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'S GOOOOOO!!!!!!
Let Me Just Show You How Easy It is
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)
- **Voice Coding:** optimized for *actions & programming* languages
  - **Default:** Everything is interpreted as a command
    (Natural language through special keywords, e.g. `say <utterance>`)
Where's the Challenge?

ONE DOES NOT SIMPLY
TALK TO A COMPUTER
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!
My Job is Data Analysis
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
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.

```python
settings():
    key_hold = 150.0
    key_wait = 20.0
    speech.engine = 'wav2letter'
```
```sql
mode: user.sql
mode: command
and code.language: sql

active language: "SQL is working"
select: "SELECT "
select star: "SELECT * "
distinct: "DISTINCT "
from: "FROM "
select (all | star | asterisk) from: "SELECT + FROM "
where: "WHERE "
and: "AND "
not: "NOT "
or: "OR "
in: "IN "
as: "AS "
group by: "GROUP BY "
having: "HAVING "
like: " LIKE "
limit: "LIMIT "
order by: "ORDER BY "
descending: " DESC "
between: " BETWEEN "
dot id: ".id"
is not null: " IS NOT NULL "
is null: " IS NULL "
inner join: " INNER JOIN 
on: " ON 
case when then end: "CASE WHEN THEN END "
```
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.:
  
  ```csharp
  action(user.code_state_if):
  insert("if () {}")
  key(left enter up end left left left)
  ```
Handsfree Software Development

Demo 4 (Nav)

Code Navigation / grep / search

Handsfree Software Development

```javascript
import React from 'react';
import styled from 'styled-components';

function IconButton() {

}

export default IconButton;
```
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
Facial Recognition

User interface with options for facial expressions:

- face(open_mouth): key(play)
- face(open_mouth:end): key(play)
- face(raise_eyebrows): key(prev)
- face(stick_out_tongue): key(next)
- <, 137C written 3,0-1 All
- Smile:
  - Open Mouth
  - Stick Out Tongue
  - Raise Eyebrows
  - Eye Blink
  - Scrunch Nose
- Pucker Lips Outwards
- Pucker Lips Left
- Pucker Lips Right
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 + OCR

[Image of Google search engine]

Popular Voice Coding Stacks: Overview

- **Input (Hardware)**
- **Speech-to-Text**
- **API Connector**
- **Scripting Framework**
- **Custom Grammar**

![Diagram showing relationships between different voice coding tools and stacks:](https://dictation-toolbox.github.io/dictation-toolbox.org/) (accessed: January 4, 2021)

- **Eye Tracker**
- **Microphone**
- **Wav2Letter**
- **Dragon**
- **Draconity**
- **Talon**
- **knausj**
- **wolfmanstout**
- **NatLink**
- **Dragonfly**
- **Caster**
- **Vocola**
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
  - + sender/receiver for max. convenience
- **Foot pedal** for push-to-talk
- **Eye tracker**: Tobii 4C & 5 supported by Talon
  - Multi-monitor support coming (?)
Recognition

- **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

- **Multilingual 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 ...

Vosk-based German/English starter: [github.com/mqnc/talon_german](https://github.com/mqnc/talon_german)
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 ;-)
Potential Privacy Issues
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!**
Recommended Tooling & Documentation

- **Talon** (Free of Charge): [talonvoice.com](https://talonvoice.com)
  - Voice coding for Win / Linux / Mac!
  - Wiki: [talon.wiki](https://talon.wiki)
  - Speech Recognition: [github.com/facebookresearch/wav2letter](https://github.com/facebookresearch/wav2letter)

- **Paid Upgrades:**
  - Dragon Speech Recognition: [nuance.com/dragon/](https://nuance.com/dragon/)
  - Web Speech API (compatible with Talon through Chrome or Firefox)
  - Premium Support (Patreon): [patreon.com/join/lunixbochs](https://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/
Recommended Talks


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

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.)
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.

zungen 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, Gestischen, Mimik oder Gesten den Computer und die Programme steuern“, sagt er.

Thanks! So What Now?

- Slack: talonvoice.slack.com
- Ask questions! Enjoy the community!
- GI Initiative: handsfree-coding.gi.de
- Try out handsfree coding!
- Patreon: patreon.com/lunixbochs
- Support Talon Development!
- Soon also valid: wolle@uni-oldenburg.de!

Subscribe to the mailing list!

Wolfram Wingerath  wolle@baqend.com