

Bangla Automatic Speech Recognition (ASR)



Md. Abul Hasnat

Center for Research on
CRBLP
Bangla Language Processing

Center for Research on Bangla Language Processing (CRBLP),
Department of Computer Science and Engineering, BRAC University,
Dhaka, Bangladesh.

Outline

- What is ASR ?
- Status of Bangla ASR.
- How to build ASR for Bangla?
- Conclusion

What is ASR?



- I want to perform my tasks using my voice command.
- I want the computer to generate documents from my voice.

What is ASR?

- Technique to perform automatic recognition.
- Recognizing speech signals.
 - From a microphone.
 - From a telephone.
- set of words
- Two types:
 - Isolated Speech Recognition
 - Continuous Speech Recognition

What is ASR?

- Process of converting an acoustic signal to a set of words.



- Applications:
 - Commands & Control
 - Data Entry
 - Document Preparation

Speech to Text

What is ASR (research perspective)

- ASR problem for researchers:
 - Pattern Recognition.
 - Artificial Intelligent.
 - Speech processing.

- ASR for English: Since 1930

- Commercial ASR available.
 - Dragon, Philips, IBM.

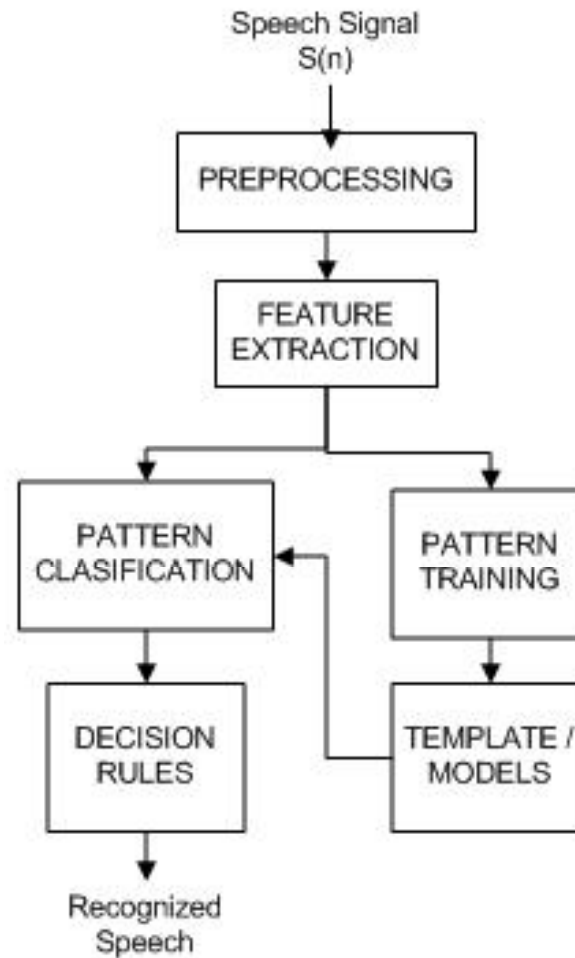
- Tools
 - HTK, JULIUS, Sphinx.

Status of ASR for Bangla

- Limited on research work, experiment and demo.
- Most recent:
 - From CSE Dept., Dhaka University
 - Limited vocabulary ASR
 - From CRBLP (2007)
 - Isolated Speech Recognizer (demo)
 - Continuous Speech Recognizer (research ongoing)

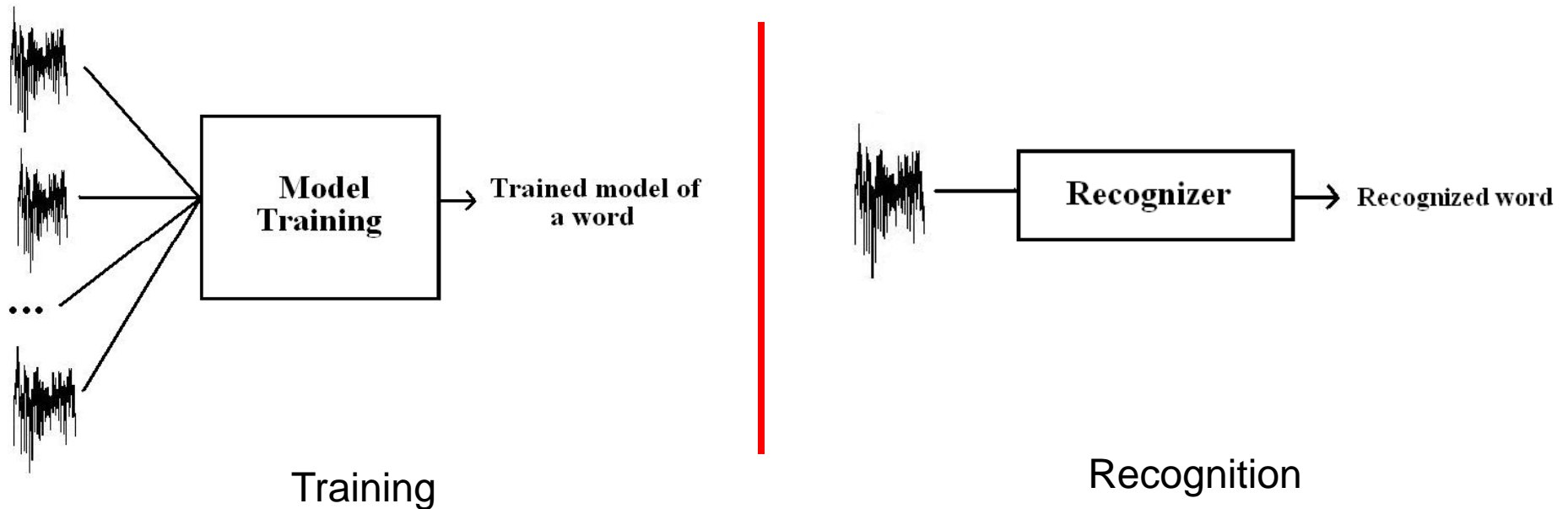
How to build an ASR for Bangla?

- Follow standard procedure to build an ASR.



How to build an ASR for Bangla? (Isolated Speech Recognition)

- ❑ Concentrate on each individual spoken word.
- ❑ Perfect word boundary detection.
- ❑ Speaker enrollment : User must provide samples.
- ❑ Limited vocabulary for Command & Control.



How to build an ASR for Bangla?

(Continuous Speech Recognition)

- Design Phoneme Set.
- Define the task Grammar (Unlimited Vocabulary).
- Pronunciation Dictionary.
- Prepare phonetically balance sentence (Training).
- Use tools to generate models for each unit
 - Monophone
 - Di-phone

How to build an ASR for Bangla?

(Continuous Speech Recognition)

□ Design Phoneme Set

| Place | | Bilabial | | Dental | | Alveolar | | Post- Alveolar | | Palatal | Velar | | Glottal |
|-------------|-----------|----------|--------------------|--------|--------------------|-----------|--------------------|----------------|--------------------|------------|---------|--------------------|---------|
| Manner | | | | | | | | | | | | | |
| Stops | voiceless | প/p/ | ফ/p ^h / | ত/t/ | থ/t ^h / | ট/t/ | ঠ/t ^h / | চ/c/ | ছ/c ^h / | | ক/k/ | খ/k ^h / | |
| | voiced | ব/b/ | ভ/b ^h / | দ/d/ | ধ/d ^h / | ড/d/ | ঢ/d ^h / | য, জ/j/ | ঝ/j ^h / | | গ/g/ | ঘ/g ^h / | |
| Nasals | | ম/m/ | | | | ন, ণ/n/ | | | | | ঙ, ঞ/ŋ/ | | |
| Trill | | | | | | র/r/ | | | | | | | |
| Flap | | | | | | ড়, ঢ়/r/ | | | | | | | |
| Fricatives | | | | | | শ, স/s/ | | | | ষ, ষ, স/ʃ/ | | | হ, ঙ/h/ |
| Lateral | | | | | | ল/l/ | | | | | | | |
| Approximant | | | | | | | | | | য়/j/ | | | |

How to build an ASR for Bangla? (Continuous Speech Recognition)

□ Define the task Grammar

sentence = subject object verb
subject = আমি | তুমি | সে
object = ভাত | কাজ | বাড়ি
verb = খাই | খাও | করি | করে | যাই | যায়

Example: Grammar

How to build an ASR for Bangla? (Continuous Speech Recognition)

□ Pronunciation Dictionary

এক ea k
দুই da u i
তিন ta i n
চার c aa r
পাঁচ p aa^ c
ছয় ch a y
সাত sh aa ta
আট aa t
নয় n a y
শূন্য s uu n n a

Example: Dictionary

How to build an ASR for Bangla?

(Continuous Speech Recognition)

- ❑ **Prepare phonetically balance sentence (Training)**
 - ❑ Create/ select a set of sentences
 - ❑ Ensure the presence of each phoneme
 - ❑ Transcribe sentences with appropriate phone label
 - ❑ Record the sentences
 - ❑ Consider variety of environments

- ❑ **In our research**
 - ❑ Total 52 sentences for phonetically balance set
 - ❑ 1814 phonemes

How to build an ASR for Bangla?

(Continuous Speech Recognition)

| | | | | | | | | |
|------|-----|-----|---|----------------|-----|----|----------------|-----|
| - | sil | 101 | খ | k ^h | da | দ | ɖ | 33 |
| অ | ɔ | 161 | গ | g | dah | ধ | ɖ ^h | 12 |
| আ | a | 217 | ঘ | g ^h | p | প | p | 39 |
| ই | i | 117 | ঙ | ŋ | ph | ফ | p ^h | 12 |
| ঈ | i: | 3 | চ | c | b | ব | b | 67 |
| উ | u | 42 | ছ | c ^h | bh | ভ | b ^h | 14 |
| ঊ | u: | 1 | জ | ʃ | m | ম | m | 36 |
| এ্যা | æ | 1 | ঝ | ʃ ^h | z | য | ɖʒ | 13 |
| এ | e | 165 | ঞ | No | r | র | r | 135 |
| ও | o | 61 | ত | t | l | ল | l | 52 |
| আঁ | ã | 5 | থ | t ^h | s | শ | ʃ | 30 |
| ইঁ | ĩ | 2 | দ | ɖ | sh | স | s | 23 |
| উঁ | ũ | 2 | ধ | ɖ ^h | h | হ | h | 28 |
| ওঁ | õ | 2 | ন | n | ra | ড় | r | 4 |
| এঁ | ẽ | 4 | ট | t | y | য় | j | 22 |
| ক | k | 102 | ঠ | t ^h | 2 | | | |

Statistics of phoneme distribution

How to build an ASR for Bangla?

(Continuous Speech Recognition)

- **Use tools to generate models for each unit**
 - Hidden Markov Model Toolkit (HTK)
 - HInit (HMM model initialization)
 - HSLab (labelling word)
 - HRest (Re-estimate the model parameters)
 - HCopy (For feature extraction)
 - HVite (Viterbi decoder : Recognizer)
 - HParse (Create a regular grammar and convert it to an intermediate form of decoding network)

Conclusion

- Concept of ASR.
- Applications of ASR.
- Status of ASR for Bangla script.
- Overview of building an ASR.

Questions

Thank You