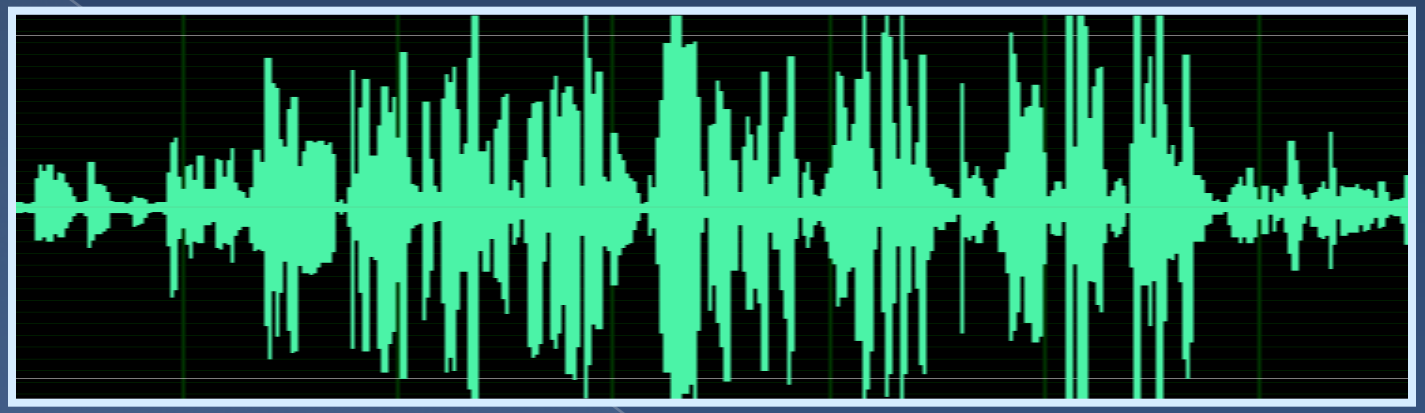
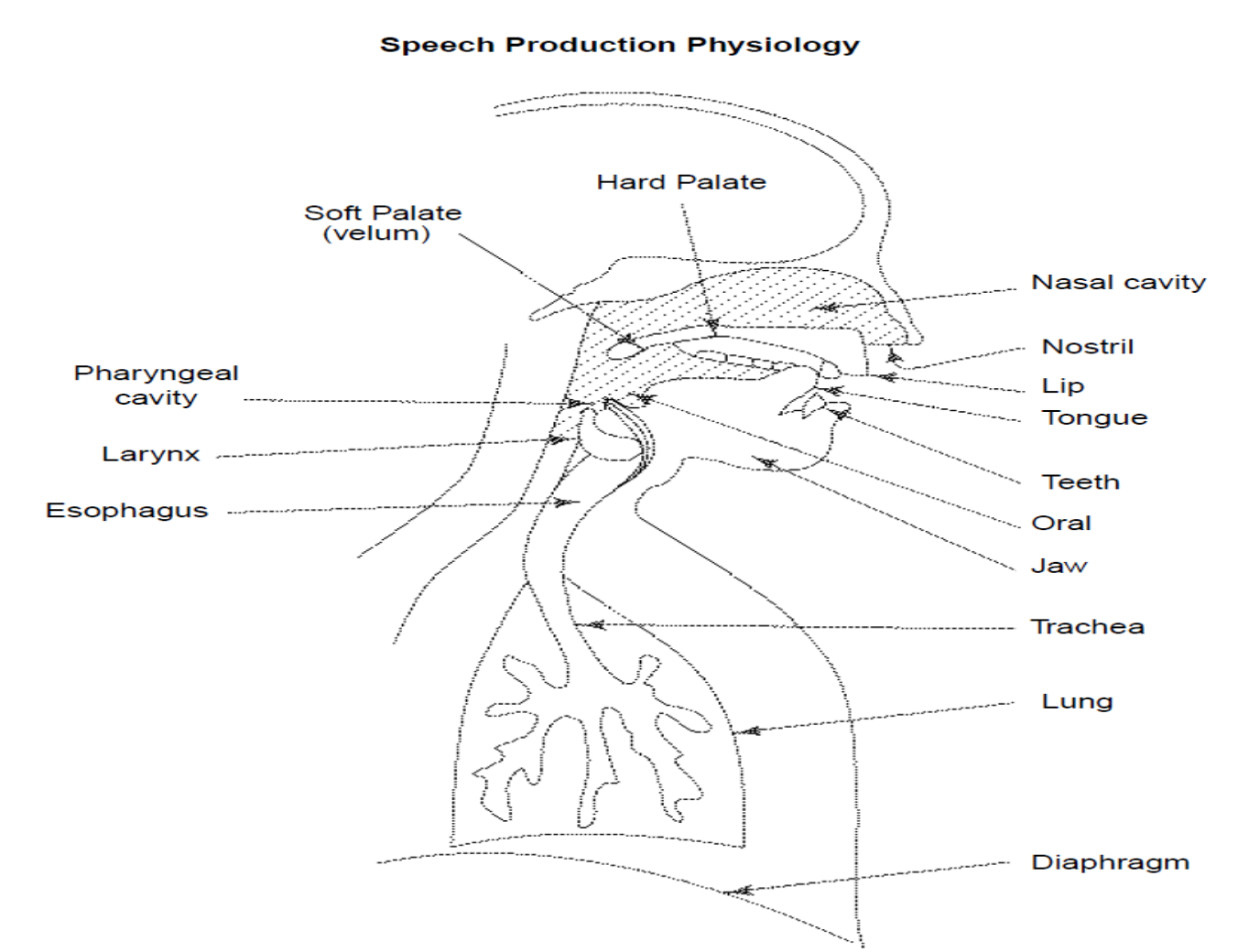
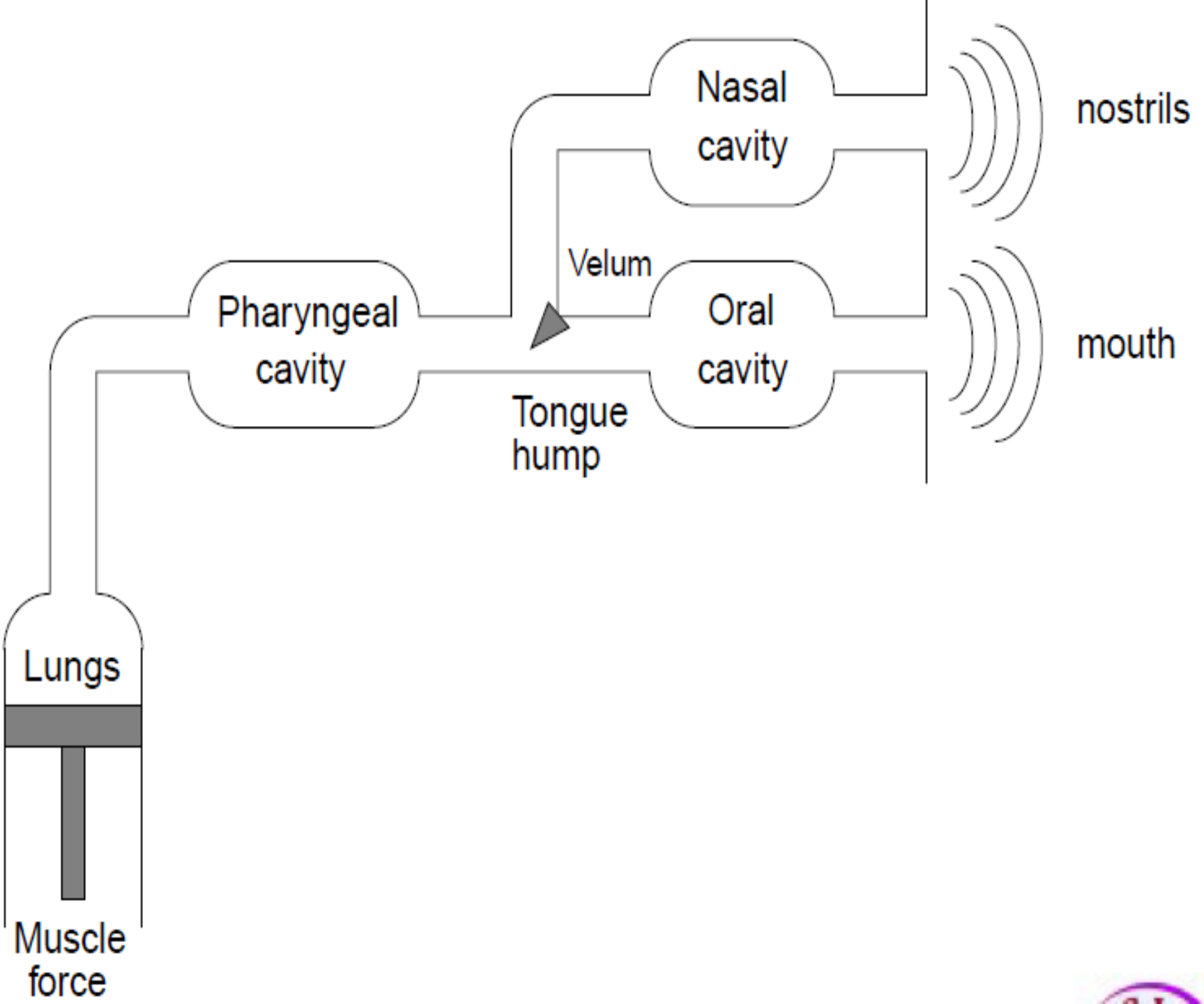
SPEECH RECGNITION



BY - HIMANSHU BHATTI
STREAM : CSE 7TH SEM
ROLL NO. : 07210102711
AIACTR

SPEECH PRODUCTION PSYSIOLOGY




Nasal
cavity
nostrils
Velum
Pharyngeal
cavity
Oral
cavity
mouth
Tongue
hump
Lungs
Muscle
force

Introduction

- What is speech recognition?

Introduction

- Speech recognition technology has recently reached a higher level of performance and robustness, allowing it to communicate to another user by talking .
- Speech Recognization is process of decoding acoustic speech signal captured by microphone or telephone ,to a set of words.
- And with the help of these it will recognize whole speech is recognized word by word .

Types of SR

- : **speaker independent** and **speaker dependent.**

- Speaker independent models recognize the speech patterns of a large group of people.

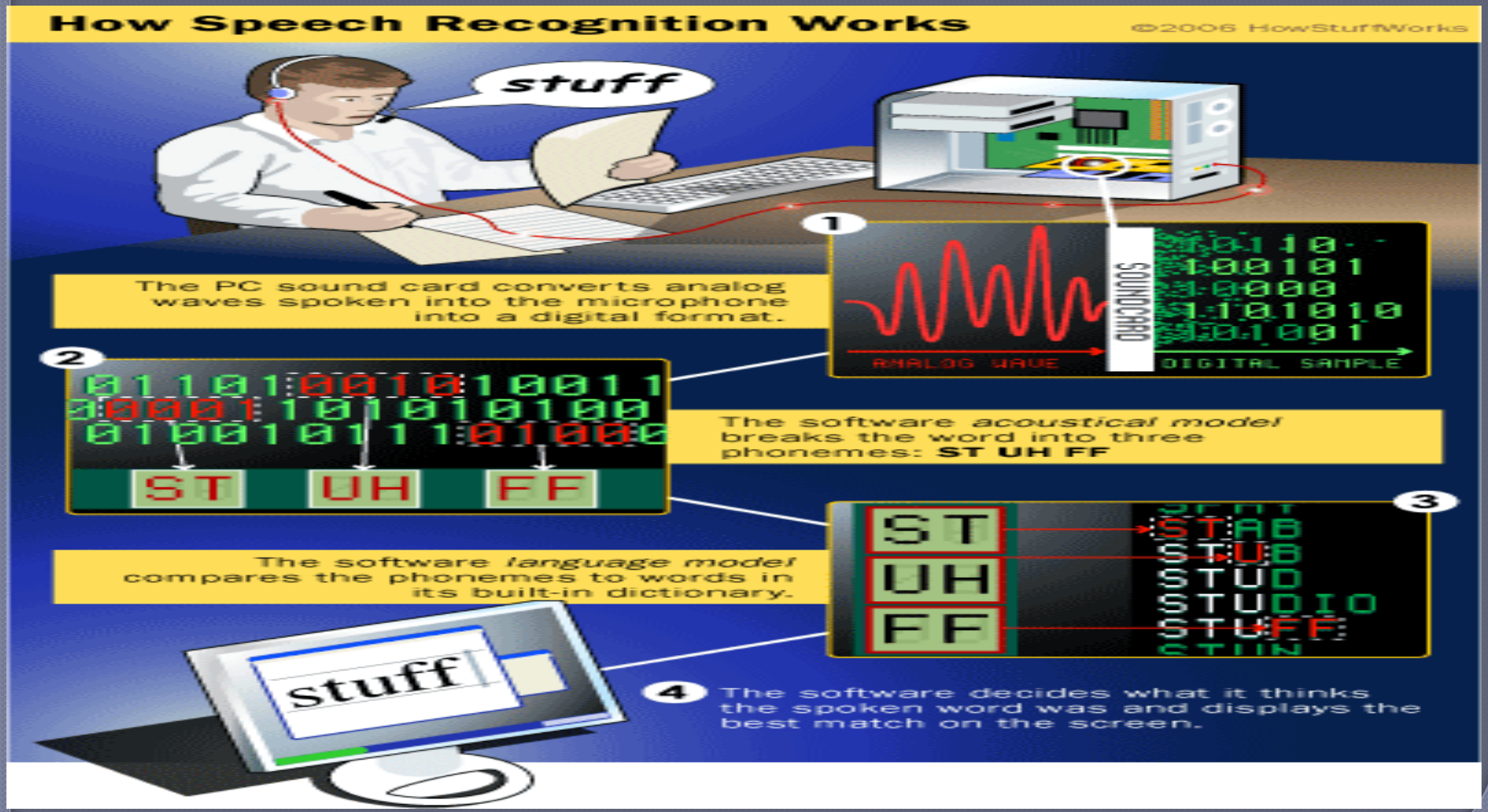
- Speaker dependent models recognize speech patterns from only one person. Both models use mathematical and statistical formulas to yield the best work match for speech. A third variation of speaker models is now emerging, called **speaker adaptive**.

- Speaker adaptive systems usually begin with a speaker independent model and adjust these models more closely to each individual during a brief training period.

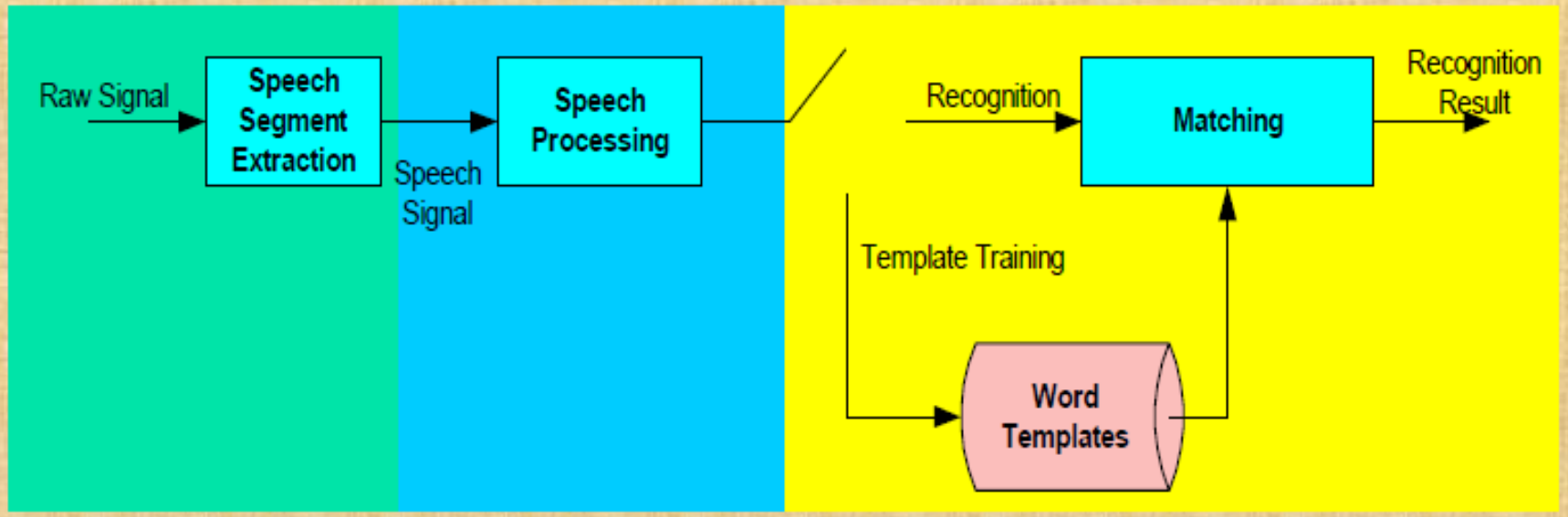
Why do we need speech recognition

- Most Natural Form Of Communication
- Differently abled people
- Illiterate
- Helplines
- Cars

How speech recognition works

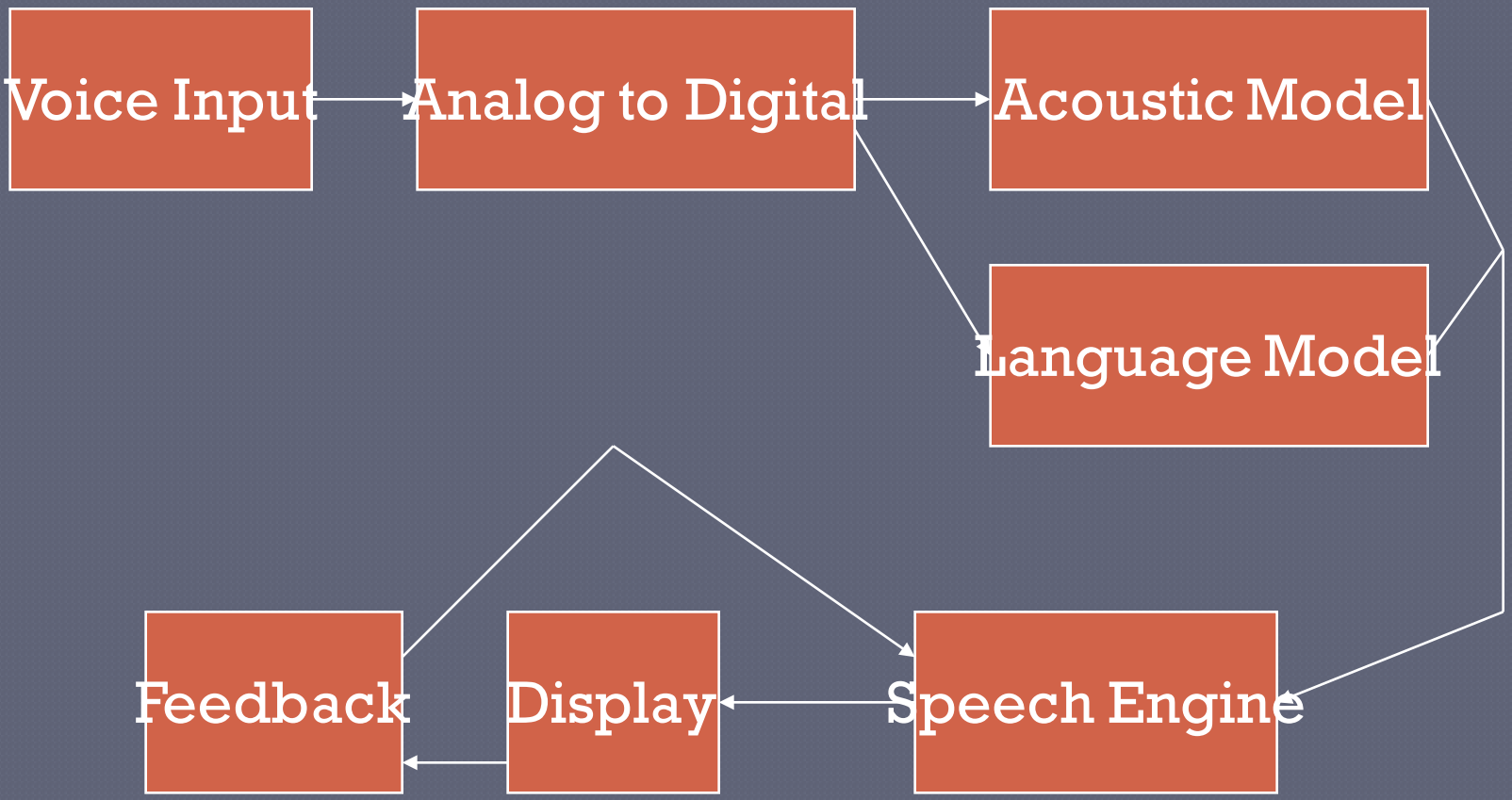


ASR System Overview



- Speech Segment Extraction
- Speech Processing and Modeling
- Pattern Recognition and Training

Recognition Flowchart presentation



Recognition Process Flow Summary

- Step 1:User Input

  The system catches user's voice in the form of analog acoustic signal.

- Step 2:Digitization

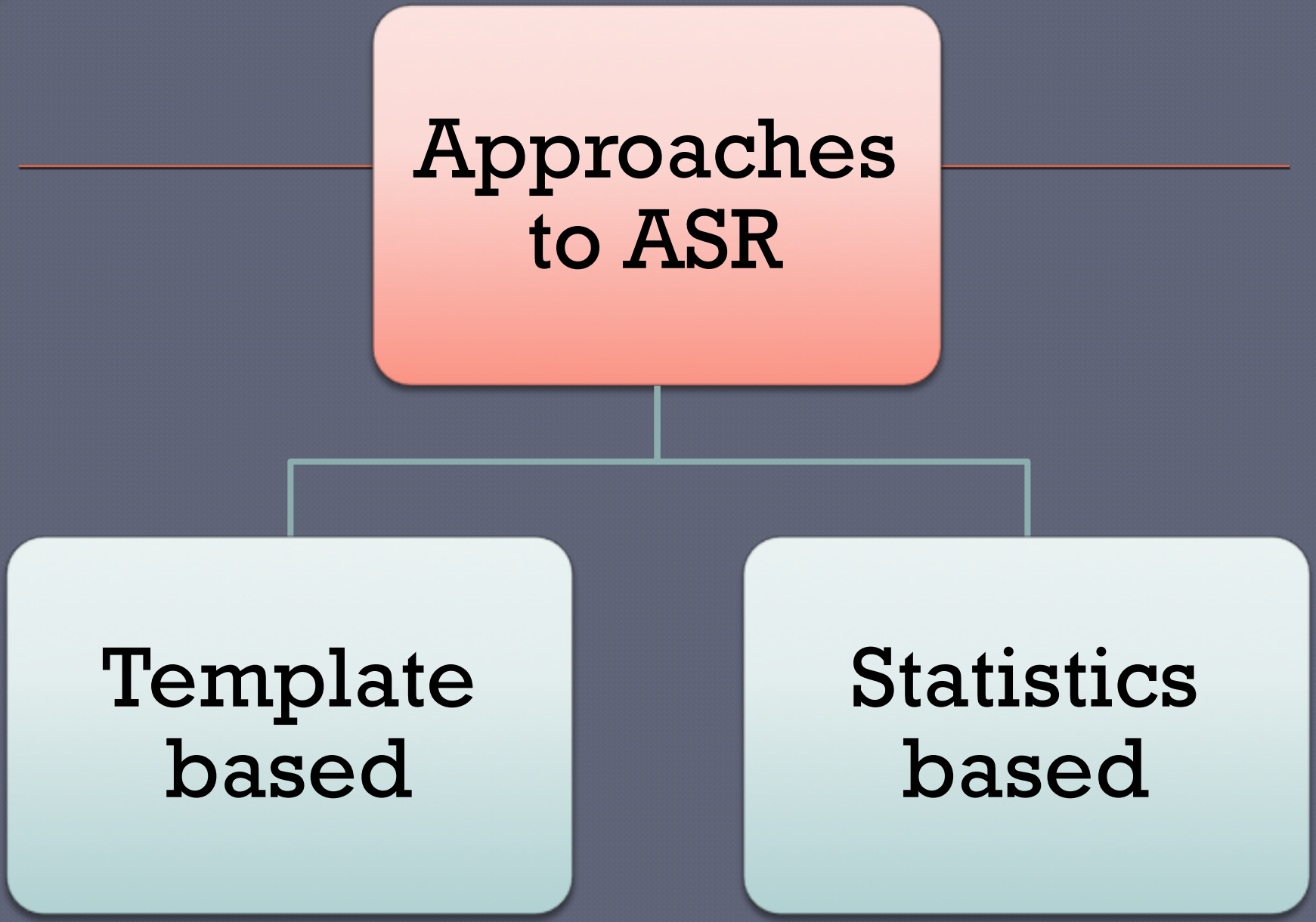
  Digitize the analog acoustic signal.

- Step 3:Phonetic Breakdown

  Breaking signals into phonemes.

Recognition Process Flow Summary

- Step 4:Statistical Modeling
  - Mapping phonemes to their phonetic representation using statistics model.
- Step 5:Matching
  - According to grammar , phonetic representation and Dictionary , the system returns an n-best list (I.e.:a word plus a confidence score)
  - **Grammar**-the union words or phrases to constraint the range of input or output in the voice application.
  - **Dictionary**-the mapping table of phonetic representation and word(EX:thu,thee→the)


Approaches to ASR
Template based
Statistics based

Template-based approach

- Store examples of units (words, phonemes), then find the example that most closely fits the input
- Extract features from speech signal, then it's "just" a complex similarity matching problem, using solutions developed for all sorts of applications
- OK for discrete utterances, and a single user

Template-based approach

- Hard to distinguish very similar templates
- And quickly degrades when input differs from templates
- Therefore needs techniques to mitigate this degradation:
  - More subtle matching techniques
  - Multiple templates which are aggregated
- Taken together, these suggested …

Statistics-based approach

- Collect a large corpus of transcribed speech recordings
- Train the computer to learn the correspondences ("machine learning")
- At run time, apply statistical processes to search through the space of all possible solutions, and pick the statistically most likely one

Statistics based approach

- Acoustic and Lexical Models
  - Analyse training data in terms of relevant features
  - Learn from large amount of data different possibilities
    - different phone sequences for a given word
    - different combinations of elements of the speech signal for a given phone/phoneme
  - Combine these into a **Hidden Markov Model** expressing the probabilities

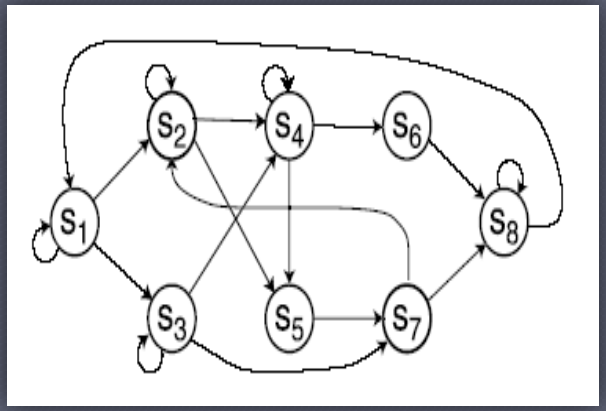
HIDDEN MARKOV MODEL (HMM)

- **Real-world has structures and processes which have (or produce) observable outputs:**
  - **Usually sequential (process unfolds over time)**
  - **Cannot see the event producing the output**

    Example: speech signals

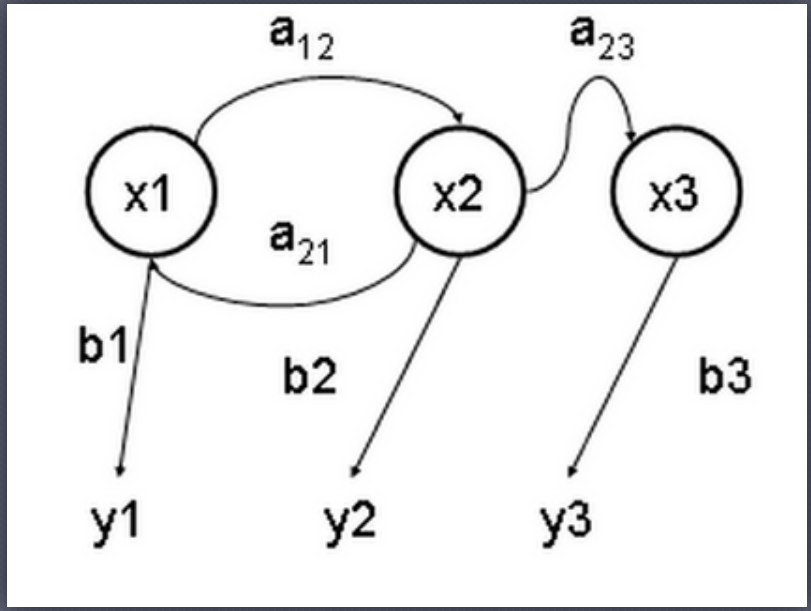
HMM Overview

- Machine learning method
- Makes use of state machines
- Based on probabilistic model
- Can only observe output from states, not the states themselves
  - Example: speech recognition
    - Observe: acoustic signals
    - Hidden States: phonemes
      (distinctive sounds of a language)

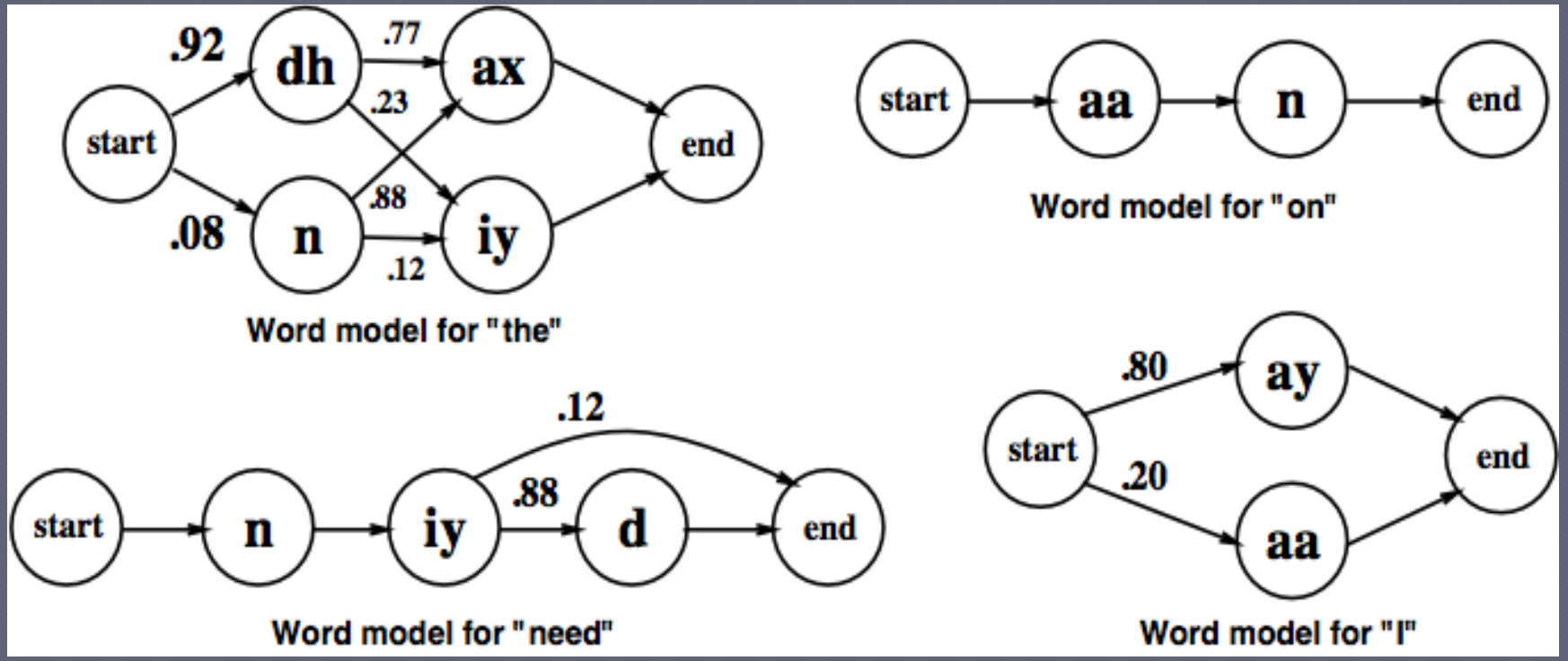


HMM Components

- A set of states (x's)
- A set of possible output symbols (y's)
- A state transition matrix (a's): probability of making transition from one state to the next
- Output emission matrix (b's): probability of a emitting/observing a symbol at a particular state
- Initial probability vector:
  - probability of starting at a particular state
  - Not shown, sometimes assumed to be 1



HMMs for some words



HMM Advantages

- **Advantages:**
  - **Effective**
  - **Can handle variations in record structure**
    - ✓ **Optional fields**
    - ✓ **Varying field ordering**

What's hard about that?

- Digitization
  - Converting analogue signal into digital representation.
- Signal processing
  - Separating speech from background noise.
- Phonetics
  - Variability in human speech.
- Phonology
  - Recognizing individual sound distinctions (similar phonemes.)
- Lexicology and syntax
  - Disambiguating homophones.
  - Features of continuous speech.
- Syntax and pragmatics
  - Interpreting features.
  - Filtering of performance errors (disfluencies).

Challenges and Difficulties of SR

Speech Recognition is still a very cumbersome problem. Following are the problem….

- **Speaker Variability**

  Two speakers or even the same speaker will pronounce the same word differently

- **Channel Variability**

  The quality and position of microphone and background environment will affect the output

Applications of Speech Recognition

- **Speech recognition applications include**
  - Voice dialling (e.g., "Call home"),
  - Call routing (e.g., "I would like to make a collect call"),
  - Simple data entry (e.g., entering a credit card number),
  - Preparation of structured documents (e.g., A radiology report),
  - Speech-to-text processing (e.g., word processors or emails), and
  - In aircraft cockpits (usually termed Direct Voice Input).

Applications

- Medical Transcription
- Military
- Telephony and other domains
- Serving the disabled

Further Applications

- Home automation
- Automobile audio systems
- Telematics

Pros of Speech Recognition

- Faster than “hand-writing”.
- Allows for better spelling, whether it be in text or documents.
- Helpful for people with a mental or physical disability .
- Hands-free capability .

Cons of Speech Recognition

- No program is 100% perfect
- Factors that affect the accuracy of speech recognition are: slang, homonyms, signal-to-noise ratio, and overlapping speech
- Can be expensive depending on the program

References

- **http://en.wikipedia.org/wiki/Speech_recognition**
- **https://www.scribd.com/doc/130376790/Speech-Recognition**
- "Speaker Independent Connected Speech Recognition- Fifth Generation Computer Corporation". Fifthgen.com.
- **http://books.google.co.in/books?hl=en&lr=&id=iDHgboYRzmgC&oi=fnd&pg=PA1&dq=speech+recognition+papers+publications&ots=jb6NESTrjF&sig=oMKROIXccSgEyMGOZmi5lkToJvM#v=onepage&q=speech%20recognition%20papers%20publications&f=false**
- **http://www.speechrecognition.com**
- **https://www.google.co.in/?gfe_rd=cr&ei=GbHdU9f1MtKAoAOW64GADg&gws_rd=ssl**

THANK YOU.....!!