

# Voice

**Speaking with Your Computer**

**Craig Buchek**

# Part 1

# Part 1

- Text-to-Speech (TTS)
- Dictation
- Voice Control

# Text-to-Speech (TTS)

# Demo: SAM

- **SAM** (Software Automatic Mouth)
- 1982 by **Don't Ask Software**
- Commodore C64, Apple II, Atari 400/800
- **Text-To-Phoneme**
  - HEH3LOW, MAY3 NEY3M IHZ KRAY5G
- Phoneme-To-Speech
- Macintalk - **1984 Macintosh demo**

# Text-to-Speech

- Reading text aloud
- Accessibility
  - Screen readers
    - Visually impaired
    - Apple VoiceOver

# Text-to-Speech

- Voice assistants
  - Siri, Alexa, Google Assistant, etc
- Navigation
  - Turn-by-turn directions
- Audio books
- Automated phone systems (IVR)

# Text-to-Speech

- Real-time translation
- Voice cloning
- Multi-modal generative AI



# Demo: MacOS Live Speech

- Control Center
  - Accessibility > Hearing > Live Speech

# Demo: MacOS Spoken Content

- System Settings
  - Accessibility > Spoken Content
    - select a System Voice
      - Alex, Samantha, etc
    - enable Speak Selection
    - add custom Pronunciations
- Adds a "Speech" menu item to the context menu

# Dictation

# Dictation

- Speech recognition
- Voice input to text
- Natural Language Processing (NLP)

# Speech Recognition

- Engines
- Models

# Speech Recognition

- Language
  - Vocabulary
- Accuracy
- Speed
  - Latency
- Speaker independence

# Recognition Engines

- [CMU Sphinx](#) (free)
- [Whisper](#) (OpenAI) (free)
- [DeepSpeech](#) (Mozilla) (free)
- Mac Voice Control engine
- [Flashlight ASR](#) (Facebook AI Research) (free)
  - formerly [WAV2Letter](#)

# Recognition Engines

- Picovoice (\$\$\$\$)
- Amazon Transcribe
- IBM Watson
- Google Cloud Speech-to-Text
- Microsoft Azure Speech Service
- Microsoft Dictate



# Sphinx

- Original version
  - 1998
- CMU Sphinx
  - Sphinx4 (Java)
- PocketSphinx (Android)

# Recognition Models

- Conformer-2
  - AssemblyAI
- Conformer-D
  - used by Talon
- Vosk (free)
- Kaldi (free)
- Zamia (free)
- Julius (free)

# Demo: MacOS Dictation

- Built-in dictation feature
  - MacOS, iOS, iPadOS
- System Settings > Keyboard > Dictation
  - Shortcut: Press Control key twice

# Voice Control

# Voice Control

- System commands
- Application control
- Custom commands
- Productivity benefits

# Demo: Apple Voice Control

- System Settings > Accessibility > Motor > Voice Control
  - Enable Voice Control
  - Vocabulary
    - Custom words and phrases
  - Commands

# Demo: Apple Voice Control

- "Show Commands"
- "Show Numbers"
- "Show Grid"
- "Command Mode"
- "Dictation Mode"
- "Spelling Mode"

## Demo: Apple Voice Control

- "Click <menu> menu"
- "Click <button>"
- "Click and hold mouse"
- "Search for <phrase>"
- "Scroll <direction>"



# Demo: Apple Voice Control

- "Press <key> key"
- "New line"
- "Type <phrase>"
- "Type <letters>"
- "<emoji> emoji"

## Demo: Apple Voice Control

- "Insert <phrase> before/after <phrase>"
- "Replace <phrase> with <phrase>"
- "Select <phrase>"
- "Extend selection 2 words"
- "Capitalize/lowercase/uppercase that"
- "Put quotes around that"

# Demo: Apple Voice Control

- Demo

# Talon

- Talon Voice Control
  - Free + Premium
  - MacOS, Windows, Linux
  - Python
- Talon Community
- Talon Slack

# Talon - Speech Engine

- Needs a speech recognition engine
  - Conformer D (free)
    - part of Talon, but separate download
  - Conformer b108 (free)
  - **Whisper?** (premium)
  - Dragon Dictate (Windows, \$\$\$)
  - Wav2letter (free)
    - now Flashlight Automatic Speech Recognition
    - from Facebook AI Research
  - **Vosk?** (free)

# Talon User Files

- maps words to commands/actions
- Talon Community
  - previously "knausj\_talon"
- [https://talon.wiki/integrations/talon\\_user\\_file\\_sets/](https://talon.wiki/integrations/talon_user_file_sets/)

# Specialized Talon User Files

- Mouse movement
- Vim
- Web browsers
- UI navigation
- Text snippets
- AXKit (macOS accessibility API)
- Cursorless (VS Code)

# Talon Demo - Help

- "help active"
  - show commands available in active application
- "help search <term>"
- "help context <term>","help context"
- "help alphabet"
- "help symbols"
- "help format"



# Talon Demo - Modes

- Command mode
  - "command mode"
- Dictation mode
  - "dictation mode"
  - "say <phrase>", "sentence <phrase>", "title <phrase>", "kebab <phrase>"
- Sleep mode
  - "talon sleep", "talon wake", "go to sleep", "wake up"
- Voice typing mode
- Spelling mode?
  - "spell <word>", "spell that"

## Talon Demo - Mouse

- "control mouse"
- "touch", "righty", "mid click"
- "command click", "shift click"
- "dub click", "duke", "trip lick"
- "drag", "end drag"
- "mouse grid, 3, 1", "grid off"

# Talon Demo - Navigation

- "focus chrome"
- "go 2 words right"
- "tail", "head", "go line start"
- "go top"
- "go 3 down"
- "go 12"
- "go page down" (or just "page down" to use key binding)
- "go way down"
- "find it"
  - "next one"
- "file save"

## Talon Demo - Navigation

- "file save"
- "indent", "indent less"
- "yes I am sure", "cancel" (dialogs)
- "desk 2", "desk next", "window move desk 3"
- "navigate comma" moves after the next "," on the line
- "navigate before five" moves before the next "5" on the line
- "navigate left underscore" moves before the previous "\_" on the line
- "navigate right after second plex" moves after the second "plex" on the line
- "navigate phrase hello world" moves after the first "hello world" on the line
- "big word neck/pre", "small word neck/pre"

## Talon Demo - Editing

- "new line"
- "undo that", "redo that", "repeat that 4 times"
- "copy that", "cut that"
- "paste that", "paste match"
- "clone that"
- "new line above"
- "select that", "before that", "extend that"
- "select all"?
- "nope that", "scratch that"
- "spell that alpha bravo ...", "spell that title case alpha bravo ..."

# Talon Demo - Formatting

- "help reformat"
- "abbreviate ..."
- "dubstring snake that"
- "title how's it going"

# Talon Demo - Voice Typing

- "air bat cap"
- "help modifiers"
- "command shift 5" (screenshot)
- "press enter"
- "press right shift twice"?
- "press function"?
- "arrow", "dub arrow"
- "spamma" (comma + space)
- "emoji poop", "emoji thumbs up", "emoji happy", "emoji crying"
- "emoticon laughing", "emoticon sad", "emoticon wink"

## Talon Demo - Text Snippets

- "round", "round that" (parens)
- "box"
- "diamond"
- "curly"
- "twin"
- "quad"
- "skis"
- "percentages"
- "pad"



# Talon Demo - Dictation

- "help dictation"
- "ignore" (ignore next phrase, to talk to someone)
- "recent list"
- "phones that", "phones 3rd word left"
- "date insert", "timestamp insert UTC"

## Talon Demo - Media

- "volume up/down/mute"
- "media play/pause"
- "media next/previous"
- "microphone show", "microphone pick <name>"

## Talon Demo - Terminal

- "lisa"
- "katie up"
- "go work"?
- "disk free human"
- "change owner", "change mode recurse"
- "make dear", "work dear"
- "sort human"
- "SSH", "pseudo"
- "translate", "unique"
- "word count lines"

# Talon - Macros

- "help macros"
- "macro list"
- "macro record", "macro stop"
- "macro save as <name>"
- "macro play <name>"

# Talon Demo - Customization

- "customize alphabet"
- "customize additional words"
- "customize words to replace"
- ~/.talon
- .talon files
- Python

## Talon Demo - App Specific

- "help app"
- "help app vs code"?

# AXKit

- [AXKit](#) (free)
- Uses macOS accessibility API
- "talon copy menu select"
  - generate Talon script to run menu item under cursor
    - manually add to your talon files
- "document open"
- "document open in VS Code"

# Numen

- [Numen](#) (free)
- Voice control
- Linux only
- [demo](#)
- `$NUMEN_MODEL`



# Part 2

# AI and Voice

# AI & Voice

- Large Language Models with voice
- Voice generation/synthesis
- Voice cloning
- Future possibilities



# Part 2

## Part 2

- AI and Voice
- Coding with Voice

# AI & Voice

# AI & Voice

- Large Language Models with voice
- Voice generation/synthesis
- Voice cloning



# AI & Voice

- Natural Language Processing (NLP)
  - Machine Learning (ML)
- Natural Language Understanding (NLU)
  - Meaning, Context, Intent, Sentiment
  - Semantic Analysis:
    - Named Entity Recognition (NER)
    - Part-of-Speech (POS) tagging
    - Dependency Parsing
    - Text Classification
  - Ontology
    - Relationships between words and phrases

# AI Assistants

- Apple Siri
- Amazon Alexa, Echo, Dot
- Google Assistant
- ChatGPT voice mode
- Microsoft Cortana?
- Samsung Bixby?

# Voice Cloning

- Apple Personal Voice
- Voice-overs
- Movie dubbing
- Virtual assistants

# Apple Personal Voice

- Creates a synthesized version of your voice
  - "Create a voice that sounds like you"
- Uses ML
  - Records 150 phrases
  - Processes overnight
- Live Speech
  - System Settings > Accessibility > Speech > Live Speech
- Augmented speech apps

# Voice Generation

- Multi-modal generative AI
  - LLMs

# Coding with Voice

# Coding with Voice

- VS Code
- GitHub Copilot
- Cursor
- Cursorless (demo)

## Demo: VS Code + Talon

- "bar extensions", "bar search", "bar explore"
- "focus editor", "panel terminal", "panel problems"
- "show settings json", "show settings folder json"
- "wrap switch"
- "file hunt <file name>", "save ugly"
- "file copy path", "file rename", "file move", "file clone"
- "suggest show", "hint show"
- "definition show", "definition peek", "references find"
- "format selection"
- "refactor rename"



## Demo: VS Code + Talon

- "fold that", "unfold that", "fold those", "fold two"
- "select less", "select more/this"
- "join lines"
- "curse undo"
- "select word", "skip word"
- "select line", "skip line"
- "preview markdown"
- "jump line 12"?
- "command palette"?
- "multi cursor add below"?, "multi cursor disable"?
- "camel that", "snake that", "kebab that", "dub string", "dunder", "hammer"