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?

Jeff A. Bilmes
http://melodi.ee.washington.edu/v
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)

Jeff A. Bilmes    http://melodi.ee.washington.edu/v
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.)
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

Jeff A. Bilmes
http://melodi.ee.washington.edu/v
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-confusible 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.

![Diagram of tongue height and advancement with vowel sounds and mapping arrows]

Jeff A. Bilmes  http://melodi.ee.washington.edu/v
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.

Jeff A. Bilmes  http://melodi.ee.washington.edu/v
Vocal Joystick Drawing

http://melodi.ee.washington.edu/vj

Jeff A. Bilmes
VoiceDraw

Jeff A. Bilmes  http://melodi.ee.washington.edu/v
Vocal Joystick: Toy 3D Robotic Arm

Control of a Robotic Arm with the Vocal Joystick:

Introducing the VoiceBot

http://ssli.ee.washington.edu/vj/
Summary and the Future

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.

Jeff A. Bilmes  
http://melodi.ee.washington.edu/v