Music: Project!

By Group 7: b01902028 周書禾 b01902032 江東峻 b01902096 曾偉綸

2016.6.21 MSAR final project

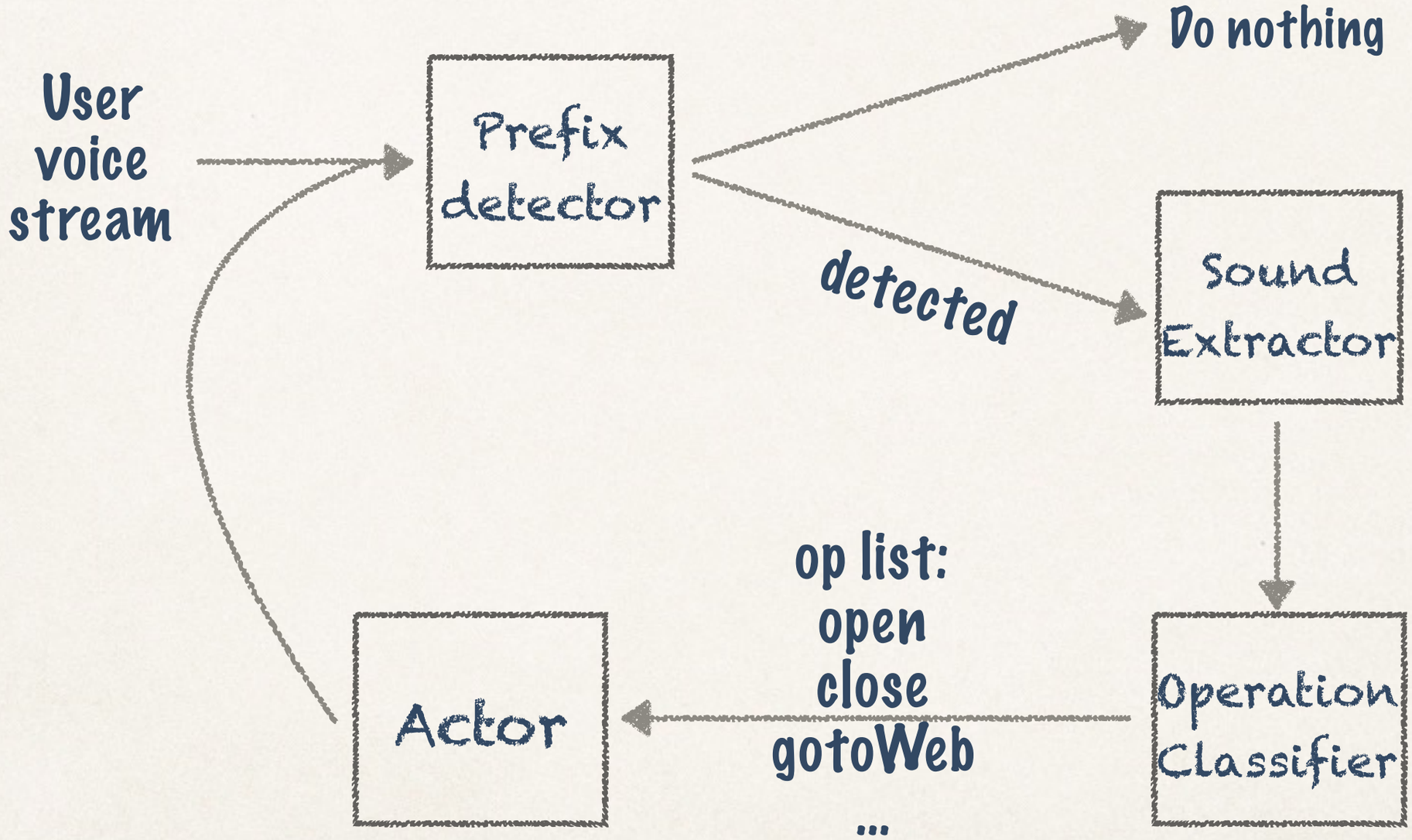
Goal

- Detect real-time voice stream
- Voice commands for simple instructions
- Without Speech Recognition
- User-defined voice for instructions

System Structure

- Prefix detector
- Operation Sound Extractor
- Operation classifier
- Actor

Control Flow Graph



Prefix detector

- A daemon program detects sound continuously
- if detected, then call *Operation Sound Extractor*

Operation Sound Extractor

- Called by *prefix detector*
- Segment the operation sound after the prefix
- Then pass the segmentation to *operation classifier*

Operation classifier

- Classify the segmentation
- According to each operation, detect the parameter

Actor

- Receive the operation id and enforce the operation
- Operation including: open, close, switch … etc
- Implemented by pyAutoGUI

Implementation

- Sound extractor
- Operation Classifier
- Operation Actor

Sound Extractor

- A thread records background sound

- Check the background sound chunk for about every 0.2 seconds, and a chunk has about 2 seconds.

- Use Endpoint detection technique to extract the sound, prefix and operations alike.

Operation Classifier

- Audio files normalized by volume (same mean)
- Extract features mfcc and melspectrogram
- Distance: summing the DTW distances of the two features.
- Operate only if it is the one with the smallest distance and smaller than a pre-set threshold.

Operation Actor

- accept instructions and have corresponding response
- use keyboard and mouse control API
- some default actions: open, kill, switchDesktop, switchWindow, switchTabWindow, music control, open url, …, and other self-define action.

DEMO