

Toy Spoken Dialogue System – using Speech APIs

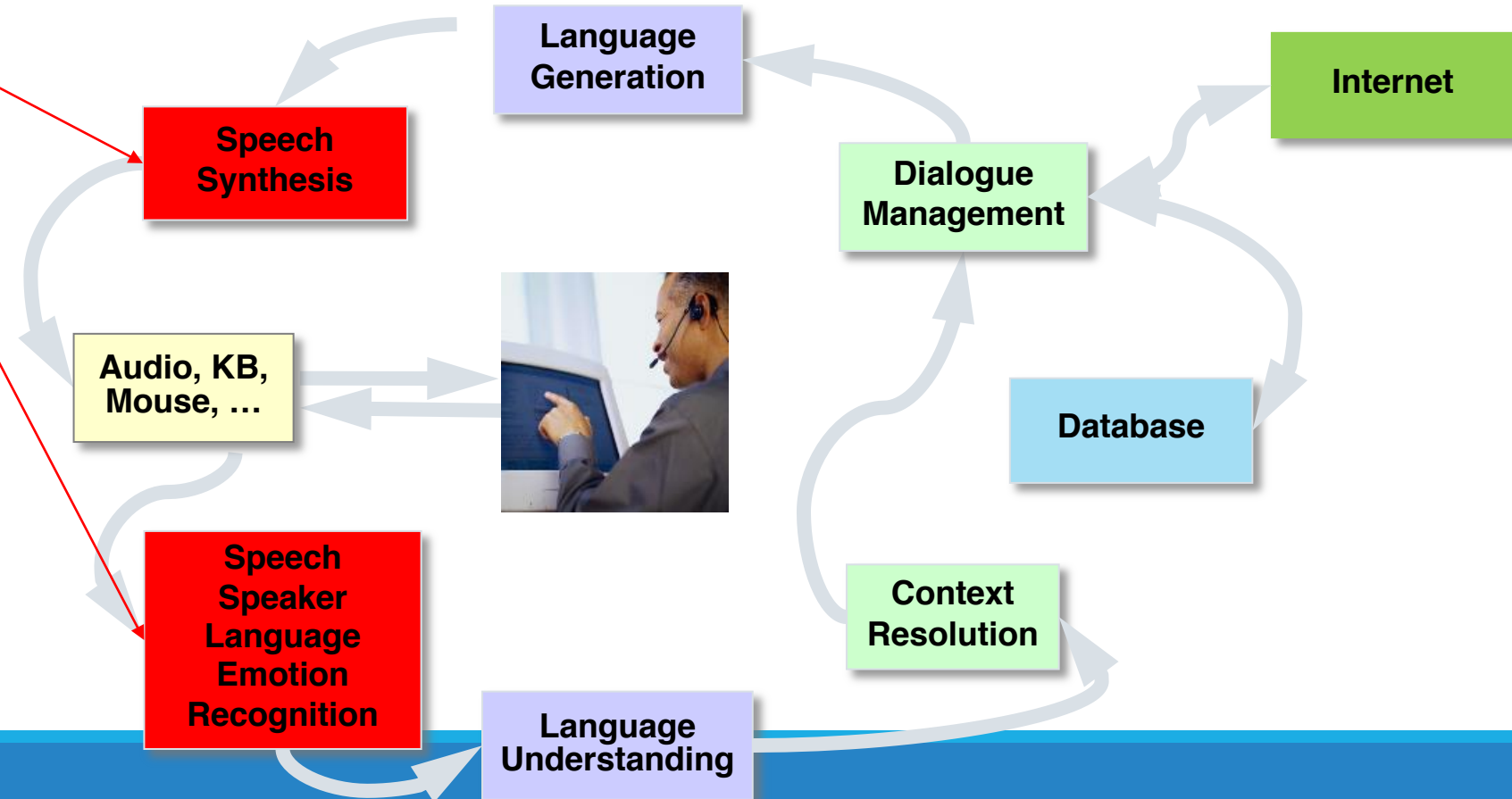
Yuan-Fu Liao

National Taipei University of Technology

Spoken Dialogue System

Speech Recognition

Speech Synthesis



SpeechRecognition API

Library for performing speech recognition, support:

- CMU Sphinx (works offline)
- **Google Speech Recognition** (default test key)
- Google Cloud Speech API
 - “GOOGLE_CLOUD_SPEECH_CREDENTIALS” required (non-free)
- Wit.ai
- Microsoft Azure Speech
- Houndify API
- IBM Speech to Text
- Snowboy Hotword Detection (works offline)

Google Speech API

```
asr.recognize_sphinx(audio)
asr.recognize_google(audio)
asr.recognize_google_cloud(audio, credentials_json=GOOGLE_CLOUD_SPEECH_CREDENTIALS)
asr.recognize_wit(audio, key=WIT_AI_KEY)
asr.recognize_bing(audio, key=BING_KEY)
asr.recognize_houndify(audio, client_id=HOUNDIFY_CLIENT_ID, client_key=HOUNDIFY_CLIENT_KEY)
asr.recognize_ibm(audio, username=IBM_USERNAME, password=IBM_PASSWORD)
```

- Speech Recognition

- import speech_recognition as sr
- asr = sr.Recognizer()
- with sr.AudioFile(human_sound_file) as source:
 - audio_data = asr.record(source)
 - text = asr.recognize_**google**(audio_data, language = 'zh-TW', show_all = False)



- Speech Synthesis

- from gtts import gTTS
- toSpeak = gTTS(text = '大奶微微：'請找伯恩', lang = 'zh-TW', slow = False)

Toy Conversation

- Speech Synthesis
 - robot_speaks('你好')
- Recording
 - audio=get_audio()
- Speech Recognition
 - text = get_text(audio)
- Dialogue Management
 - process_text(text)

```
[14] # Robot Greeting
robot_sound_file = robot_speaks("What's your name, Human?")
print(robot_sound_file)
Audio(robot_sound_file, autoplay=True)
```

```
Person : What's your name, Human?
robot-2.mp3
```

▶ 0:01 / 0:01 ————— 🔊 ⋮

```
[17] # Human Speak
human_sound_file = get_audio()
```

```
human-2.wav
```

Saving the recording... pls wait!

▶ 0:00 ————— 🔊

```
[18] # Speech Recognition
text = get_text(human_sound_file)
print(text)
```

```
Justin
```

```
[19] # Robot Speak
name = text
robot_sound_file = robot_speaks("Hello, " + name + '.' + " " + "What could I do for your?")
print(robot_sound_file)
Audio(robot_sound_file, autoplay=True)
```

```
Person : Hello, Justin. What could I do for your?
robot-3.mp3
```

▶ 0:03 / 0:03 ————— 🔊 ⋮

Toy Question Answering

- Recording
 - `audio = get_audio()`
- Speech Recognition
 - `text = get_text(audio)`
- QA
 - `ans = qa_get(text)`
- Speech Synthesis
 - `robot_speaks(ans)`

```
[20] language = 'zh-TW'
      qa = {
          'iphone' : 'it is too expansive',
          'snoy' : 'it is too ugly',
          'samsung' : 'it is too dangerous',
          '大杯奶茶' : '請找伯恩Stand Up!'
      }
```

```
[21] # Human Speak
      human_sound_file = get_audio()
```

human-3.wav

Saving the recording... pls wait!

▶ 0:00 🔊

```
[22] # Speech Recognition
      text = get_text(human_sound_file)
      print(text)
```

大杯奶茶

```
[23] # output = qa.get(text.lower(), 'Please say again...')
      output = qa.get(text.lower(), '你說什麼...')
      robot_sound_file = robot_speaks(output)
      print(robot_sound_file)
      Audio(robot_sound_file, autoplay=True)
```

Person : 請找伯恩Stand Up!
robot-4.mp3

▶ 0:02 / 0:02 🔊 ⋮