

11-737 Multilingual NLP

Speech Lectures



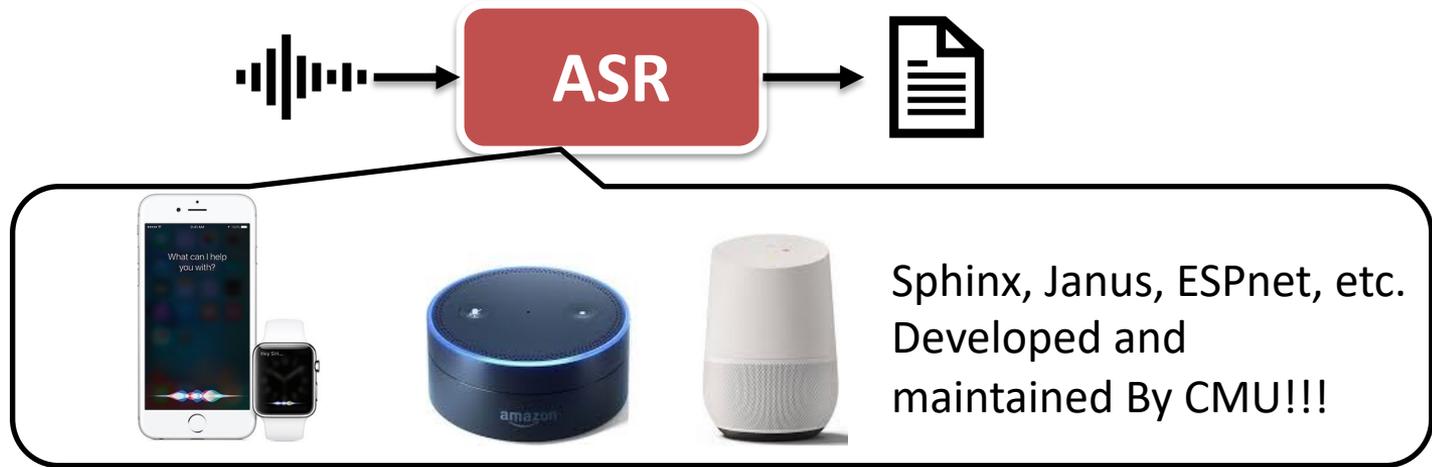
Carnegie Mellon University

Language Technologies Institute

Schedule

- **Class 13, 3/1:** Speech -- *Watanabe*
 - [L] Wave Forms, Phonemes, Syllables, Tone
 - [D] Speech datasets
 - **Assignment 3 Assigned**
- **Class 14, 3/3:** Automatic Speech Recognition -- *Watanabe*
 - [T] Automatic Speech Recognition
 - [M] ASR models
- **Class 15, 3/8:** Sequence-to-Sequence Models for Speech Processing -- *Watanabe*
 - [M] Connectionist Temporal Classification
 - [M] Attention-based models for ASR/TTS
- **Class 16, 3/10:** Text-to-speech -- *Black*
 - [T] Text-to-speech
 - [M] TTS models
- **Class 17, 3/15:** Multilingual ASR and TTS -- *Watanabe*
 - [M] Models for multilingual ASR and TTS
- **Assignment 3 Due 3/21**

Automatic Speech Recognition (ASR)



Widely used in many applications!

Speech Synthesis (TTS: Text to Speech)



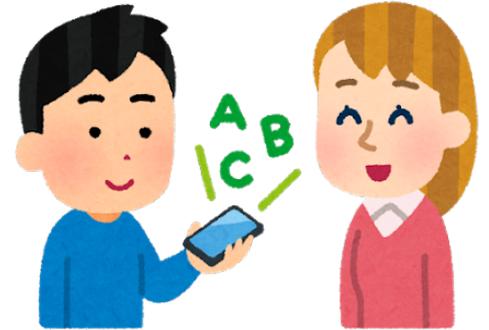
Inverse problem of ASR

Remarks

- The course mainly covers high-level explanations and system descriptions of **ASR**, **TTS**, and related technologies
 - If you want to know more about them, please consider to take “11-751 Speech Recognition and Understanding” and “11-492 Speech Processing” 😊
- **Most of ASR and TTS technologies are studied with major languages** (English, Chinese, German, French, Japanese, etc.)
 - Rich resources, accumulated knowhow, marked priority
- **What can you learn? The lectures will focus toward how to build ASR/TTS systems in *any* language**
 - In Assignment 3, we will ask you to pick up one language and build an ASR system based on ESPnet: <https://github.com/espnet/espnet>
 - Focus a bit more on end-to-end ASR

One of the ultimate goals of human language technologies:

Speech to Speech Translation



Combining **ASR** + **machine translation** + **TTS**

Not directly covered in the lecture but all core technologies are covered in this lecture