What You Say is What You Get

Handsfree Coding in 2022

Informatik 2022
Sep. 28, 2022
Wolle

Carl von Ossietzky
Universität
Oldenburg

It's Simple, Really!

The requirements:

- **Microphone**: Every notebook has one!
- **Speech Recognition Software (SR)**: Included in Windows since 2007!
- **Voice Command Execution**: Available in every SR software!

Where‘s the Challenge?

Where's the Challenge?

WSR, Dragon, …

- **Automatic Speech Recognition (ASR)**: optimized for natural languages
  1. Signal processing extracts features from audio recording
  2. Acoustic model recognizes phonemes
  3. Language model finds a matching sequence of words:

  → Default: Every utterance is interpreted as (spoken) text

  (Commands only through special keywords)

Dragonfly, Talon, …

- **Voice Coding**: optimized for actions & programming languages

  → Default: Everything is interpreted as a command

  (Natural language through special keywords, e.g. `say <utterance>`)

I Am Wolle

I'm data engineer,
not an ASR or HCI expert!

Research:
- Stream Processing
- Real-Time Databases
- NoSQL & Cloud Systems
- ...

+

Practice:
- Web Caching
- Big Data Analytics
- Anger Management
- ...

Universität Hamburg
DER FORSCHUNG | DER LEHRE | DER BILDUNG

Carl von Ossietzky
Universität
Oldenburg

Baqend


Look,
No Hands!

Handsfree Coding

- **Different behavior** for different semantics, for example:
  - C#: `funky test funk` → `private void testFunk()`
  - JavaScript: `funky test funk` → `function testFunk()`
- **Intuitive IDE shortcuts** such as
  - `"run code"` instead of `<shift-f10>`
  - `"find usage"` instead of `<ctrl-alt-f7>`
- **Powerful templates**, `e.g.:`

```
action(user.code_state_if):
    insert("if () {}")
    key(left enter up end left left left)
```

Handsfree Coding: Talon

Josh W. Comeau. Hands-free Coding, Josh W. Comeau's Blog (2020)

Snappy Noise Control With Parrot

- Available on GitHub: github.com/chaosparrot/parrot.py
- Noise-controlled actions with latency <50ms
- Workflow
  1. Record sounds
  2. Train model for recognition
  3. Map sounds to actions
- Compatible (and recommended in combination with) with other tooling:
  - Often used with Project IRIS (eye tracking)
  - Can be used to produce **Talon-compatible** models

Custom noises for your Talon grammar!

Eye Tracking & Noise Recognition

- **Calibration** for adjusting your eye tracker to your current position
- **Noises** for actions (e.g. clicking & right-clicking):
  - Extremely low latency (<50ms)
  - Talon currently supports `*pop*` & `*hiss*`
  - Custom noise models available va Parrot

github.com/chaosparrot/parrot.py

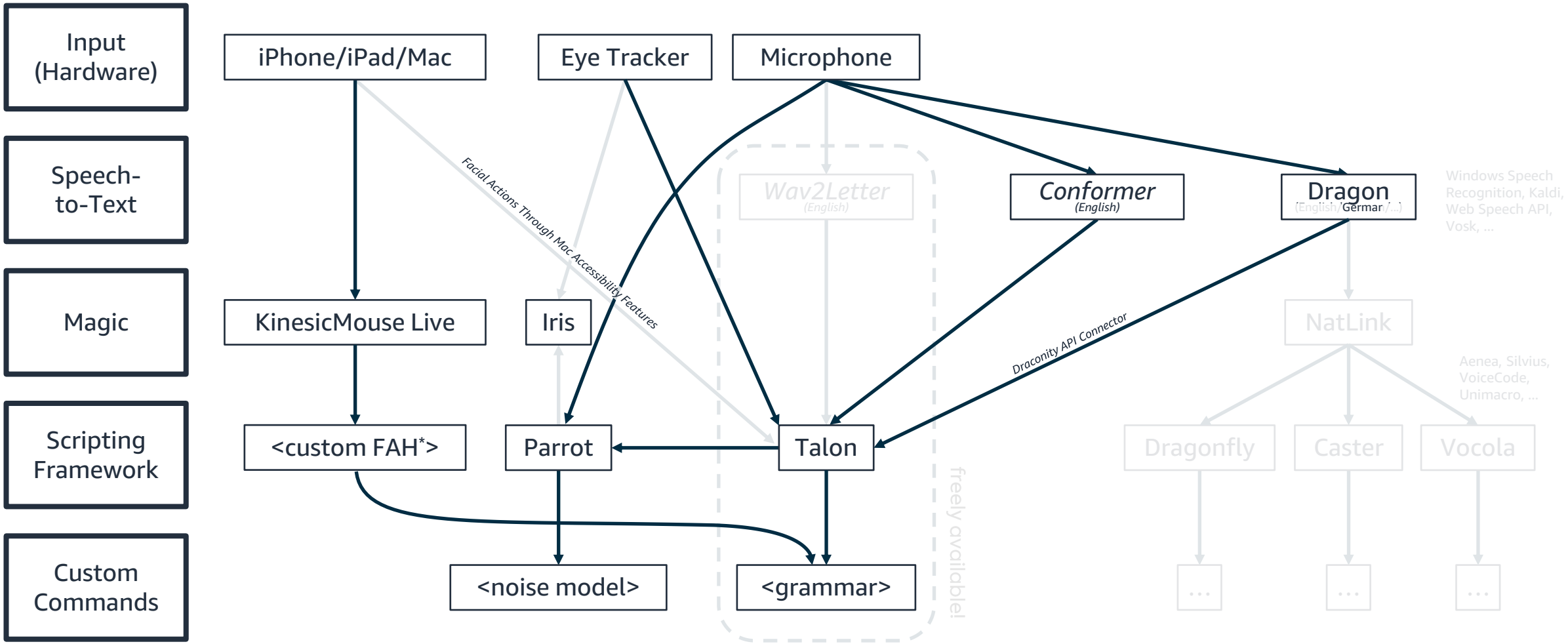
- **Different Modes** for convenience:
  - Zoom: (1) `*pop*` for zooming, (2) `*pop*` for clicking
  - Head tracking: eye gaze (jumps) + head movement (adjustment)
- **Debug** mode & camera overlay

Handsfree Gaming: Eyes + Face + Voice/Noise

wolle.science/twitch


The Base Setup

Popular Handsfree Coding Stacks: Overview



*Facial Action Handling
Please note that this overview is NOT complete: On every level, there are MANY other options!

This overview was inspired by:
https://dictation-toolbox.github.io/dictation-toolbox.org/ (accessed: January 4, 2021)

Multi-Computer Setup

Why This is Still Worth All the Hassle

Productivity

- Speed up input-heavy tasks
- Faster navigation through easy-to-remember shortcuts

Convenience

- Intuitive interfaces
- Relieve your hands

Accessibility

Compensate handicaps:

- Injuries (e.g. broken hand)
- Repetitive stress injury (RSI)
- Cubital Tunnel Syndrome
- …

General Awesomeness

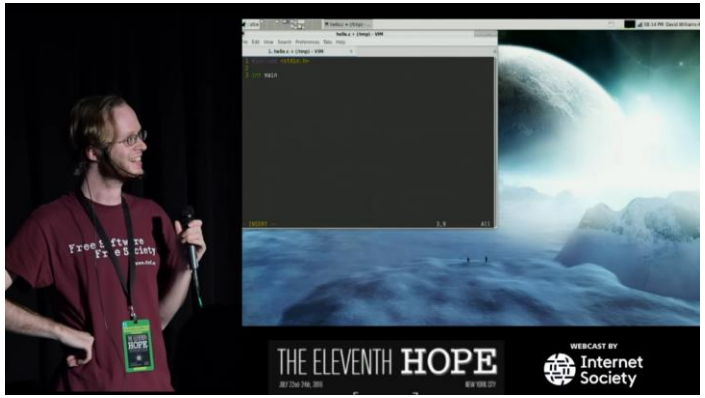
- Talk to your computer!!!

Helpful Resources & Outlook

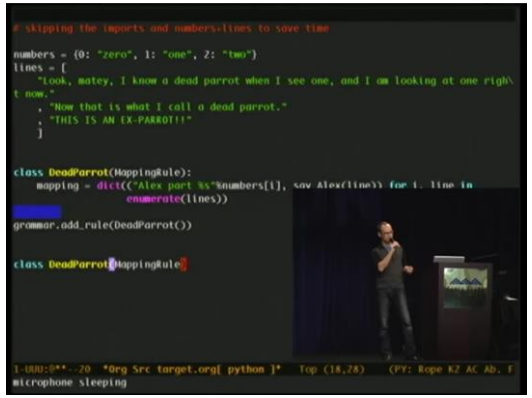
Recommended Talks

Emily Shea. Voice Driven Development: Who needs a keyboard anyway?, Strange Loop (2019)

Boudewijn Aasman. Coding by Voice with Dragonfly, PyGotham (2018)



David Williams-King. Coding by Voice with Open Source Speech Recognition, The Eleventh Hope (2016)



Tavis Rudd. Using Python to Code by Voice, PyCon US (2013)

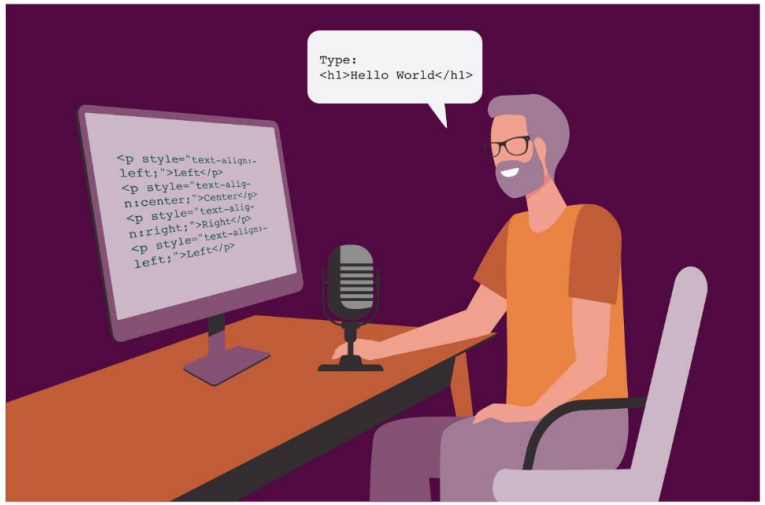
Closing Recommendations

- **Keep it simple**: Prioritize ease-of-use over efficiency at the start
  (in particular: get used to an existing grammar before optimizing it)

- **Keep it reasonable**: Try to find use cases that make sense for you
  (e.g.: I'm not giving this talk handsfree, since I can use my index finger)

- **Keep it in mind**: Handsfree coding might save you one day
  (revisit this talk when you struggle with RSI, broken hand, etc.)

Read the Article!



Softwareentwicklung ohne Maus und Tastatur

Sprechen ist das neue Klicken

Dr. Wolfram Wingerath, Michaela Gebauer

Für die Bedienung des Computers brauchte man viele Jahre Maus und Tastatur – heute kann man mit Sprache, Gestik und Mimik sogar programmieren.

zung des Computers ganz ohne Einsatz ihrer Hände."

Wolle ist 33 Jahre alt, Data Engineer und erprobt seit mehr als zehn Jahren Eingabemethoden zur Softwareentwicklung ohne Maus und Tastatur. Inzwischen setzt er fast ausschließlich auf Handsfree Coding, da er damit effizienter arbeitet. „Dadurch muss ich mir keine kryptischen Shortcuts mehr merken und kann ganz bequem mit Sprache, Geräuschen, Mimik oder Gestik den Computer und die Programme steuern", sagt er.

Beim Handsfree Coding spielt das Voice Coding eine zentrale Rolle. Hierbei wird Quellcode per Spracheingabe erstellt. Voice Coding ist jedoch nicht mit handelsüblicher Software zur automatischen Spracherkennung (Automatic Speech Recognition, ASR) vergleichbar. Es gibt zwar einige offensichtliche Parallelen zum Diktieren von Textnachrichten. Mit Standardsoftware zur Spracherkennung kann man aber nicht ohne Weiteres effizient programmieren, da ASR auf die Interpretation und Synthese einer konkreten natürlichen Sprache ausgelegt ist. Sie verwendet dafür jeweils spezifische Modelle, Grammatiken und Optimierungen bei der Ausgabe, etwa, wenn sie automatisch Satzzeichen einfügt oder Substantive großschreibt. Bei typischer ASR-Software sind Befehle stets mit einem Schlüsselwort einzuleiten und durch Sprechpausen abzuschließen. Während sich so einfache Tastenaktionen umsetzen lassen – etwa mit der Aussage „press Enter" zum Drücken der Eingabetaste –, ist die Ausführung von komplexen Aktionen oder Aktionssequenzen eher beschwerlich und ineffizient.

Wolfram Wingerath, Michaela Gebauer: Sprechen ist das neue Klicken, iX 9/2021 (https://wingerath.cloud/2021/ix)

Thanks! So What Now?

Slack
talonvoice.slack.com

Ask questions!
Enjoy the community!

Subscribe to the mailing list!

GI Initiative
handsfree-coding.gi.de

Try out handsfree coding!

Patreon
patreon.com/lunixbochs

Support Talon Development!

Wolfram Wingerath wolle@uol.de

Carl von Ossietzky
Universität
Oldenburg