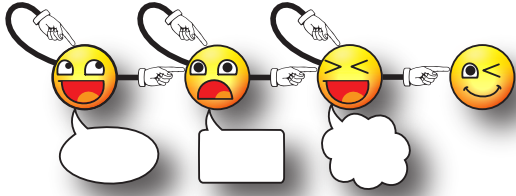


Spoken Word Recognition

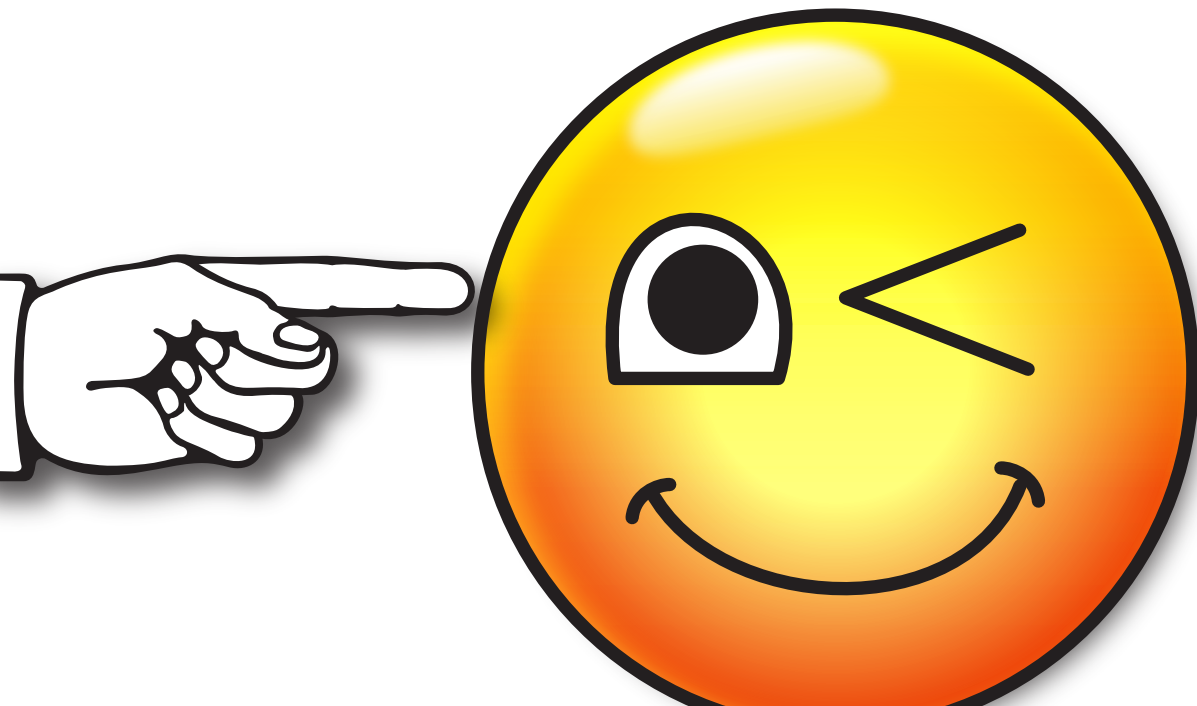


Takagi Kazuyuki

takagi@uec.ac.jp



10 NOVEMBER 2024



メディア情報学実験
メディア情報検索・認識（音声）
～単語音声認識～

高木一幸 takagi@uec.ac.jp

2024 年／令和 06 年 10 月 10 日版

目次

第 1 章	はじめに	1-1
1.1	音声認識とは	1-1
1.1.1	音声認識の難しさ [28]	1-1
1.1.2	音声認識の歴史	1-2
1.2	単語音声認識とは	1-3
1.3	概要	1-4
1.4	必須条件	1-5
1.5	日程	1-6
1.6	レポート	1-7
1.6.1	形式・期限	1-7
1.6.2	課題一覧	1-7
1.6.3	穴埋め問題について	1-7
1.7	計算機実習の準備	1-8
1.7.1	ファイルのコピー	1-8
1.7.2	環境設定の追加	1-8
1.7.3	シェル環境についての注意	1-8
1.7.4	エディタについての注意	1-8
第 2 章	音声	2-1
2.1	音声生成のしくみ	2-2
2.1.1	音源—フィルタ理論	2-2
2.1.2	声道モデル・母音の生成	2-3
2.1.3	子音の生成	2-5
2.2	日本語の音素	2-7
2.2.1	母音	2-7
2.2.2	子音	2-11
第 3 章	スペクトル分析	3-1
3.1	短時間フレーム分析	3-1
3.2	高域強調	3-1
3.3	窓関数	3-4
3.4	離散フーリエ変換	3-5
3.5	メル周波数	3-5
3.6	フィルタバンク分析	3-5
3.7	メル周波数ケプストラム係数 (MFCC)	3-7

3.8	特徴抽出はデータ圧縮	3-7
3.9	スペクトル分析例	3-9
3.10	実習	3-10
3.10.1	フィルタバンク分析	3-10
3.10.2	MFCC 分析	3-11
3.10.3	音声収録の練習	3-13
3.10.4	レポート (第 1 週)	3-25
第 4 章	統計的音声認識	4-1
4.1	枠組み	4-1
4.2	基本原理	4-2
第 5 章	隠れマルコフモデル (HMM)	5-1
5.1	基本 HMM の定式化	5-1
5.2	HMM の 3 つの問題	5-2
5.3	確率計算法	5-3
5.3.1	前向きパス (Forward) アルゴリズム	5-4
5.3.2	後ろ向きパス (Backward) アルゴリズム	5-6
5.3.3	最適状態系列推定 (Viterbi アルゴリズム)	5-6
5.3.4	パラメータ推定 (Baum-Welch アルゴリズム)	5-8
5.4	音声認識のための HMM	5-10
5.4.1	一方向性 HMM	5-10
5.4.2	連続 HMM	5-11
5.4.3	スケーリング (scaling)	5-14
5.5	学習から認識までの手順	5-14
5.6	実習	5-15
5.6.1	Forward アルゴリズム	5-16
5.6.2	Backward アルゴリズム	5-17
5.6.3	Baum-Welch アルゴリズム	5-18
5.6.4	多次元ガウス確率密度関数	5-25
5.6.5	レポート (第 2 週)	5-26
第 6 章	単語音声認識	6-1
6.1	はじめに	6-1
6.2	単語 HMM	6-2
6.3	単語音声認識実験の概要	6-3
6.3.1	基本構成	6-3
6.3.2	音声区間検出	6-5
6.4	単語 HMM の学習	6-7
6.5	実習	6-8
6.5.1	認識用設定ファイルの作成	6-9
6.5.2	単語音声データの作成	6-10

6.5.3	スペクトル分析	6-21
6.5.4	単語 HMM の学習	6-21
6.5.5	学習の検証	6-24
6.5.6	On-The-Fly 単語音声認識	6-25
6.5.7	プログラム構成	6-30
6.5.8	レポート (第 3 週)	6-31
第 7 章 おわりに		7-1
付 録 A 数字単語音声のスペクトル分析実例		A-1
付 録 B 実習用ファイル		B-1