What You Say is What You Get

Handsfree Coding in 2024

Data Science Meetup, Hamburg, Germany
November 25, 2024
Wolle

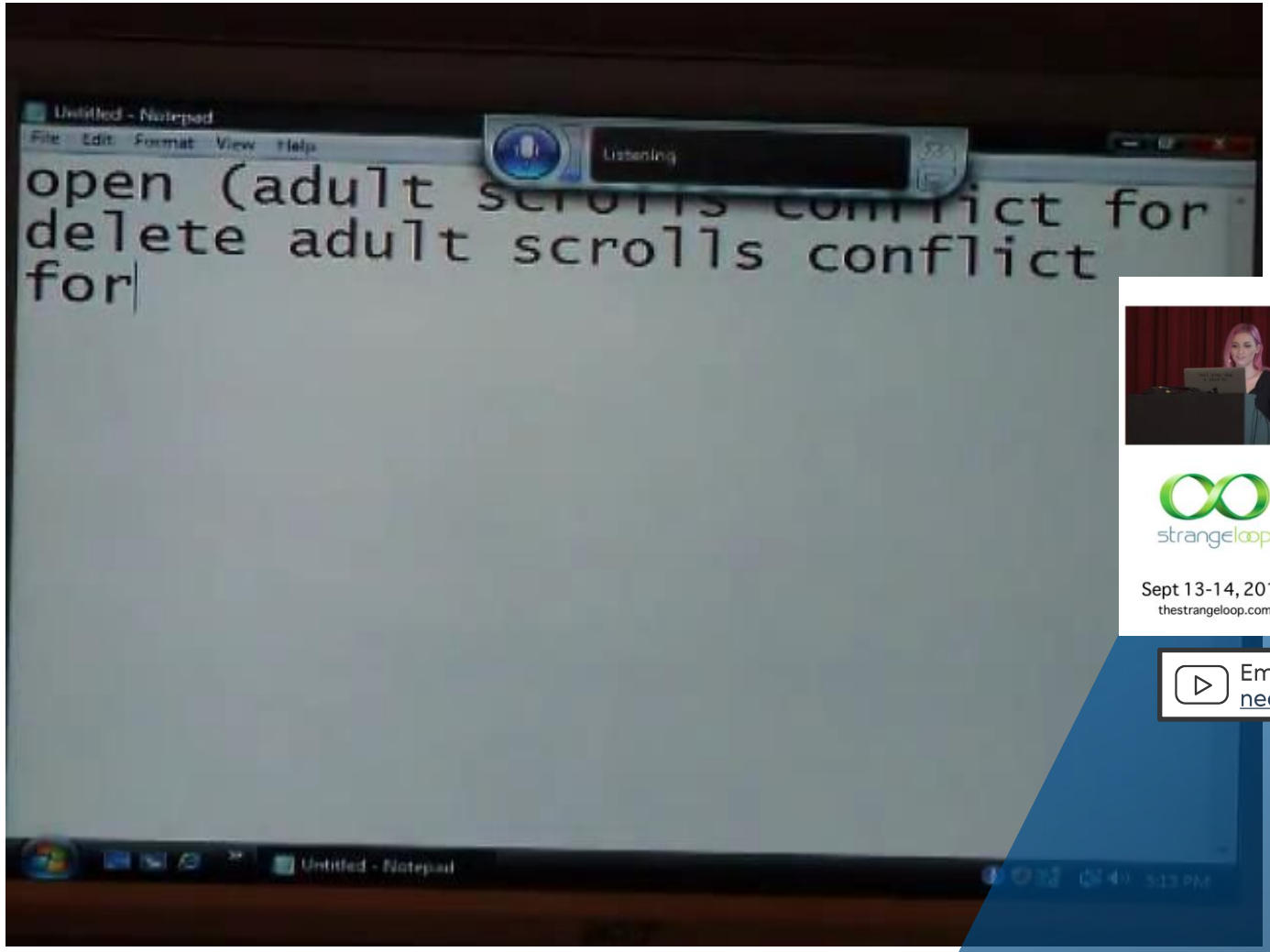
Videos & Slides Available at https://wolle.science

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!

Let Me Just Show You How Easy It is

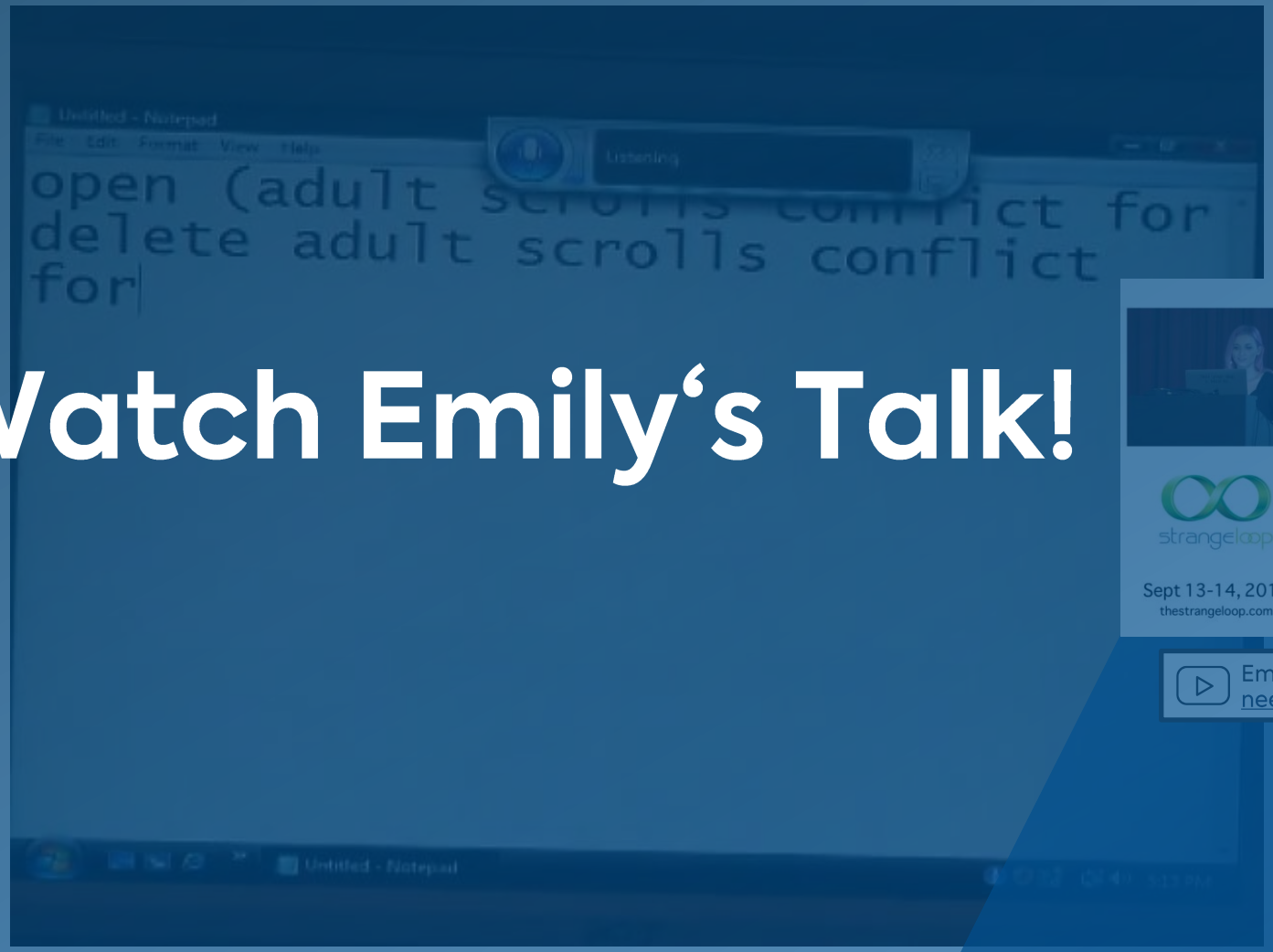


Emily Shea. Voice Driven Development: Who needs a keyboard anyway?, Strange Loop (2019)

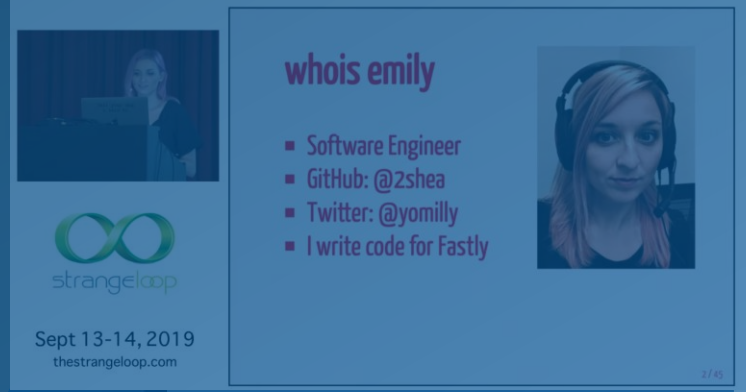
scrubadub1. Windows Vista Speech Recognition Tested - Perl Scripting, YouTube, 2007

Idea to use this video blatantly stolen from: Emily Shea. Voice Driven Development: Who needs a keyboard anyway?, Strange Loop, 2019

Let Me Just Show You How Easy It is



Go Watch Emily's Talk!



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

scrubadub1. Windows Vista Speech Recognition Tested - Perl Scripting, YouTube, 2007

Idea to use this video blatantly stolen from: Emily Shea. Voice Driven Development: Who needs a keyboard anyway?, Strange Loop, 2019

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)

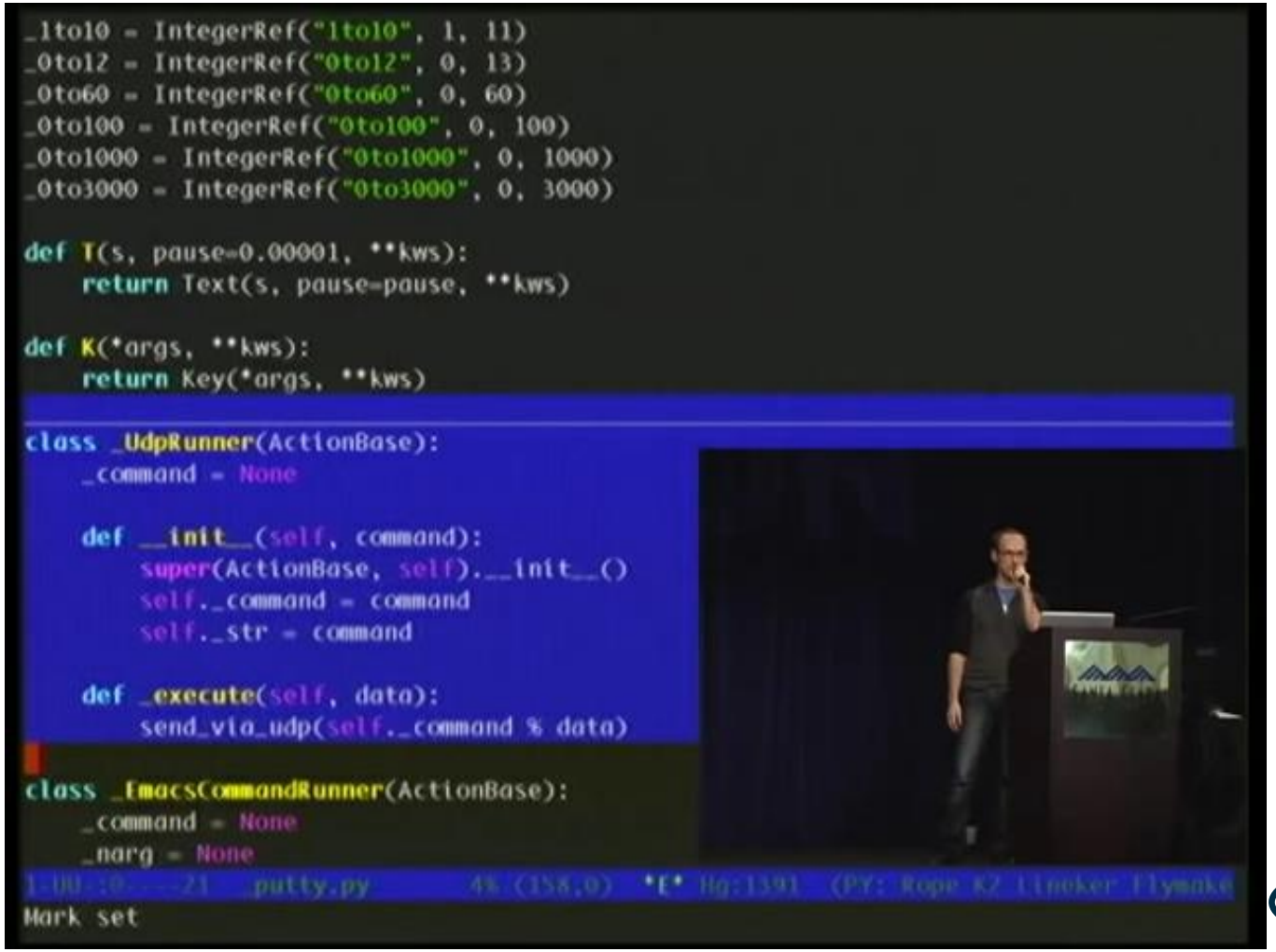
Talon, Dragonfly ...

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

  → Default: Everything is interpreted as a command

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

Handsfree Coding: How It Actually Looks



Using Dragonfly!

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

I Am Wolle

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

Wolfram Wingerath
Data Science

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

+

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

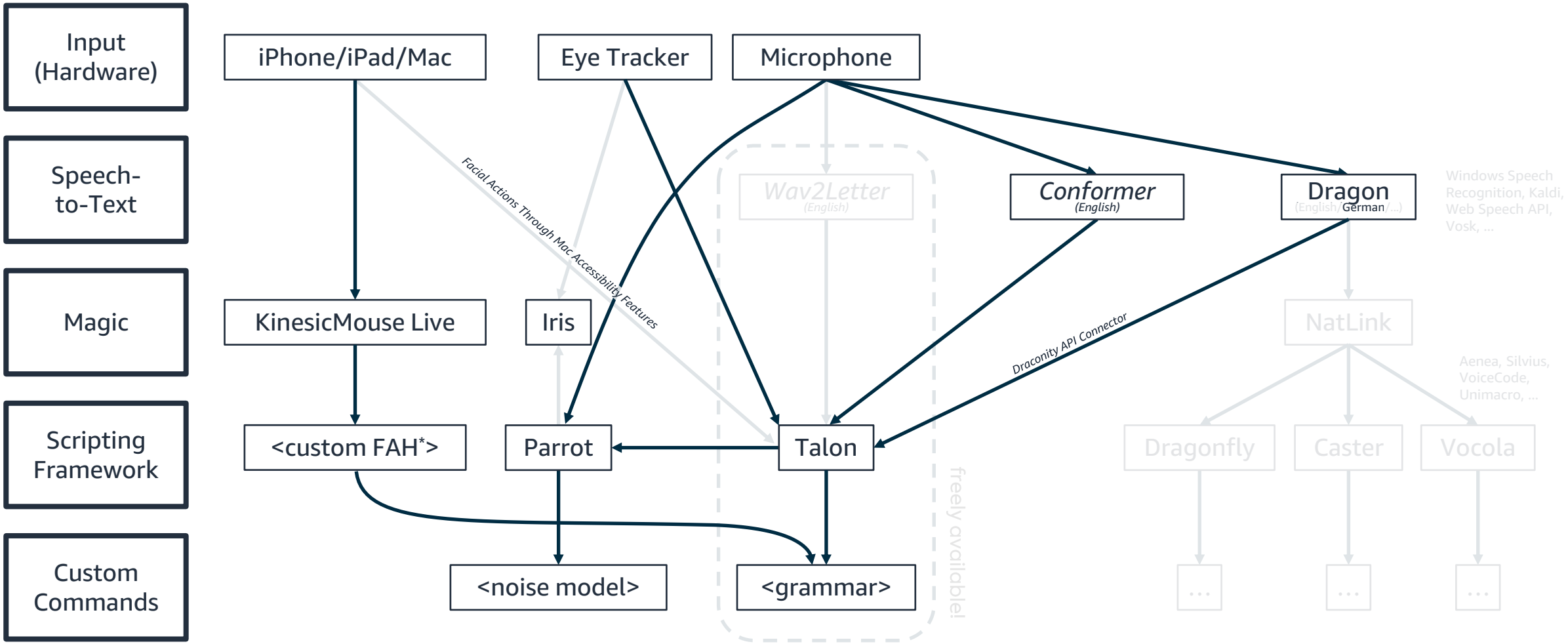
Universität Hamburg
DER FORSCHUNG | DER LEHRE | DER BILDUNG

Carl von Ossietzky
Universität
Oldenburg

Baqend

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

Look,
No Hands!

Handsfree Gaming: Eyes + Face + Voice/Noise

wolle.science/twitch

Pitfalls &
Challenges

Recognition Accuracy Issues

- **Microphone** determines accuracy!
  - *Build quality*: built-in < gaming headset < stage mic
  - *Positioning*: consistent, close to your mouth, away from all noise
  - *Mixed bag: Noise canceling* via hardware or software (e.g. RTX Voice)
- **Environment**: Minimize noise for you and annoyance for others!
  - Suspend ASR / mute mic accordingly (e.g. via push-to-talk pedal)
- **Homophones** should be avoided, e.g. through:
  - Grammar optimization to avoid ambiguity
  - Clear pronunciation

Potential Privacy Issues

- **Watch Your Tongue**: Passwords & confidential info may be leaked …
  - … through plain acoustics (beware *eavesdroppers*!)
  - … as they are stored your *command history*!
  - … to involved third parties (e.g. with *Web Speech*)
- **Watch Your Transmitter**: Wireless solutions are often not encrypted!
- **Watch Your Eyes**: Your eye movement may give away a lot

  → perhaps avoid continuous eye tracking ;-)

<insert eye tracking challenge joke here>

Workflow & Anger Management Issues

- **Beware the Trolls**: Having an audience generally does not help!
  - → Prepare to hear „Format C“ from your colleagues a lot
- **Keep your calm**: Shouting at the computer will not help, either!
  - → Stay in your neutral voice, even when raging inside …
- **Avoid Voice Strain**: Find a comfortable way to speak A LOT!
  - → e.g. use your natural voice & drink a lot of tea
- **Command chaining**: Anticipate what is going to happen!
  - → Practice, practice, practice!

General Issues

- **Multilanguage support** is still in ist infancy
  - → Non-English language models all have their problems
  - → Designing command libraries for different languages means effort
- **Complex setup** with many moving parts:
  - → Random stuff sometimes just happens, get used to it!
  - → Fallback to manual input sometimes necessary …
- **MACHINE LEARNING!!!**
  - → Models often reflect typical issues (data bias, data quality issues, …)
  - → Sometimes you have to just hope for the best …

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

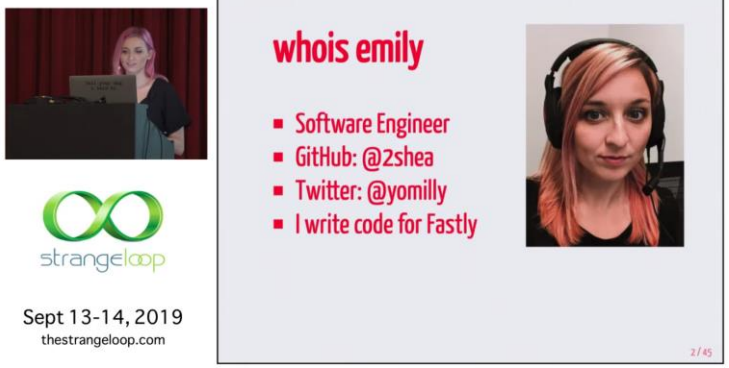
Tooling Recommendations (Incomplete!)

- **Talon** (Free of Charge): talonvoice.com / talon.wiki
  - Voice coding for Win / Linux / Mac!
  - Starter Grammar (English): github.com/knausj85/knausj_talon
- parrot.py (noise control): github.com/chaosparrot/parrot.py
- Cursorless (code editing for VSCode): github.com/cursorless-dev
- Rango (handsfree browsing): github.com/david-tejada/rango
- Paid Upgrades:
  - Talon Premium Support: patreon.com/join/lunixbochs
  - Dragon Speech Recognition: nuance.com/dragon/

Alternatives: Speech Recognition

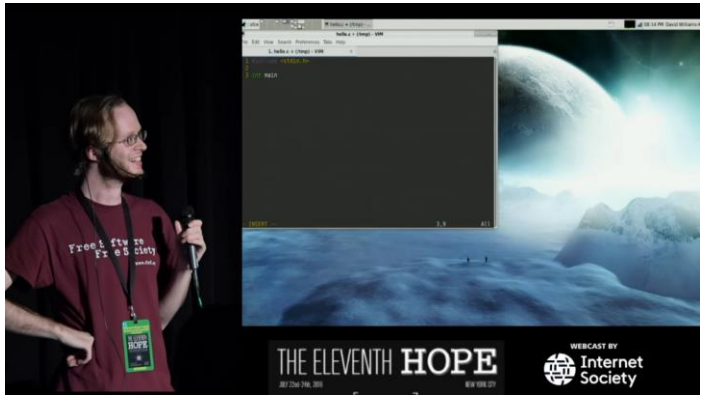
- Speech Recognition
  - WSR (Windows Speech Recognition): Built into Windows
  - Kaldi: github.com/kaldi-asr/kaldi
  - Vosk (ASR on mobile devices!): github.com/alphacep/vosk-api
  - Web Speech API (compatible with Talon through Chrome or Firefox)
- Scripting:
  - NatLink: sourceforge.net/p/natlink/
  - Dragonfly: github.com/dictation-toolbox/dragonfly
  - Caster: github.com/dictation-toolbox/Caster
  - Vocola (Voice Command Language): vocola.net

Recommended Talks

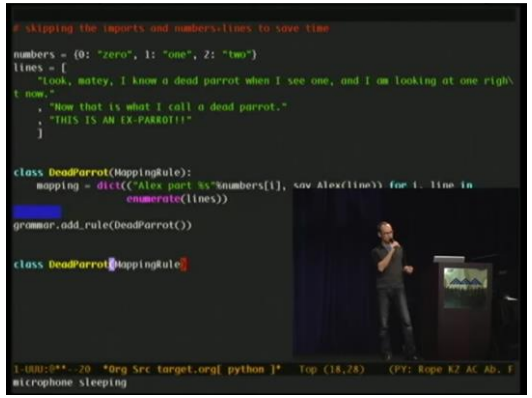


Emily Shea. Voice Driven Development: Who needs a keyboard anyway?, Strange Loop (2019)

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

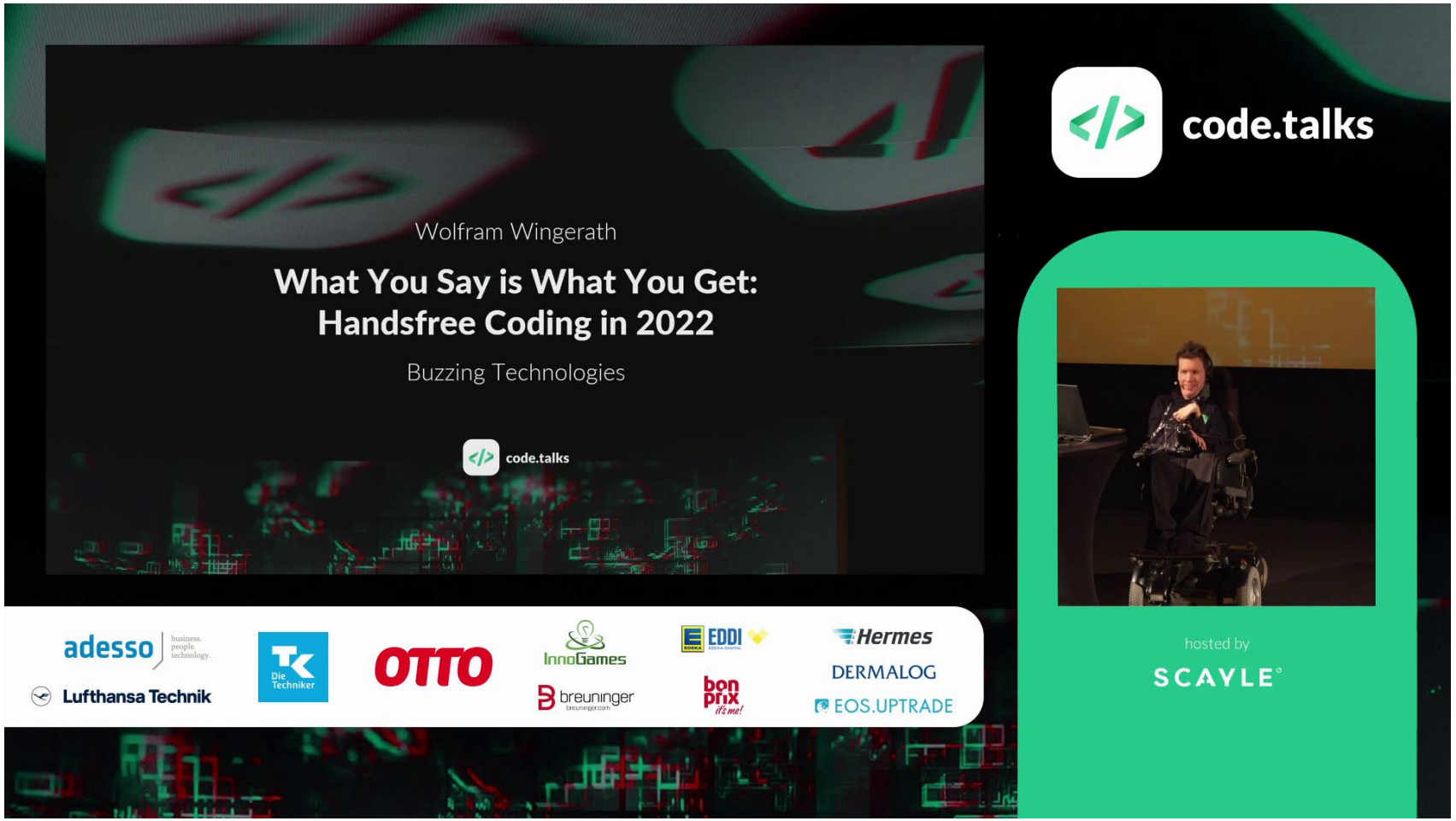


David Williams-King. Coding by Voice with Open Source Speech Recognition, The Eleventh Hope (2016)



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

CodeTalks Video

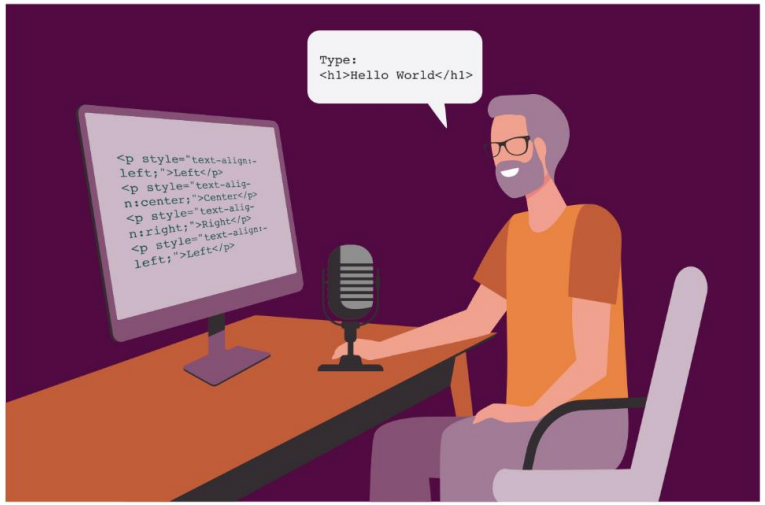


Wolfram Wingerath. What You Say is What You Get: Hands-Free Coding in 2022, CodeTalks (2022)

Articles & Blogs

- Emily Shea: whalequench.club/
  - Talon user
  - Very good starter instructions
- James Stout: handsfreecoding.org/
  - Dragonfly user
  - Huge collection of relevant blog posts
- Josh W. Comeau (2020): joshwcomeau.com/blog/hands-free-coding/
- Dusty Phillips (2020): dusty.phillips.codes/2020/02/15/on-voice-coding/
- Max Gravenstein (2018): medium.com/hubabl/handsfree-fe70980f36b/

Heise 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)

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.)

Thanks! So What Now?

Slack
talonvoice.slack.com

Join the community!

Subscribe to the mailing list!

GI Initiative
handsfree-coding.gi.de

Try out handsfree coding!

Patreon
patreon.com/lunixbochs

Support Talon Development!

Videos & Slides Available at https://wolle.science

Wolfram „Wolle“ Wingerath wolle@uol.de