

What You **Say** is What You Get

Handsfree Coding in 2021

DBDC / Munich, Germany

Wolfram Wingerath

Dec. 7, 2021

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



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**?

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)
- **Voice Coding:** optimized for actions & programming languages
 - Default: Everything is interpreted as a command
(Natural language through special keywords, e.g. `say <utterance>`)

Caster, Talon, ...

Where's the **Challenge**?



Outline: What This Talk is Going to Cover

1

My Personal Background

As data engineer at Baqend, I
use handsfree coding every day.

2

Demo & Usage Examples

Handsfree coding is awesome
and can be useful for everyone!

3

Setup & Best Practices

No-cost base setup with optional
upgrades (e.g. for eye tracking).

4

How to Get Started

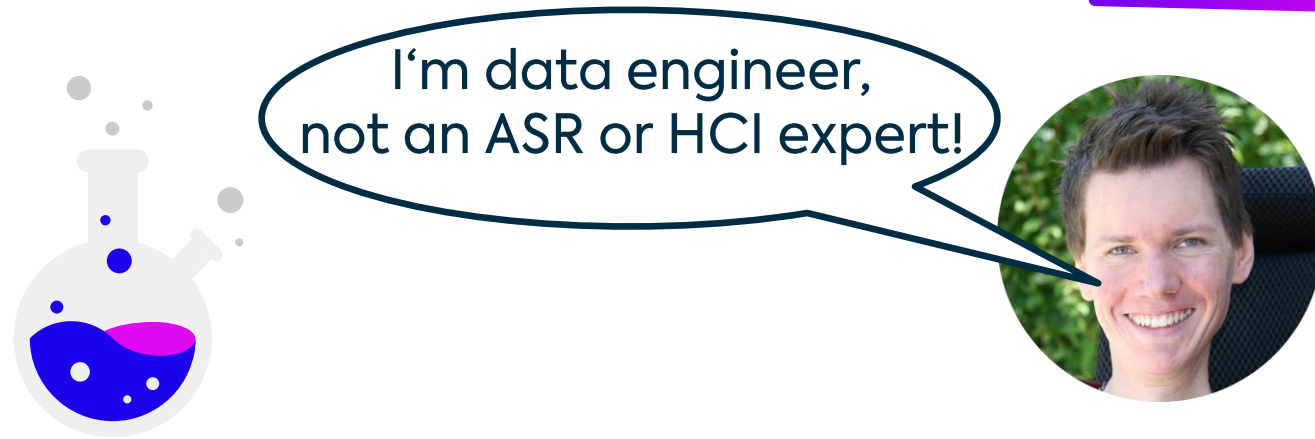
Videos, blogs, articles, support &
community – engage now!



The background image shows a person's hands using test equipment. A Tektronix TDS 3054 oscilloscope displays two waveforms on its screen. Next to it is a LEADER LV 5800 logic analyzer with a multi-screen display showing various digital signal parameters like 'Signal', 'TDS Pos', 'TDS Len', and 'Count'. The entire scene is overlaid with a semi-transparent purple filter.

My Job is Data Analysis

I Am Wolle



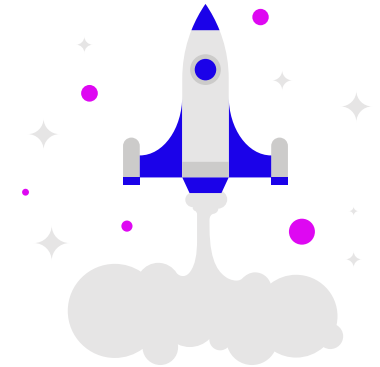
Research:

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



Practice:

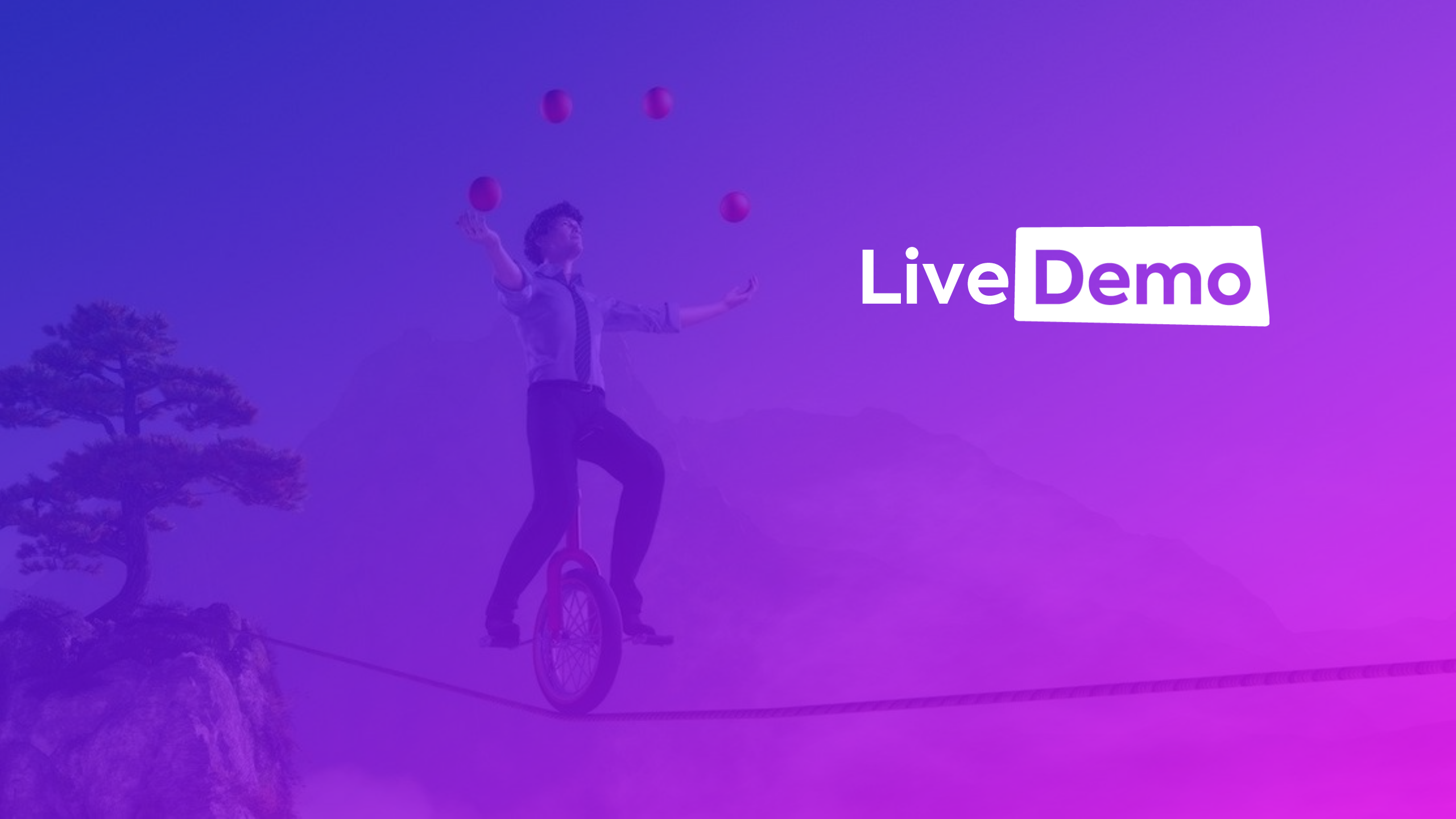
- Web Caching •
- Big Data Analytics •
- Anger Management •
- ... •



Universität Hamburg

DER FORSCHUNG | DER LEHRE | DER BILDUNG





Live Demo

Look,
No Hands!



Basic Voice **Control** (1/3)

- **Actions & Symbols:**

- Numbers, brackets, etc.: one space paren bang slash → 1_(!/
- Modifiers, e.g.: shift touch **or** hold alt **or** release control

- **Spelling through a phonetic alphabet:**

- NATO alphabet: alpha bravo charlie delta echo → abcde
- Optimized alphabet: air bat cam drum each → abcde

- **Typing:**

- Dictation: say hey comma you → hey, you
- Type recognized text: phrase hey comma you → hey comma you

Basic Voice **Control** (2/3)

- **Navigation, e.g.:**
 - **Cursor movement:** go left **or** go way up **or** home
 - **Selection:** mark left **or** mark way up **or** mark all
- **Formatters, e.g.:**
 - **Camel case:** camel hey you → heyYou
 - **Dashed:** kebab hey you → hey-you
 - **Dotted:** dotted hey you → hey.you
 - **Without whitespace:** smash hey you → heyyou
 - **Uppercase:** uppercase hey you → HEY YOU

Basic Voice **Control** (3/3)

- **Command History & Help:**

- Manage history: `command history` **or** `command history clear`
- Show & hide help: `help active` **or** `help close`

- **Application management, e.g.:**

- Launch & focus: `launch Firefox` **or** `focus Firefox`

- **Command Management for efficiency, e.g.:**

- Repetition: `one third` → `111`

- Chaining: `{paren close paren}{go left}{say hi}` → `(hi)`
1.) type: `()` 2.) move cursor: `←` 3.) dictate: `hi`

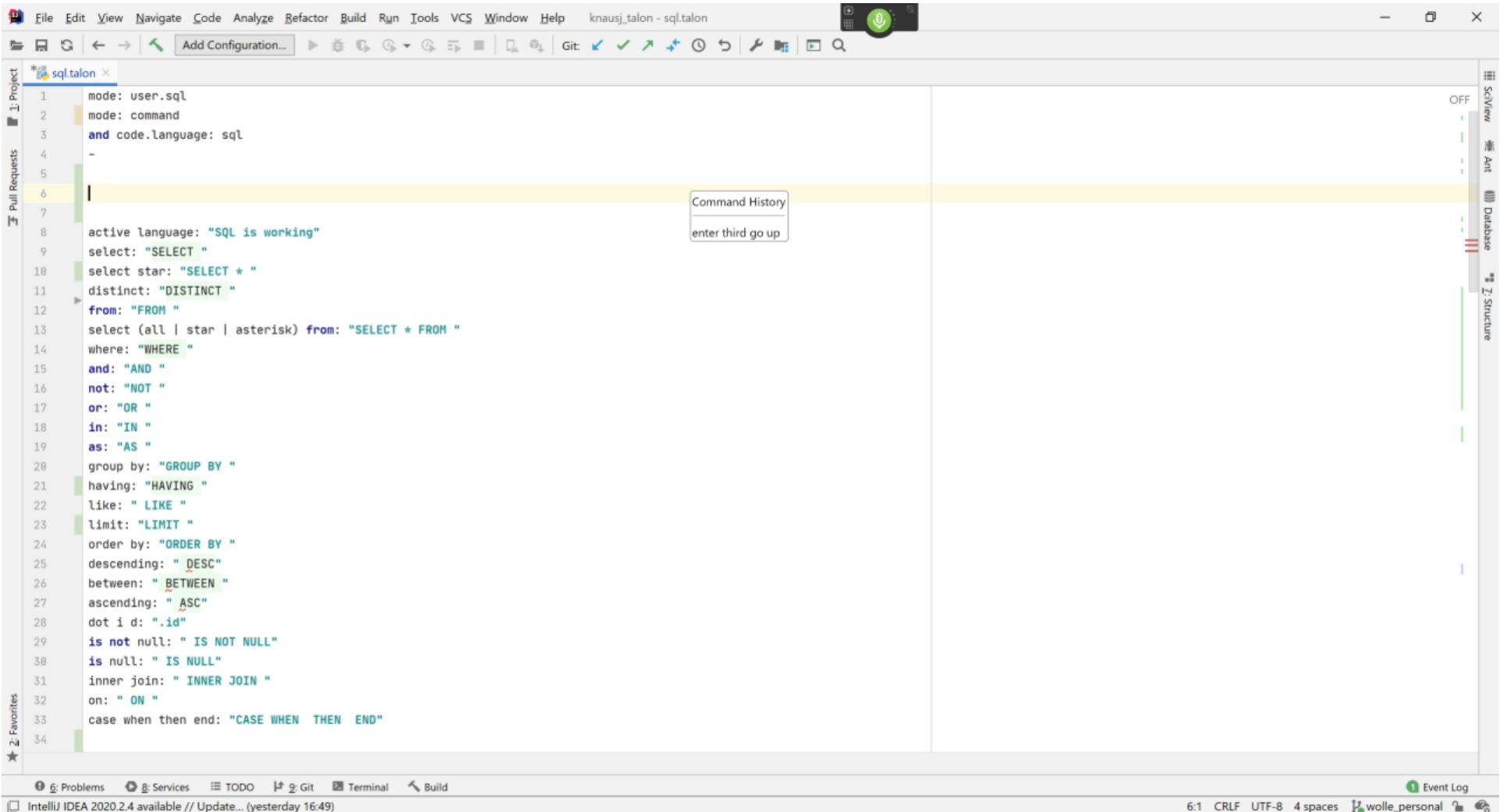
Dynamic Scripting

- **On-the-fly prototyping** for your grammar!
 - Python for live reloading
 - Voice coding inception: Work in the grammar with the grammar
- **Customization** of format and spelling:
 - Example: `volleh` → `Wolle`
- **Settings**, delays, speech engine, etc.

```
settings():  
    key_hold = 150.0  
    key_wait = 20.0  
    speech.engine = 'wav2letter'
```



Dynamic Scripting



Handsfree Software **Development**

- **Different modes** for different languages, 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 Software Development



The screenshot shows a terminal window with a dark background. At the top, the text "Demo 4 (Nav)" is displayed in yellow. Below it, the text "Code Navigation / grep / search" is shown in white. At the bottom of the terminal, a command prompt is visible with the text "1-00-:0** -20 target.org All (1,0) *E* (Org K2 Hl Ab. Fill)".



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

Handsfree Software Development

```
1  import React from 'react';  
2  import styled from 'styled-components';  
3  
4  function IconButton() {  
5  
6  }  
7  
8  export default IconButton;  
9
```

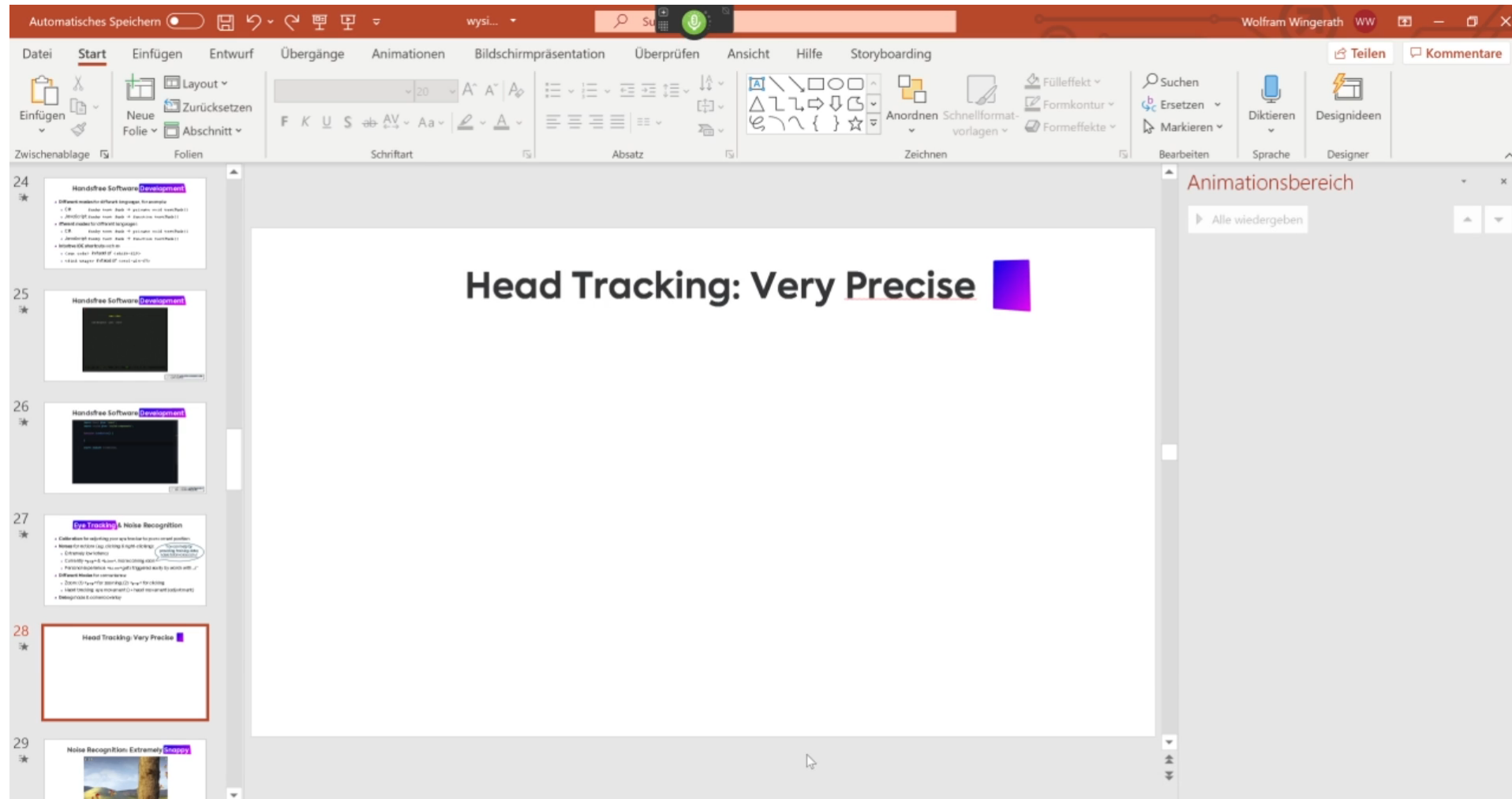


Eye Tracking & Noise Recognition

- **Calibration** for adjusting your eye tracker to your current position
- **Noises** for actions (e.g. clicking & right-clicking):
 - Extremely low latency
 - Currently **pop** & **hiss**, more coming soon
 - Personal experience: **hiss** gets triggered easily by words with „s“
- **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

You can help by
providing training data:
noise.talonvoice.com/

Head Tracking: Very **Precise**



Noise Recognition: Extremely **Snappy**

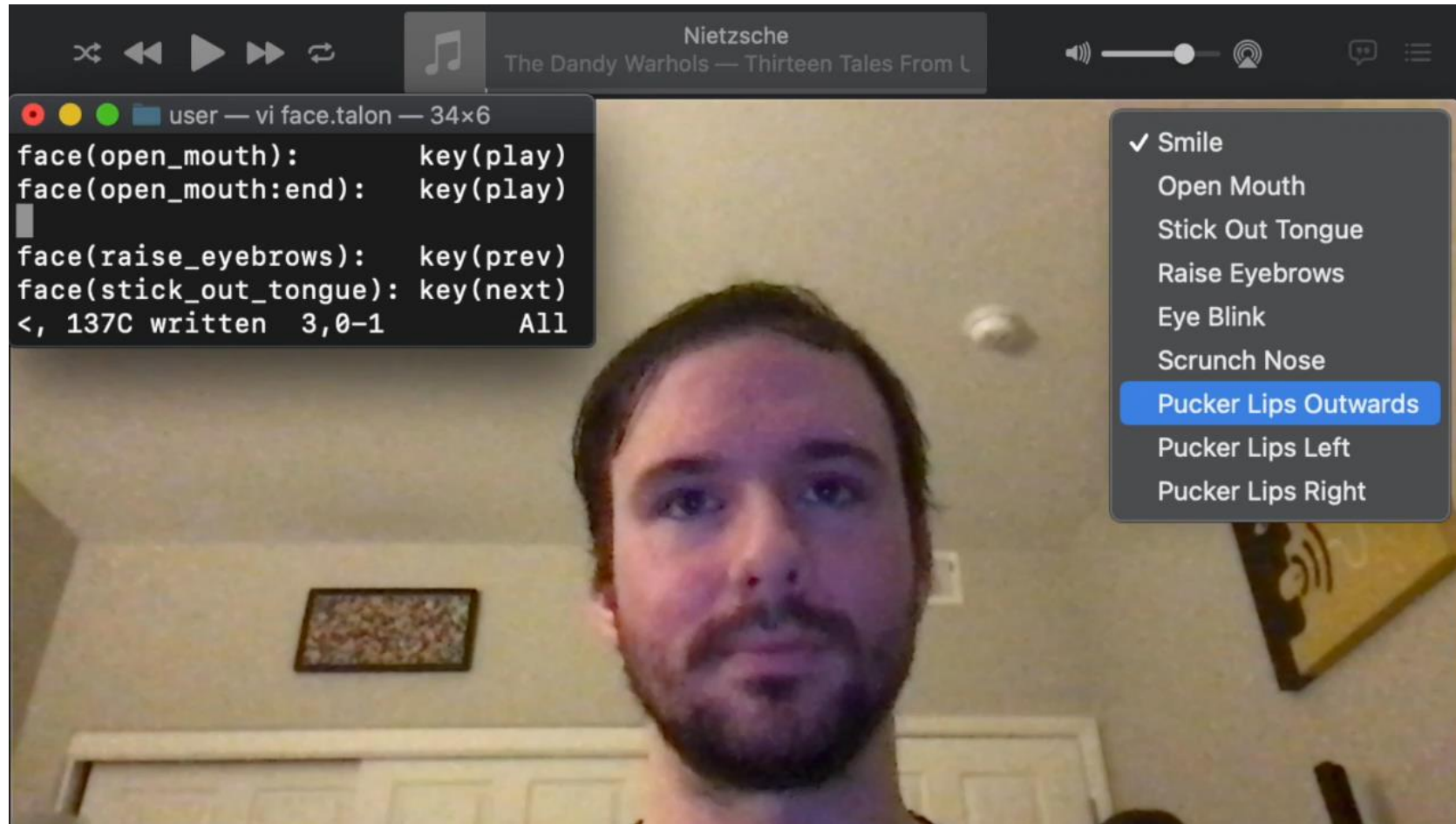


More Noise Recognition

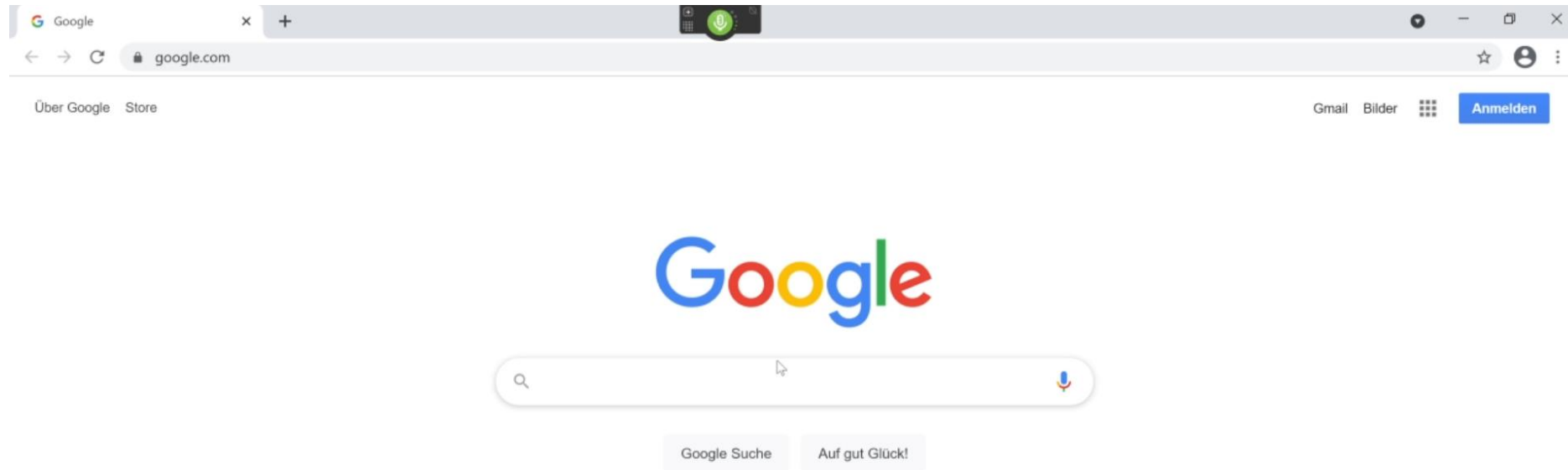


Cian Dowd. [Speed Runners Hands-Free](#),
YouTube (2021)

Facial Recognition



Zoom Mode: Even more Precise



Handsfree Browsing

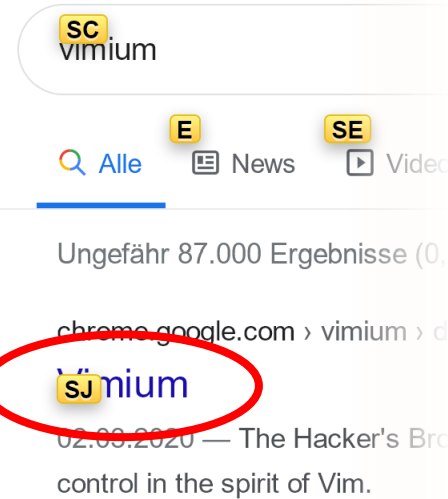
+ **Vimium** browser extension: vimium.github.io/

Example navigation without keyboard:

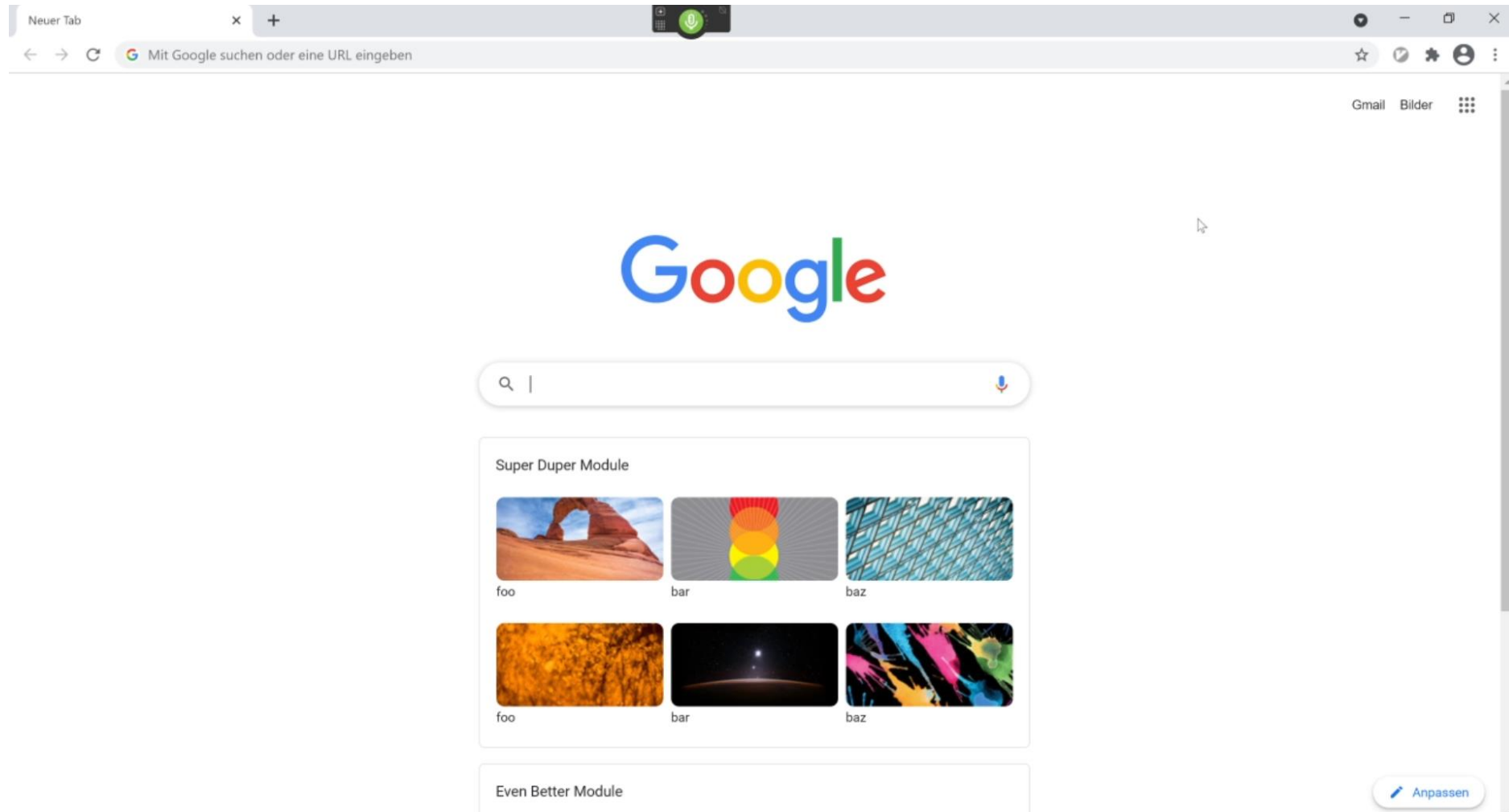
1. Show clickable links by pressing the `f` key
2. Press `sj` keys to click on Vimium link

+ **Talon** for Voice + eye + noise control

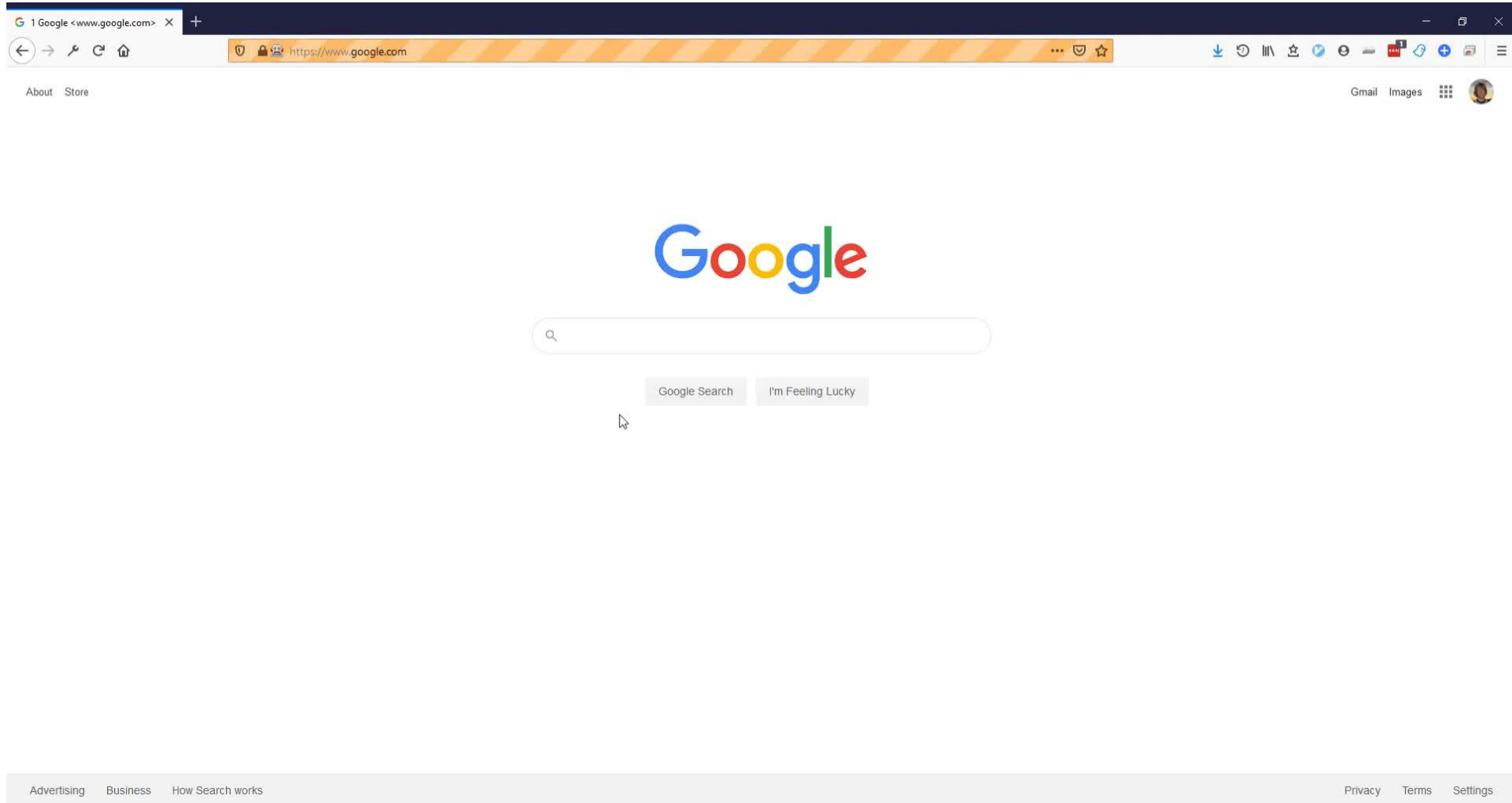
- Simplify shortcuts with easy-to-remember utterances
- Select links with your eyes



Handsfree Browsing



Hot New Stuff: Eye Tracking +

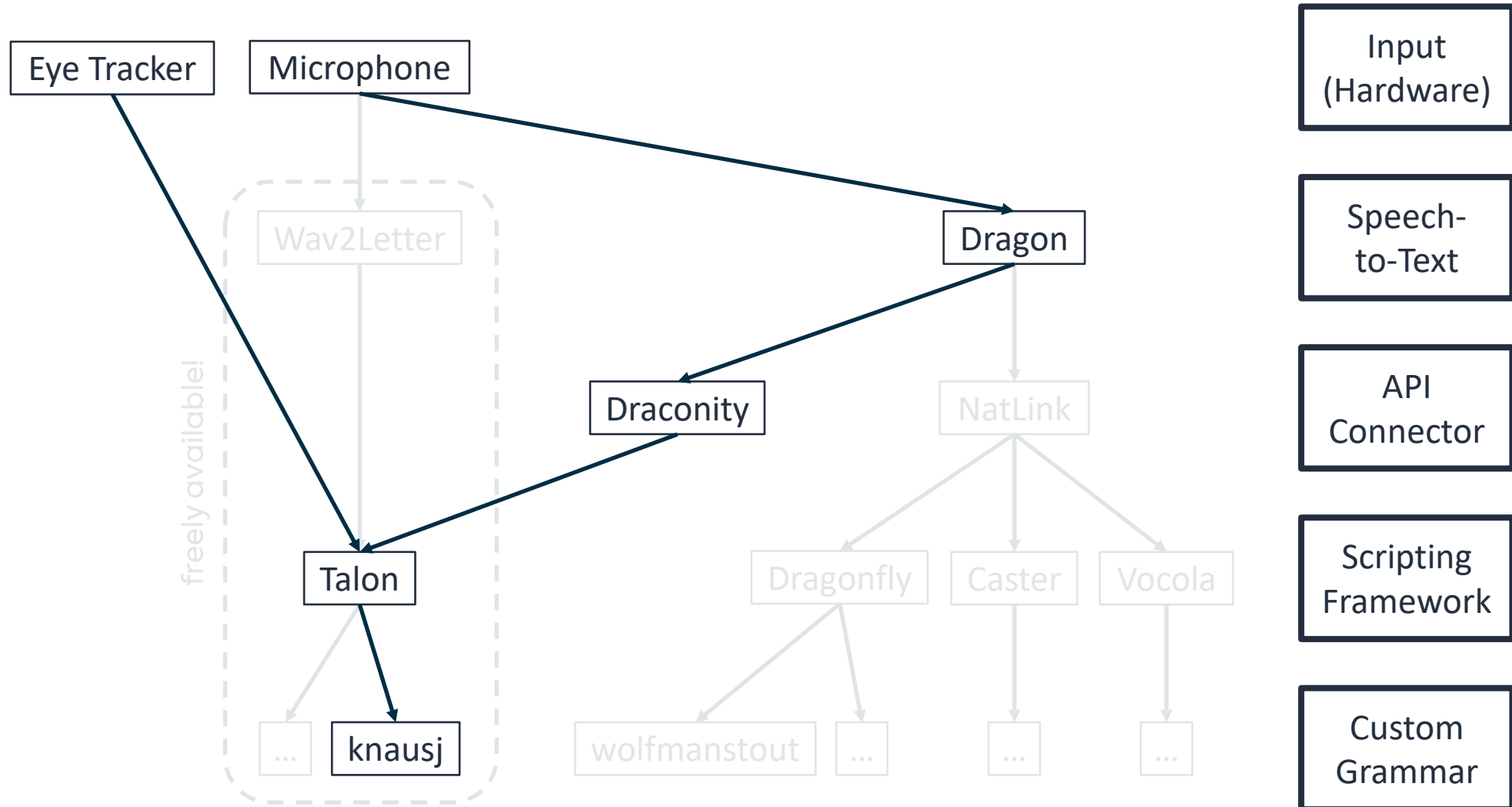


James Stout. [Say what you see: Efficient UI interaction with OCR and gaze tracking](#). Hands-Free Coding Blog (2020)

A professional microphone is positioned in the foreground, centered vertically. Behind it, a laptop screen is visible, displaying a webpage with a grid of images. The entire scene is overlaid with a semi-transparent purple gradient. The text 'The Base Setup' is centered over the microphone and screen.

The Base Setup

Popular Voice Coding **Stacks**: Overview



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

A woman with blonde hair is shown from the chest up, wearing a black headset microphone. She has both hands raised in the air, palms facing forward, in a gesture of presentation or enthusiasm. She is wearing a dark-colored top with a small, light-colored pattern. The background is a blurred office or meeting room with wooden paneling and a laptop visible on the right. The entire image is overlaid with a semi-transparent purple gradient.

Upgrades & Add-Ons

Supplementary Equipment

- **High-quality (!) microphone:** Get one now!
 - A wired (or good Bluetooth!) headset
 - Steno-mask: For noisy places & special looks
 - XLR headset for maximum accuracy



Sennheiser MB Pro 1/2



DPA 4188

→ + sender/receiver for max. convenience

- **Foot pedal for push-to-talk**



- **Eye tracker:** Tobii 4C & 5 supported by Talon

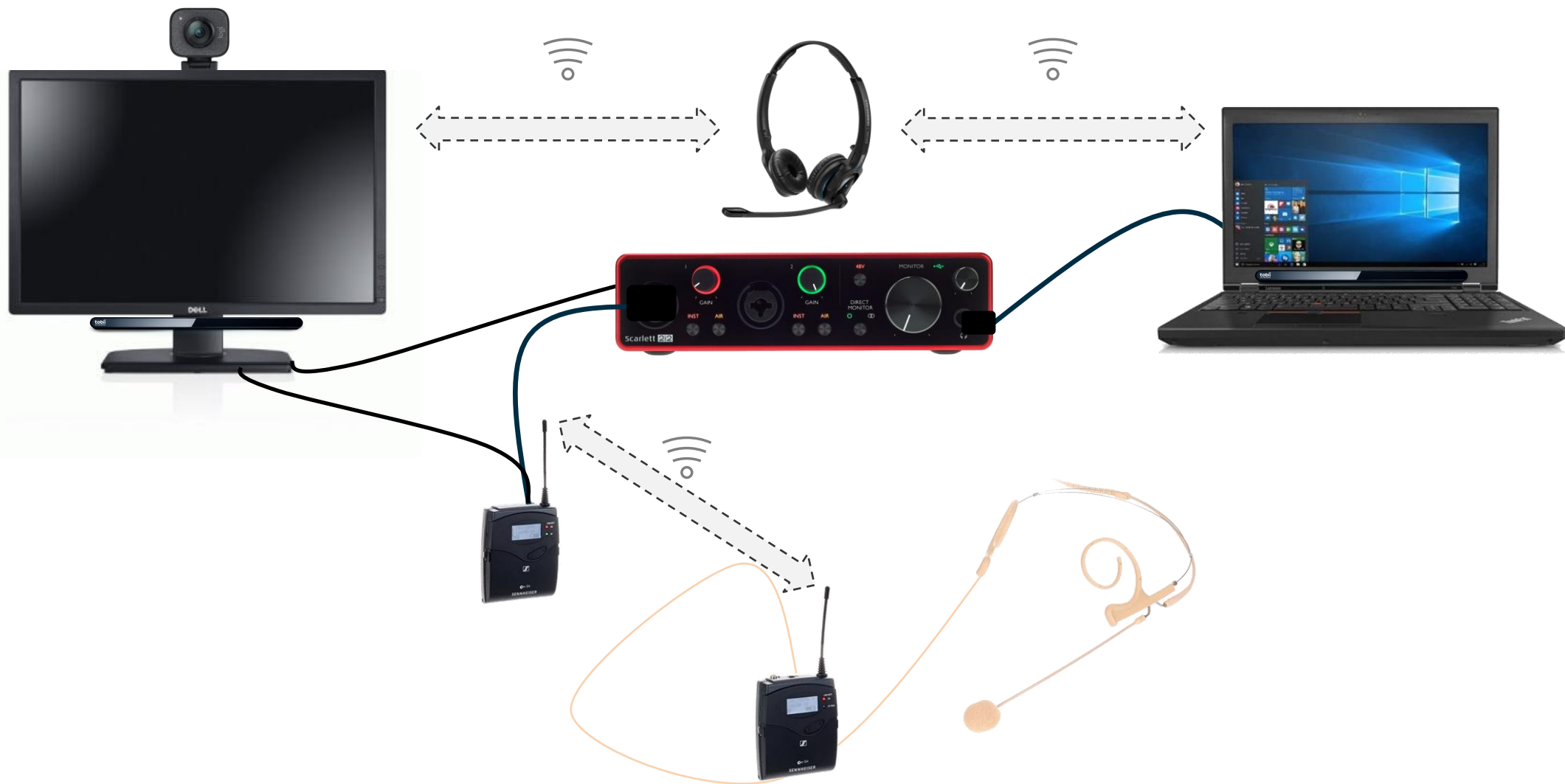


Sennheiser EW 112P G4 E-Band

- Multi-monitor support coming (?)



Multi-Computer Setup



A photograph of a person tripping over a banana peel on a sidewalk. The person is wearing a black shirt, red shorts, and blue sneakers. They are in the middle of a fall, with one leg bent and the other extended. A banana peel is on the ground near their foot. The background is a plain, light-colored wall. The entire image has a purple overlay.

Pitfalls & Challenges

Recognition Accuracy

- **Microphone** determines accuracy!
 - *Build quality*: built-in < gaming headset < stage mic
 - *Noise canceling* via hardware or software (e.g. RTX Voice)
 - *Positioning*: consistent, close to your mouth, away from all noise
- **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

General Issues

Vosk-based German/English starter:
github.com/mqnc/talon_german

- **Multilanguage support** is still in its infancy
 - Switching Dragon language profiles takes around 30-60 seconds ...
 - Using the command library with different languages is awkward ...
- **Eye tracker stability** is an issue as it disconnects sporadically
 - You will have to restart Talon from time to time ...
- **Random crashes** are rare, but do occur from time to time:
 - Fallback to manual input sometimes necessary ...

Workflow & Anger Management

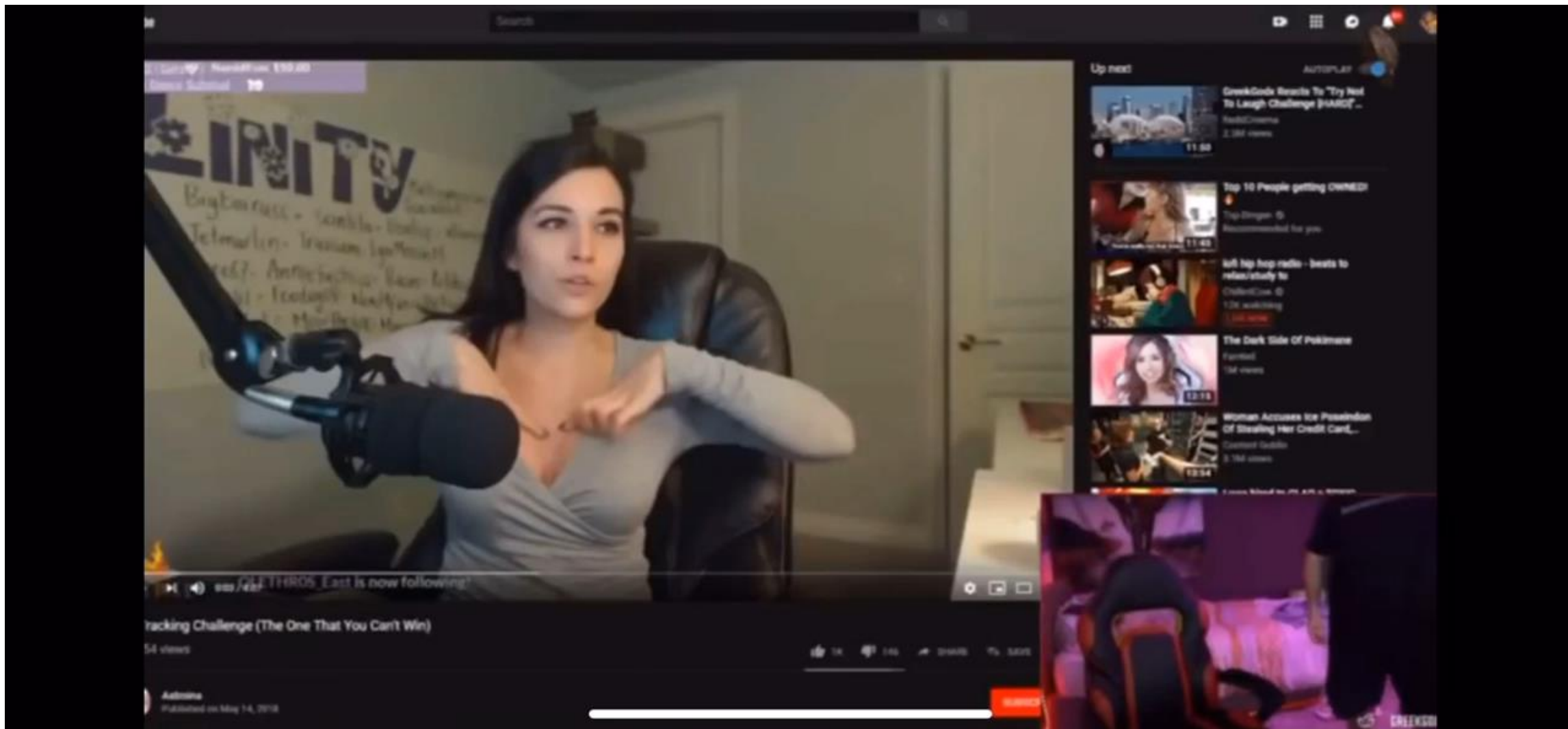
- **Beware the Trolls:** Having an audience generally does not help!
 - Prepare to hear „Format C“ from your colleagues a lot
- **Keep your calm:** Shouting at the computer will not help, either!
 - Stay in your neutral voice, even when raging inside ...
- **Avoid Voice Strain:** Find a comfortable way 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!

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 lame joke here>

Potential Privacy Issues



Greekgodx. [Eye Tracking Challenge](#), Twitch (2019)

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!!!

The **Hoff** approves!



A person with long hair is seen from behind, looking out at a harbor. In the background, several large cranes are visible against a sunset sky. The water reflects the light from the cranes and the sky. The overall scene is peaceful and contemplative.

Helpful Resources & Outlook

Recommended Tooling & Documentation

- **Talon** (Free of Charge): talonvoice.com
 - Voice coding for Win / Linux / Mac!
 - Wiki: talon.wiki
 - Speech Recognition: github.com/facebookresearch/wav2letter
 - Starter Grammar (English): github.com/knausj85/knausj_talon
- Paid Upgrades:
 - Dragon Speech Recognition: nuance.com/dragon/
 - Web Speech API (compatible with Talon through Chrome or Firefox)
 - Premium Support (Patreon): patreon.com/join/lunixbochs



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
- Scripting:
 - NatLink: sourceforge.net/p/natlink/
 - Dragonfly: github.com/dictation-toolbox/dragonfly
 - Caster: github.com/dictation-toolbox/Caster
 - Vocola (Voice Command Language): vocola.net

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/

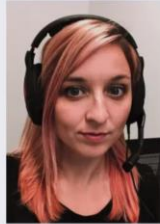
Recommended Talks



Sept 13-14, 2019
thestrangeloop.com

whois emily

- Software Engineer
- GitHub: @2shea
- Twitter: @yomilly
- I write code for Fastly



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



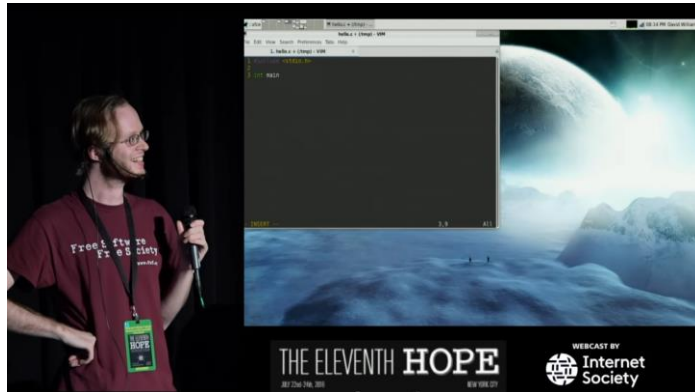
Dragonfly

Core Features

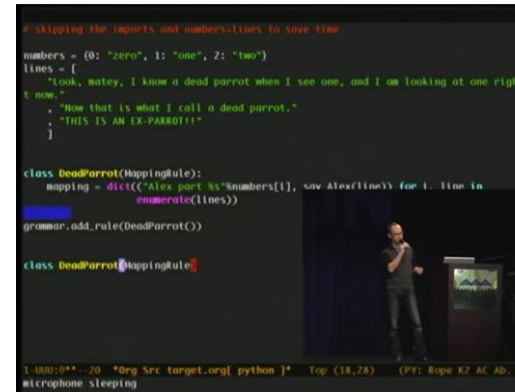
- Language Object Model
- Support for multiple speech recognition
 - Default supports DNS and WSR
- Built-in action framework
 - Knows how to talk to you



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!

REPORT | SOFTWAREENTWICKLUNG



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](https://wingerath.cloud/2021/ix), 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!

Soon also valid: wolle@uni-oldenburg.de

Wolfram Wingerath

wolle@baqend.com