深蓝学院
shenlanxueyuan.com

Speech Algorithms: from Theory to Practice

An Overview

Xiangang Li, Guoguo Chen

Outline

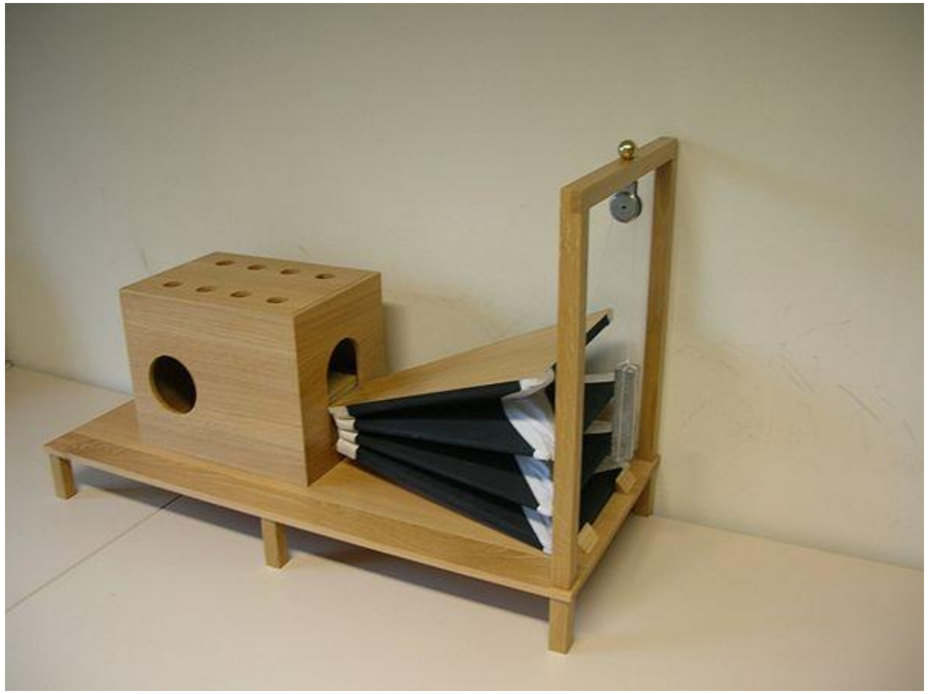
- A brief history of speech algorithms
- Course goals
- Course outlines
- Demo: a simple speech recognition system

Outline

- A brief history of speech algorithms
- Course goals
- Course outlines
- Demo: a simple speech recognition system

A brief history: the “machines”

- Wolfgang von Kempelen's speaking machine – 18th century

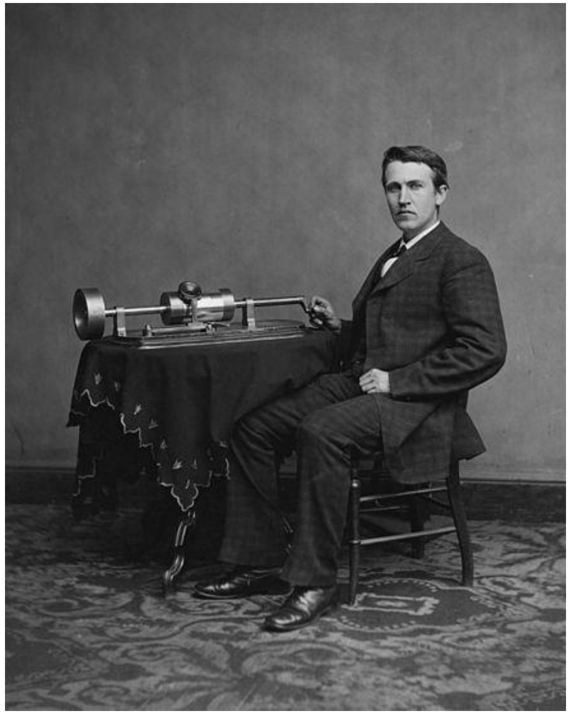


A replica of Kempelen's speaking machine, built 2007–09 at the Department of Phonetics, Saarland University, Saarbrücken, Germany

A brief history: the “machines”

- Wolfgang von Kempelen's speaking machine – 18th century

- Thomas Edison’s phonograph – 19th century



Thomas Edison with his second phonograph, photographed by Levin Corbin Handy in Washington, April 1878

A brief history: the “machines”

- Wolfgang von Kempelen's speaking machine – 18th century
- Thomas Edison’s phonograph – 19th century
- “Radio Rex” the commercial toy – 1910s

Radio Rex from 1910s - The first speech recognition commercial toy

A brief history: entering the modern era

- Audrey created in 1952 - First known and documented speech recognizer

A team at Bell Labs designs the Audrey, a machine capable of understanding spoken digits.

A brief history: entering the modern era

- Audrey created in 1952 - First known and documented speech recognizer

- Template based models were heavily used

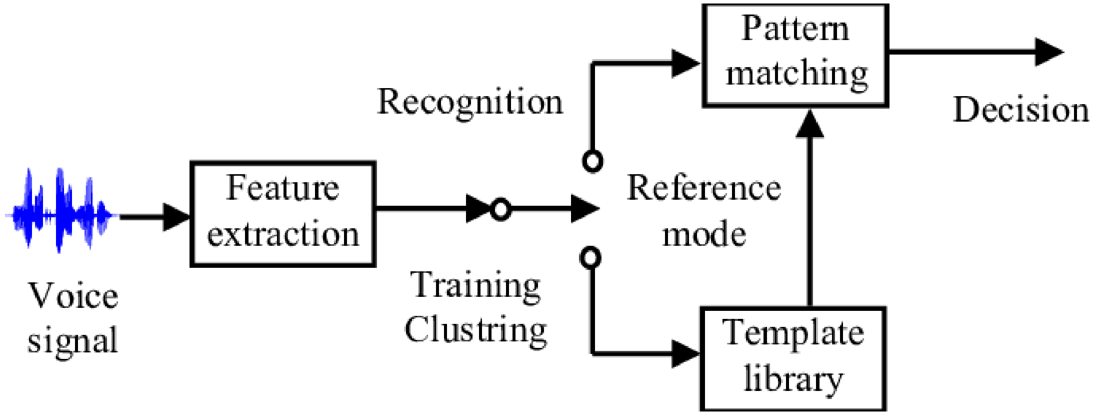


Diagram of template based models.
Photo credit: https://www.researchgate.net/

A team at Bell Labs designs the Audrey, a machine capable of understanding spoken digits.

A brief history: statistical models

- Speech recognition in 1970s – “assists” from information theory

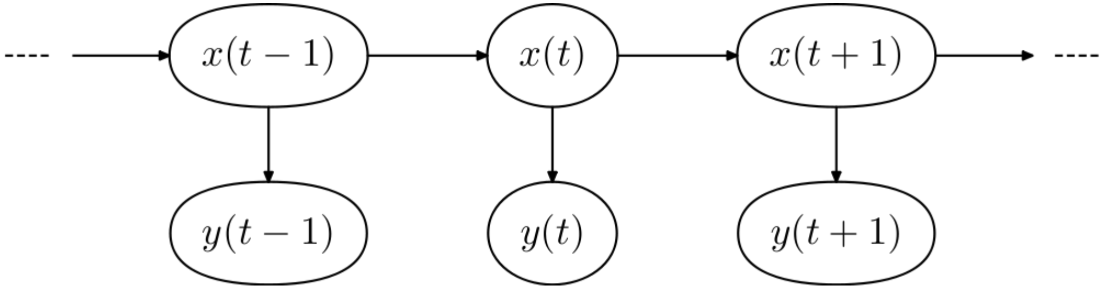
Frederick Jelinek, pioneer in statistical speech recognition.

“Airplanes don’t flap their wings.”

“Every time we fire a phonetician/linguist, the performance of our system goes up.”

A brief history: statistical models

- Speech recognition in 1970s – “assists” from information theory
- Speech recognition in 1980s – the rise of Hidden Markov Models and Gaussian Mixture models



Hidden Markov Models

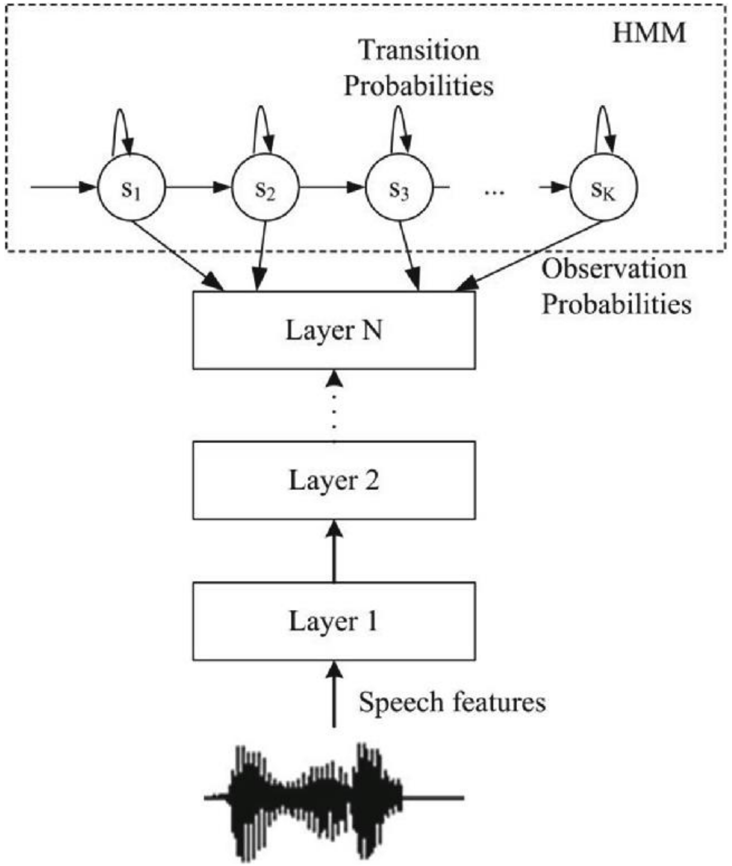
Frederick Jelinek, pioneer in statistical speech recognition.

“Airplanes don’t flap their wings.”

“Every time we fire a phonetician/linguist, the performance of our system goes up.”

A brief history: neural network models

- Neural network models in 1980s – “hybrid” model

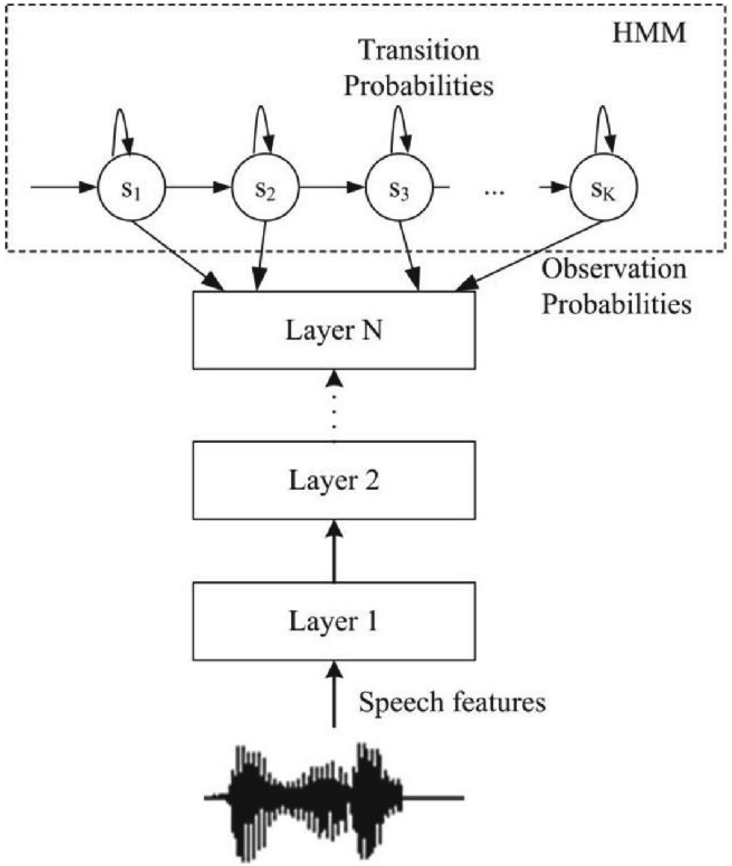


Architecture of the DNN-HMM hybrid system.
Photo credit: https://www.researchgate.net/

A brief history: neural network models

- Neural network models in 1980s – “hybrid” model

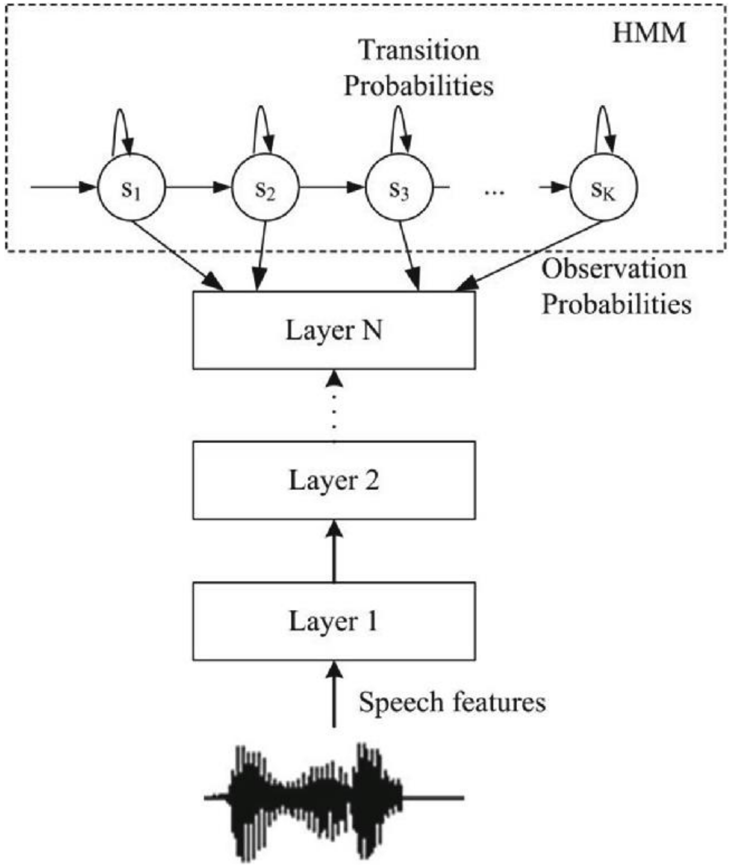
- Neural network models in 2010s – Microsoft researchers made hybrid model work for speech recognition



Architecture of the DNN-HMM hybrid system.
Photo credit: https://www.researchgate.net/

A brief history: neural network models

- Neural network models in 1980s – “hybrid” model
- Neural network models in 2010s – Microsoft researchers made hybrid model work for speech recognition
- Neural network models after 2014 – Google researched proposed end to end neural network models



Architecture of the DNN-HMM hybrid system.
Photo credit: https://www.researchgate.net/

A brief history: applications

- 1990, Dragon Dictate by Dragon Systems

James and Janet Baker, founders of Dragon Systems, a pioneering voice recognition technology company. Photo credit: New York Times

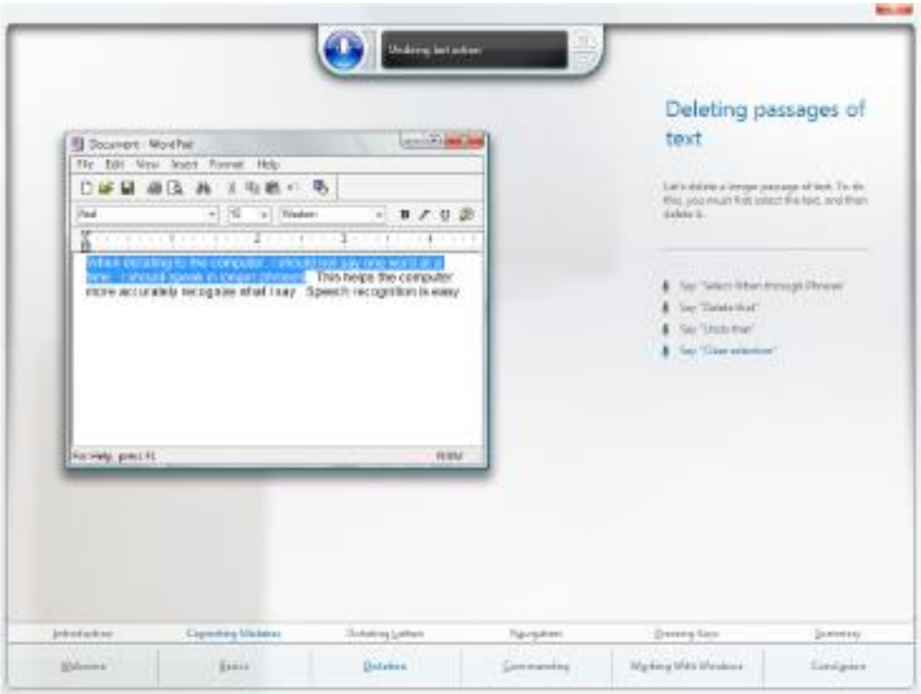
A brief history: applications

- 1990, Dragon Dictate by Dragon Systems
- 1997, Dragon NaturallySpeaking by Dragon Systems

Dragon Naturally Speaking is still on market.

A brief history: applications

- 1990, Dragon Dictate by Dragon Systems
- 1997, Dragon NaturallySpeaking by Dragon Systems
- 2000s, voice input on Windows, Mac OS X, etc.

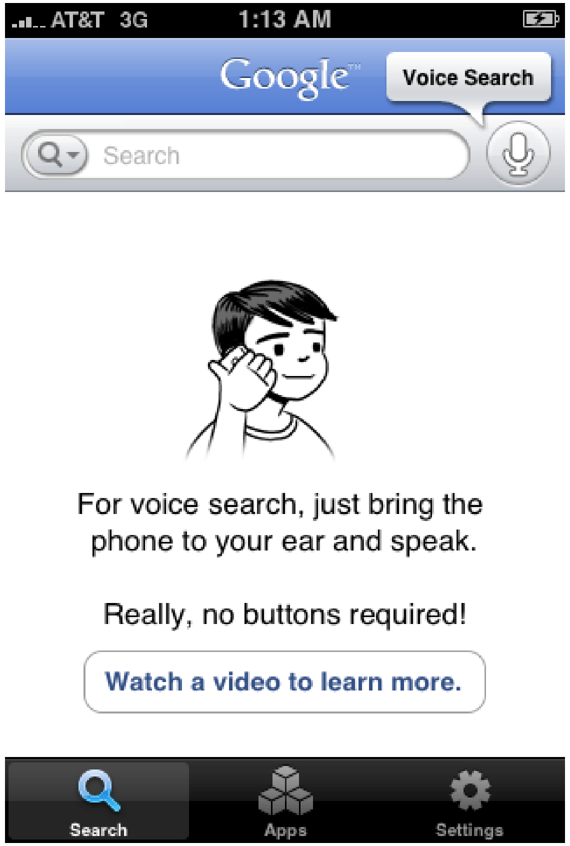


When speaking to the computer, I should not say one word at a time. I should speak in longer phrases. This helps the computer more accurately recognize what I say. Speech recognition is easy

The tutorial for Windows Speech Recognition in Windows Vista.

A brief history: applications

- 1990, Dragon Dictate by Dragon Systems
- 1997, Dragon NaturallySpeaking by Dragon Systems
- 2000s, voice input on Windows, Mac OS X, etc.
- 2008, voice search by Google



For voice search, just bring the phone to your ear and speak.

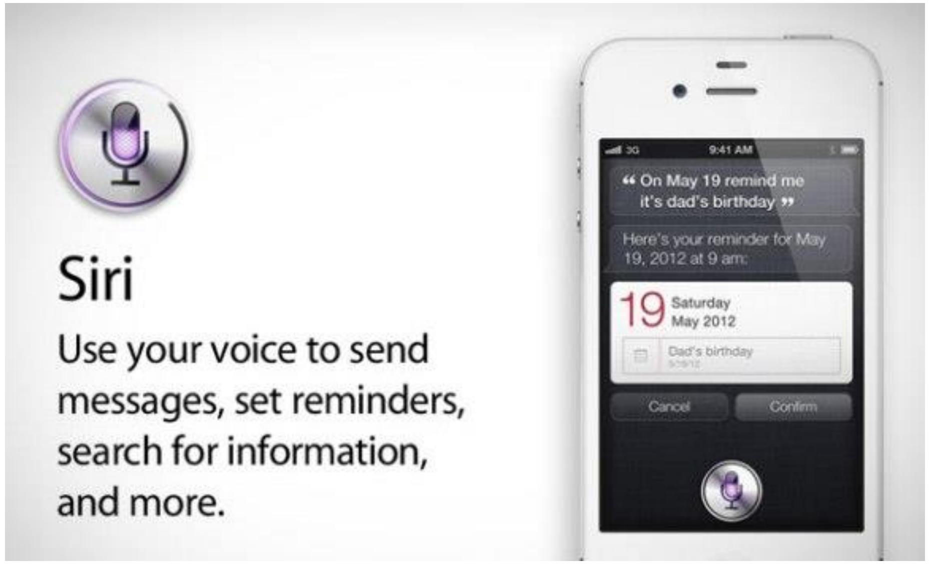
Really, no buttons required!

Watch a video to learn more.

In 2008, Google rolled out voice search on iOS devices.

A brief history: applications

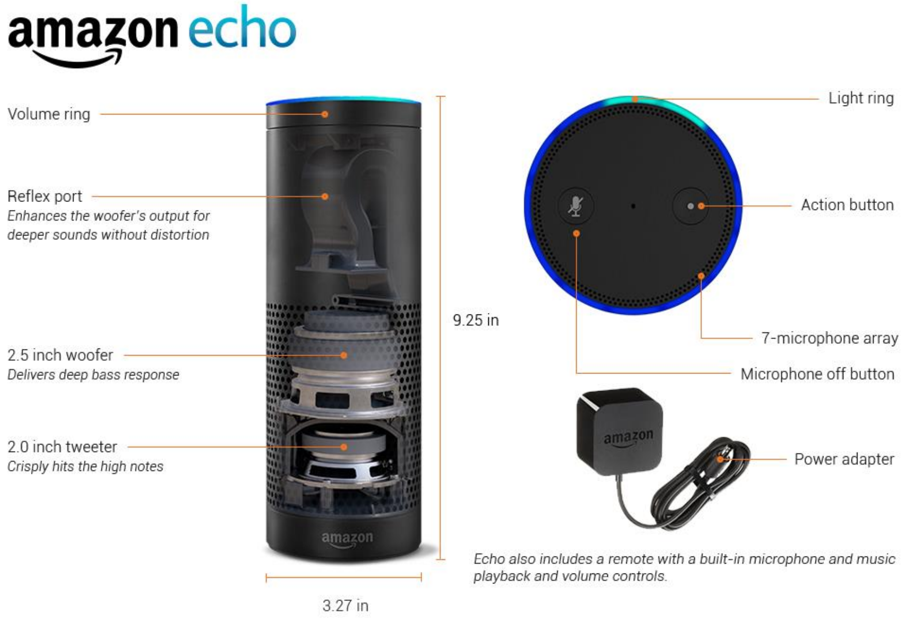
- 1990, Dragon Dictate by Dragon Systems
- 1997, Dragon NaturallySpeaking by Dragon Systems
- 2000s, voice input on Windows, Mac OS X, etc.
- 2008, voice search by Google
- 2011, Siri on iPhone 4S by Apple


Siri
Use your voice to send messages, set reminders, search for information, and more.
9:41 AM
“On May 19 remind me it's dad's birthday”
Here's your reminder for May 19, 2012 at 9 am:
19 Saturday May 2012
Dad's birthday
Cancel
Confirm

In 2011, Apple released Siri, the personal voice assistant on iPhone 4S.

A brief history: applications

- 1990, Dragon Dictate by Dragon Systems
- 1997, Dragon NaturallySpeaking by Dragon Systems
- 2000s, voice input on Windows, Mac OS X, etc.
- 2008, voice search by Google
- 2011, Siri on iPhone 4S by Apple
- 2014, Echo by Amazon
- … …


amazon echo
Volume ring
Reflex port
Enhances the woofer's output for deeper sounds without distortion
2.5 inch woofer
Delivers deep bass response
2.0 inch tweeter
Crisply hits the high notes
amazon
9.25 in
3.27 in
Light ring
Action button
7-microphone array
Microphone off button
amazon
Power adapter
Echo also includes a remote with a built-in microphone and music playback and volume controls.

In 2014, Amazon released Echo, a speaker with far-field voice assistant.

A brief history: applications

- 1990, Dragon Dictate by Dragon Systems
- 1997, Dragon NaturallySpeaking by Dragon Systems
- 2000s, voice input on Windows, Mac OS X, etc.
- 2008, voice search by Google
- 2011, Siri on iPhone 4S by Apple
- 2014, Echo by Amazon
- … …

The surge of smart speakers in China

Outline

- A brief history of speech algorithms
- Course goals
- Course outlines
- Demo: a simple speech recognition system

Course goals

- Basic theories of speech algorithms

Course goals

- Basic theories of speech algorithms
- Practical issues in speech applications

Course goals

- Basic theories of speech algorithms

- Practical issues in speech applications

- Hands-on exercises

Outline

- A brief history of speech algorithms
- Course goals
- Course outlines
- Demo: a simple speech recognition system

Course outlines

- Speech recognition
- Wake word detection
- Speaker recognition
- Speech synthesis

Course outlines: speech recognition

- The task: transcribe human voice into text
  - Clean v.s. noisy
  - Close talk v.s. far-field
  - Reading v.s. spontaneous
- The applications
  - Call center
  - Voice search
  - Voice input
  - Voice assistant
  - … …

Course outlines: speech recognition

- What you will learn
  - GMM-HMM speech recognition systems
  - DNN-HMM speech recognition systems
  - End-to-end speech recognition systems
  - Attention/Transformer based speech recognition systems
  - Unsupervised learning for speech recognition
  - Speech recognition production systems
  - On-device speech recognition

Course outlines: wake word detection

- The task: detect given keywords
  - Single keyword
  - Voice commands
  - Low computation resource
  - High accuracy
- The applications
  - Smart speakers
  - Voice assistants
  - Car applications
  - … …

Course outlines: wake word detection

- What you will learn
  - Template based wake word detection
  - HMM based wake word detection
  - DNN based wake word detection
  - Practical tricks in production system

Course outlines: speaker recognition

- The task: identify speaker from human voice
  - Speaker identification (1:N match)
  - Speaker verification (1:1 match)
- The applications
  - Recommendations (music, information, etc)
  - Payment
  - Identification
  - … …

Course outlines: speaker recognition

- What you will learn
  - I-vector based speaker recognition
  - D-vector based speaker recognition
  - X-vector based speaker recognition
  - Speaker recognition and face recognition
  - Practical tricks in production system

Course outlines: speech synthesis

- The task: generate human voice from text
  - Synthesis
  - Voice clone

- The applications
  - Voice assistants
  - Audio books
  - Toys
  - … …

Course outlines: speech synthesis

- What you will learn
  - Traditional speech synthesis systems
  - WaveNet speech synthesis system
  - Tacotron speech synthesis system
  - Practical tricks in production system

Outline

- A brief history of speech algorithms
- Course goals
- Course outlines
- Demo: a simple speech recognition system

Demo

Please follow the instructor.

Thanks!