Speech Recognition

BY

Charu joshi

Introduction

- What is Speech Recognition?
  - also known as *automatic speech recognition* or *computer speech recognition* which means understanding voice by the computer and performing any required task.

- Where can it be used?

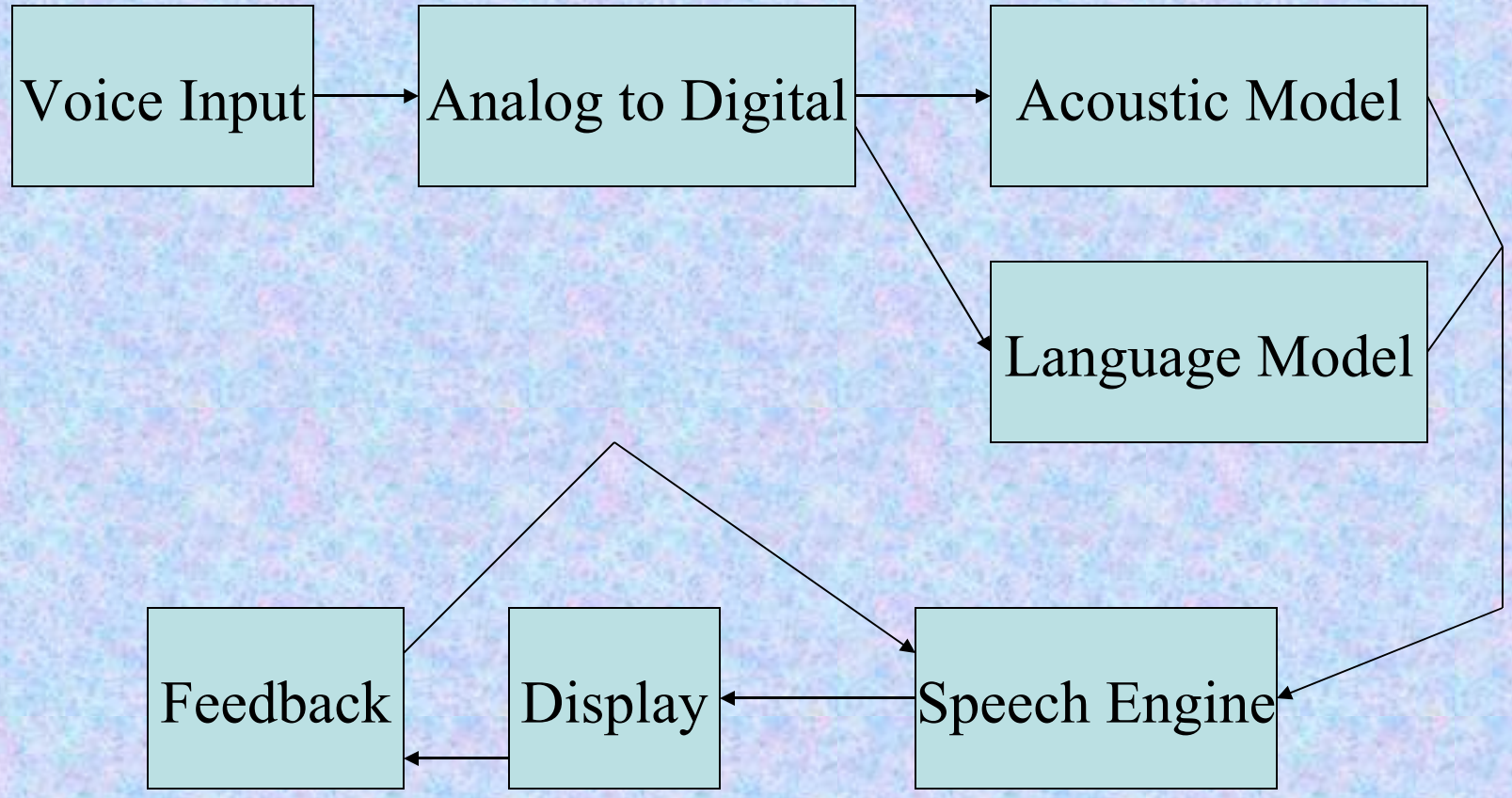
  - Dictation

  - System control/navigation

  - Commercial/Industrial applications

  - Voice dialing

Recognition



- Acoustic Model
  - An acoustic model is created by taking audio recordings of speech, and their text transcriptions, and using software to create statistical representations of the sounds that make up each word. It is used by a speech recognition engine to recognize speech.
- Language Model
  - Language modeling is used in many natural language processing applications such as speech recognition tries to capture the properties of a language, and to predict the next word in a speech sequence.

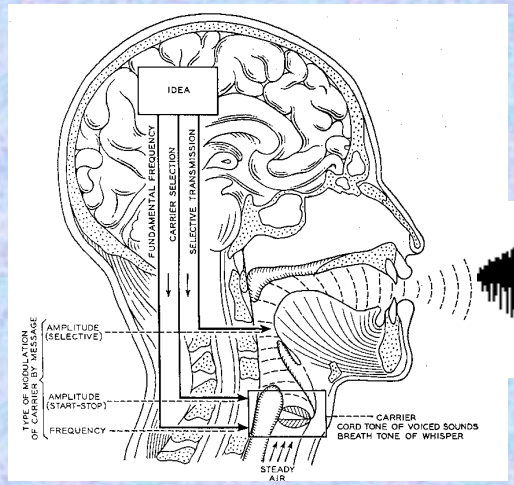
TYPES OF VOICE RECOGNITION

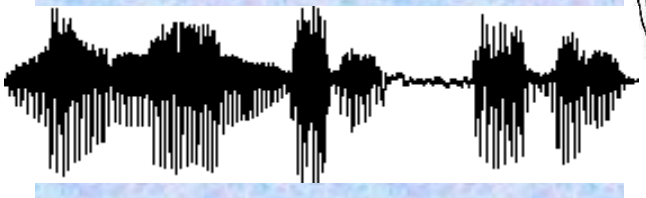
- There are two types of speech recognition. One is called *speaker-dependent* and the other is*speaker-independent* . Speaker-dependent software is commonly used for dictation software, while speaker-independent software is more commonly found in telephone applications.
- Speaker-dependent software works by learning the unique characteristics of a single person's voice, in a way similar to voice recognition. New users must first "train" the software by speaking to it, so the computer can analyze how the person talks. This often means users have to read a few pages of text to the computer before they can use the speech recognition software.

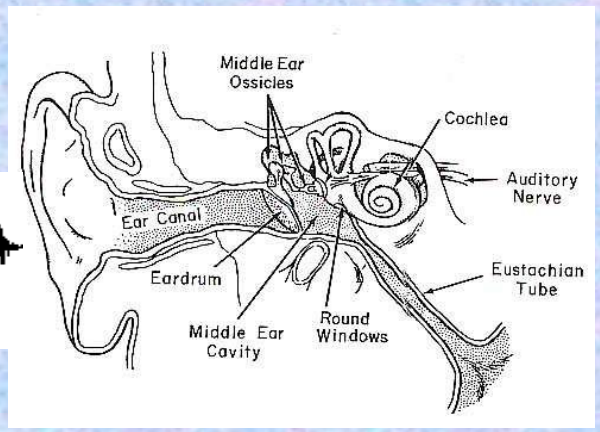
TYPES OF VOICE RECOGNITION

- Speaker-independent software is designed to recognize anyone's voice, so no training is involved. This means it is the only real option for applications such as interactive voice response systems — where businesses can't ask callers to read pages of text before using the system. The downside is that speaker-independent software is generally less accurate than speaker-dependent software.
- Speech recognition engines that are speaker independent generally deal with this fact by limiting the grammars they use. By using a smaller list of recognized words, the speech engine is more likely to correctly recognize what a speaker said.

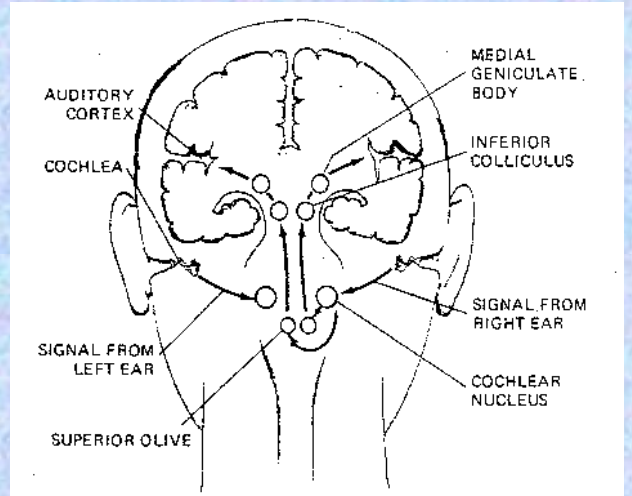
How do humans do it?



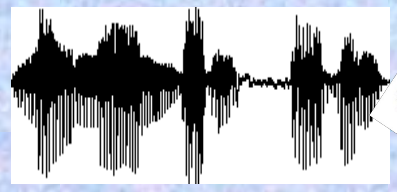




- Articulation produces
- sound waves which
- the ear conveys to the brain
- for processing

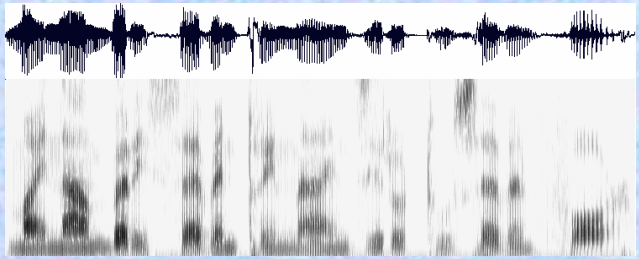


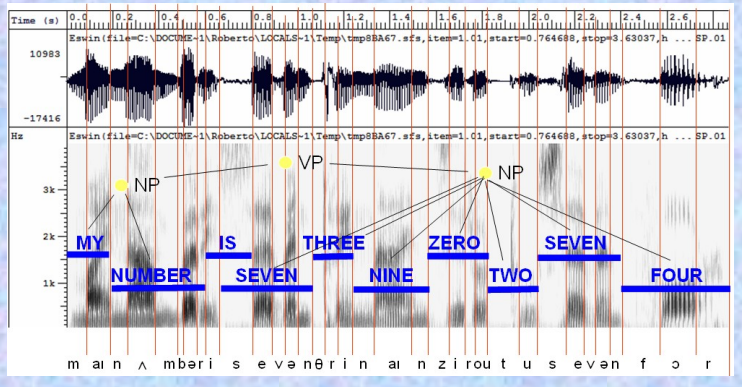
How might computers do it?



Acoustic waveform

Acoustic signal

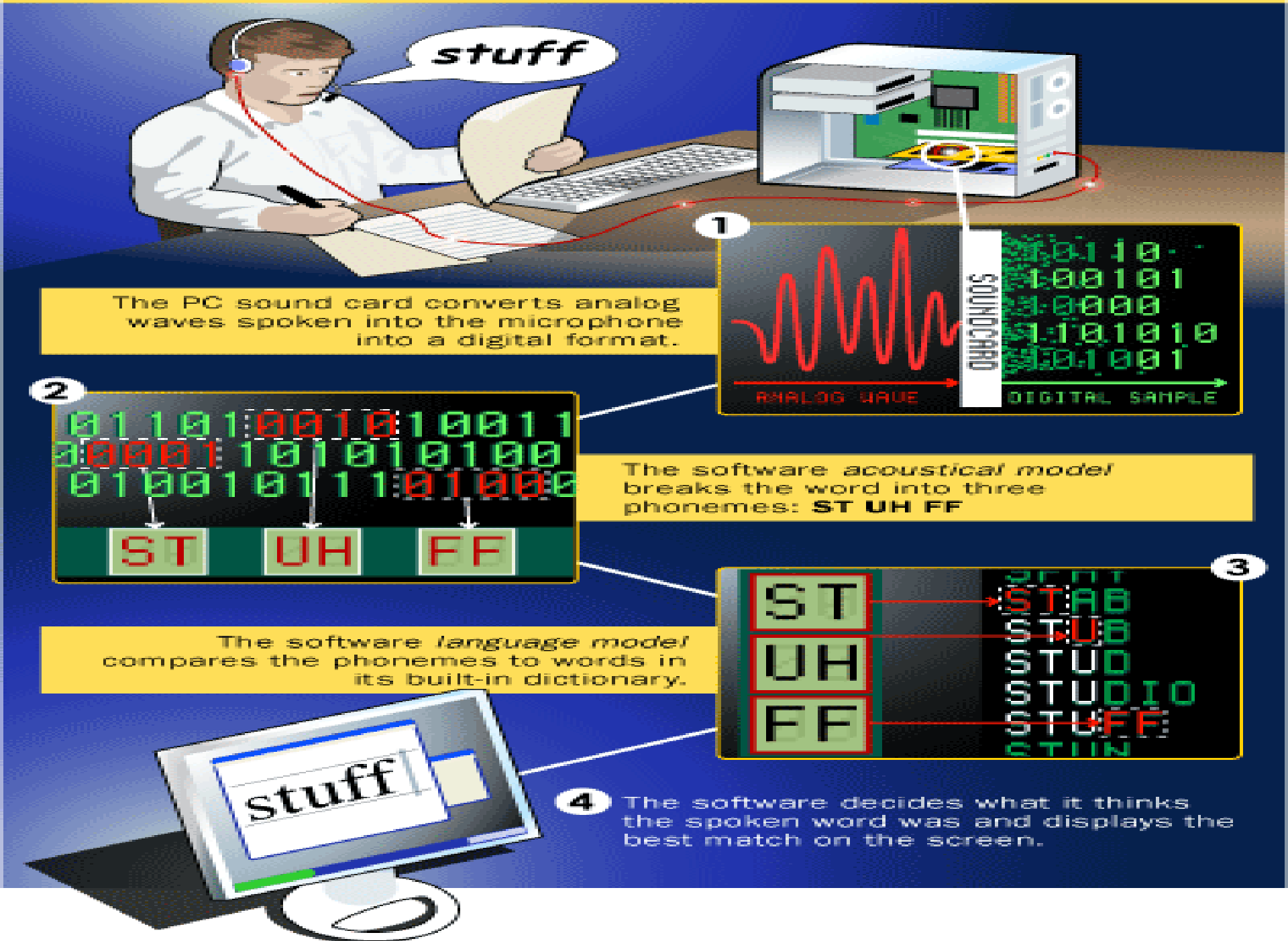




Speech recognition

- Digitization
- Acoustic analysis of the speech signal
- Linguistic interpretation

How Speech Recognition Works



The PC sound card converts analog waves spoken into the microphone into a digital format.

The software *acoustical model* breaks the word into three phonemes: **ST UH FF**

The software *language model* compares the phonemes to words in its built-in dictionary.

4 The software decides what it thinks the spoken word was and displays the best match on the screen.

DIFFERENT PROCESSES INVOLVED

- Digitization
  - Converting analogue signal into digital representation
- Signal processing
  - Separating speech from background noise
- Phonetics
  - Variability in human speech
- Phonology
  - Recognizing individual sound distinctions (similar phonemes)
  - is the systematic use of sound to encode meaning in any spoken human language
- Lexicology and syntax
  - **Lexicology** is that part of linguistics which studies *words*, their nature and meaning, words' elements, relations between words,words groups and the whole lexicon. .

DIFFERENT PROCESSES INVOLVED(CONTD.)

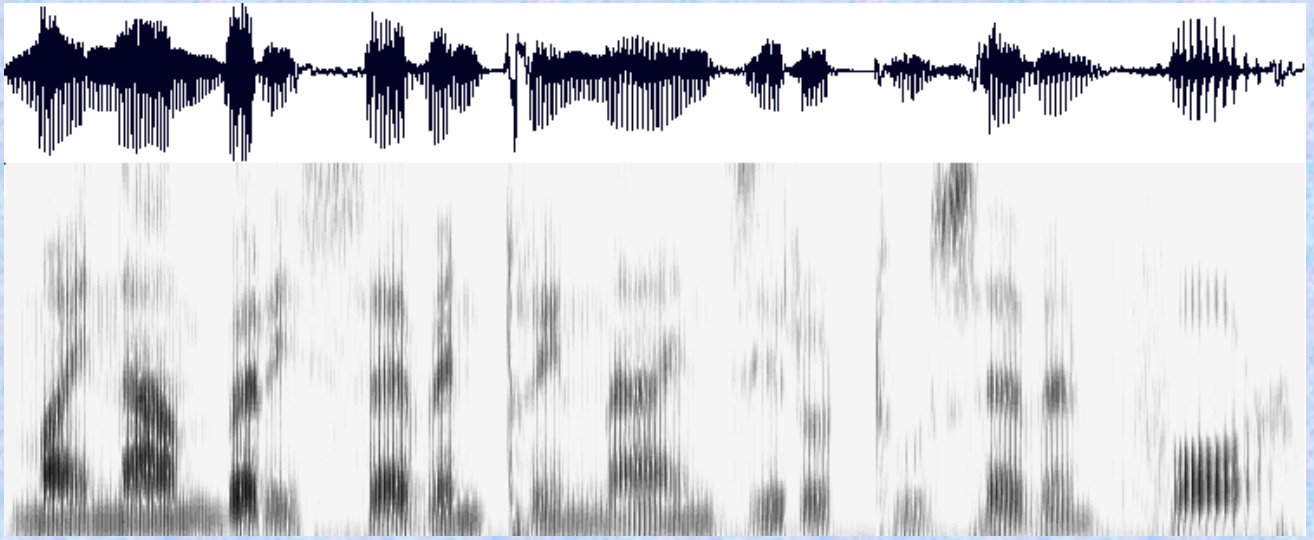
- Syntax and pragmatics
  - Semantics tells about the meaning
  - Pragmatics is concerned with bridging the explanatory gap between sentence meaning and speaker's meaning

Digitization

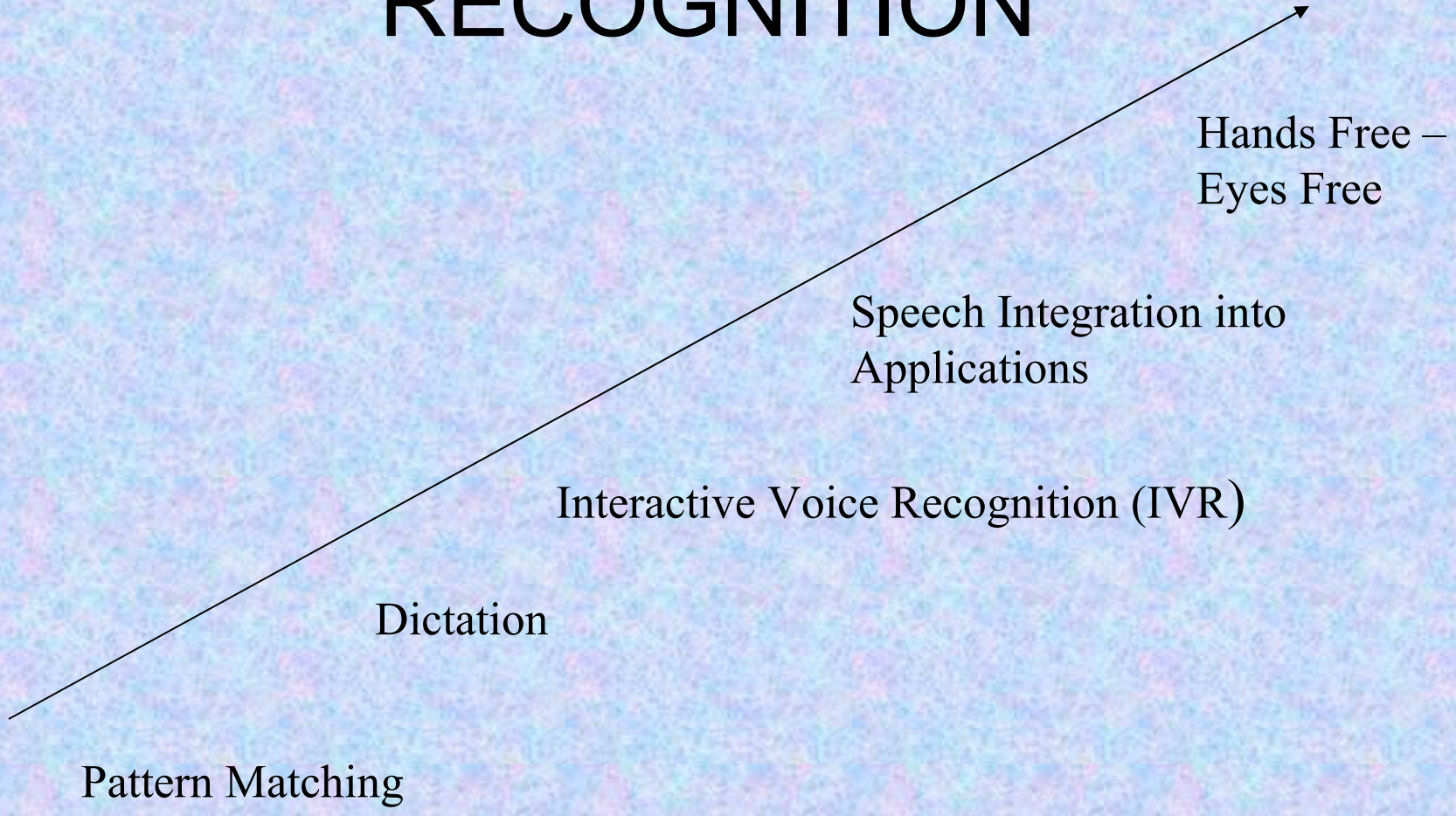
- Analogue to digital conversion
- Sampling and quantizing
  - ✓ Sampling is converting a continuous signal into a discrete signal
  - ✓ Quantizing is the process of approximating a continuous range of values
- Use filters to measure energy levels for various points on the frequency spectrum
- Knowing the relative importance of different frequency bands (for speech) makes this process more efficient
- E.g. high frequency sounds are less informative, so can be sampled using a broader bandwidth (log scale)

Separating speech from background noise

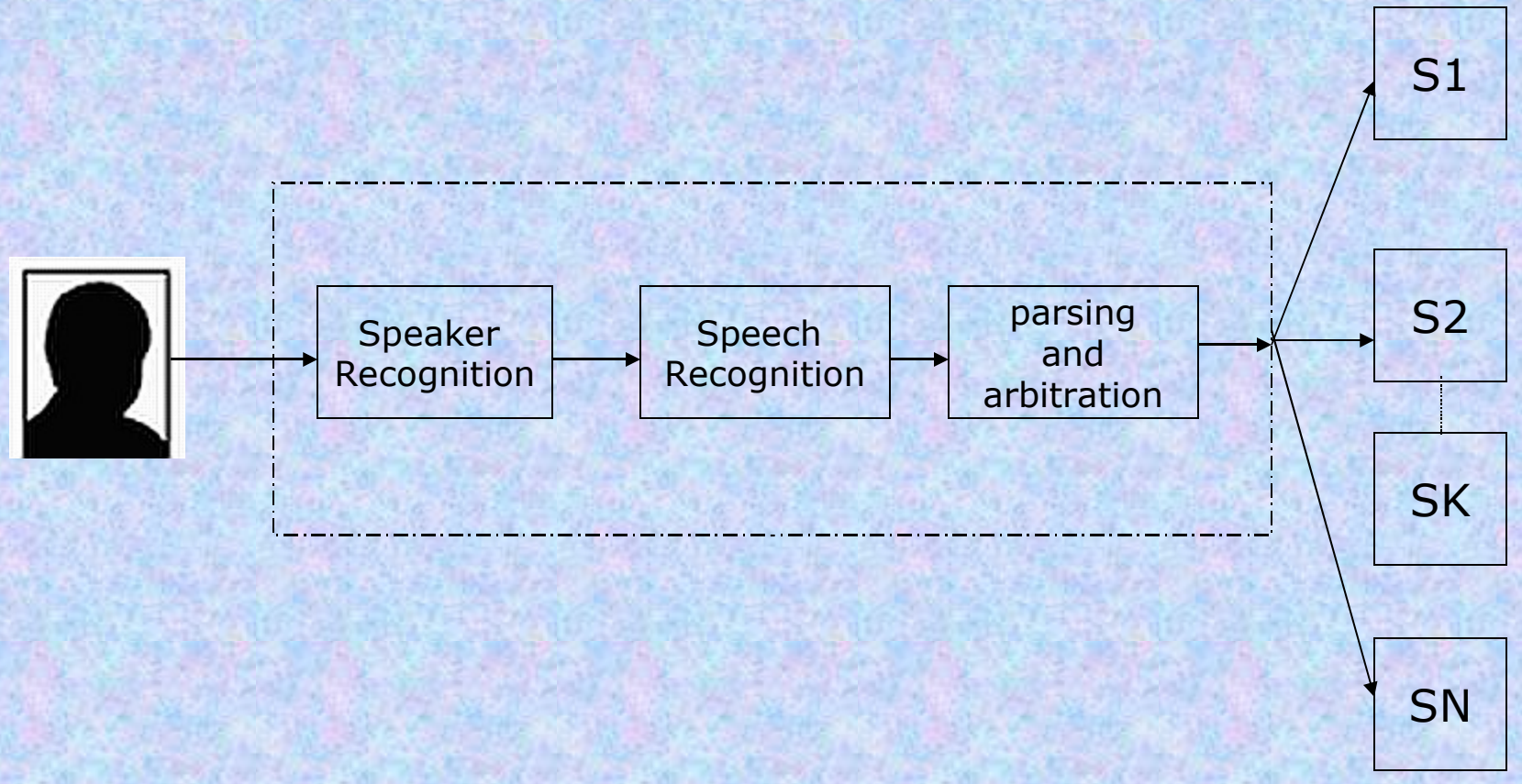
- Noise cancelling microphones
  - Two mics, one facing speaker, the other facing away
  - Ambient noise is roughly same for both mics
- Knowing which bits of the signal relate to speech

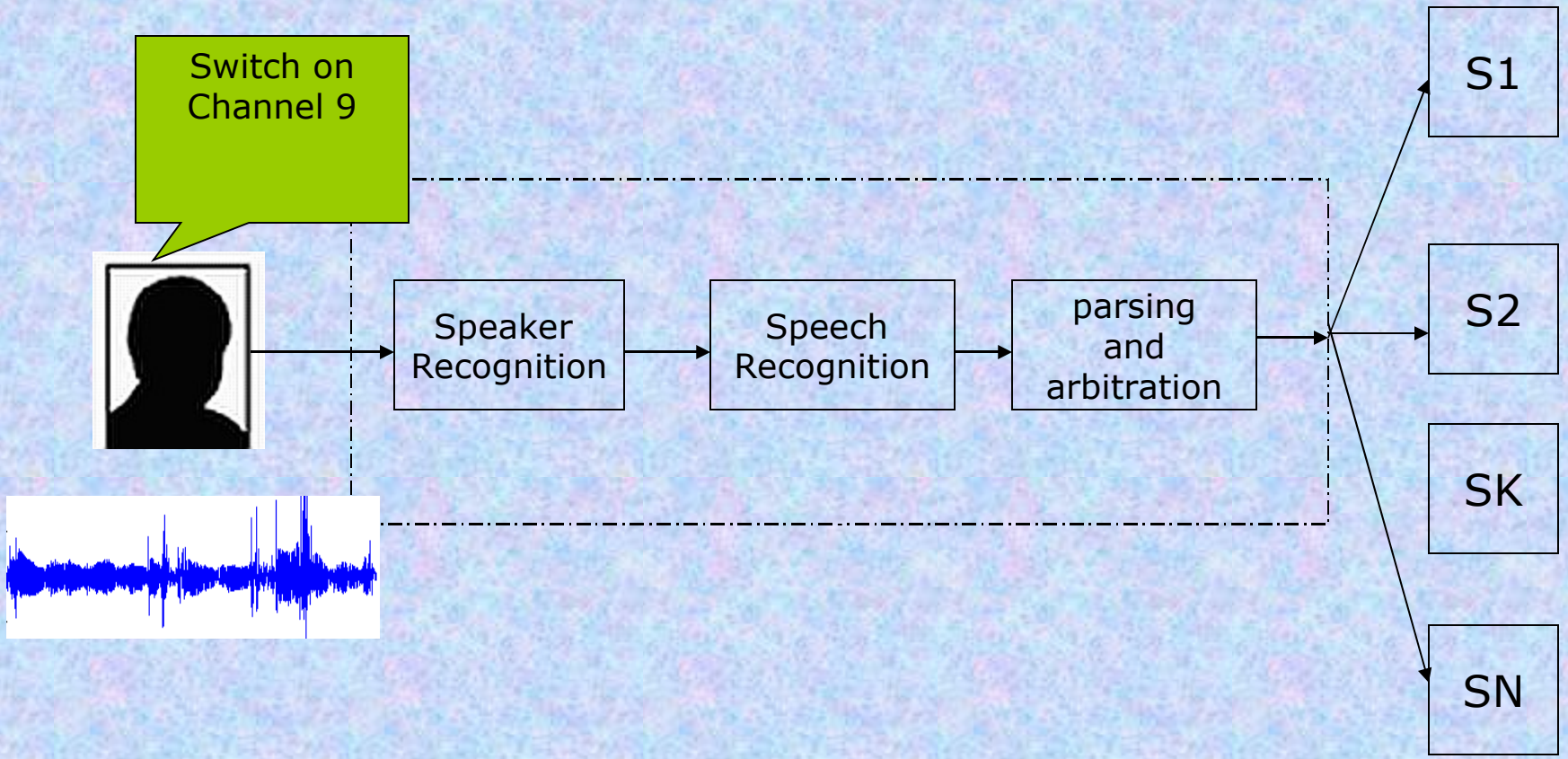


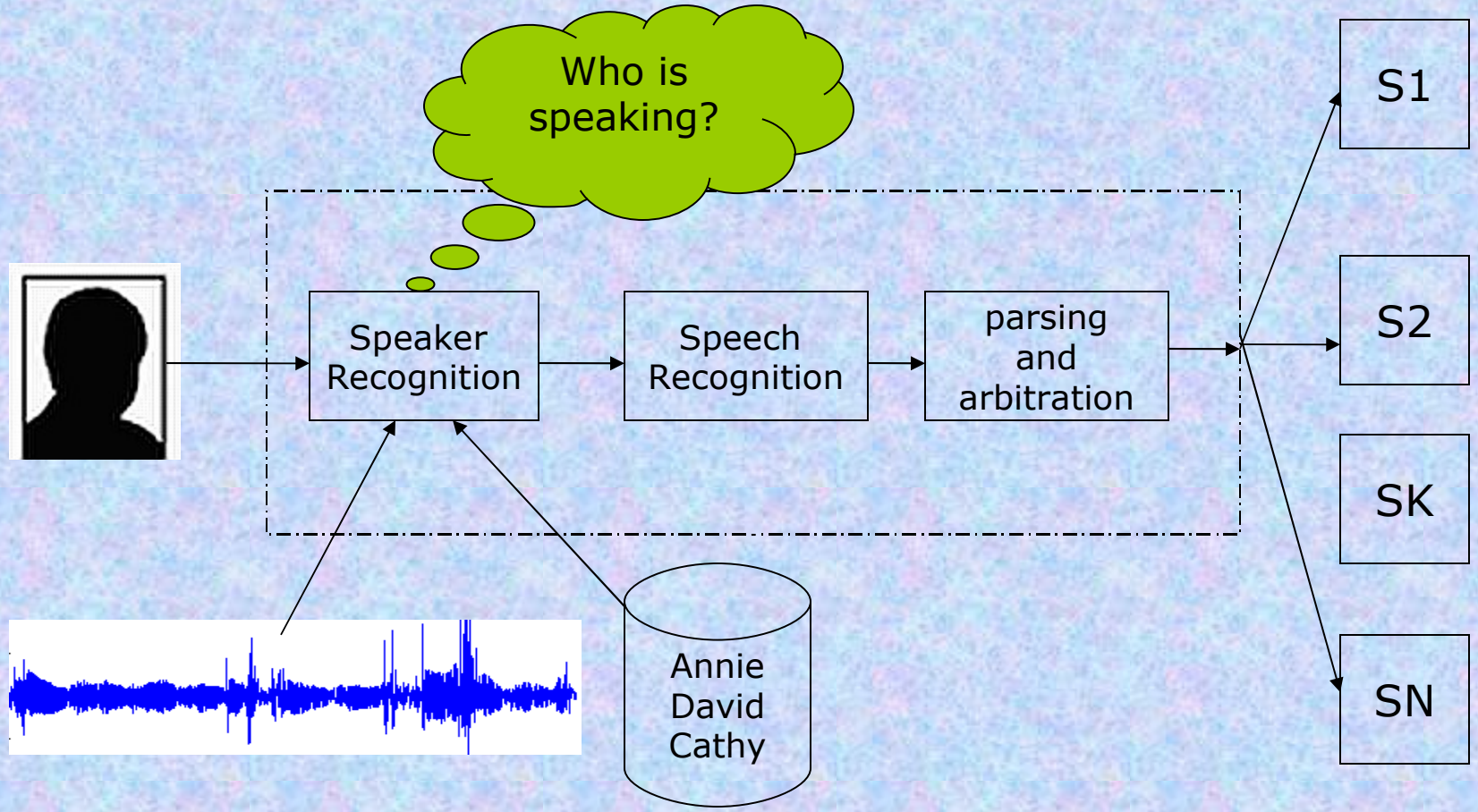
EVOLUTION OF VOICE RECOGNITION



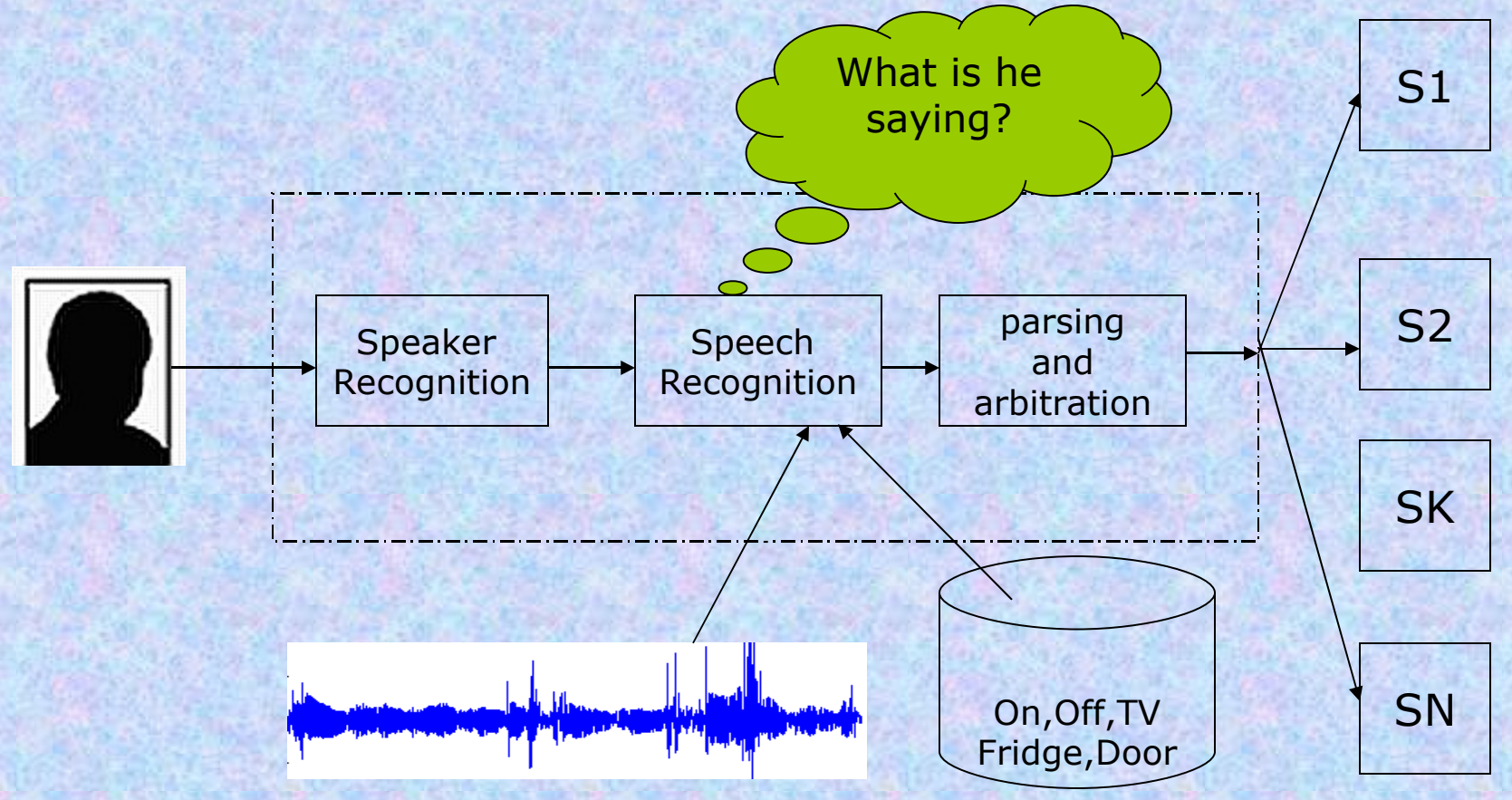
Process of speech recognition



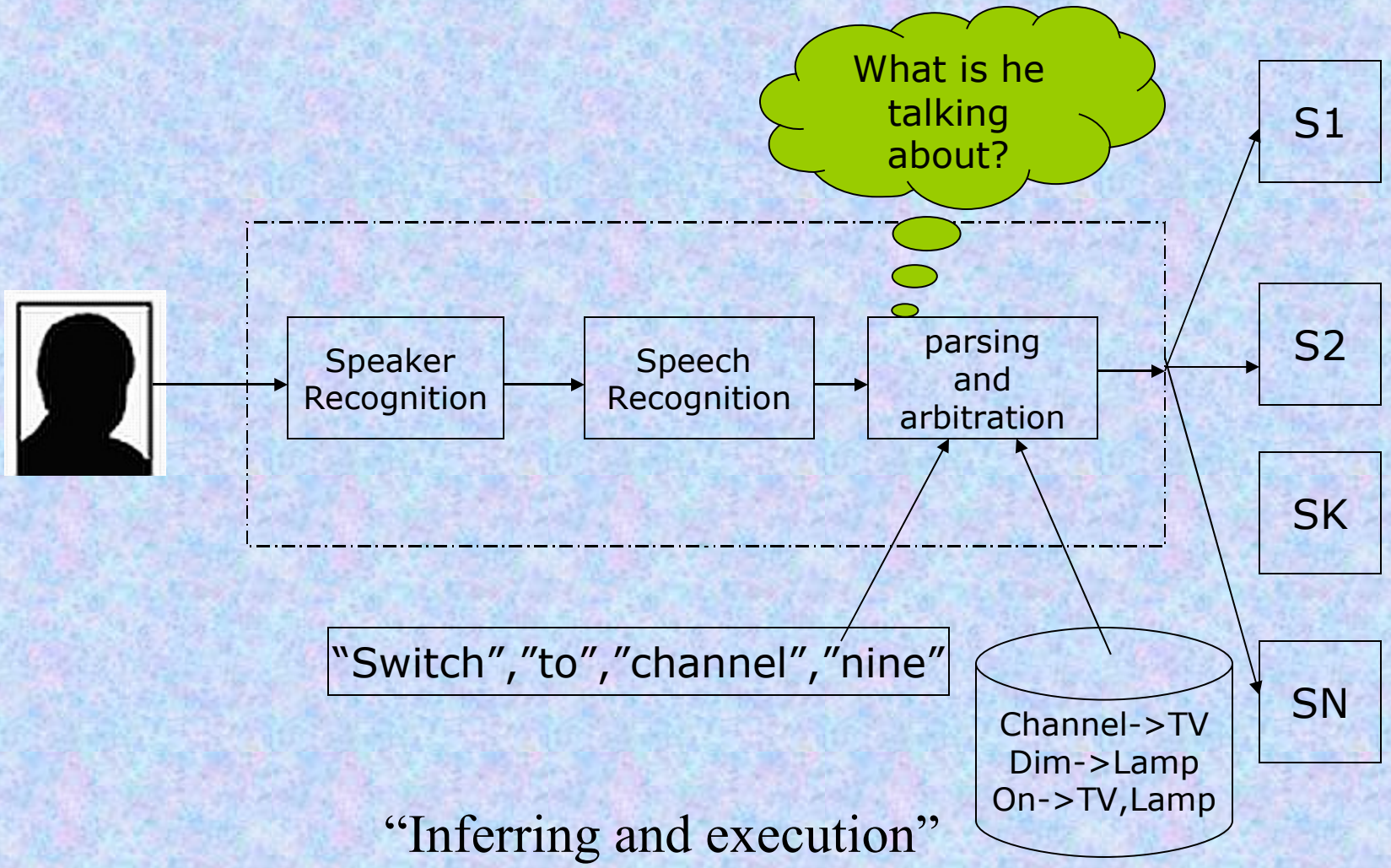

Switch on
Channel 9
Speaker
Recognition
Speech
Recognition
parsing
and
arbitration
S1
S2
SK
SN



"Authentication"

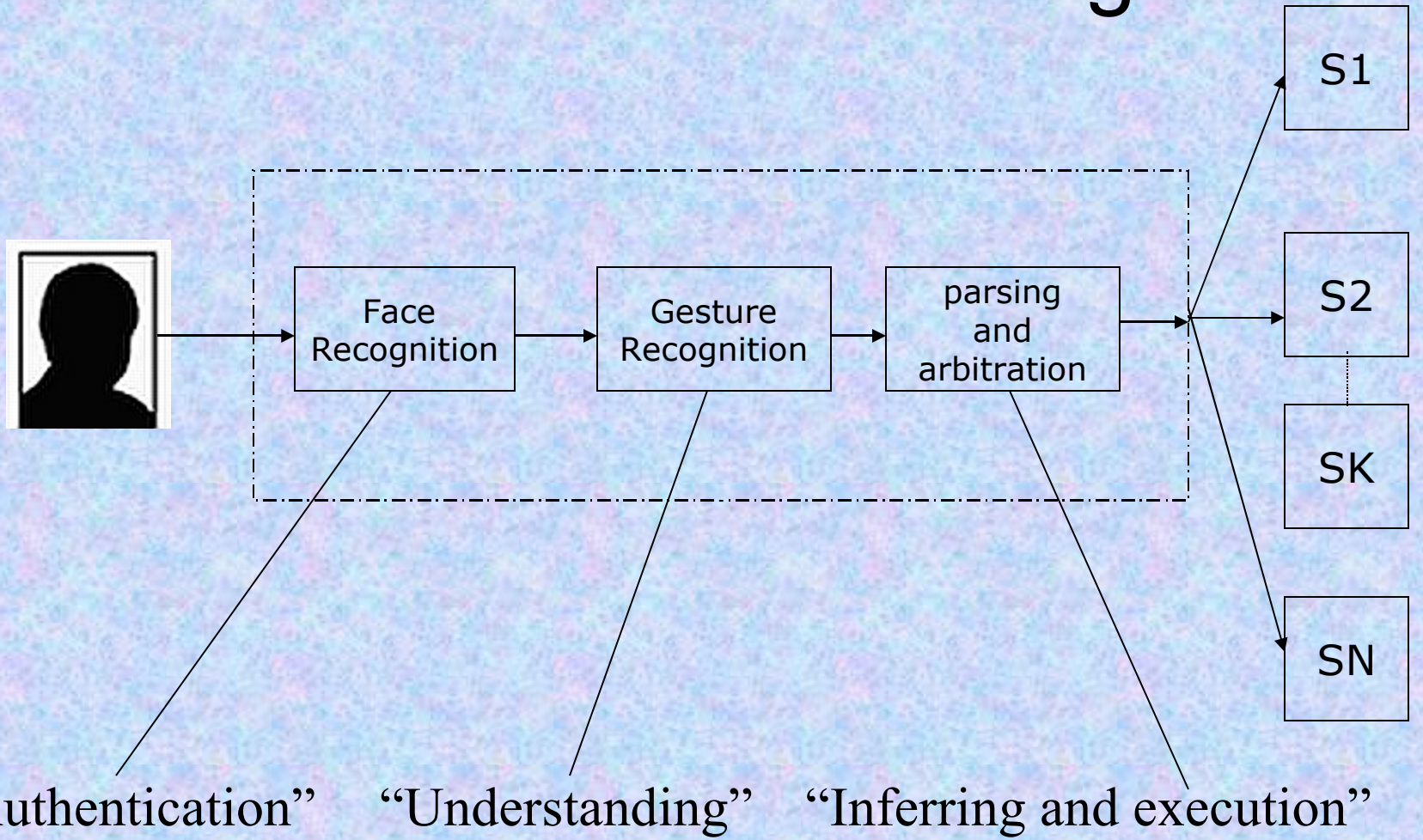


"Understanding"



"Inferring and execution"

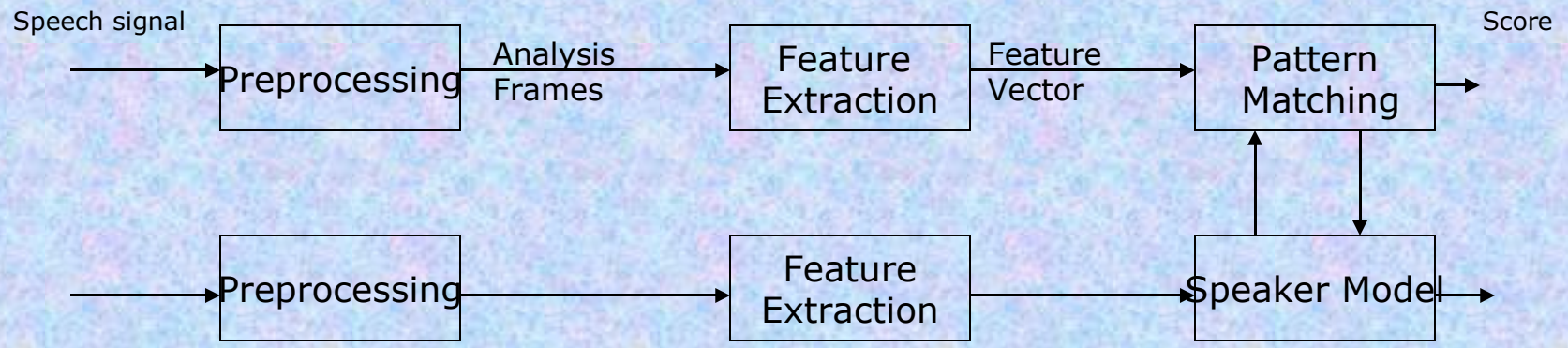
Framework of Voice Recognition



Speaker Recognition

- Definition
  - It is the method of recognizing a person based on his voice
  - It is one of the forms of biometric identification
- Depends of speaker specific characteristics.

Generic Speaker Recognition System



ADVANTAGES

- Advantages
  - People with disabilities
  - Organizations - Increases productivity, reduces costs and errors.
  - Lower operational Costs
  - Advances in technology will allow consumers and businesses to implement speech recognition systems at a relatively low cost.
    - Cell-phone users can dial pre-programmed numbers by voice command.
    - Users can trade stocks through a voice-activated trading system.
    - Speech recognition technology can also replace touch-tone dialing resulting in the ability to target customers that speak different languages

DISADVANTAGES

- Difficult to build a perfect system.
- Conversations
  - Involves more than just words (non-verbal communication; stutters etc.
  - Every human being has differences such as their voice, mouth, and speaking style.
- Filtering background noise is a task that can even be difficult for humans to accomplish.

Future of Speech Recognition

- Accuracy will become better and better.
- Dictation speech recognition will gradually become accepted.
- Small hand-held writing tablets for computer speech recognition dictation and data entry will be developed, as faster processors and more memory become available.
- Greater use will be made of "intelligent systems" which will attempt to guess what the speaker intended to say, rather than what was actually said, as people often misspeak and make unintentional mistakes.
- Microphone and sound systems will be designed to adapt more quickly to changing background noise levels, different environments, with better recognition of extraneous material to be discarded.