

# Data Mining and Machine Learning

## Application of HMMs for Automatic Speech Recognition – Introduction

Assignment Project Exam Help

<https://powcoder.com>

Add WeChat powcoder

Peter Jančovič



UNIVERSITY OF  
BIRMINGHAM

# Objectives

- Introduce automatic speech recognition
- Understand why speech recognition is difficult
  - Continuity
  - Variability <https://powcoder.com>
  - Confusability
  - Effects of accent
- Speech recognition terminology



# Why is speech recognition difficult?

Intuitively...

- Meaning is represented by sentences
- A Sentence is a sequences of words
- A Word is a sequences of phonemes
- ...
- This view of speech is based on text
- But speech is NOT just “*Acoustic text*”



# Speech is....

- Continuous
  - “*We were away a year ago*”
- Variable **Assignment Project Exam Help**
  - “*bread and butter*” or “*brembudder*”**<https://powcoder.com>**
- Ambiguous
  - “*The grey tape can fix that leak*”
  - “*The great ape can fix that leek*”
  - “*The great ape can fix that league*”
  - “*The great tape can. Fix that’ll eek!*”



# Speech is Continuous

file=we\_were\_away.sfs speaker=MJR token=we were away a year ago



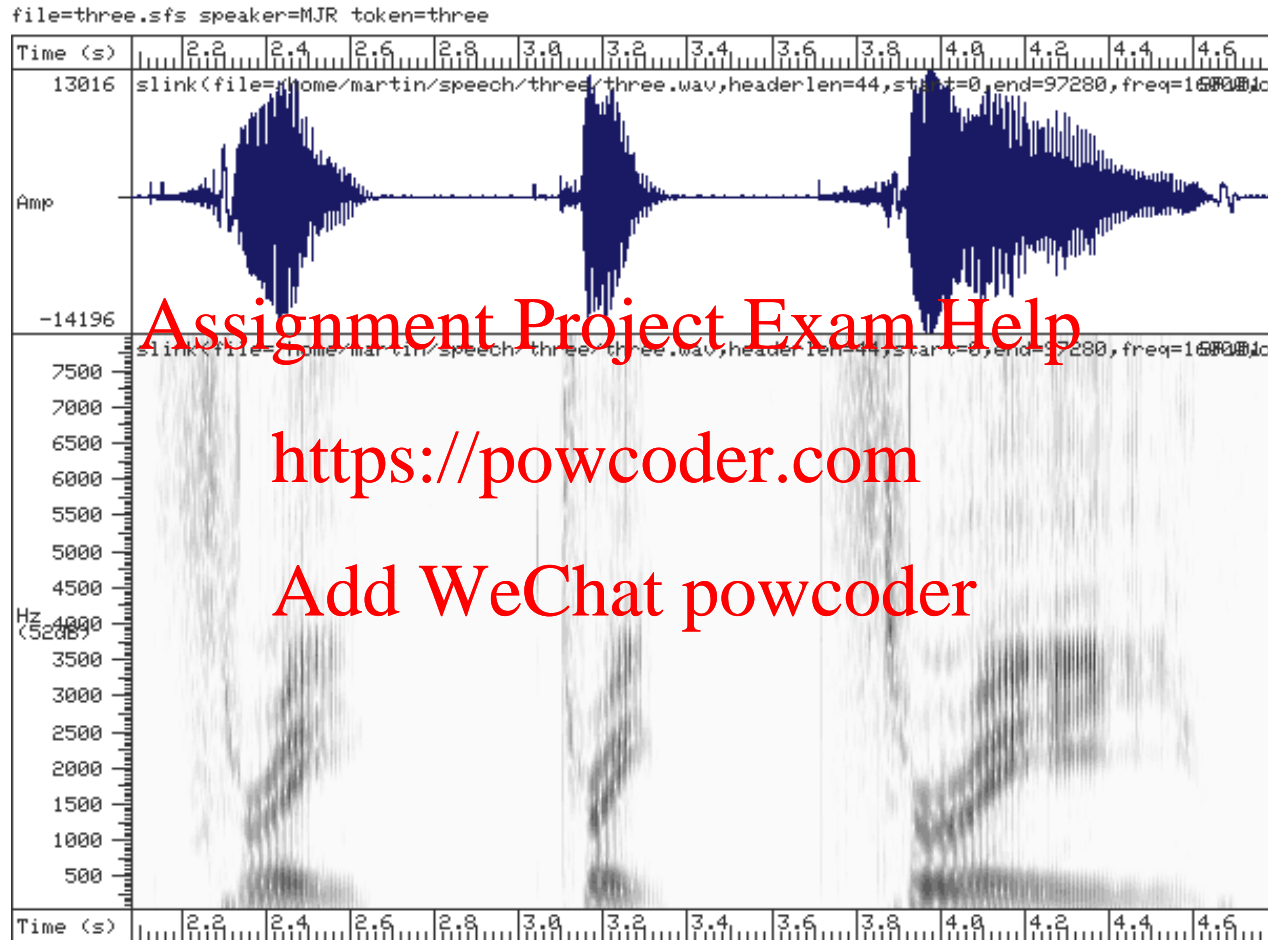
Assignment Project Exam Help

<https://powcoder.com>

Add WeChat powcoder



# Variability



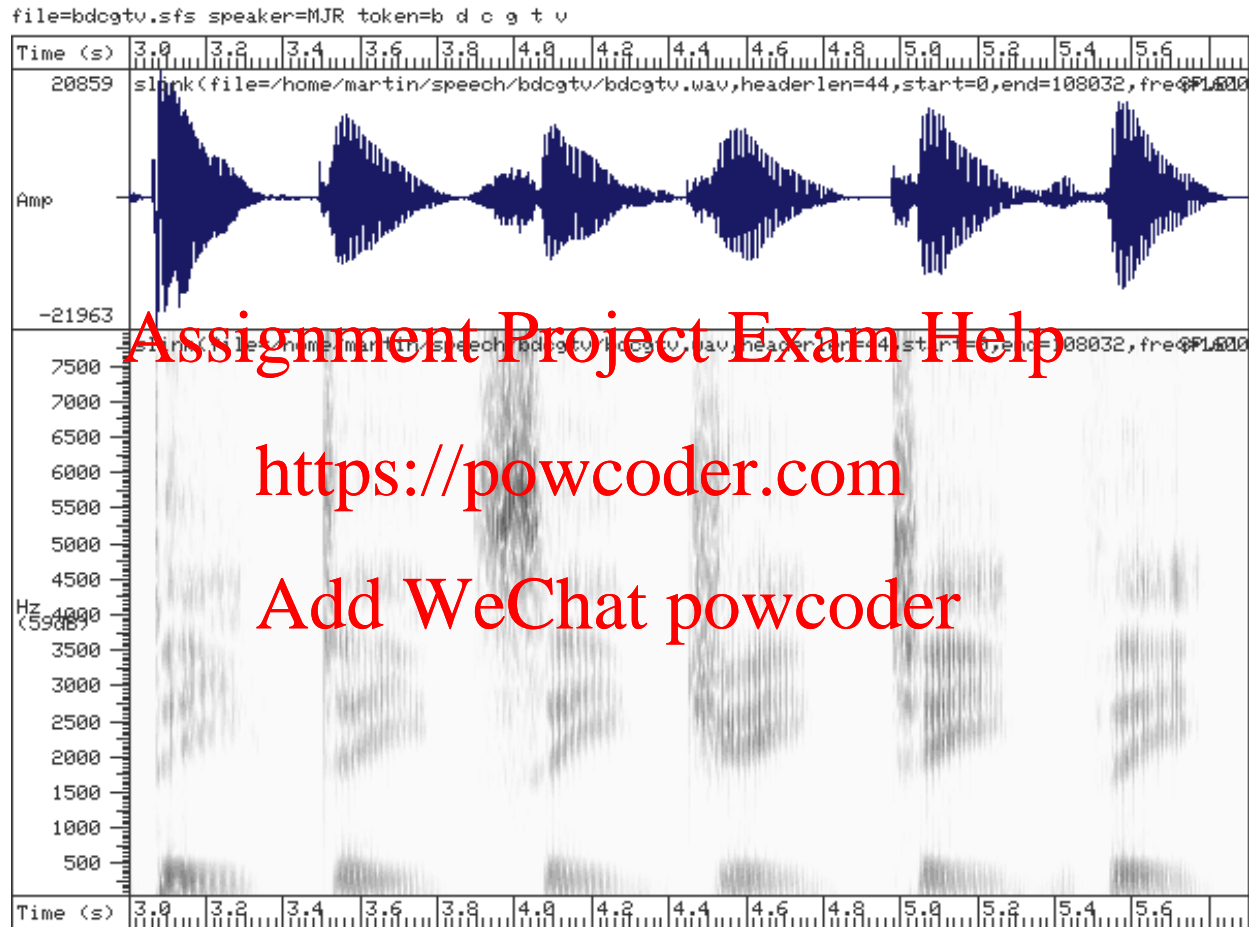
Assignment Project Exam Help

<https://powcoder.com>

Add WeChat powcoder



# Confusability



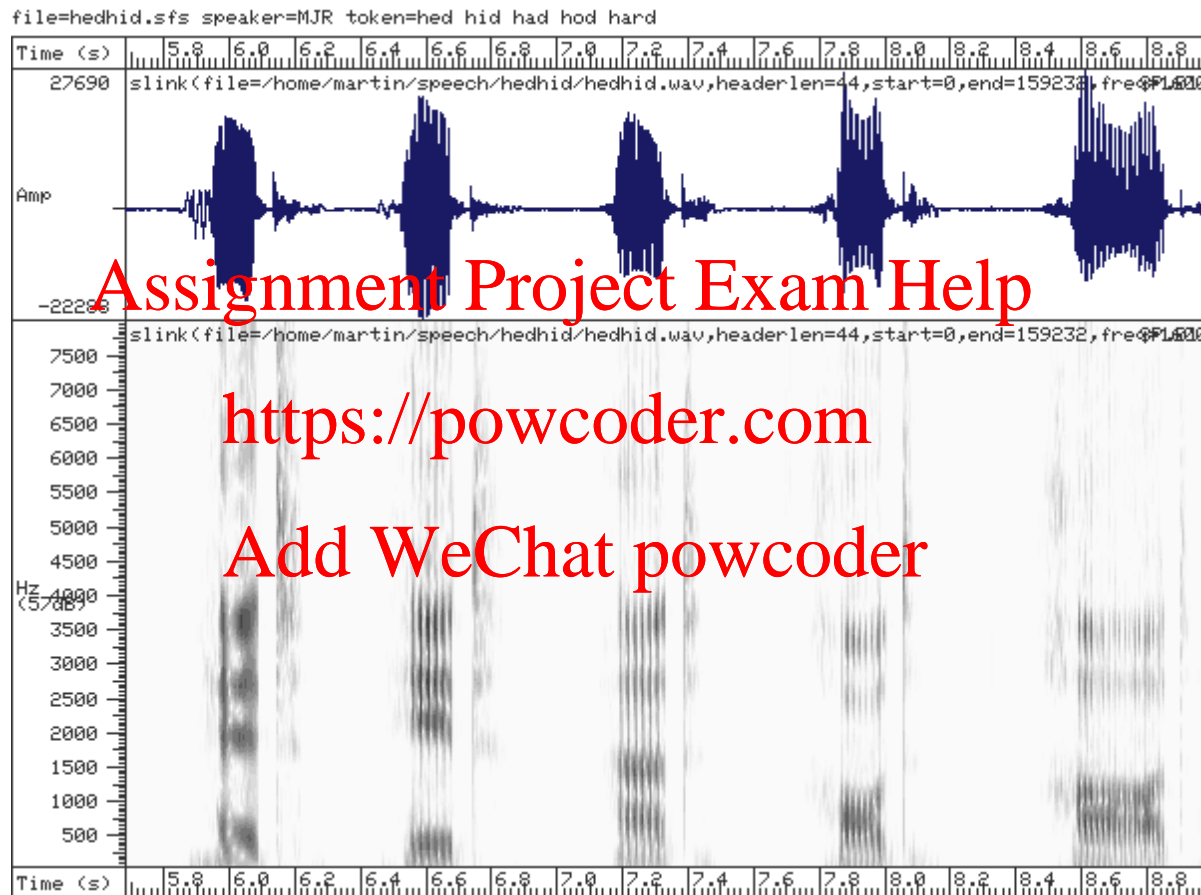
Assignment Project Exam Help

<https://powcoder.com>

Add WeChat powcoder



# English Vowels : /h\_d/



Assignment Project Exam Help

<https://powcoder.com>

Add WeChat powcoder





# “*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 *duration* of the vowel /i/ may be a more important cue to recognition!

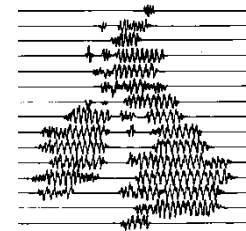


# ABI – Accents of the British Isles

- Corpus of recordings of 15 different accents of British English

- 300+ subjects
- Approx. 20 minutes of speech per subject
- 20+ subjects (10m, 10f) per accent
- Each subject born in town, lived there all of his or her life, parents born in town
- Funded by 20/20 Speech
- Up to £2K (academic) or £20K (industry)

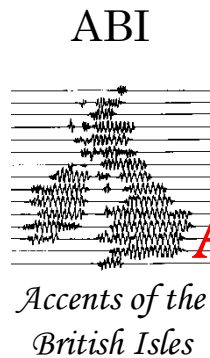
ABI



*Accents of the  
British Isles*



# ABI



Lowestoft

Elgin

Glasgow

Ulster

Denbigh



Assignment Project Exam Help

<https://powcoder.com>



## ■ ABI II

Add WeChat powcoder

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



# 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, lexical (words), syntactic (grammar), semantic and domain-specific knowledge
- Most successful approach to date is based on a combination of **hidden Markov models (HMMs) with (deep) neural networks**



# Speech Recognition Terminology

- Basic problem in speech recognition is **variability**
- Early attempts to solve problem by **removing** it
- **Speaker variability**
  - **Speaker-dependent** speech recognition systems train on, and subsequently recognise, a **single speaker**
  - **Multiple-speaker** systems work for a particular **population** of speakers
  - **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.



# Terminology (Continued)

- Another source of variability is co-articulation between words
- **Isolated word** recognition systems require the user to leave **gaps** between words
- **Connected speech recognition** systems recognize isolated phrases or sentences
- **Continuous speech recognition** systems recognize continuous speech.

Assignment Project Exam Help

<https://powcoder.com>

Add WeChat powcoder

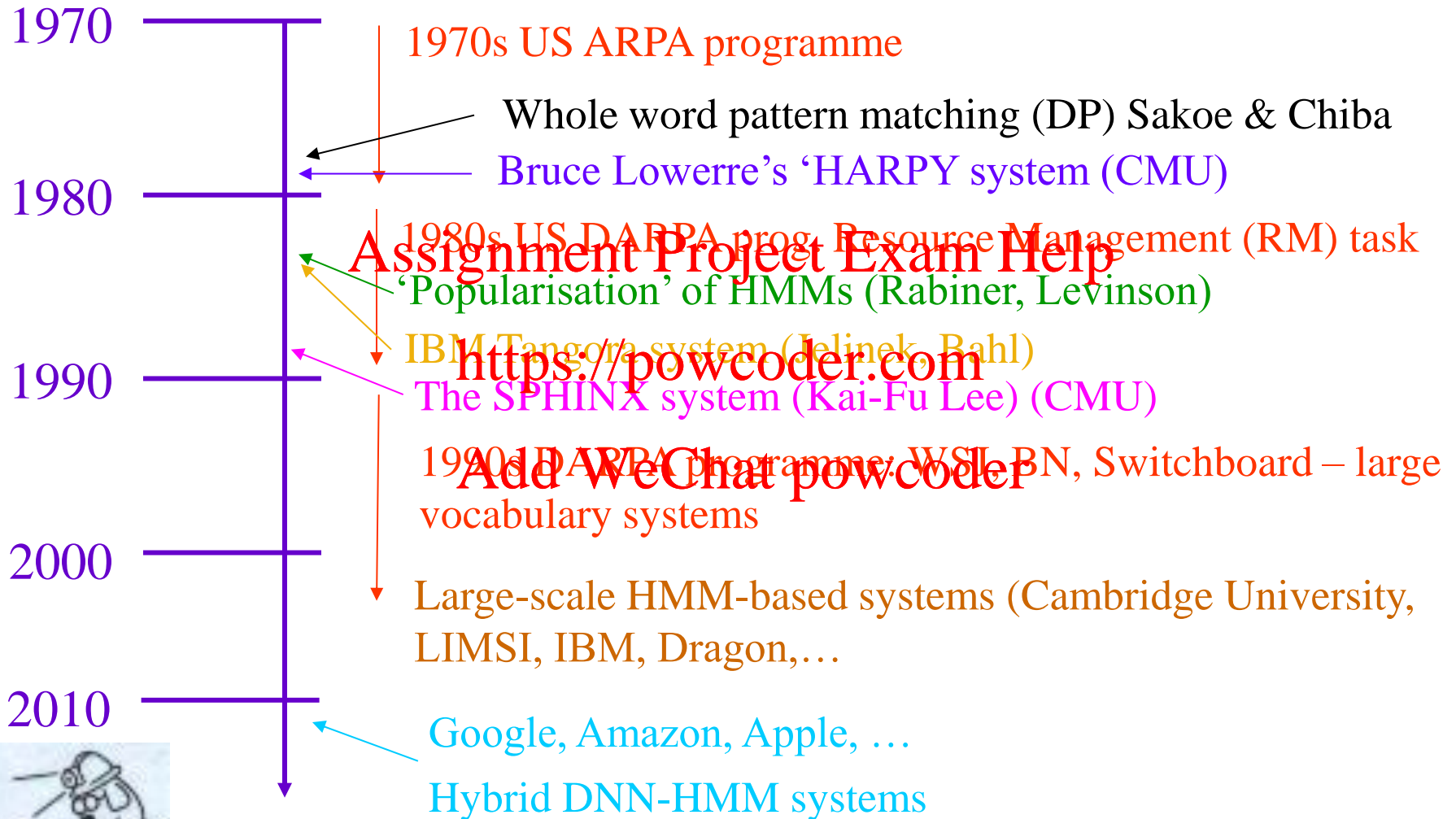


# Vocabulary Size

- Another important issue is vocabulary size
- **Small vocabulary** systems work with vocabularies of *10-100* words
- **Medium vocabularies** comprise around *100* to *5,000* words
- **Large Vocabulary Continuous Speech Recognition (LVCSR)** systems can cope with 60,000 words, while
- **Unlimited vocabulary** systems have no vocabulary size limitation



# Historical perspective



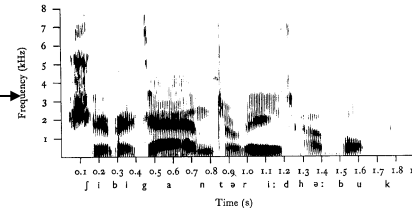


# Phoneme-HMM Speech Recogniser

Speech signal



Front-end  
signal  
processing



Phoneme  
HMM store

N-gram  
language model

Pronunciation  
dictionary

Application  
Compiler

Viterbi Decoder

Optimal word sequence  
Or N-Best List,



# Summary

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

Assignment Project Exam Help

- Speech recognition terminology

<https://powcoder.com>

- Historical perspective

Add WeChat powcoder

