

Speech Recognition

M.Tech. Artificial Intelligence, Second Year, NMIMS

By,

Bilal Hungund, Data Scientist, Halliburton

Exciting recent developments have disrupted this field



Amazon Alexa +
Alexa Prize
2014



Neural TTS voice cloning
2017



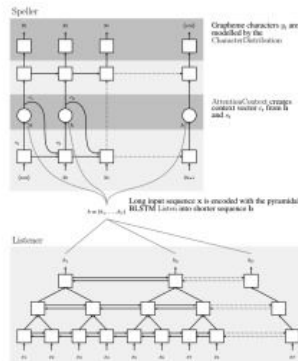
Apple
Siri
2011



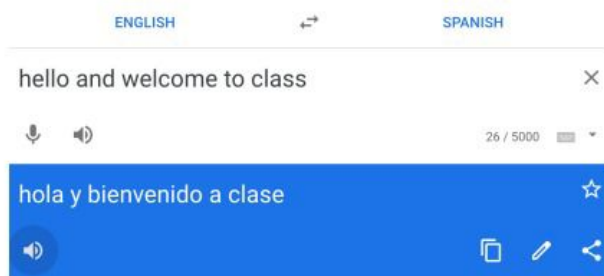
Google
Assistant
2016



Microsoft
Cortana
2014



End-to-end neural becomes SOTA
2015 - present



Realtime speech-speech translation
2020

Youtube Sources

Speech and its Types

- What is speech and speech processing
- Digitization and recording of speech
- Acoustic and articulatory phonetics

Speech: Means of communication or expression of thought and feelings by articulating sounds

Why speech recognition?

Aspects of speech :

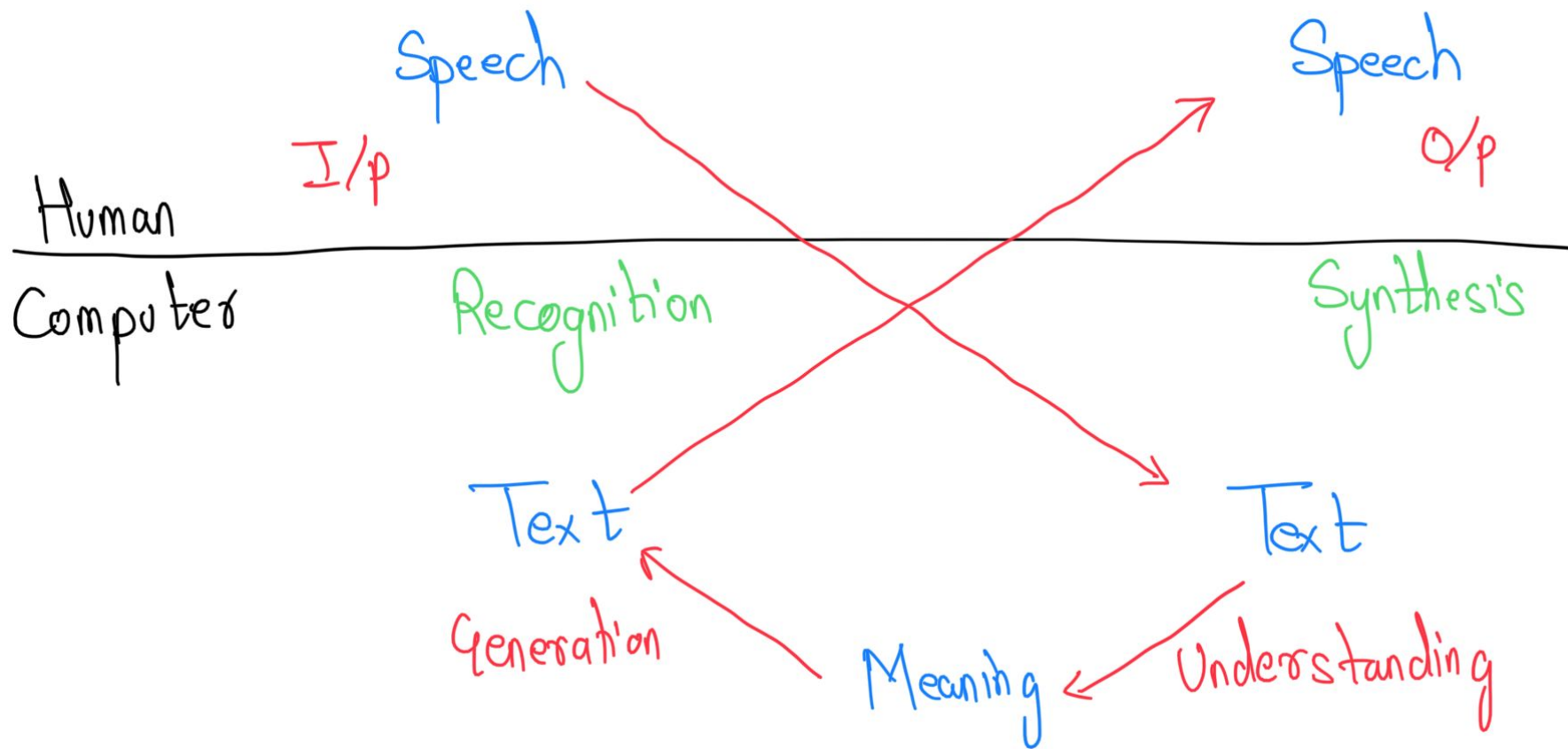
1) Perceptual Variability
Language, dialects, accents

2) Intricate Variability
Speech signal, amplitude, pitch,
duration

Difference between speech & other data

→ Not contains Invariant Units

→ Produce in a non-linear way



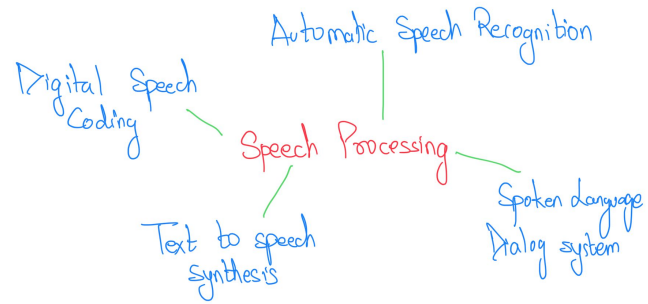
Categories of speech sounds

- Voicing: Vocal cords will be vibrating
- Frication: Constriction and Turbulence in vocal tracts
- Plosion: Air pressure built and released in a sudden noisy burst

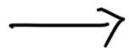
Speech Processing

Study of speech signals and methods of processing

Intersection between DSP and NLP



Raw or
Tagged
Text



Text
Analysis

- Document Structure
- Text Normalization
- Linguistic Analysis



Phonetic
Analysis

- Homograph disambiguation
- Grapheme-Phoneme Conversion



Prosodic
Analysis

- Pitch
- Duration



Speech
Analysis

- Voice Rendering

SYNTHESIS

Speech Production

It produce through coordinated movements of articulators

Spectral Characteristics

→ Source

→ Vocal Tract shape: Access resonator

→ Radiation of Sound

Filters for vocal
cords

Phonological Variation

↳ Individual Differences

↳ Environmental Factors

↳ Diversity of Languages

↳ Real World Issues

Vowels — Semi Vowels — Nasals — Fricatives — Stops

Acoustics Phonetics :

Physical properties of speech . Sound waves
(Frequencies, amplitudes, durations)

→ Waveform : Analyze Acoustic properties

→ Tuning Fork : Air Fluctuations & Oscillations

Articulatory Phonetics

→ Initiation → Phonation → Articulation