m_2) + \cdots + D(m_{k+1},T)\}$$

$$(1 \leq m_1 < m_2 < \cdots < m_k < T)$$

上式を満足する単語系列, $w(1,m_1), w(m_1+1,m_2), \cdots, w(m_k+1,T)$が認識結果の単語系列である。この最小化問題を動的計画法で効率よく計算する。

$$D_0 = 0$$

$$D_n = \min_{m=1,\cdots,n} \{D_{m-1} + D(m,n)\}$$

Level Building 法: 両端点フリーDP法の概念

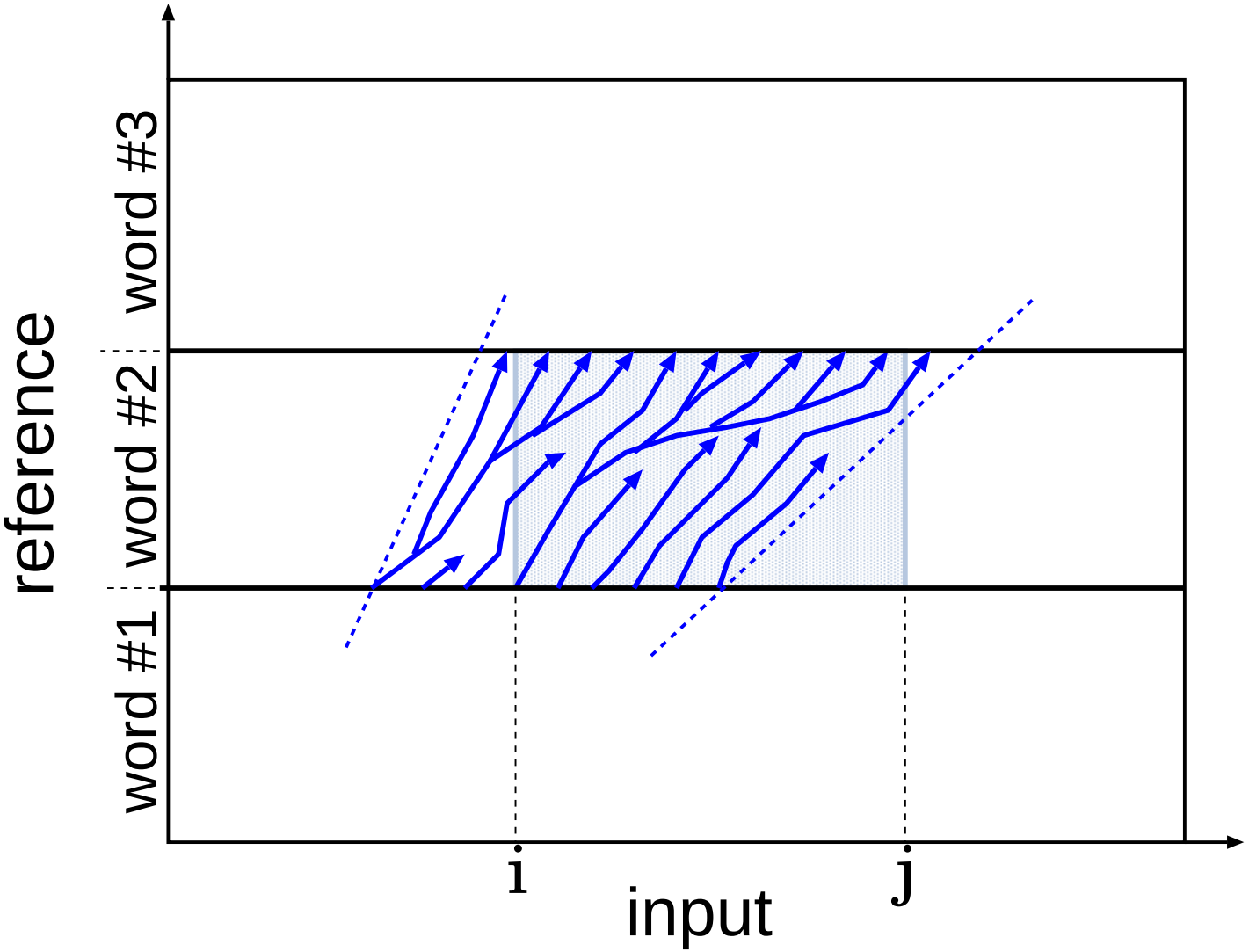
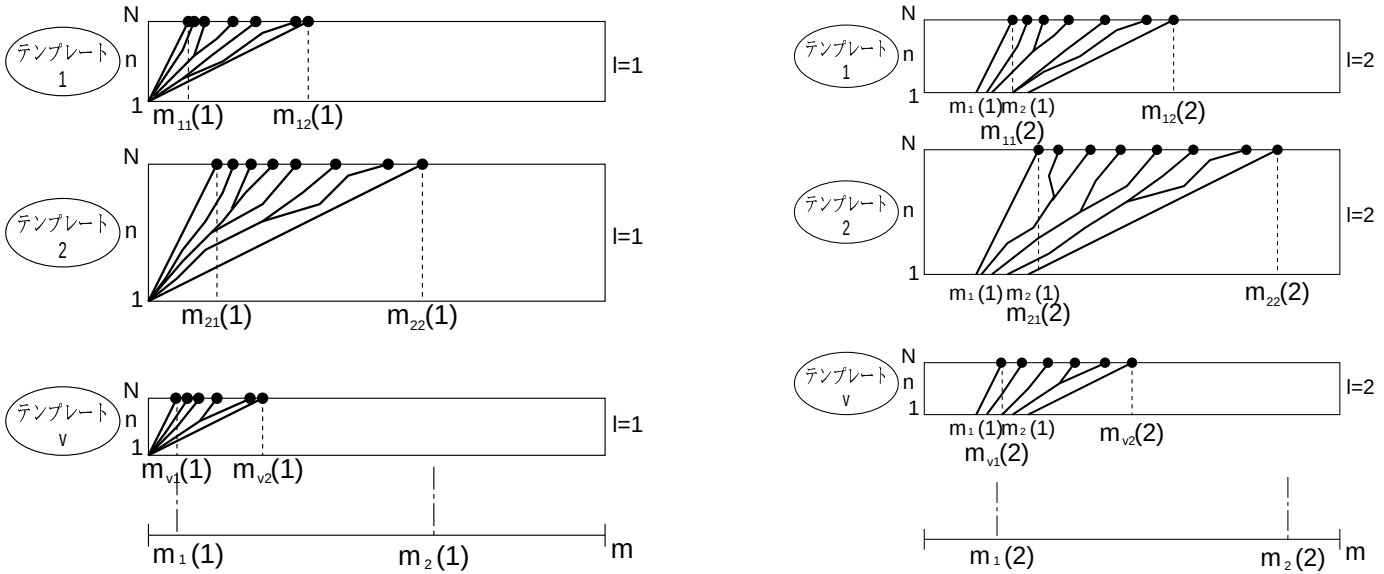


図5. 端点フリーDP – 始終端を固定せず、ある区間について最適経路が求まる。

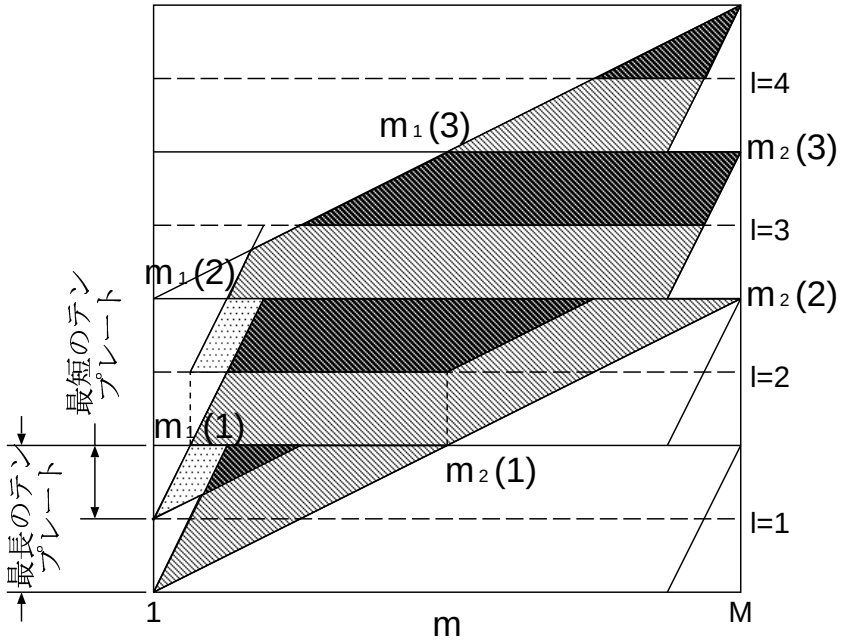
Level-Building法

主に連続数字認識に用いられた (Bell Labs: C. Myers & L. Rabiner)

1. まず1桁目(first level)で入力音声の開始時点を始点とし，すべての単語テンプレートについて終端自由のDPマッチングを行う．
2. 2桁目(second level)以降は，前の桁で得られた終端点までの累積距離を初期値として両端点のDPマッチングにより，累積距離を求める．
3. これをくり返して許される最大の桁数までの累積距離の中で最小となる単語系列が認識結果となる．



Level-Building法

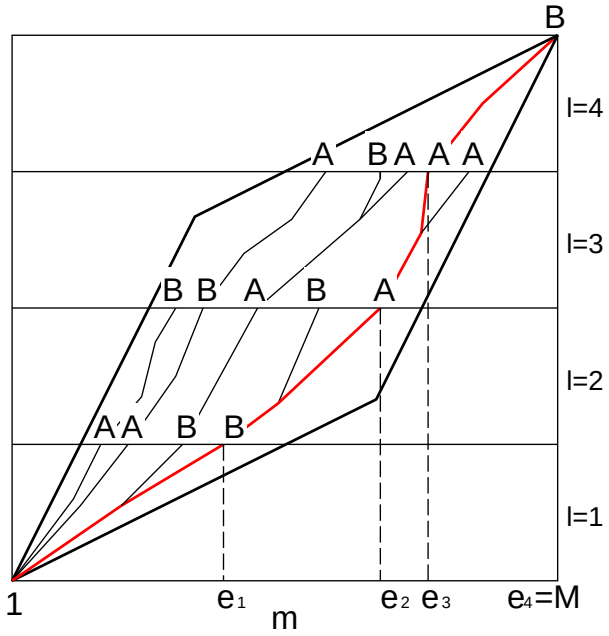


経路が最終到達点に達した後のバックトラック処理により，最適な単語系列を得ることができる．バックトラック処理に必要な情報は，

1. バックポインタ
2. 単語(テンプレート)

◆取り得る経路の領域

左図は，4-level の Level-Building 法の，最長のテンプレートに対する DP 領域と最短のテンプレートに対する DP 領域を表す．



Level Building 法: 複数の単語の接続

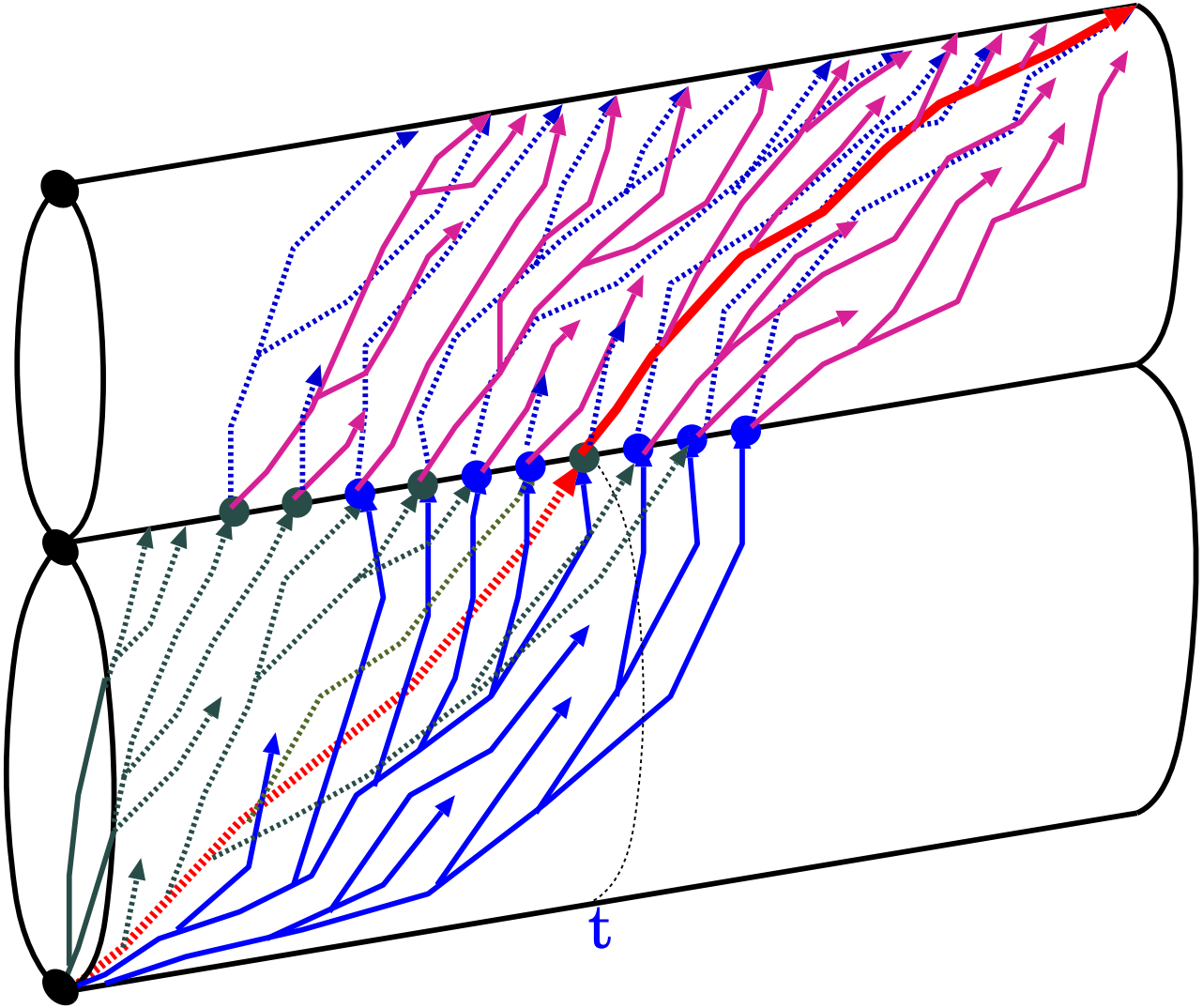


図6. 接続点で距離累積値が最も小さい単語のみ残す。その上に次のレベルを両端点フリーDPにより建増しする。

One-pass DP 法: 複数の単語の接続

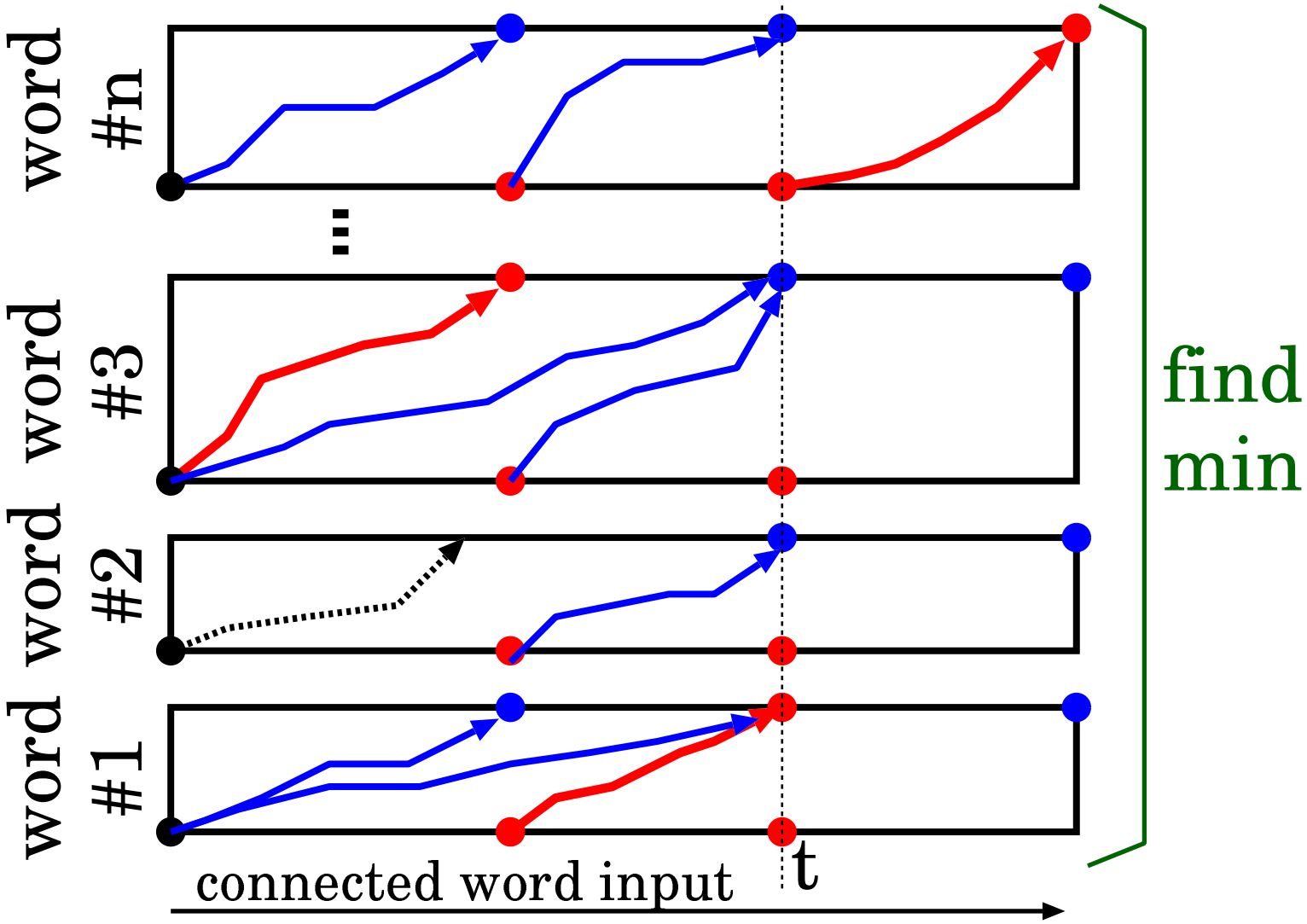


図7. 接続点で距離累積値が最も小さい単語のみ残す。

One-Pass DP法

i:入力音声の時間フレーム
k:テンプレート
j:テンプレートの時間フレーム
(テンプレートkのフレーム長は$J(k)$)

経路Wを以下のようにおく．lは経路番号を表す．

$$\boldsymbol{W} = (w(1), w(2), \cdots, w(l), \cdots, w(L))$$

ただし，$w(l)$は格子点$(i(l), j(l), k(l))$を表すものとする．
連続単語音声認識は，以下のような最小化問題となる．

$$\min_{\boldsymbol{W}} \sum_{l} d(w(l))$$

One-Pass DP法

■ テンプレート内でのパス選択

$$w(l-1) \in \{(i-1,j,k),(i-1,j-1,k),(i,j-1,k)\}$$

$$D(i,j,k) = d(i,j,k) + \min\{D(i-1,j,k), D(i-1,j-1,k), D(i,j-1,k)\}$$

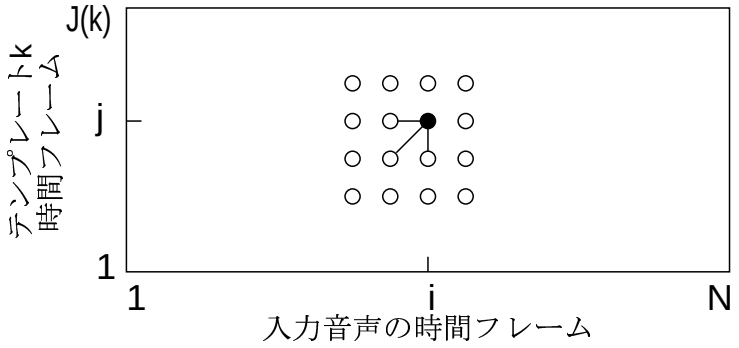


図8. テンプレート内での最適パス選択

■ テンプレート間でのパス選択

$$w(l-1) \in \{(i-1,1,k); (i-1,J(k^*),k^*) : k^* = 1,\cdots,K\}$$

$$D(i,1,k) = d(i,1,k) + \min\{D(i-1,1,k), D(i-1,J(k^*),k^*) : k^* = 1,\cdots,K\}$$

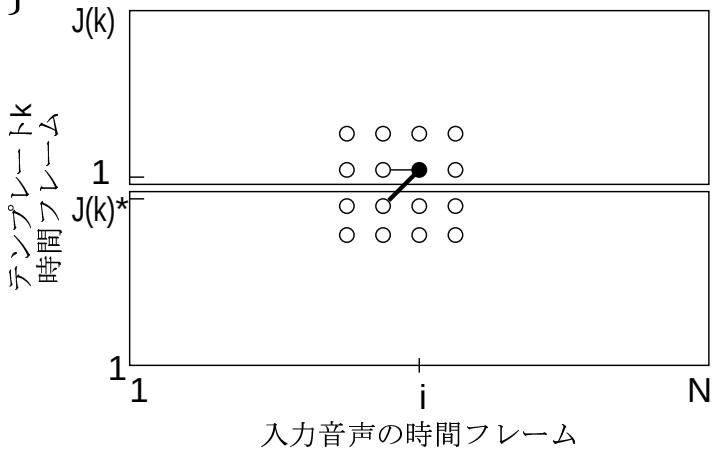


図9. テンプレート間での最適パス選択

One-Pass DP法

One-Pass DP法を用いた連続音声認識アルゴリズム

- **初期値** $D(1,j,k) = \sum_{n=1}^{j} d(1,n,k)$
- **累積距離の逐次計算**
  1. $i = 2, \cdots, N$ について
     2〜5を繰り返す
  2. $k = 1, \cdots, K$ について
     3〜5を繰り返す
  3. $D(i,1,k)$: テンプレート間パスによる累積距離
  4. $j = 2, \cdots, J(k)$ について
     5を繰り返す
  5. $D(i,j,k)$: テンプレート内パスによる累積距離
- **バックトラック処理**
  - バックポインタ
  - 単語(テンプレート)

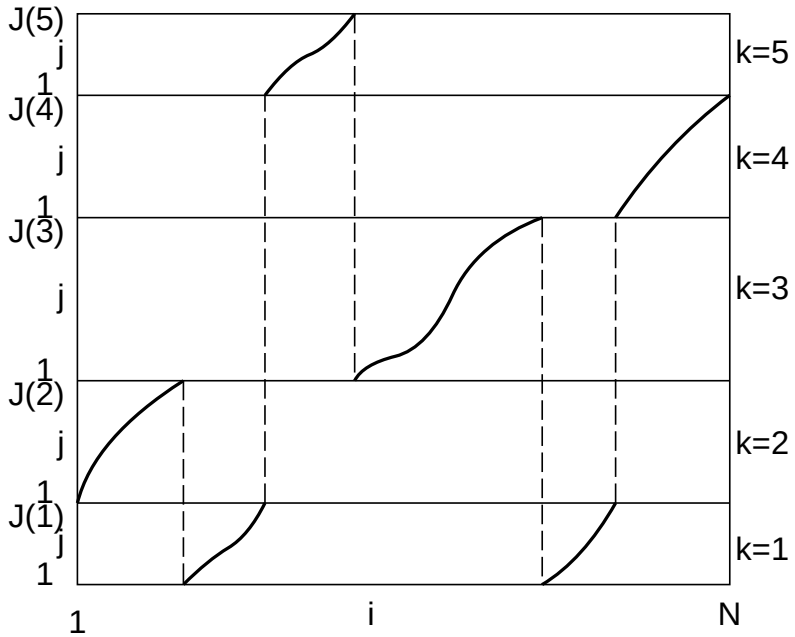


図10. One-Pass DP法の概念図

One-Pass DP法: Viterbiアルゴリズムと比較

- テンプレートの時間フレーム → 状態
- 入力音声の時間フレーム → 状態遷移の回数
- 累積距離 → 対数尤度

上のように置き換えると，Viterbiアルゴリズムとなる．Viterbiアルゴリズムでは，前向き確率$\alpha(q,t)$の最大値を以下のように求める．

$$\alpha(q,t) = \max_j \{\alpha(p,t-1)a_{pq}b_q(y_t)\}$$

ただし，a_{pq}は状態pからqへの状態遷移確率，$b_q(y_t)$は状態qにおいてy_tを出力する確率である．上式を対数尤度を用いると以下のようになる．

$$\log\alpha(p,t) = \log b_q(y_t) + \max_j \{\log\alpha(p,t-1) + \log a_{pq}\}$$

One-Pass DP法のテンプレート内での累積距離は以下の式で求められる

$$D(i,j,k) = d(i,j,k) + \min\{D(i-1,j,k), D(i-1,j-1,k), D(i,j-1,k)\}$$

One-Pass DP法: Viterbiアルゴリズムと比較

両者において，

- $\log b_q(y_t)$ **と** $d(i,j,k)$
- $\{\log \alpha(p,t-1)+\log a_{pq}\}$ **と** $\{D(i-1,j,k), D(i-1,j-1,k), D(i,j-1,k)\}$

が対応していることになる．

※また，テンプレート間での遷移則は言語モデルに対応する．

二段DP法 / LB法 / One-Pass DP法 の比較

- 一般に、二段DPは計算量が多い。一段目ですべての単語、すべての始点からDP計算。
- Level-Building と One-pass DP は、繰り返しループの内外が異なるだけ。
- Level-Building は、発声が終わらなければ行えない。One-pass DP は実時間で演算が進められる。
- One-pass DP は、同一語彙の任意語数連結の場合にさらに効率が高いが、その場合は語数制御ができない。
- One-pass DP は HMM の Viterbi アルゴリズムと同等。