

※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。

第4回 音声処理実験 SP-2

距離最小パターン選択による単語音声認識

参考文献:

[荒木 2007] 荒木雅弘, “フリーソフトで作る音声認識システム,” 森北出版, 2007.

1

1

SP-2 目次

4-1
音声認識とパターン認識



3

4-2
音声の基礎知識



11

4-3
パワースペクトルと距離

教科書「応用数学」pp. 49-56



19

2

2

4-1 音声認識とパターン認識



4-1

4-2

4-3

3

3

(狭義の)音声認識 … 1/2

- 音声テキスト(文字列)に変換する



- 情報理論の問題であれば？

- **圧縮**: 符号系列のエントロピーを削減
→ 符号間の辞書があれば**圧縮(認識)**ができるのでは？
- 必要な辞書: 長い符号から短い符号への変換規則

音声信号

特定テキスト

→ “長い符号”の“完全一致”は実現できるの？

4

4

(狭義の)音声認識 … 2/2

(アプローチ)辞書の逆引きによる音声認識

1. 入力の長い符号 \tilde{y} と“類似した”長い符号 y を辞書から探す
2. 辞書から長い符号 y に変換できる短い符号 x を選ぶ
 → 音声信号 → テキスト

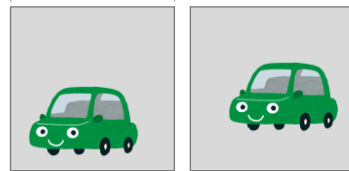
■ “類似した(似ている)”とは？

- ・ 2者の間の距離が短い

■ 音と音の“距離”とは？

- ・ 画像なら2枚の画像で画素値を比較すれば距離が測れた(ホントに？ [右図])

縦横 W ピクセル



2枚の距離は？

第4回 (SP-2) → 特徴量に基づく音の距離の計算

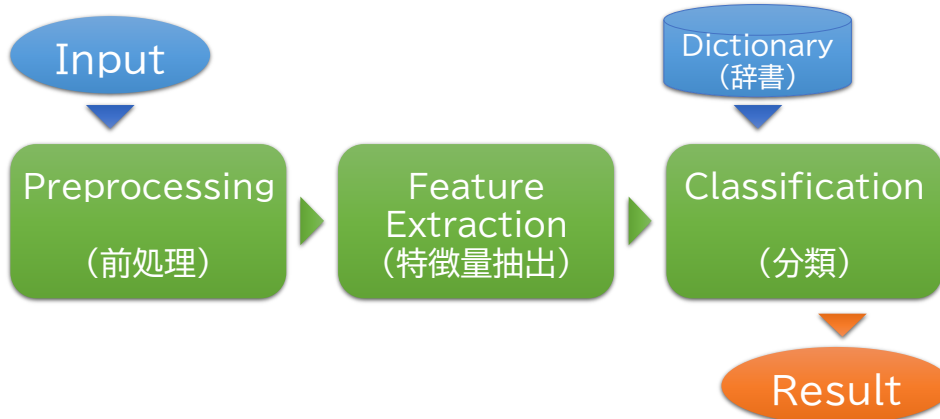
第5回 (SP-3) → 時間変動に頑健な音の距離の計算

5

5

パターン認識 Pattern Recognition

- 入力パターンが属するクラスを見つける問題
- 分類器 (Classifier) は「分ける」ことしかできない
 - ・ 分類した結果に何らかの意味を与えて「認識」

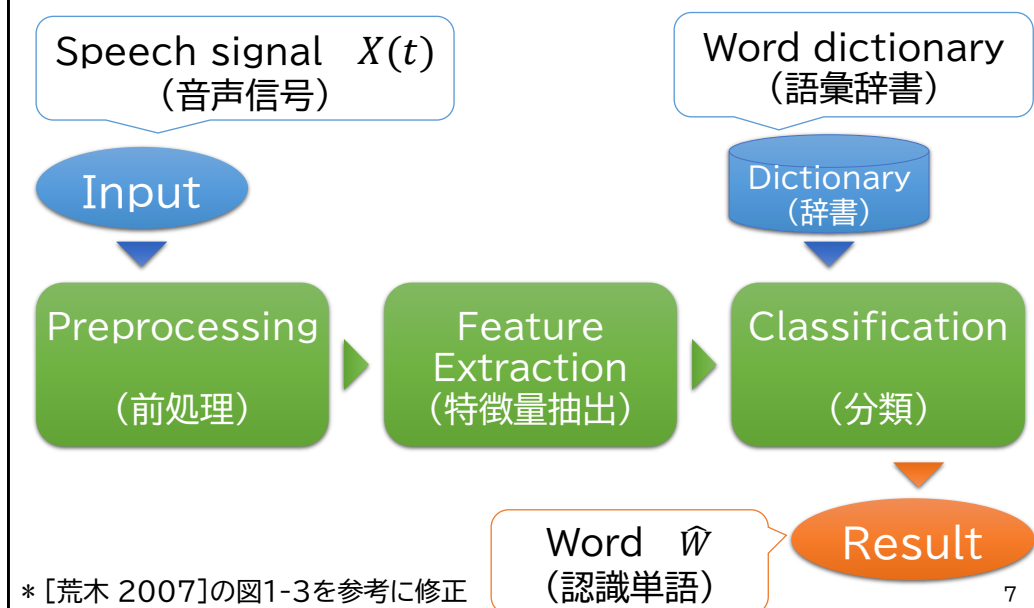


* [荒木 2007]の図1-3を参考に修正

6

6

音声認識(単語音声認識の場合)



7

本講義の実装方針

■ 前処理 (Preprocessing)

- 本講義では考えないことにする
- 一般には雑音除去などをして「きれい」な音にする

■ 特徴量抽出 (Feature extraction)

- SP-2: パワースペクトル(SP-1/問題2)を用いる
- SP-3: スペクトログラム(SP-1/問題3)を用いる

■ 分類 (Classification)

- 最近傍法(Nearest Neighbor)を用いる

8

8

最近傍法 Nearest Neighbor

- クラスごとに参照パターンを1つ以上用意する

$$\mathcal{C}(z_1) = \text{○} \quad \mathcal{C}(z_2) = \text{□} \quad \mathcal{C}(z_3) = \text{○} \quad \mathcal{C}(z_4) = \text{△} \quad \dots$$

- 入力データ x に最も近い参照パターン z_k が属するクラスを, 入力データのクラスとみなす

$$\mathcal{C}(x) = \mathcal{C} \left(\arg \min_{z_k} D(x, z_k) \right)$$

- 近い → 距離を計算する必要がある！
- 最も近い → **全部**調べる必要がある？ → “探索”

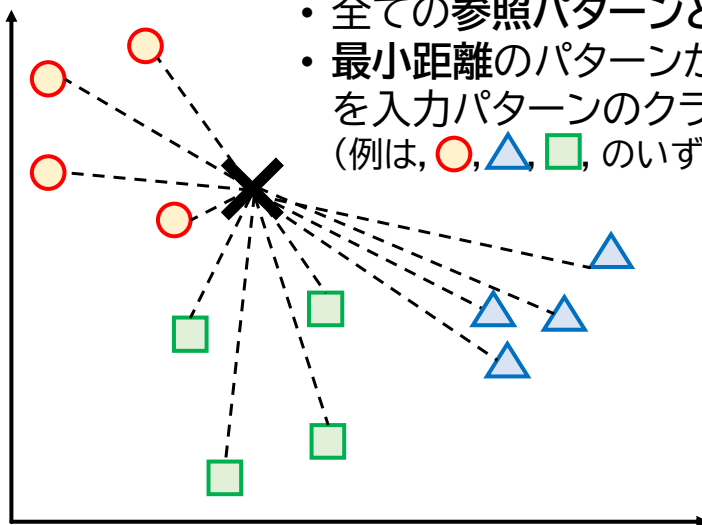
9

9

例題4-1: Nearest Neighbor

全探索による解法

- 全ての参照パターンとの距離を計算
- 最小距離のパターンが属するクラスを入力パターンのクラスと判定
(例は, ○, △, □, のいずれか)



発展的考察の
参考: k-NN

10

10

4-2 音声の基礎知識



4-1

4-2

4-3

11

11

音声の構成単位

■ **音節**: ひらがなの単位とほぼ同じ

- 例: かわいい → 「か」「わ」「い」「い」
- 例: しよっぱい → 「しよっ」「ぱ」「い」

■ **音素**: ローマ字のアルファベットとほぼ同じ

- アルファベットと区別するために / / で囲って表記する
- 例: かわいい → /k/, /a/, /w/, /a/, /i:/
- 例: しよっぱい → /sh/, /o/, /Q/, /p/, /a/, /i/

■ **モーラ**: 俳句で考えるような拍とほぼ同じ

- 例: かわいい → 「か」「わ」「い」「い」
- 例: しよっぱい → 「しよ」「っ」「ぱ」「い」

12

12

参考

本講義での音素の表記

■ 基本的にヘボン式のローマ字で書く

• 忘れがちな例

- し → /sh/ /i/ ち → /ch/ /i/ じ(ち) → /j/ /i/
- つ → /ts/ /u/ ふ → /f/ /u/ (ずい → /z/ /i/)

■ 特殊な音

• 促音(そくおん)

- ずごっく → /z/ /u/ /g/ /o/ /Q/ /k/ /u/

• 撥音(はつおん)

- ぎゃん → /gy/ /a/ /N/

• 拗音(ようおん)

→ 子音にくっつける(例: /gy/)

• 長音(ちょうおん)

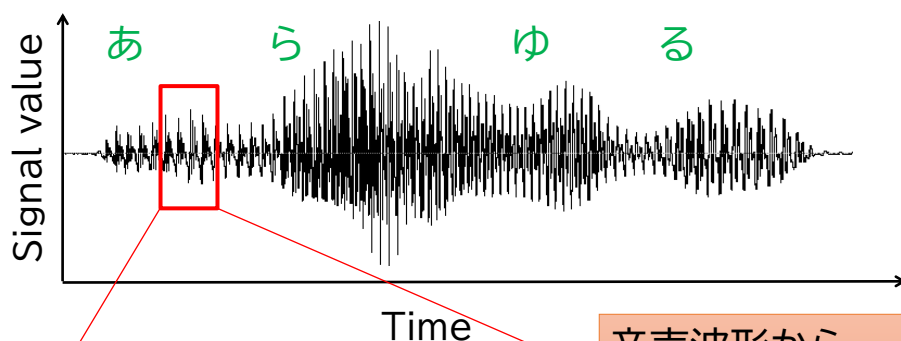
→ 母音にくっつける(長母音)

- ベーーす → /b/ /e:/ /s/ /u/

13

13

音声波形の時間的特徴 … 1/2

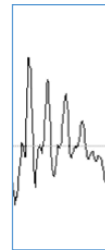
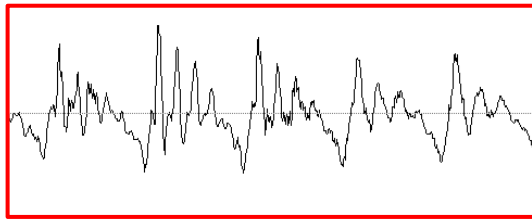


音声波形から
一定の長さの波形を
抜き出して拡大

14

14

音声波形の時間的特徴 … 2/2



■特徴的な波形の繰り返しが現れる

(例: 左上の波形は右上の波形を5個つなげたような波形)

- ① 形状 … 「音素」の種類に相当
- ② 繰り返し回数 … 「音素」を発声する長さに相当
- ③ 長さ(幅) … 基本周波数(声の高さ)に相当

15

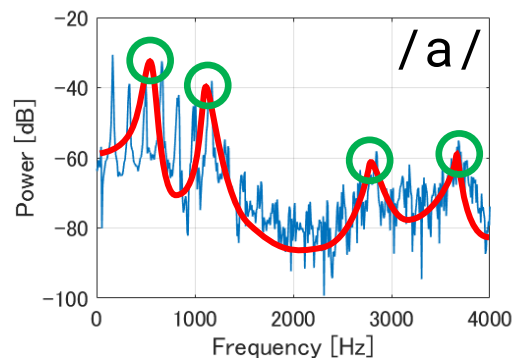
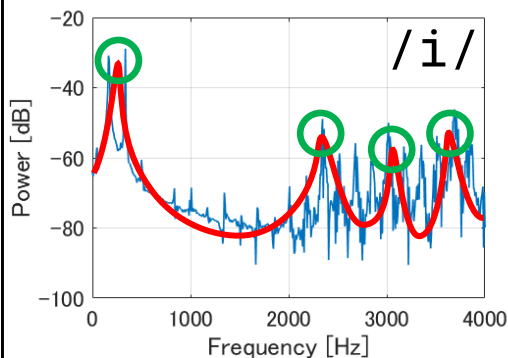
15

母音の対数パワースペクトル

※ 微細構造は、
青線から赤線
を引いた成分

■微細構造 → 幅が「声の高さ」により変化

■包絡 → 概形が「音素」により変化



参考: 包絡のピークを**フォルマント**と呼ぶ
低域から順に第1フォルマント(F_1), 第2フォルマント(F_2), …

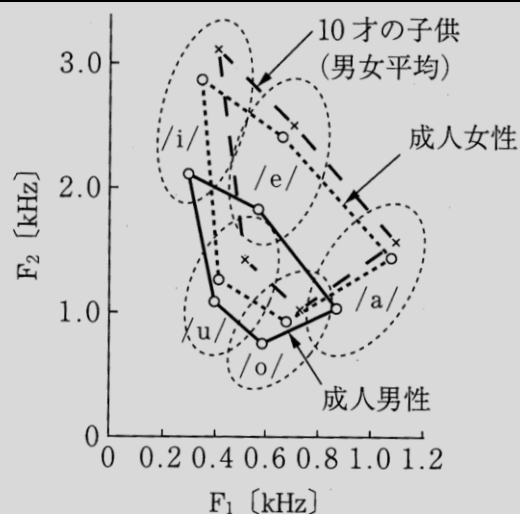
16

16

参考

母音の知覚*

- 第1フォルマントと第2フォルマントの周波数を散布図として作図
- 日本語の5母音では右図のような5角形になる



日本語の母音ホルマント Nakagawa (1980)

図 4.13 母音の知覚と F_1 , F_2

* 鈴木陽一他, “音響学入門,” 図4.13, p. 96, コロナ社, 2011.

17

17

※これは、ページ数合わせの白紙ページです。

18

18

4-3 パワースペクトルと距離

教科書「応用数学」 pp. 49-56



4-1

4-2

4-3

19

19

パワースペクトルと距離

- パワースペクトル推定をおこない、2つの波形のパワースペクトル間の誤差の二乗和(の根)を計算

- 誤差の二乗和

$$D_{euc}(\mathbf{u}, \mathbf{v})^2 = \sum_i (u_i - v_i)^2$$

- 誤差の二乗和の根

$$D_{euc}(\mathbf{u}, \mathbf{v}) = \sqrt{D_{euc}(\mathbf{u}, \mathbf{v})^2}$$

※ 本講義では通常のパワースペクトルの誤差ではなく、
対数パワースペクトルの誤差を用いることにする。

20

20

参考

距離とユークリッドノルム

■ ベクトル \mathbf{u}, \mathbf{v} の内積教科書 2.2節
pp. 49 - 50

$$(\mathbf{u}, \mathbf{v}) = \mathbf{u}^\top \mathbf{v} = \sum_i u_i v_i \quad (2.65)$$

■ ベクトル \mathbf{u} のユークリッドノルム

$$\|\mathbf{u}\| = \sqrt{(\mathbf{u}, \mathbf{u})} = \sqrt{u_1^2 + u_2^2 + \cdots} \quad (2.66)$$

■ ベクトル \mathbf{u}, \mathbf{v} の差のユークリッドノルム

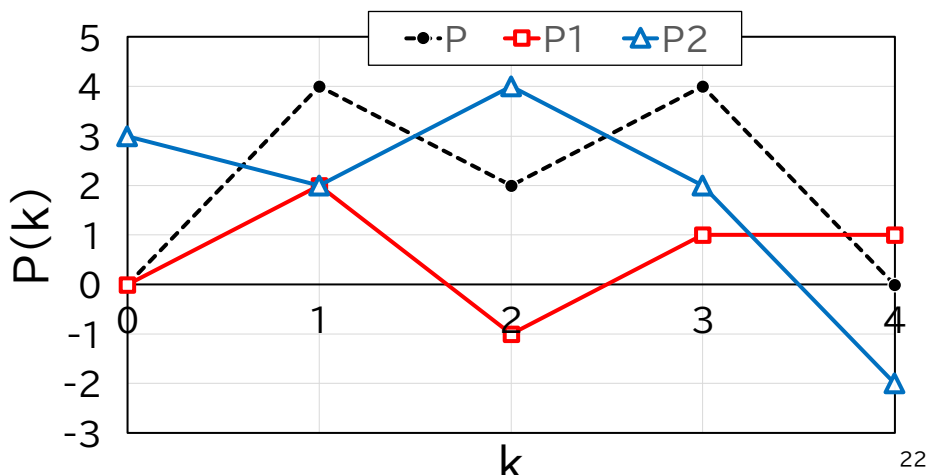
$$\begin{aligned} \|\mathbf{u} - \mathbf{v}\| &= \sqrt{(\mathbf{u} - \mathbf{v}, \mathbf{u} - \mathbf{v})} \\ &= \sqrt{(u_1 - v_1)^2 + (u_2 - v_2)^2 + \cdots} \end{aligned}$$

21

例題4-2: 距離の計算例

以下の距離として誤差の二乗和の根を計算せよ.

1. P1とPの間の距離 2. P2とPの間の距離



22

例題4-2:距離の計算例(解答例)

$$\begin{aligned} D_{euc}(\mathbf{P}_1, \mathbf{P})^2 &= (0 - 0)^2 + (2 - 4)^2 + (-1 - 2)^2 \\ &\quad + (1 - 4)^2 + (1 - 0)^2 \\ &= 23 \end{aligned}$$

$$\begin{aligned} D_{euc}(\mathbf{P}_2, \mathbf{P})^2 &= (3 - 0)^2 + (2 - 4)^2 + (4 - 2)^2 \\ &\quad + (2 - 4)^2 + (-2 - 0)^2 \\ &= 25 \end{aligned}$$

$$\therefore D_{euc}(\mathbf{P}_1, \mathbf{P}) = \sqrt{23} \quad D_{euc}(\mathbf{P}_2, \mathbf{P}) = \sqrt{25}$$

23

23

※これは、ページ数合わせの白紙ページです。

24

24