Data Mining and Machine Learning

Applicati r
Automati gnition –
Introduction

Peter Jančovič

UNIVERSITY OF BIRMINGHAM

Data Mining and Machine Learning

Objectives

- Introduce automatic speech recognition

- Understand why speech recognition is difficult

  – Continuity

  – Variability

  – Confusability

  – Effects of accent

- Speech recognition terminology

Data Mining and Machine Learning

UNIVERSITY OF BIRMINGHAM

Why is speech recognition difficult?

Intuitively...

- Meaning is represented by sentences

- A Sentence i

- A Word is a

- ...

- This view of speech is bas

- But speech is NOT just "*Acoustic text*"

UNIVERSITY OF
BIRMINGHAM

Data Mining and Machine Learning

Speech is....

- Continuous
  - *"We were away a year ago"*

- Variable

  - *"bread an* ”

- Ambiguous

  - *"The grey tape can fix that"*
  - *"The great ape can fix that leek"*
  - *"The great ape can fix that league"*
  - *"The great tape can. Fix that'll eek!"*

Data Mining and Machine Learning

UNIVERSITY OF BIRMINGHAM

Speech is Continuous

Assignment Project Exam Help

https://eduassistpro.github.io/

Add WeChat edu_assist_pro

Data Mining and Machine Learning

UNIVERSITY OF
BIRMINGHAM

Variability

Assignment Project Exam Help

https://eduassistpro.github.io/

Add WeChat edu_assist_pro

UNIVERSITY OF
BIRMINGHAM

Confusability

Assignment Project Exam Help

https://eduassistpro.github.io/

Add WeChat edu_assist_pro

Data Mining and Machine Learning

UNIVERSITY OF
BIRMINGHAM

English Vowels : /h_d/

Assignment Project Exam Help

https://eduassistpro.github.io/

Add WeChat edu_assist_pro

Data Mining and Machine Learning
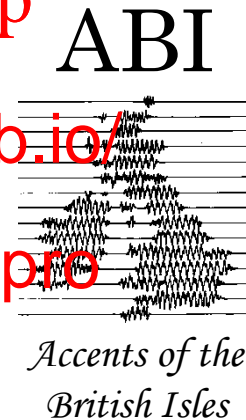
UNIVERSITY OF
BIRMINGHAM

"League" or *"Leek"*?

- *"league"* = / l i g /

- *"leek"* = / l i k /

- Difference appears to be in the final consonant.

  - /g/ is *voiced*
  - /k/ is *unvoiced*

- But in natural fluent speech, the length of the vowel /i/ may be a more important cue to recognition!
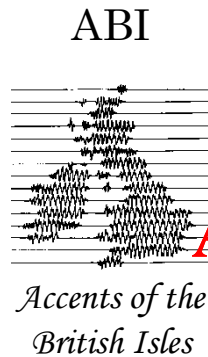
UNIVERSITY OF BIRMINGHAM

ABI – Accents of the British Isles

- Corpus of recordings of 15 different accents of British English
  - 300+ subjects
  - Approx. 20 m
  - 20+ subjects (
  - Each subject born in town, lived his or her life, parents born in to
  - Funded by 20/20 Speech
  - Up to £2K (academic) or £20K (industry)

Assignment Project Exam Help

https://eduassistpro.github.io/

Add WeChat edu_assist_pro

ABI

Accents of the British Isles

Data Mining and Machine Learning

UNIVERSITY OF BIRMINGHAM

ABI

ABI

Accents of the British Isles

Lowestoft 🔊

Elgin 🔊

Glasgow 🔊

Ulster 🔊

D

- ABI II
    - Extended corpus
    - Systematic study of effect of accent on recogniser performance
    - Systematic study of acoustic correlates of accent

Data Mining and Machine Learning

UNIVERSITY OF BIRMINGHAM

Approaches to Speech Recognition

- Many approaches to speech recognition have been tried in past

- Researchers in early days believed there was insufficient information in the acoustic data to recognise speech, and that additional sources of **knowledge** were necessary

  - acoustic-phonetic, tactic (grammar), semantic and ge

- Most successful approach to-da is a combination of **hidden Markov models (HMMs) with (deep) neural networks**

Data Mining and Machine Learning

UNIVERSITY OF **BIRMINGHAM**

Speech Recognition Terminology

- Basic problem in speech recognition is **variability**

- Early attempts to solve problem by **removing** it

- **Speaker variability**

  Assignment Project Exam Help

  - **Speaker-dep** speech recognition systems train on, and subsequently https://eduassistpro.github.io/

  - **Multiple-spe** icular **population** of speakers Add WeChat edu_assist_pro

  - **Speaker Independent** systems work for **any speaker**, with no implicit or explicit training

  - **Speaker adaptive** systems automatically adapt to a new speaker. E.G: begin with a speaker-independent system, and then adapt the system to a particular speaker to obtain a speaker-dependent system.

Data Mining and Machine Learning

UNIVERSITYOF
BIRMINGHAM

Terminology (Continued)

- Another source of variability is co-articulation between words

- **Isolated word** recognition systems require the user to leave **gaps** between words

- **Connected spe** Assignment Project Exam Help ecognize isolated phrases or sente https://eduassistpro.github.io/

- **Continuous speech recognitio** Add WeChat edu_assist_pro cognize continuous speech.

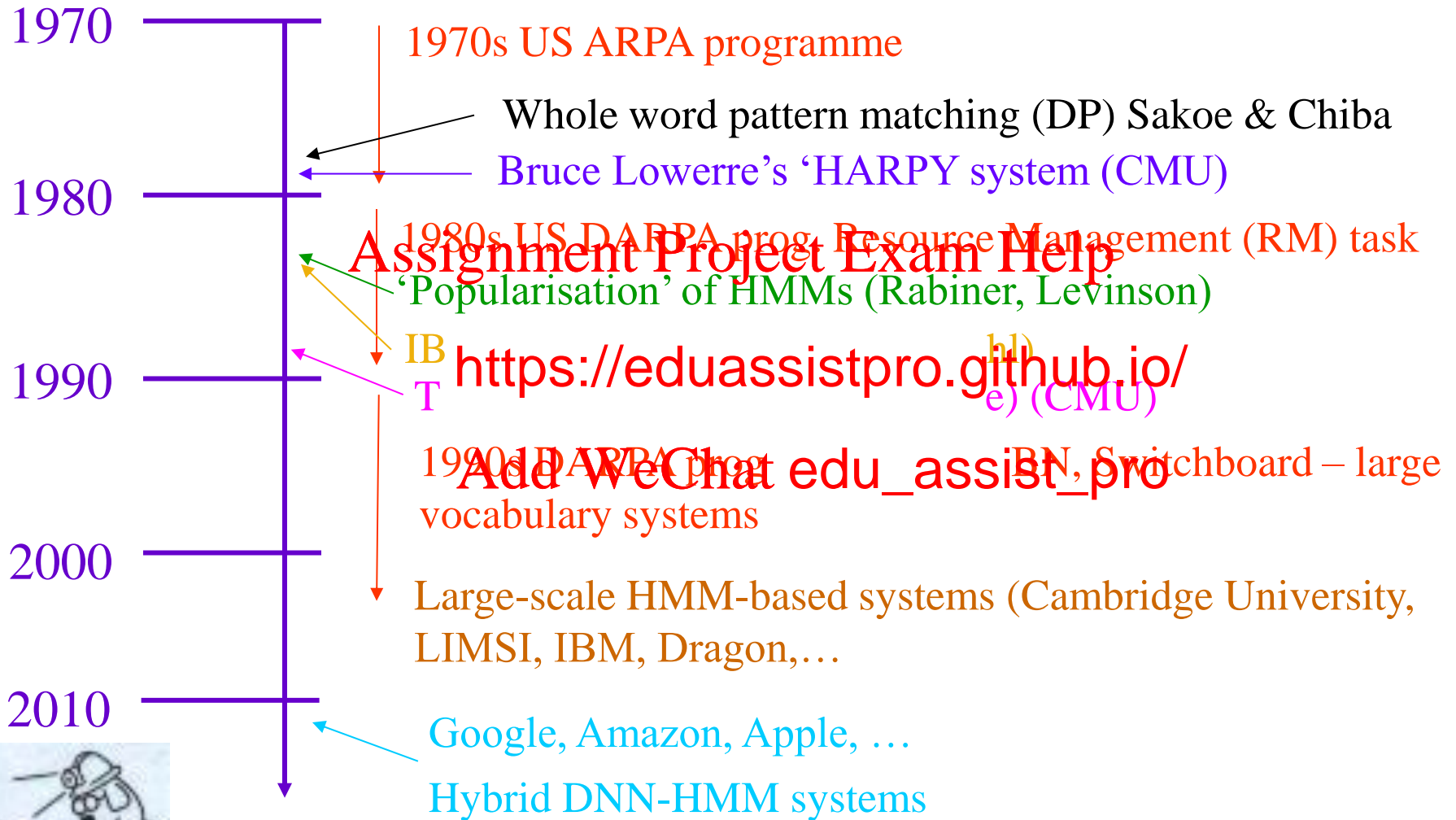Data Mining and Machine Learning

UNIVERSITY OF BIRMINGHAM

Vocabulary Size
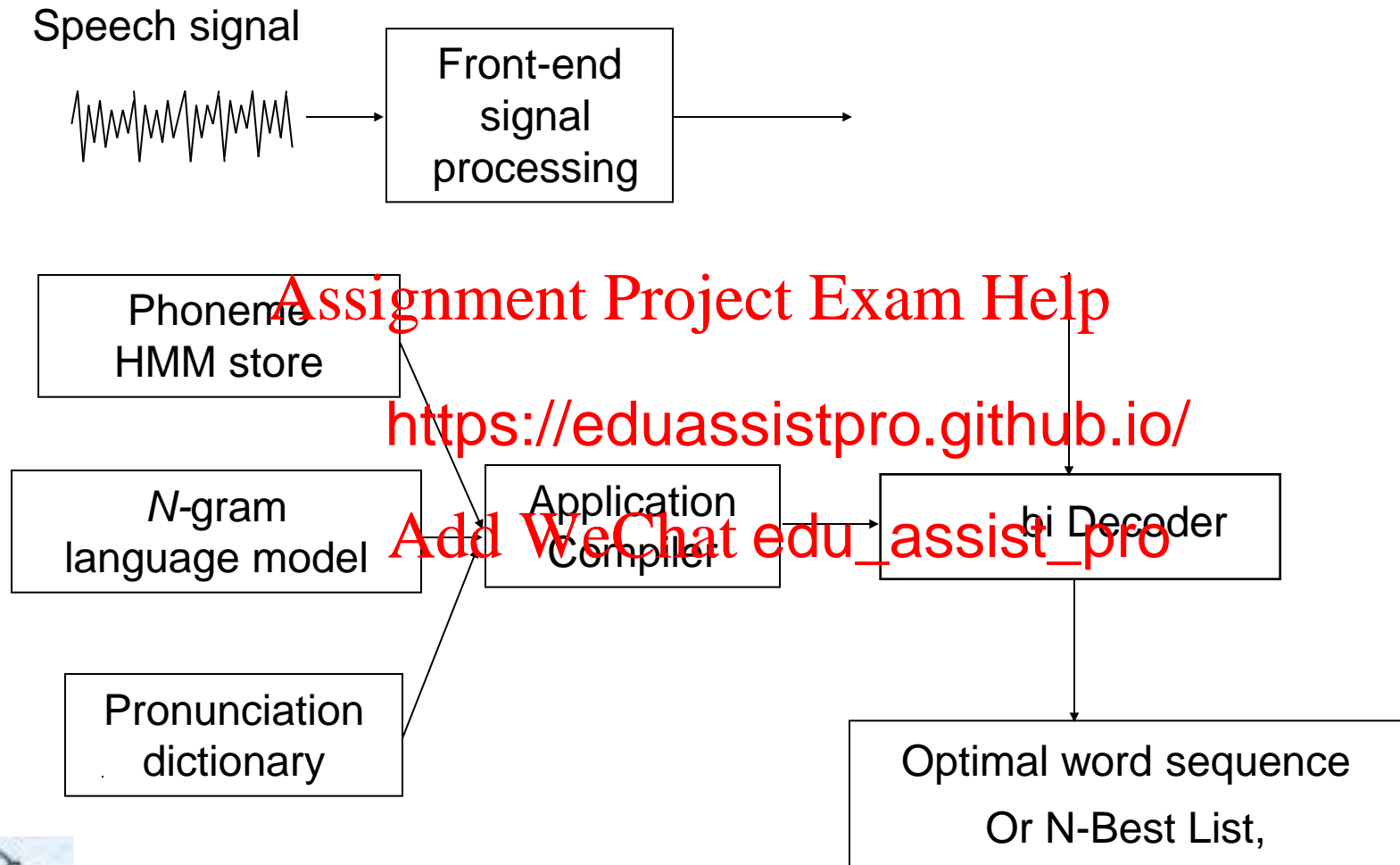
- Another important issue is vocabulary size

- **Small vocabulary** systems work with vocabularies of *10-100* words

- **Medium vocab** *00 to 5,000* words

- **Large Vocabulary Continuou** **cognition** (**LVCSR**) systems can cope wit rds, while

- **Unlimited vocabulary** systems have no vocabulary size limitation

Data Mining and Machine Learning

UNIVERSITY OF BIRMINGHAM

Historical perspective

1970 — 1970s US ARPA programme

Whole word pattern matching (DP) Sakoe & Chiba

Bruce Lowerre's 'HARPY system (CMU)

1980 — 1980s US DARPA prog. Resource Management (RM) task

Assignment Project Exam Help

'Popularisation' of HMMs (Rabiner, Levinson)

IB

https://eduassistpro.github.io/

T

e) (CMU)

1990 — 1990 DARPA host BN, Switchboard – large

Add WeChat edu_assist_pro

vocabulary systems

2000 — Large-scale HMM-based systems (Cambridge University, LIMSI, IBM, Dragon,…

2010 — Google, Amazon, Apple, …

Hybrid DNN-HMM systems

UNIVERSITY OF
BIRMINGHAM

Data Mining and Machine Learning

Phoneme-HMM Speech Recogniser

Speech signal

Front-end
signal
processing

Phoneme
HMM store

Assignment Project Exam Help

https://eduassistpro.github.io/

N-gram
language model

Application
Compiler

Add WeChat edu_assist_pro

bi Decoder

Pronunciation
dictionary

Optimal word sequence

Or N-Best List,

Data Mining and Machine Learning

Summary

- Why automatic speech recognition is difficult
  - Speech is not "acoustic text": Continuity, Variability, Confusability

- Speech reco
- Historical perspective

Data Mining and Machine Learning

UNIVERSITY OF BIRMINGHAM