What You Say is What You Get

Handsfree Coding in 2023

IGER 2023, Bamberg, Germany
July 22, 2023
Wolle

Video & 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'S GOOOOOO!!!!!!!
Let Me Just Show You How Easy It is

```
open (adult scrolls conflict for
delete adult scrolls conflict
for
```
Go Watch Emily’s Talk!
Where’s the Challenge?

ONE DOES NOT SIMPLY

TALK TO A COMPUTER
Where's the **Challenge**?

- **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!

Outline: What This Talk is Going to Cover

1. My Personal Background
   As data engineer & scientist, 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!
My Job is Data Science
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
• ...

Carl von Ossietzky
Universität Oldenburg

UH Universität Hamburg
DER FORSCHUNG | DER LEHRE | DER BILDUNG

Baqend
Look,
No Hands!
Basic Voice Control

- **Symbols, Modifiers & Navigation:**
  - Numbers, brackets, etc.: one space air bat shift one \(\rightarrow\) 1\_ab!

- **Spelling through a phonetic alphabet:**
  - NATO alphabet: alpha bravo charlie delta echo \(\rightarrow\) abcde
  - Optimized alphabet: air bat cam drum each \(\rightarrow\) abcde

- **Command Management** for efficiency, e.g.:
  - **Chaining:**
    1.) type: paren close paren \(\rightarrow\) (hi)
    2.) move cursor: go left \(\leftrightarrow\) hi
    3.) dictate: say hi \(\rightarrow\) (hi)

- **On-the-fly grammar prototyping** through Python live reloading!
mode: sleep

hello Bamberg:
key(down)
"now we have a button press and text"
• **Context-dependent behavior**, 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.:
  ```csharp
  action(user.code_state_if):
      insert("if () {}")
      key(left enter up end left left left)
  ```
import React from 'react';
import styled from 'styled-components';

import Icon from '@components/Icon';

function IconButton({ icon, children }) {
  return (
    <Wrapper>
      <Icon icon={icon} />
      {children}
    </Wrapper>
  );
}

const Wrapper = styled.button`
  background-color: var(--color-primary);
  font-size: 2rem;
  cursor: pointer;`
Handsfree Coding: Talon + Cursorless

- Available on GitHub: [github.com/cursorless-dev/](https://github.com/cursorless-dev/)
- VSCode extension
- Spoken language for **structural code editing**
  - Decorates every token on screen with a **mark**
  - Tokens can be selected via combination of mark and **scope**
  - **actions** operate on the specified tokens

- Example: `bring call vest`

  action scope mark

  (copy the function call with the marked „v“ to where my cursor is)
Handsfree Coding: Talon + Cursorless

(moving code around)
Handsfree Coding: Talon + Cursorless

(const swaping arguments)
Handsfree Coding: Talon + Cursorless

(selecting semantic entities)
Handsfree Browsing

+ **Vimium** browser extension: [vimium.github.io/](https://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

+ Natively compatible with tools like **Talon**
  
  - Simplify shortcuts with easy-to-remember utterances
  - Optional: eye tracking + noise control to select links with your gaze

+ **Rango** browser extension for Talon: [github.com/david-tejada/rango](https://github.com/david-tejada/rango)
UTTER COMMAND: WHY I REWRIT MY ENTIRE GRAMMAR

For years, I've been approach speech recognition like a hacker/ inplicant. I enjoy finding tools to learn about and applying them to situations. Over the years, I've used many different tools and systems. However, I'm now using Gaze OCR.

I recently read an article by Kim Patch about using Gaze OCR to improve the accuracy of speech recognition. The article was fascinating and I was particularly impressed with how easy it was to use and how effective it was at identifying my spoken words.

Kim described using Gaze OCR to improve the accuracy of speech recognition by placing a small device on the user's face. The device is able to track the user's gaze and use that information to improve the accuracy of the speech recognition system.

Kim also described how the device is able to track the user's gaze and use that information to improve the accuracy of the speech recognition system. The user can then use the device to control the speech recognition system by looking at specific commands.

Kim's work showed that there is a Better Way. I'm happy to say that she's done a phenomenal job at making their ideas work, but you don't have to go to all the trouble Kim's gone through. I encourage you to explore her website, but make sure you don't miss Human Machine Grammar - The Rules.

I'm really impressed with her biggest ideas — although it's also very helpful to see exactly how she implemented them. In her own words:

"I'm excited about the possibilities of using speech recognition, but I also think we need to be careful about how we use it. It's not just a matter of improving accuracy; it's also a matter of being considerate of the user. We need to make sure that we're not using speech recognition as a way to control people, but rather as a way to help them.

I'm also very interested in the possibilities of using speech recognition to help people with disabilities. I think there's a lot of potential for this technology to be used in ways that are really helpful to people who have physical disabilities.

I've been researching speech recognition for a long time, and I'm really excited to see how it's developing. I think there's a lot of potential for this technology to be used in ways that are really helpful to people.
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!
Handsfree Gaming: Noise Recognition
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 via Parrot
- **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

[github.com/chaosparrot/parrot.py]
Handsfree Gaming: Head + Eye Tracking
Handsfree Gaming: Facial Actions
Popular Handsfree Coding Stacks: Overview

Input (Hardware)
- iPhone/iPad/Mac
- Eye Tracker
- Microphone

Speech-to-Text
- Wav2Letter (English)
- Conformer (English)
- Dragonfly
- NatLink

Magic
- KinesicMouse Live
- Iris
- Parrot
- Talon

Scripting Framework
- <custom FAH>*
- <noise model>
- <grammar>

Custom Commands
- Dragon
- Conformer
- Magic

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

This overview was inspired by:
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 mic for maximum accuracy
  - + sender/receiver for max. convenience

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

- **Eye tracker**: Tobii 4C & 5 supported by Talon
  - Multi-monitor support coming (?)
Multi-Computer Setup
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 ;-)
Potential Privacy Issues
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!
- **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 ...
I just used 3,000 GPU-hours to test all 9 new OpenAI Whisper speech recognition models, two Talon acoustic models, and NVIDIA Nemo large and xlarge models. Whisper has a peculiar failure case. Here's what I think.

9:50 nachm. · 27. Sep. 2022 · Twitter Web App

Ryan Hileman @linuxbochs

Whisper sometimes exhibits what I would call "catastrophic" failures in recognition quality. I've provided some examples in the linked sheet, and I'll talk more about this downstairs.
"short context" / vocabulary tests. It places somewhere between Whisper Small and Whisper Base on those tests.

Ryan Hileman @lunixbochs - 27. Sep.
Whisper was painfully slow compared to the other models tested. I achieved much higher throughput when running my GPU tests on the largest Talon 1B model and Nemo xlarge (500M) model than any Whisper model, including Whisper Tiny (39M).

Ryan Hileman @lunixbochs - 27. Sep.
Whisper output "feels" great. It is very good at producing coherent speech, even when it is completely incorrect about what was said. While analyzing some "worst case" outputs (highest error %), I saw an audio clip of only the word "partnerships" transcribed by Whisper Large as:

"That's the end of the video. Thank you for watching. Lots of heat, December. Keep writing your comments below for new videos and feel free to contribute. If you have any questions, feel free to ask away, or make comments, post any of your comments. I'll see you next week."
"short context" / vocabulary tests. It places somewhere between Whisper Small and Whisper Base on those tests.

Ryan Hileman @lunixbochs · 27. Sep.
Whisper was painfully slow compared to the other models tested. I achieved much higher throughput when running my GPU tests on the largest Talon 1B model and Nemo xlarge (SOOM) model than any Whisper model, including Whisper Tiny (39M).

Ryan Hileman @lunixbochs · 27. Sep.
Whisper output "feels" great. It is very good at producing coherent speech, even when it is completely out of context. While analyzing some audio files I noticed that whisper had the worst performance (44%). I saw an audio clip of only the word "dataset" transcribed by Whisper as:

"That's the end of the video. Thank you for watching. Lots of heat, December. Keep up with your comments below for new videos and feel free to contribute. If you have any questions, feel free to ask away, or make comments, post any of your comments. I'll see you next week."

Thread

What you say
What you get
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!!!

*It's Awesome!*
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
Wolfram Wingerath

What You Say is What You Get: Handsfree Coding in 2022

Buzzing Technologies

[Image: Speaker presenting at an event]

[Image: Logos of sponsors]

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/
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 programmiert.

zung des Computers ganz ohne Einsatz ihrer Hände.

Wolle ist 33 Jahre alt, Data Engineer und erprobt seit mehr als zehn Jahren Eingabe- methoden 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, Gestuschen, Mimik oder Gestik den Computer und die Programme steuern“ sagt er.

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
- GI Initiative: handsfree-coding.gi.de
- Patreon: patreon.com/lunixbochs

- Ask questions!
- Enjoy the community!
- Try out handsfree coding!
- Support Talon Development!

Video & Slides Available at https://wolle.science
Wolfram Wingerath  wolle@uol.de