※本資料は課題の理解を助けるための資料であり，
厳密な説明ではない部分がある．

優先発展課題 ♫

第5回
音声処理実験 SP-3

最適経路探索を応用した単語音声認識

優先発展課題 ♫

SP-3 目次

優先発展課題 ♫

5-1
時系列データに適した距離計算方法の検討

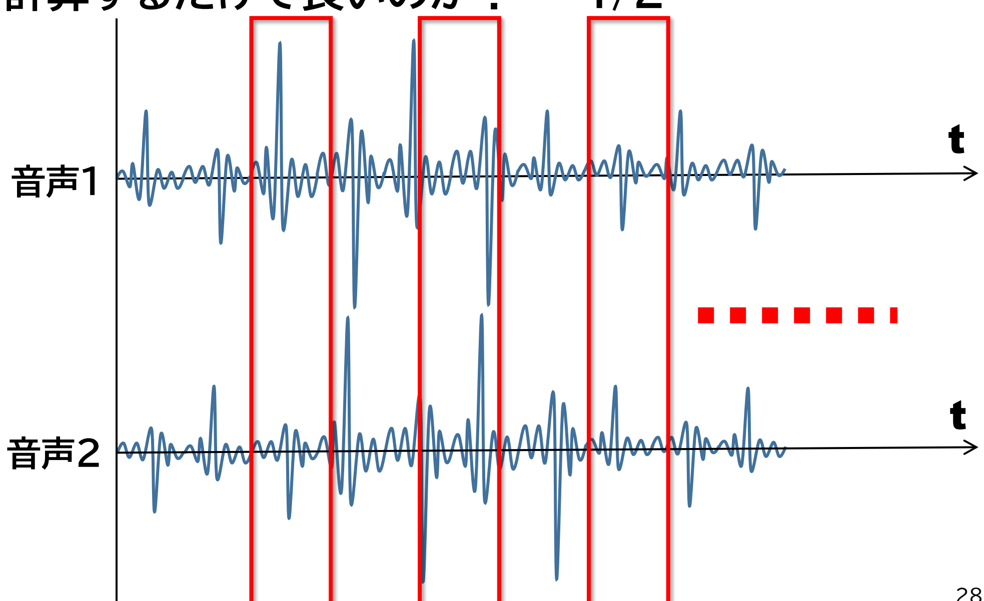
5-1
5-2
5-3

27

27

♫

フレームごとにスペクトル距離を計算するだけで良いのか？ … 1/2



28

28

フレームごとにスペクトル距離を計算するだけで良いのか？ … 2/2

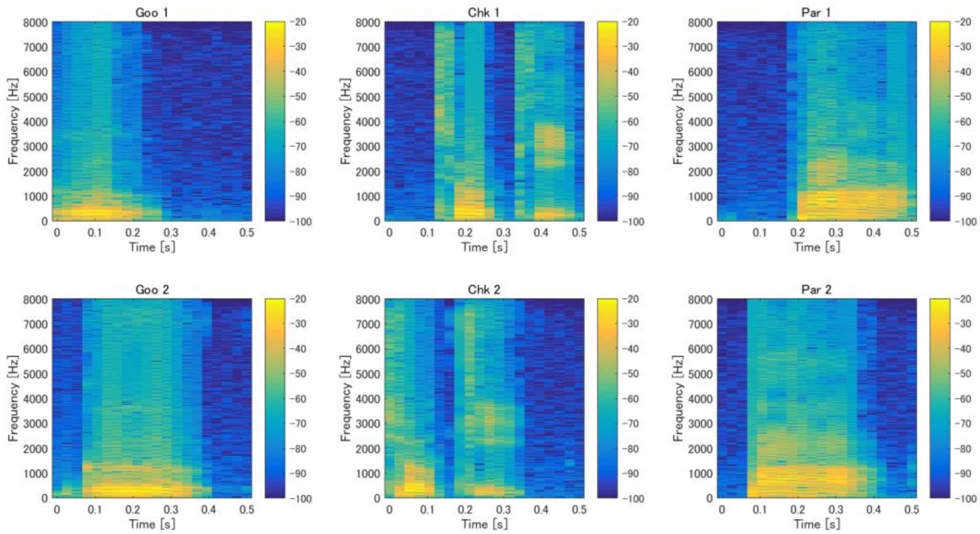


Figure: じゃんけん音声のスペクトログラム

29

29

ずれているものをどう比較する？

<u>案1</u>　統計量を代表値として比較する

- 例：合計値, 平均値, 分散, etc.

➔ 波形全体のパワースペクトルを用いる手法と本質的には変わらないのでは？

<u>案2</u>　ずれを直して比較する

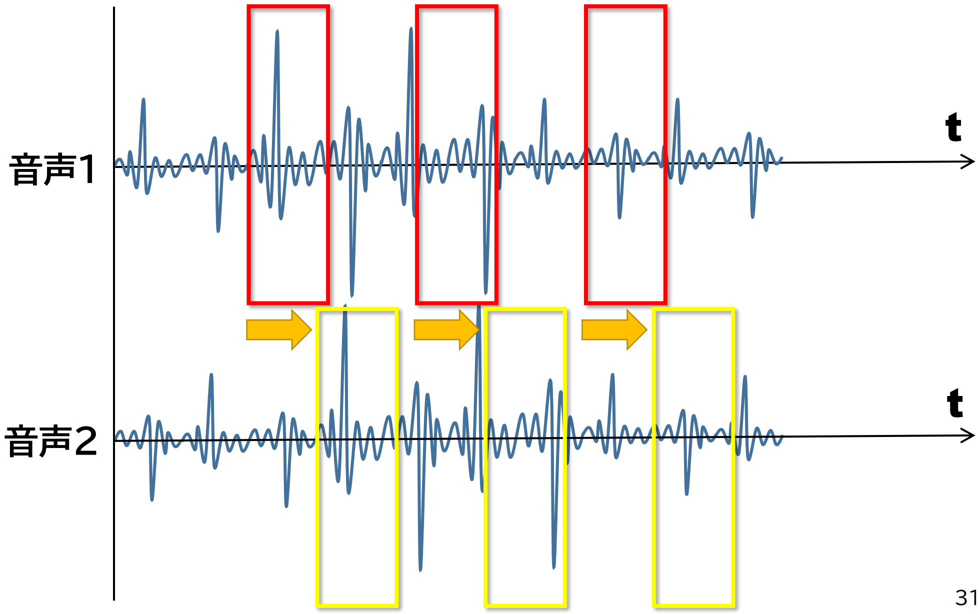
- 例：画像処理の移動・回転・拡大（アフィン変換）

➔音声信号も線形変換（写像）をすれば良いのか？

30

30

ずれを直す(案2):一律で良いのか？



31

31

参考

例題5-1:発話速度の影響

■ ゆっくりしゃべった場合と早くしゃべった場合で，スペクトログラムはどのように変化するか？
- 第3回(SP-1)の日報でおこなっているかもしれない
- 実施していない場合は，実際に録音して観察し，その違いについて考察してみるとよい

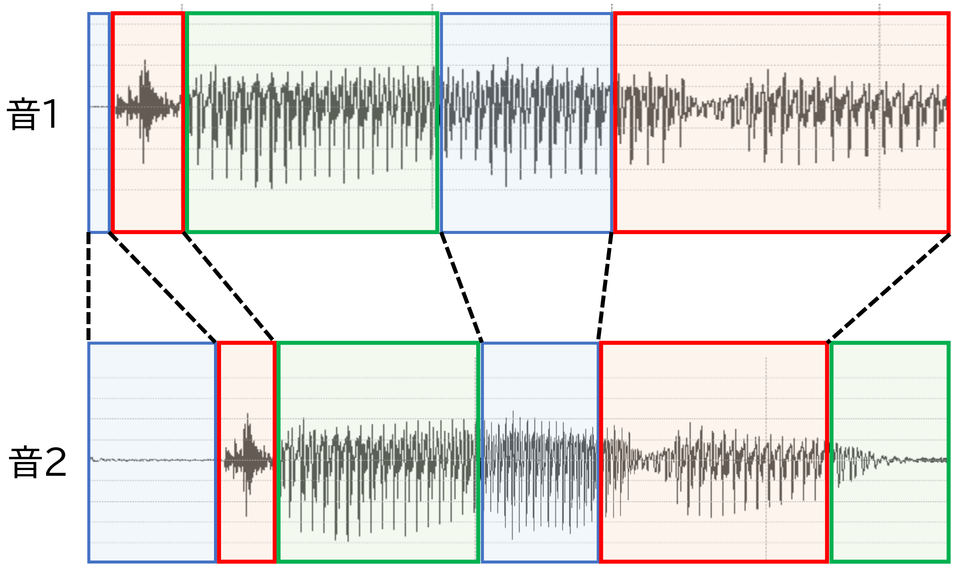
■ 例えば，
- 「かわいい」
- 「かーわーいーいー」
- 「かわいー」　　など の違いを見てみる．

32

32

時系列信号の非線形伸縮 … 1/2

※この説明では時間波形を示しているが，実際はスペクトログラムを伸縮する

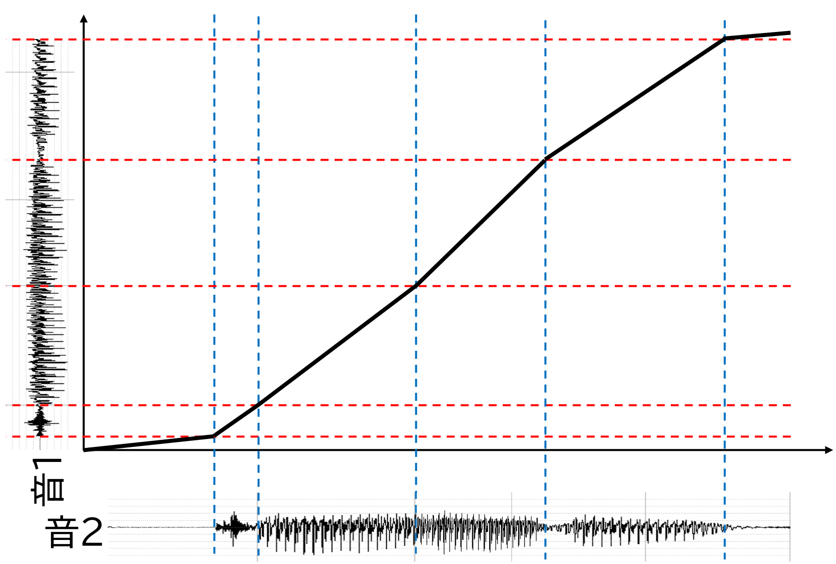


33

33

時系列信号の非線形伸縮 … 2/2

音2から音1への写像(関数)

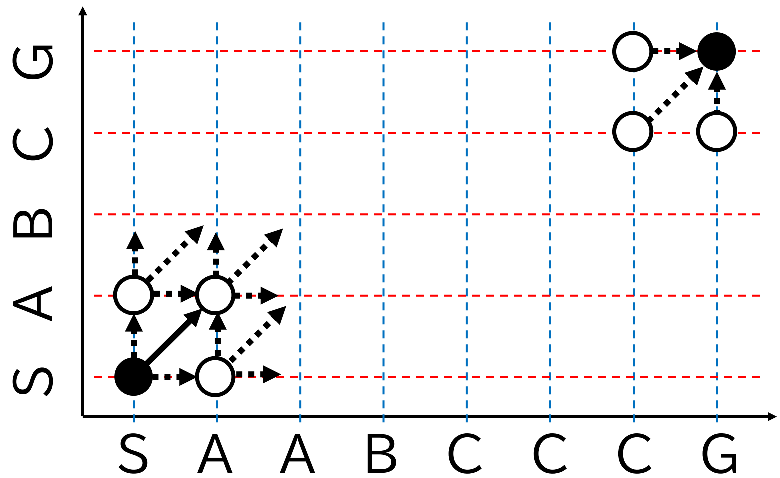


34

34

例題5-2: 非線形伸縮関数の作成

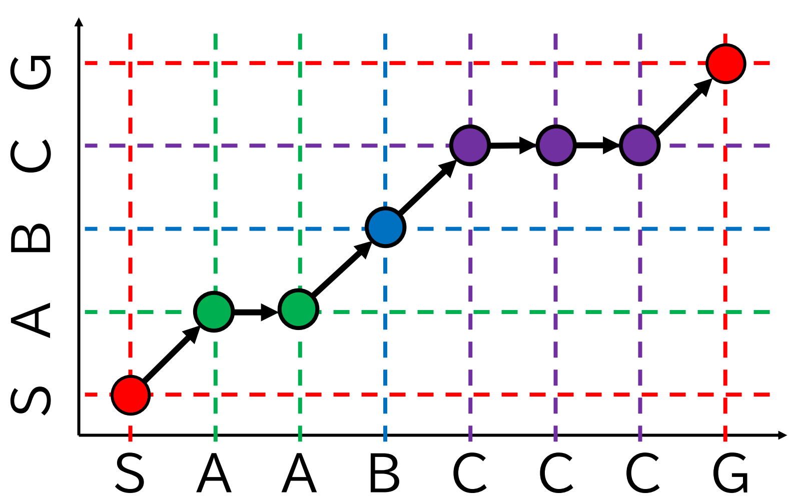
コスト最小で左下(S,S)から右上(G,G)を線で結びなさい.
- ある点からは「右」「上」「右上」にしか移動できない.
- 異なる文字が交わる点では高いコストを支払う必要がある.



35

例題5-2: 非線形伸縮関数の作成

解答例



36

優先発展課題 ♫

5-2 動的計画法による スペクトル距離

5-1
5-2
5-3

37

♫

動的計画法によるスペクトル距離 1/2

- 音声 x と音声 z で各々の時間をずらしながら比較するべきフレームを選びつつ，スペクトル距離を計算
  - $\boldsymbol{x}$ の時間軸を基準に対応する $\boldsymbol{z}$ のフレームを合わせていく
- マッチング関数 $\check{m}$:$\boldsymbol{x}$ のフレーム n から $\boldsymbol{z}$ のフレーム m へ写像

$$m = \check{m}(n)$$

- $\boldsymbol{x}$ のフレーム n と $\boldsymbol{z}$ のフレーム m の間のスペクトル距離

$$d(n, m) \geq 0, \text{where, } 1 \leq n \leq N, 1 \leq m \leq M$$

- **2音声間の距離：対応後の距離の総和の最小値**

$$D(\boldsymbol{x} \,||\, \boldsymbol{z}) \triangleq \min_{\check{m}(\cdot)} \sum_{n=1}^{N} d(n, \check{m}(n))$$

38

動的計画法によるスペクトル距離 2/2

■ 解くべき問題は何か？

➔ マッチング関数 $\check{m}(\cdot)$ に対する最小化が必要

$$\min_{\check{m}(\cdot)} \sum_{n=1}^{N} d(n, \check{m}(n))$$

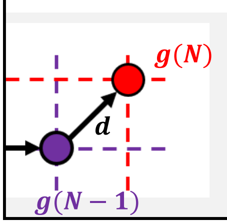
もう少し丁寧に書くと,

$$\min_{\check{m}(1),\ldots,\check{m}(N)} \sum_{n=1}^{N} d(n, \check{m}(n))$$

■ **再帰的な解法**

$$g(N) = \min_{\check{m}(\cdot)} \sum_{n=1}^{N} d(n, \check{m}(n))$$

$$= \min_{\check{m}(1),\ldots,\check{m}(N-1)} g(N-1) + d(n, \check{m}(N))$$



※ 直前の結果が無ければ前をたどっていく

39

39

参考

動的計画法によるマッチング

■ さまざまな定式化が知られている
- Dynamic Time Warping (DTW)
- DPマッチング　※DP: Dynamic Programming (動的計画法)

■ 動的計画法の概要
- 部分問題を解いて, より大きな問題を解く
  - 分割統治法(例:クイックソート)に類似
- 直前の結果を再利用する
  - 直前の結果が自明ではないなら, 直前の結果を計算
  - プログラミングでいえば再帰構造(cf. 再帰はスタック)

※ 参考書「アルゴリズム論」 8.3節 動的計画法

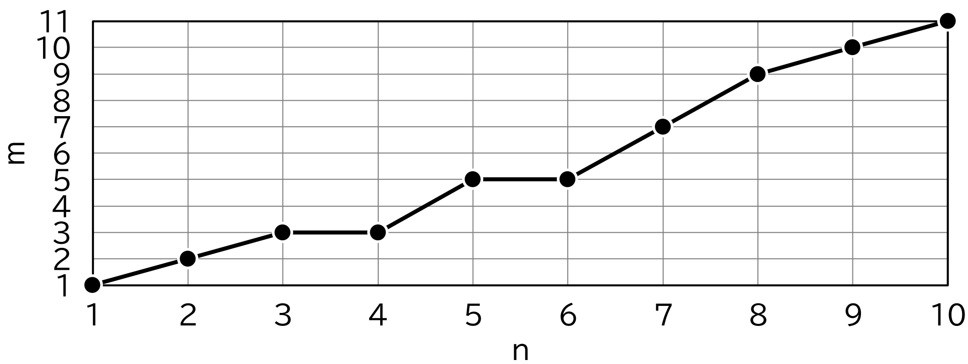
40

40

$m = \breve{m}(n)$の例

音声という“事前知識”を制約として利用

- 逆戻りするような対応は不要 ➔ 単調増加関数で良い
- 極端なスキップは避けたい ➔ 傾斜制限をつける

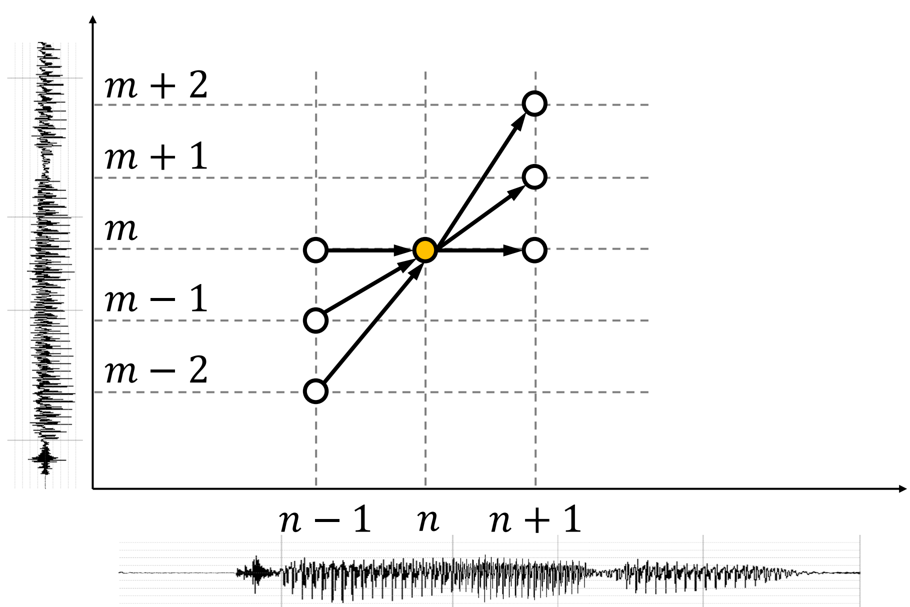


参考『フカシギの数え方』おねえさんといっしょ！ みんなで数えてみよう！
https://youtu.be/Q4gTV4r0zRs

41

41

傾斜制限の例



42

42

優先発展課題 ♫

5-3
マッチング関数の“探索”問題

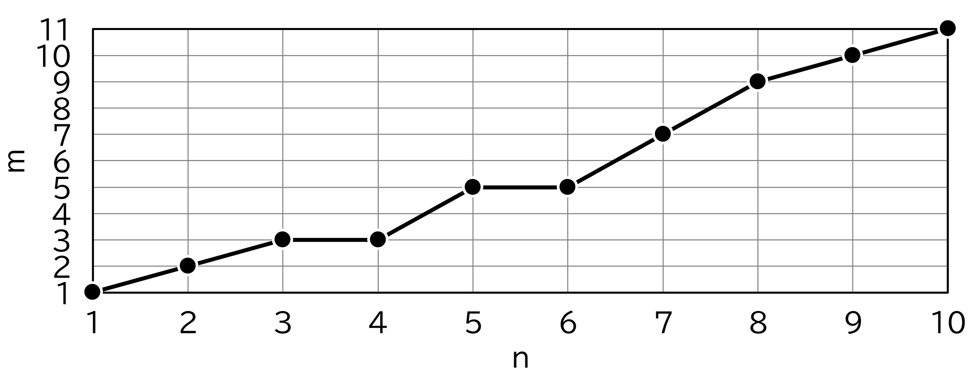
5-1
5-2
5-3

♫

マッチング関数 $m = \breve{m}(n)$ の算出 … 1/3

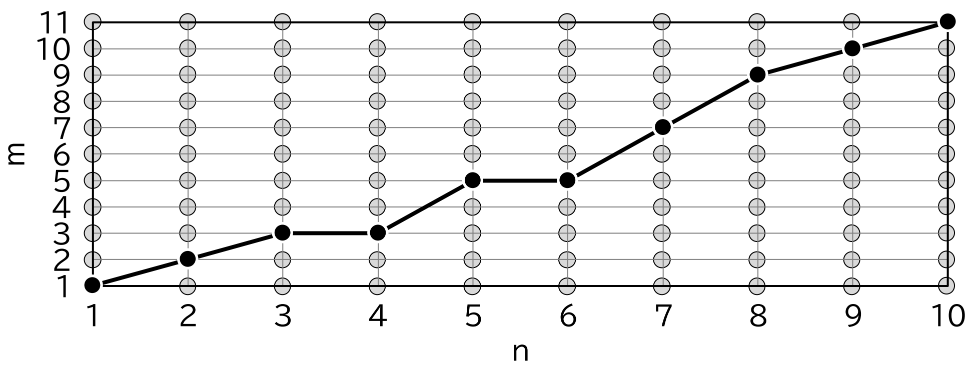
■ 道中の計算を考えてみる

- どこかで見覚えのある図ではないか？



マッチング関数 $m = \breve{m}(n)$の算出 … 2/3

■ 道中の計算を考えてみる
- どこかで見覚えのある図ではないか？

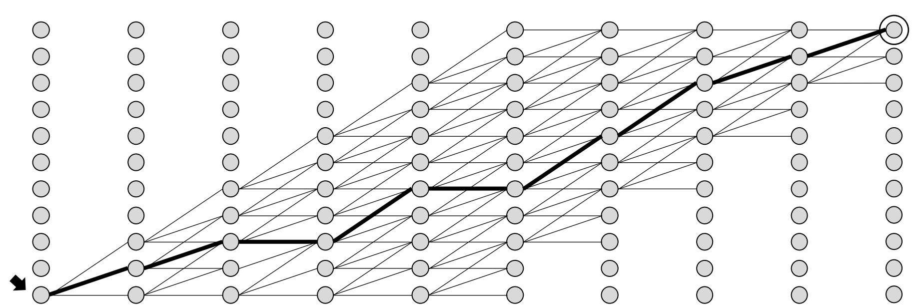


45

45

マッチング関数 $m = \breve{m}(n)$の算出 … 3/3

■ 道中の計算を考えてみる
- どこかで見覚えのある図ではないか？



46

46

マッチング関数の”探索”問題

■ グラフの最適経路探索問題としても解けそう！
- **初期ノード：** 先頭フレーム同士の対応位置とする
- **最終ノード：** 最終フレーム同士の対応位置とする
- **コスト：** 音声フレーム間の二乗距離とする

➔マッチング関数は最適経路そのものでは？

■ 最適経路の累積コストは,
フレーム間の二乗距離の和に相当する(はず)
➔累積コストが2音声の距離として使えそう？

47

47

参考

探索の効率化

■ 典型的な方針はいくつか考えられる

1. アルゴリズムの工夫で無駄な計算を省く
   - 原理的に高速な方法
   - 事前計算による方法
2. 音声としての事前知識をヒューリスティクスやグラフの制約として与える
3. 並列計算で解ける問題に分割する

48

48