J.A.R.V.I.S

Li Fang Chan
Katrin Jasim

Table of Contents

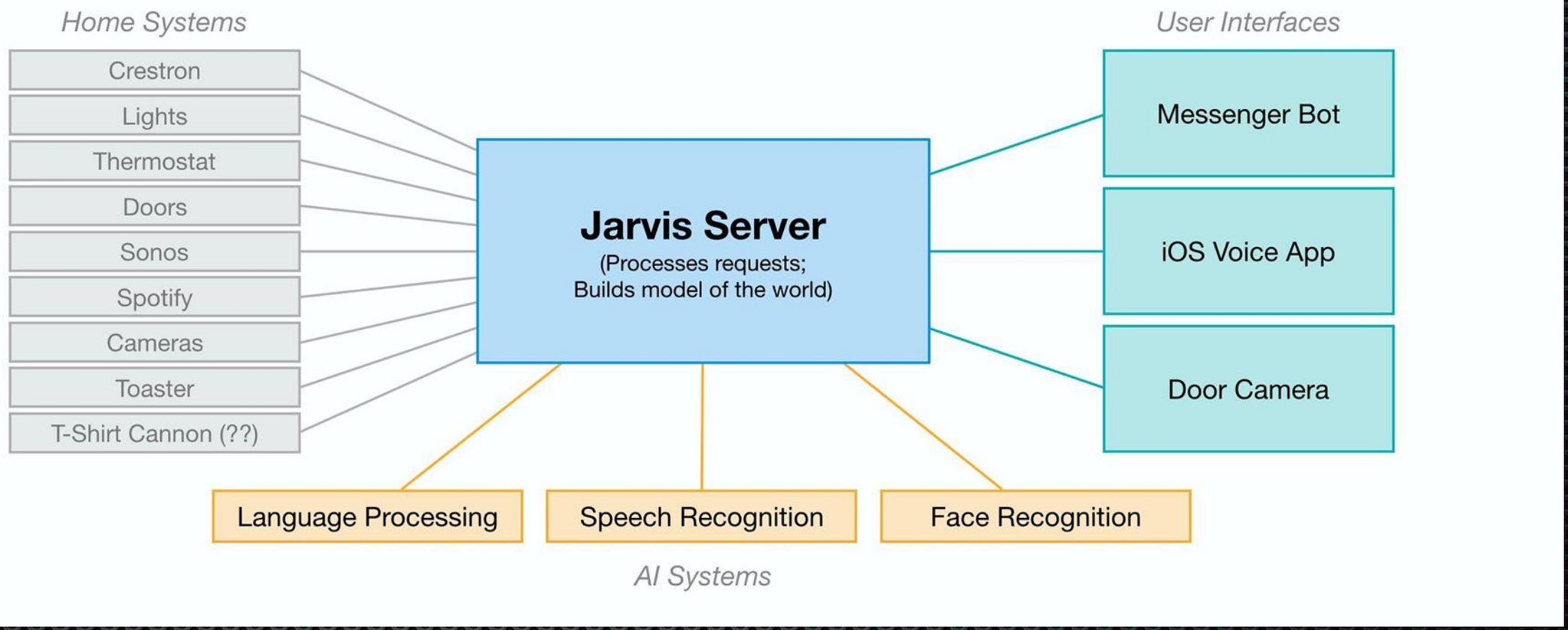
- Introduction
  - J.A.R.V.I.S (Iron Man)
  - Jarvis (Mark Zuckerberg)
- Natural Language Processing
  - Natural Language Understanding
    - Speech Recognition
    - Hidden Markov Models and Viterbi Algorithm
- Voice recognition
- Facial Recognition
- Sources

J.A.R.V.I.S (Iron Man / Tony Stark)

- Just A Rather Very Intelligent System
- was a natural-language user interface computer system
- Possesses complex human-like capabilities and becomes increasingly human-like with time
- Speech and voice recognition, and complex motion detection form the foundation of J.A.R.V.I.S

Jarvis (Mark Zuckerberg)

- Similar to J.A.R.V.I.S
- Based on Natural Language Processing, Speech and Voice Recognition, and Facial recognition
- Communication via Jarvis Bot (Messenger bot) and Jarvis Server (Computer)
- Controls home
- Continuously learns his tastes and patterns and improves based on that

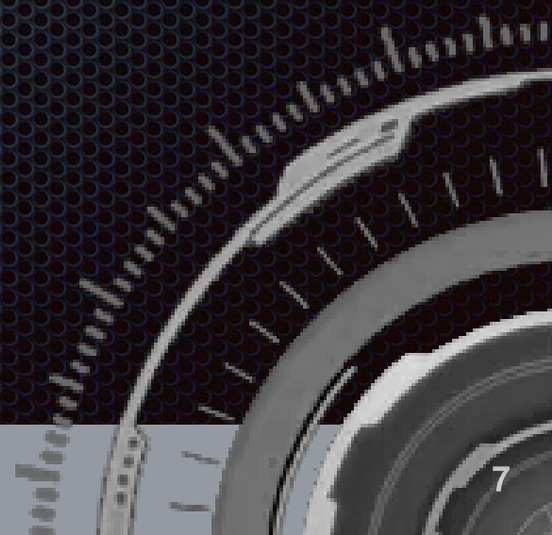

Home Systems
Crestron
Lights
Thermostat
Doors
Sonos
Spotify
Cameras
Toaster
T-Shirt Cannon (??)
Jarvis Server
(Processes requests;
Builds model of the world)
User Interfaces
Messenger Bot
iOS Voice App
Door Camera
Language Processing
Speech Recognition
Face Recognition
AI Systems

NLP : Natural Language Processing

- branch of artificial intelligence that helps computers understand, decode and manipulate human language
- Natural Language Understanding (NLU) and Natural Language Generation (NLG)

NLU: Natural Language Understanding

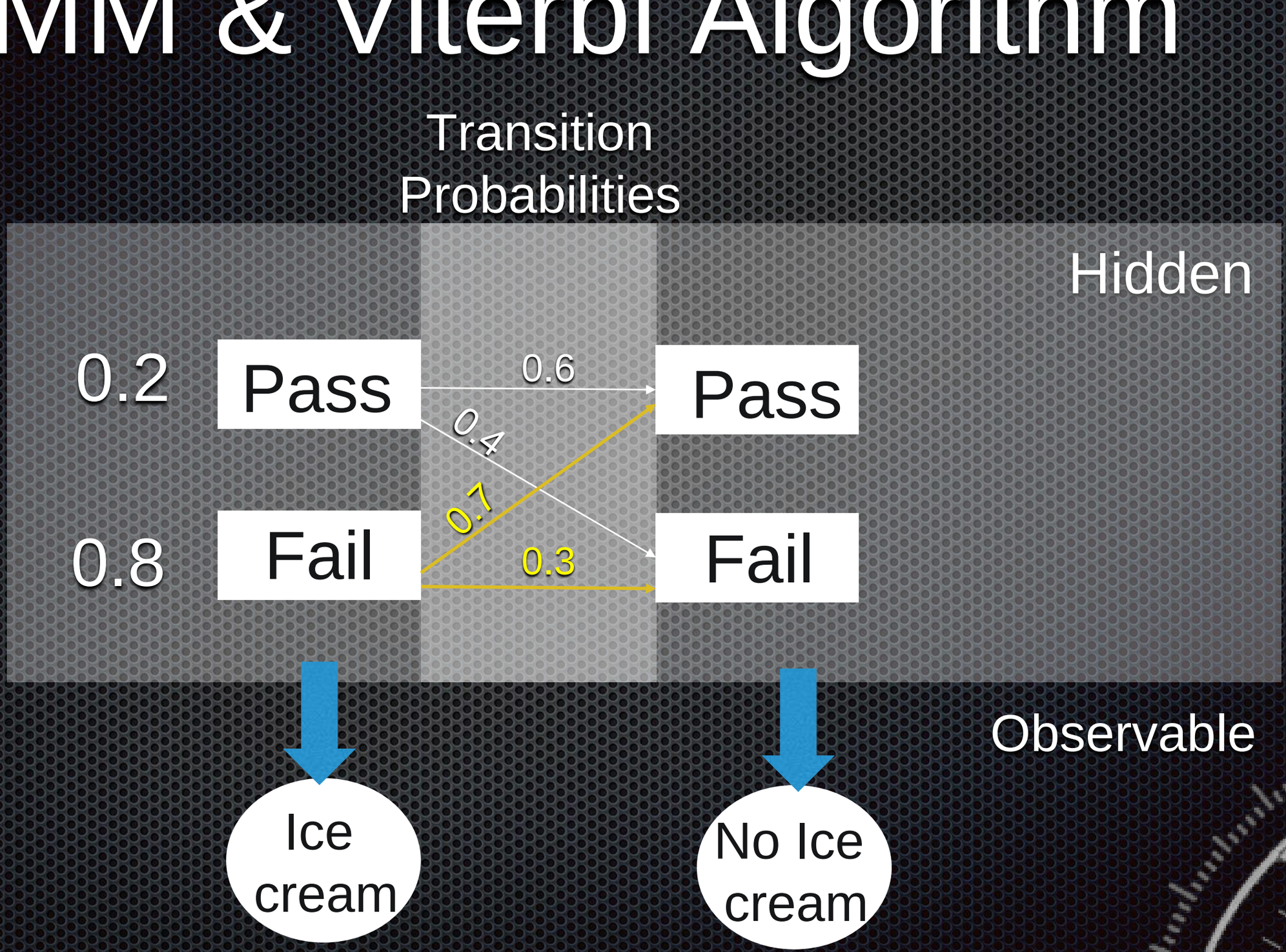
- Step 1: natural language converted into artificial language (Speech Recognition)
- Step 2: NLU tries to understand meaning of the text
- Mostly based on Hidden Markov Models



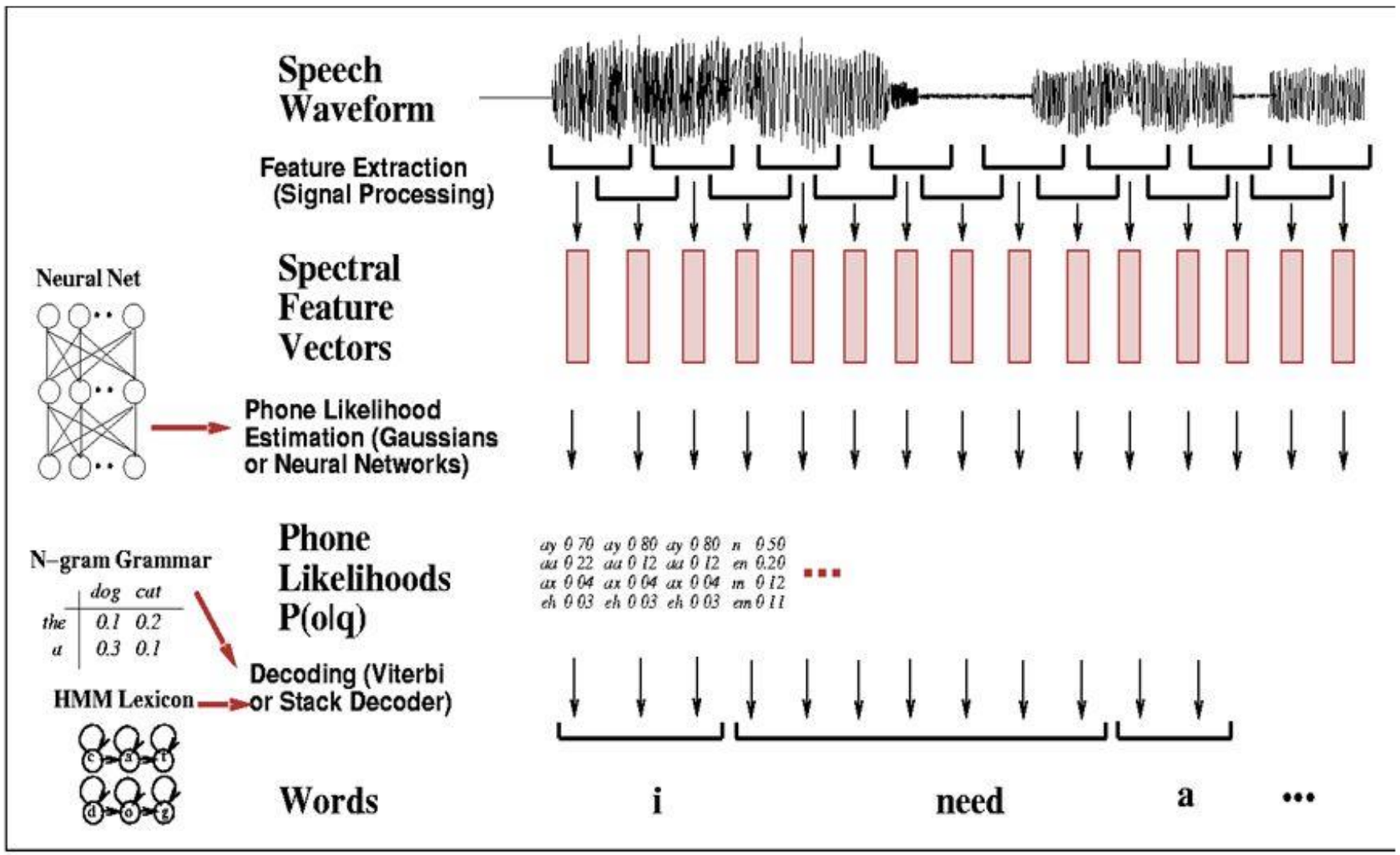
NLU – Step 1
Speech Recognition

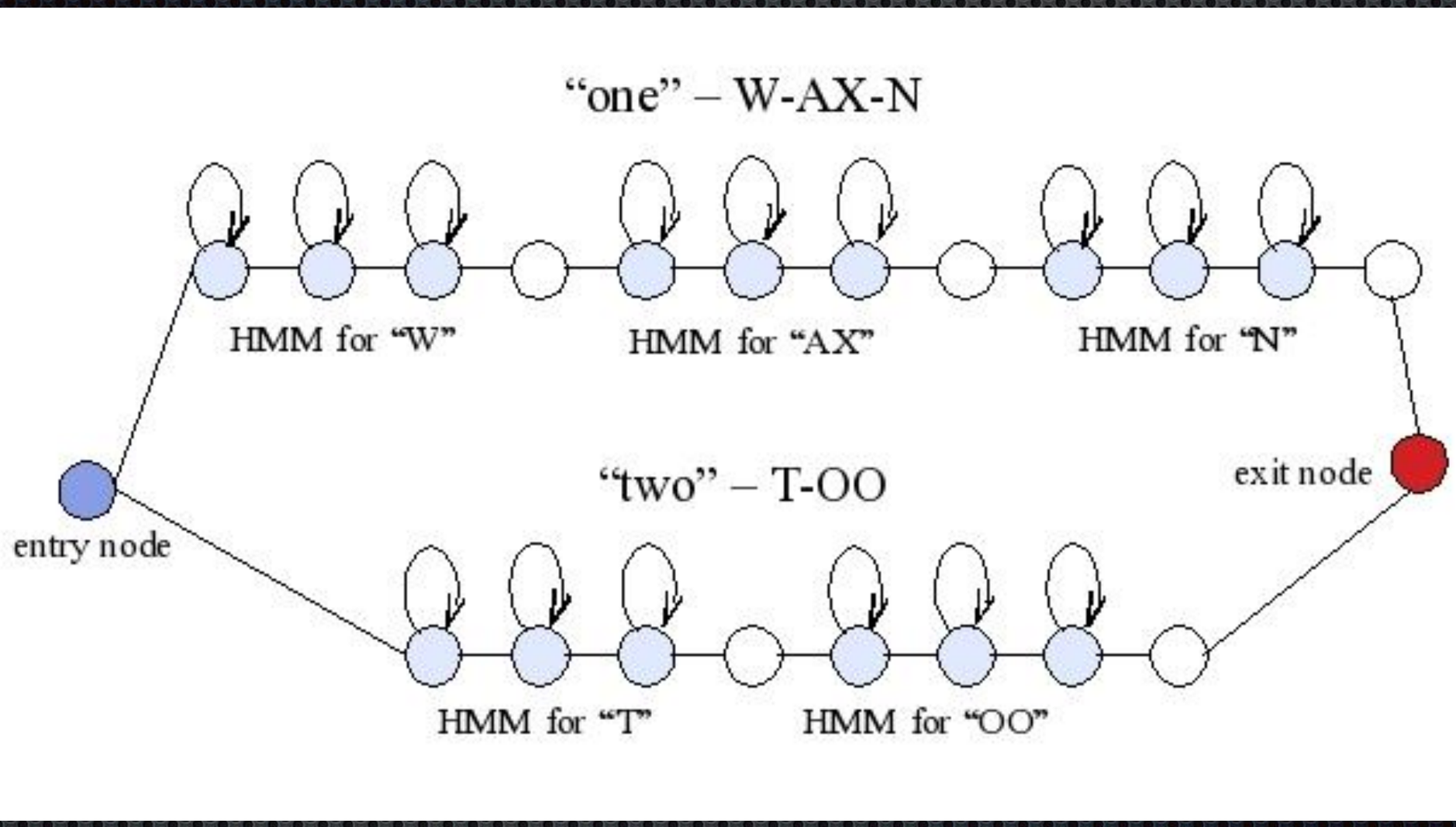
- Transforms sound waves into data
- Identify what is being said (syntax)
- Identify what is meant (semantics)
- Hidden Markov Models
- Viterbi algorithm – finds most likely sequence of hidden states

HMM & Viterbi Algorithm



Schematic Architecture for a (simplified) Speech Recognizer




"one" – W-AX-N
HMM for "W"
HMM for "AX"
HMM for "N"
entry node
exit node
"two" – T-OO
HMM for "T"
HMM for "OO"

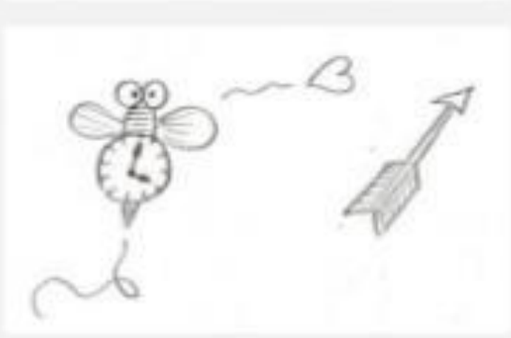
NLU – Step 2 Understanding

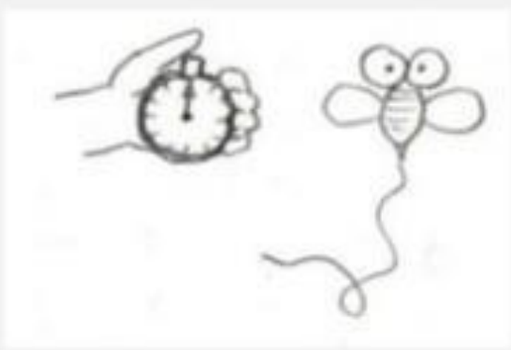
- different NLU systems use different techniques
- generally:

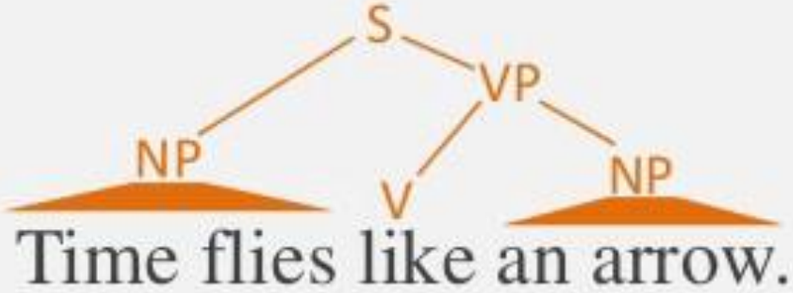
1. Part-of-Speech tagging (POS): computer must understand what each word is – subject, verb, noun etc

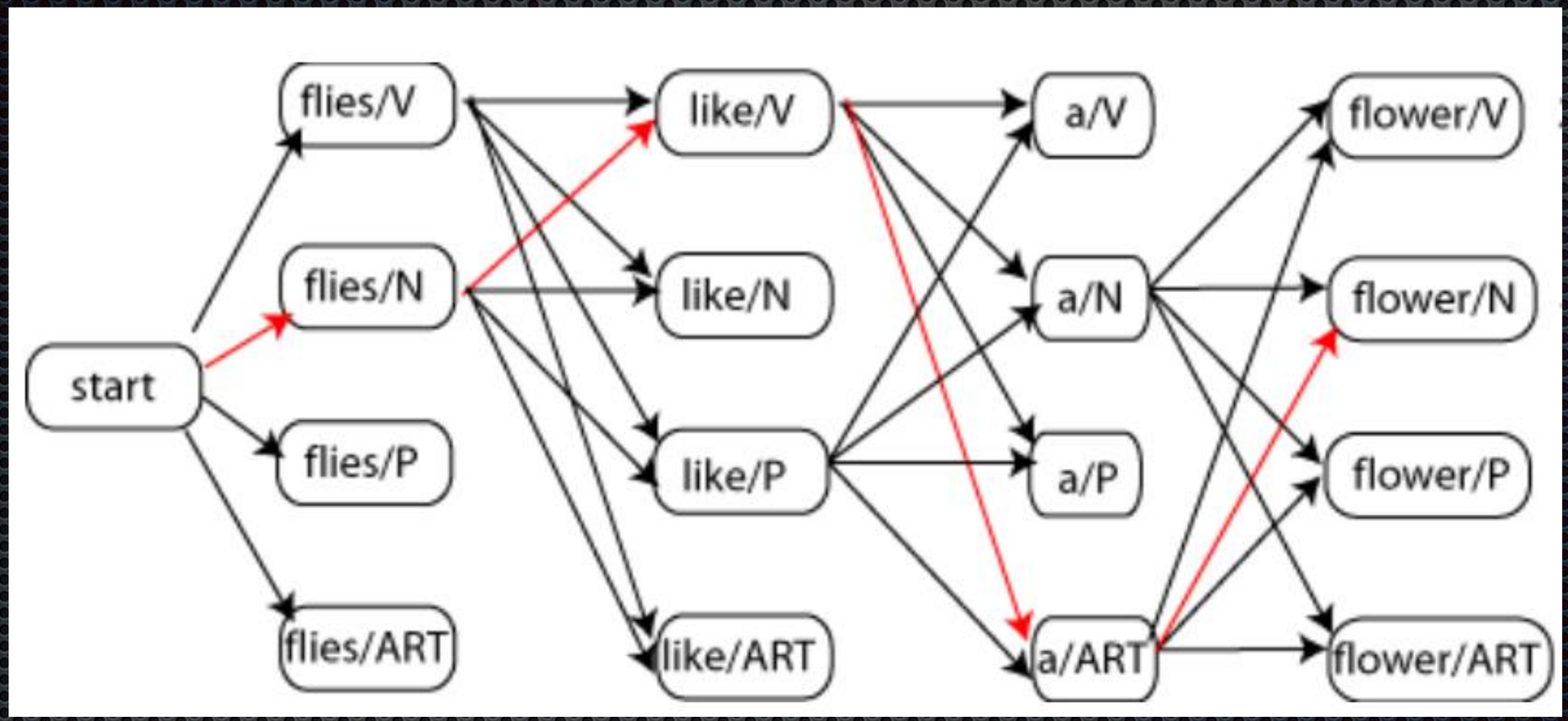
2. Modern NLP algorithms use statistical machine learning to apply a set of grammar rules to the natural language and determine the most likely meaning behind what was said (HMMs)

Syntactic Ambiguity








start
flies/V
flies/N
flies/P
flies/ART
like/V
like/N
like/P
like/ART
a/V
a/N
a/P
a/ART
flower/V
flower/N
flower/P
flower/ART

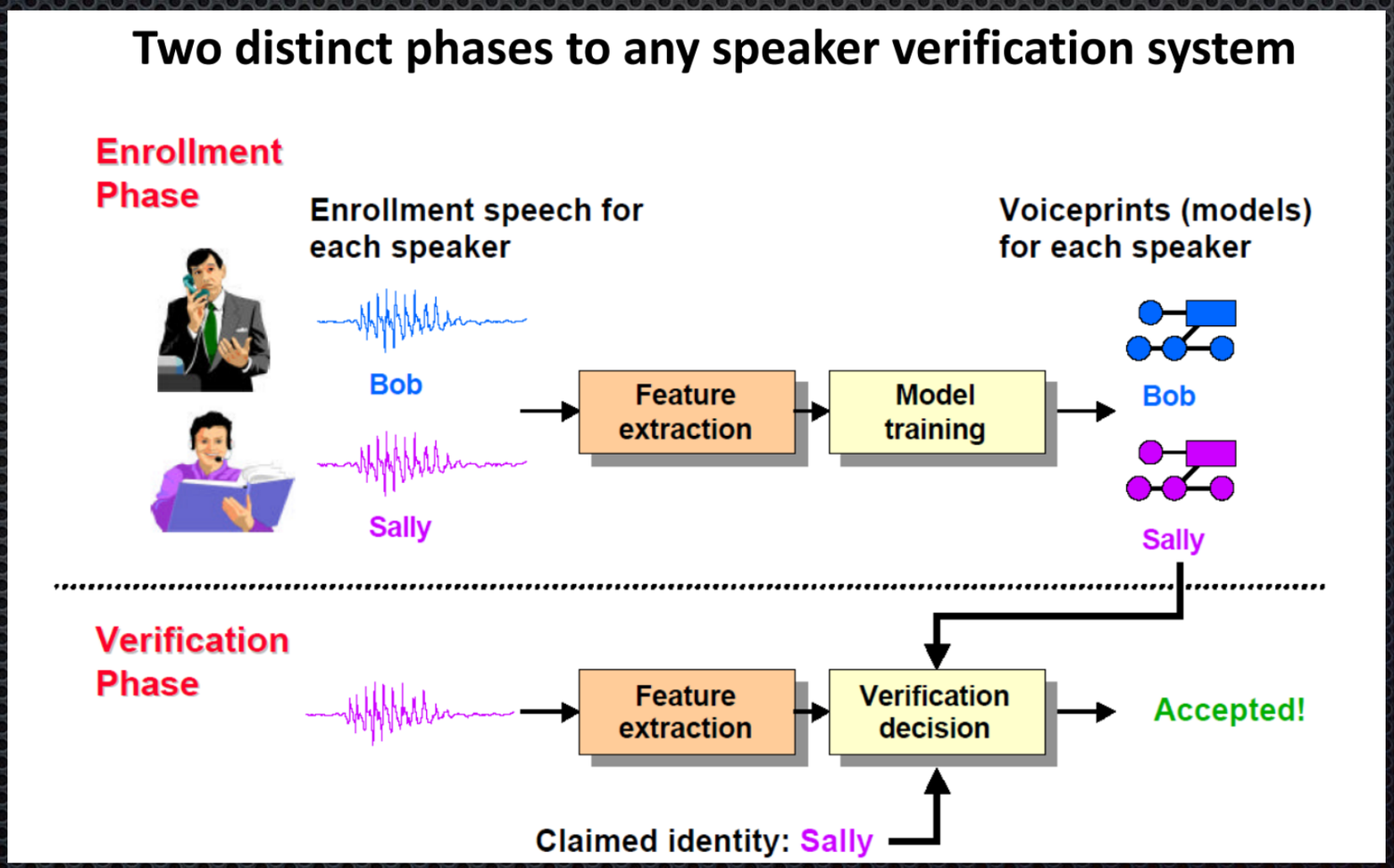
Pope Francis' Baby Steps on Gays - TIME

content.time.com › magazine › article

12 Aug 2013 · The Pope's Baby Step on Gays . Francis' encouraging words show how far the Roman Catholic Church still has to go. By Gene ...

Voice recognition

- Jarvis can tell who is talking (digital profile of a persons speech)

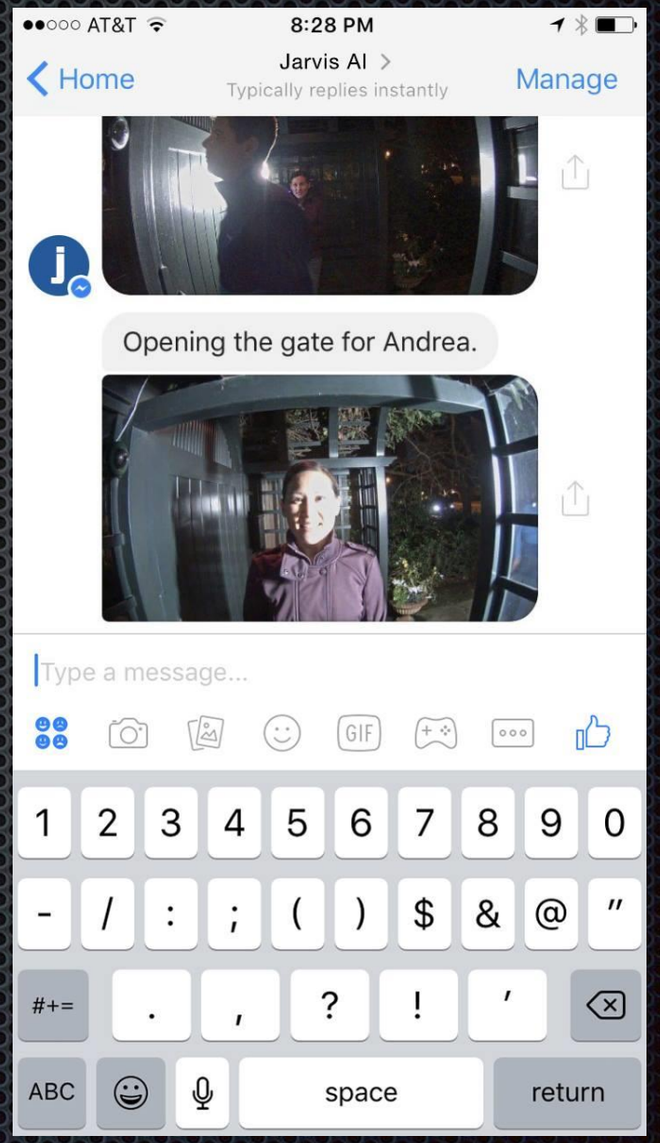


Voice Recognition

- Low level features - nasality, breathiness or roughness
- High level features - rhythm, intonation, idiolect
- Pattern matching - comparison of the extracted frames with known speaker models
- Assign extracted features to HMM states

Facial recognition

- Jarvis knows who it can let into the house



Facial recognition

- identifying or verifying a person from a digital image or a video frame from a video source
- human face has 80 nodal points
- data is stored as "face print"
- 3D recognition provides more accuracy than 2D counterpart
- difficult to fool

Facial recognition

- Jarvis' face recognition system based on DeepFace (97.25% accuracy)
- numerous cameras
- simple server continuously watches the cameras:
  - Step 1: face detection
  - Step 2: face recognition
- information is compared with a database of known faces to find a match

Sources

https://www.sas.com/en_us/insights/analytics/what-is-natural-language-processing-nlp.html

https://hackernoon.com/building-a-facial-recognition-pipeline-with-deep-learning-in-tensorflow-66e7645015b8

https://blog.neospeech.com/what-is-natural-language-processing/

http://www.ll.mit.edu/mission/communications/ist/publications/aaas00-dar-pres.pdf

http://research.cs.tamu.edu/prism/lectures/sp/l16.pdf

https://www.engineersgarage.com/articles/speech-recognition

http://www.cse.unsw.edu.au/~billw/cs9414/notes/nlp/ambiguity/ambiguity-2009.html

http://www.cs.tut.fi/~sgn24006/PDF/L09-HMMs-in-speech-recognition.pdf

https://www.facebook.com/notes/mark-zuckerberg/building-jarvis/10154361492931634/

http://dipayandev.blogspot.com/2012/03/speech-recognition-by-java-sphinx4.html

https://devopedia.org/speech-recognition

http://www.biometric-solutions.com/speaker-recognition.html