The Vocal Joystick:
Voice-based Continuous Control of Electro-mechanical Devices

Jeff Bilmes

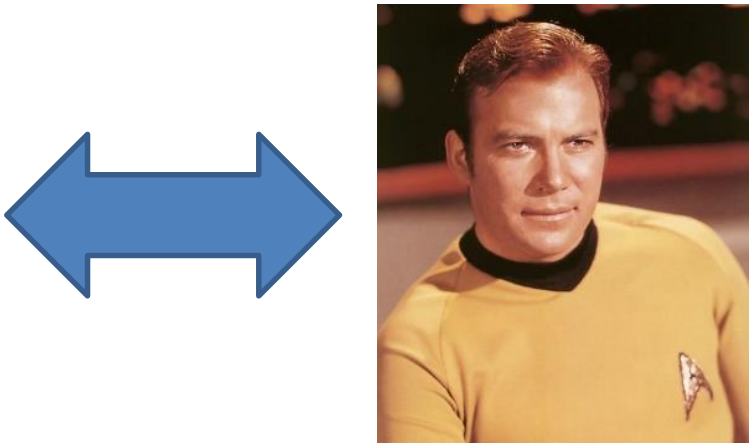
http://melodi.ee.washington.edu/~bilmes

University of Washington, Seattle

Department of Electrical Engineering

A Speech Mouse

- Can you use speech to do what a mouse does?
- Can you use speech to control what a joystick can control?



??

The Vocal Joystick

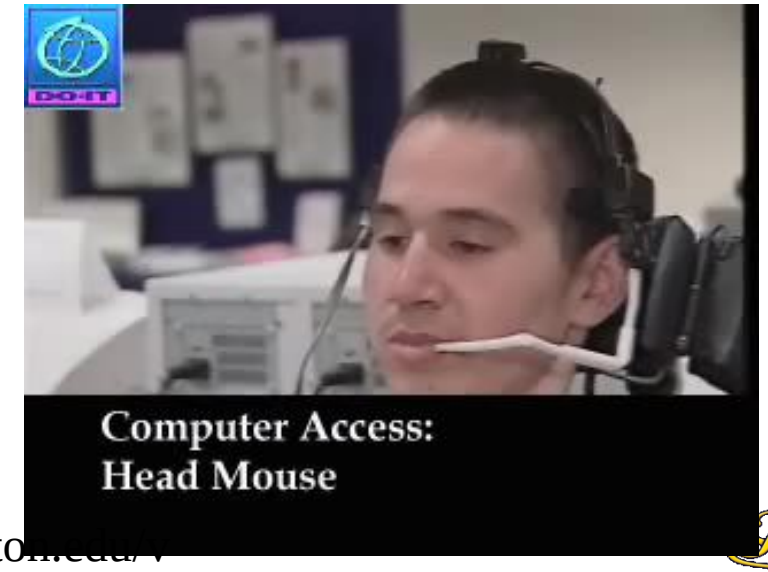
- The Vocal Joystick: Use the voice to produce real-time continuous control signals to control standard computing devices and robotic arms.
- The analogy of a joystick:
  - small number of discrete commands (button presses) for simple tasks, modality switches, etc.
  - multiple simultaneous continuous degrees of freedom to be controlled by continuous aspects of your voice (e.g., pitch, amplitude, vowel-quality, vibrato)

Motivation

- Significant population of individuals with poor (or no) motor abilities, but have good use of their voice.
  - Motor impairments since the time of birth
  - Accidents (car/bicycle accidents, sports injuries)
  - Veterans & war injuries
- Many devices exist for their use (sip-and-puff switches (similar to Morse code), head-tracking mice, eye-tracking mice, etc.)

eye-tracking mouse.

head-mouse video



Issues with existing technology

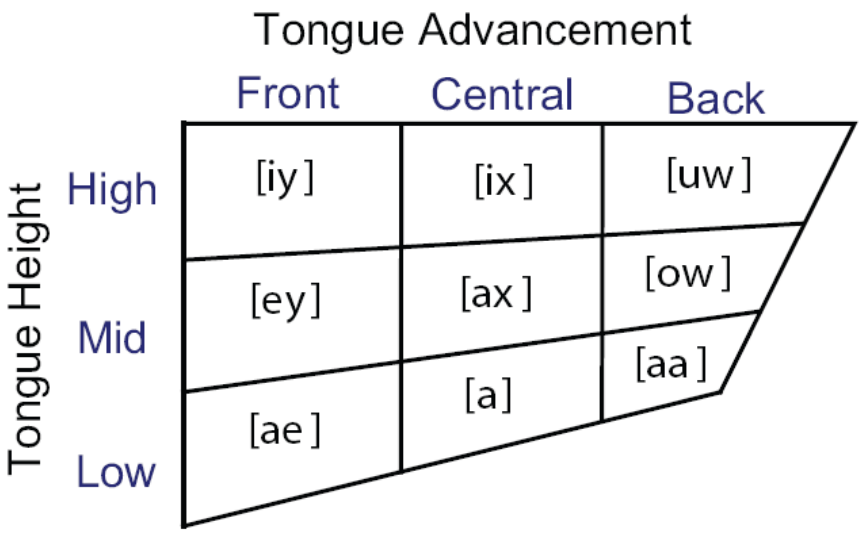
- Expensive, requiring special purpose hardware
- Not be most efficient (leading to user frustration)
- Invasive (BCI neural sensors) or noisy (BCI skull sensors)
- Standard speech-recognition non-ideal for continuous control (e.g., mouse-movement, robotic limb control). Imagine: “move-left”, “move-up”, etc.
- When voice-based, it might not use the full capabilities of the human voice
  - reduced communication bandwidth
  - users with (even not quite) full voice control can do more

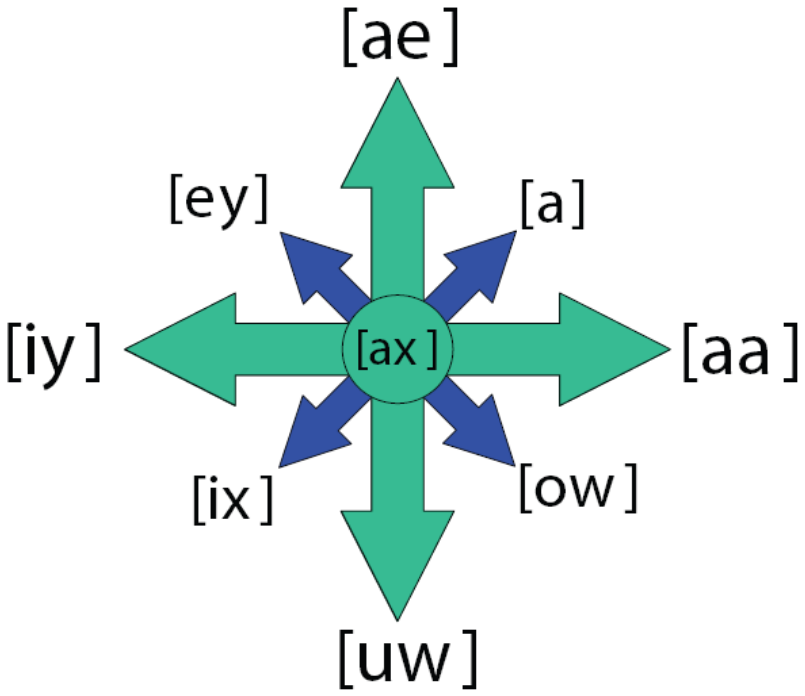
Vocal Joystick Design Goals

- easy to learn and remember (by the user)
  - keep cognitive load at a minimum
- easy to speak (reduce vocal strain)
- easy to recognize (as noise-robust and non-confusable as possible)
- exploitive: use full capabilities of human vocal apparatus
- universal (attempt to use vocal characteristics that minimize the chance that regional languages/dialects preclude its use)
- complementary: can be used jointly with existing speech-recognition
- computationally cheap: leave enough computational headroom for other important applications to run.
- Infrastructure: standard hardware, microphone + computer
- Infrastructure: like a library, easy to incorporate into applications.
- "Individualizable": can be individually configurable

Vocal Joystick Mouse: Mapping

- Standard mice map physical space to physical space.
- Here, we must map vocal tract articulatory change to physical space

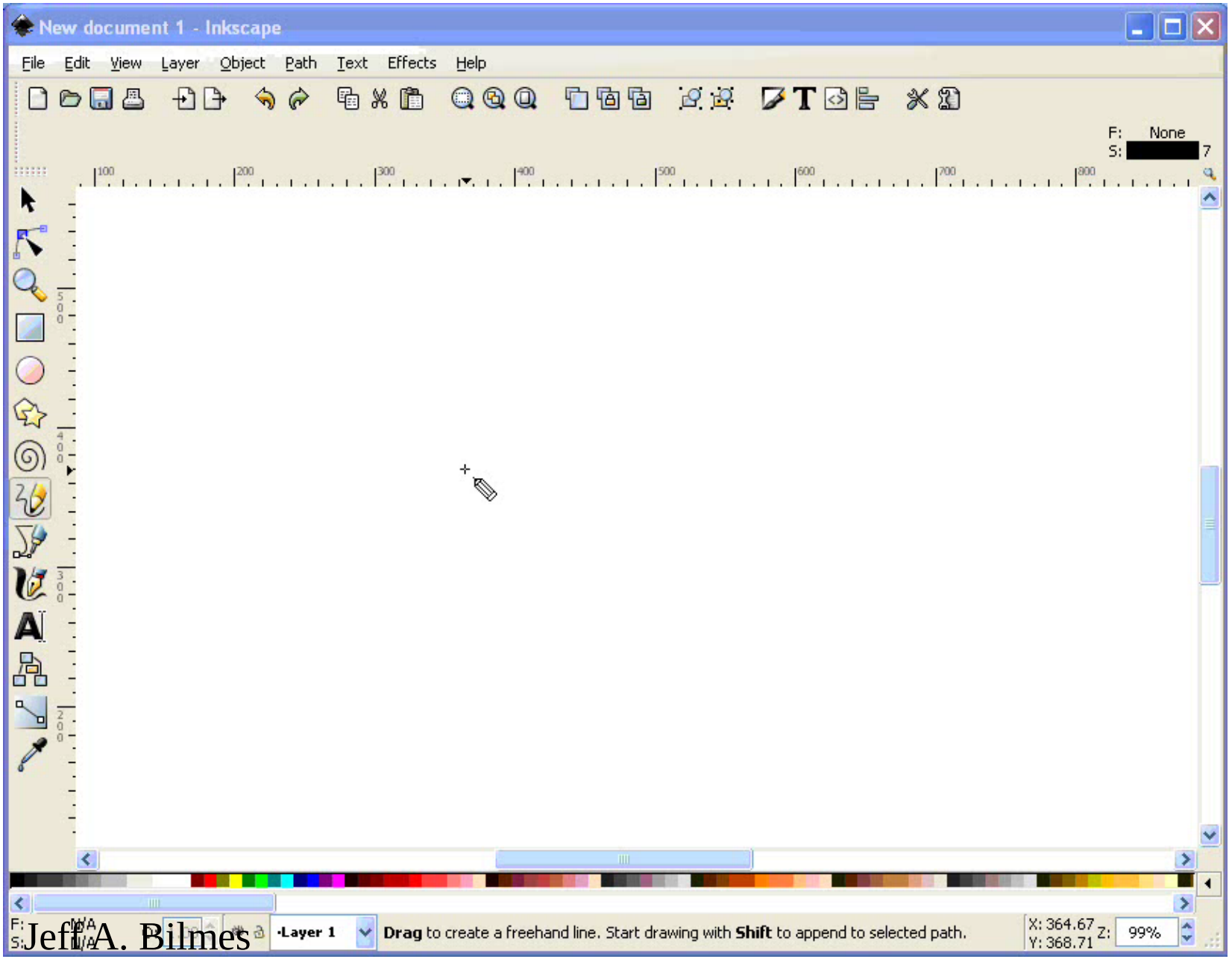




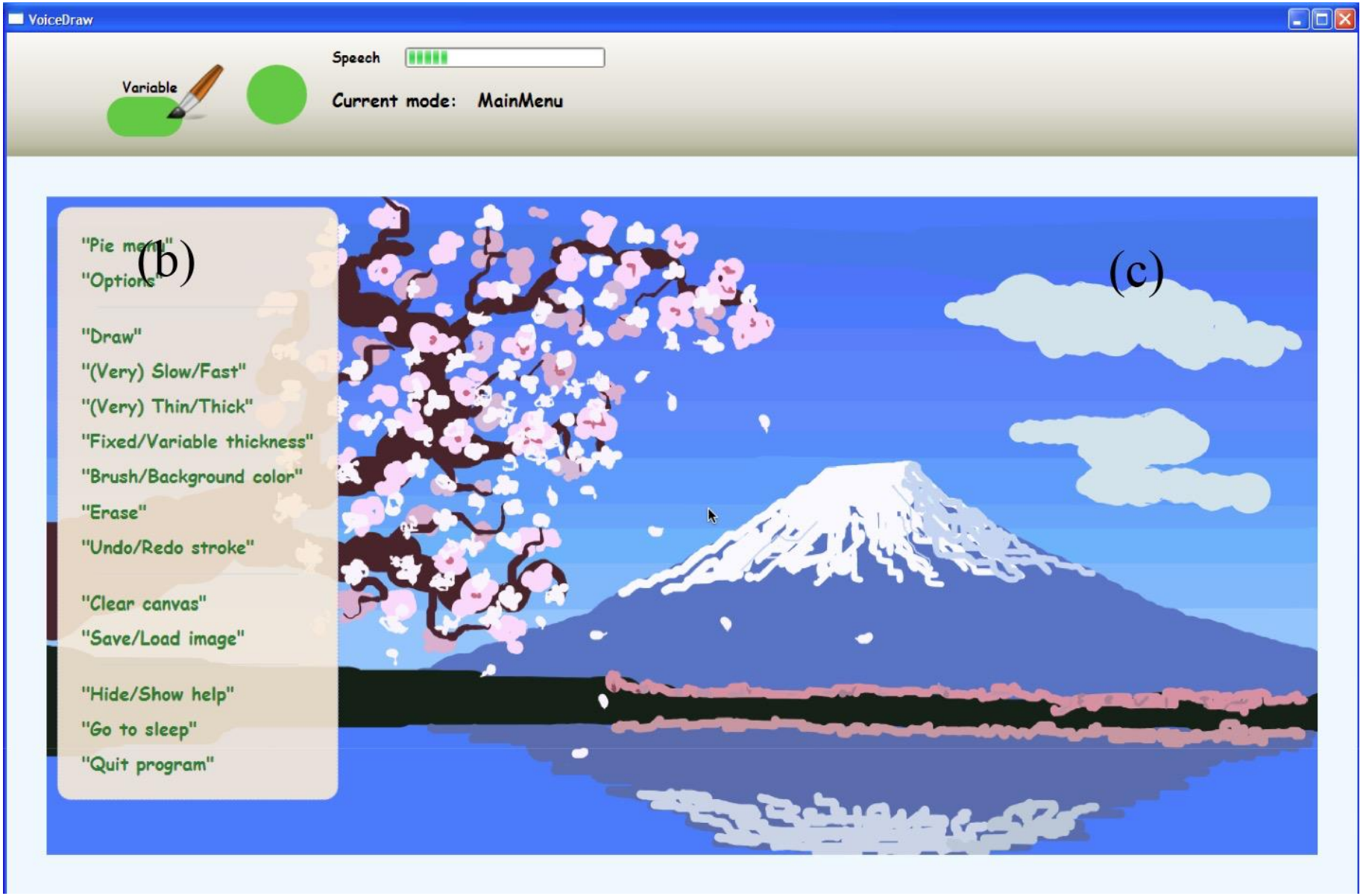
The VJ-Mouse and VoiceBot

- The VJ-mouse and VJ-VoiceBot
  - Research mostly concentrated on a VJ-controlled mouse (which is still quite general).
  - Allows us to perform a variety of tasks on a standard WIMP desktop (mouse movement and mouse clicks, and thus web browsing, slider control, some video games, Dasher typing, etc.)
  - VoiceBot: shows a simple voice-controlled robotic arm.

Vocal Joystick Drawing



VoiceDraw


VoiceDraw
Variable
Speech
Current mode: MainMenu
(b)
"Pie menu"
"Options"
"Draw"
"(Very) Slow/Fast"
"(Very) Thin/Thick"
"Fixed/Variable thickness"
"Brush/Background color"
"Erase"
"Undo/Redo stroke"
"Clear canvas"
"Save/Load image"
"Hide/Show help"
"Go to sleep"
"Quit program"
(c)

Vocal Joystick: Toy 3D Robotic Arm

Summary and the Future

1. Voice-based human-computer interface for individuals with motor impairments.
2. Continuous aspects of the human voice to affect continuous movement in on-screen devices and simple robots
3. Long-term goal: voice-control complex robotic systems, use full vocal capabilities

1. Long-term goal: voice-control complex robotic systems, use full vocal capabilities, real-time high-dimensional continuous outputs, hyper-smart assisted control.