The VoiceBot: A Voice Controlled Robot Arm

Brandi House, Jonathan Malkin, Jeff Bilmes
Department of Electrical Engineering, University of Washington
{bhouse,jsm,bilmes}@ee.washington.edu

ABSTRACT
We present a system whereby the human voice may specify continuous control signals to manipulate a simulated 2D robotic arm and a real 3D robotic arm. Our goal is to move towards making accessible the manipulation of everyday objects to individuals with motor impairments. Using our system, we performed several studies using control style variants for both the 2D and 3D arms. Results show that it is indeed possible for a user to learn to effectively manipulate real-world objects with a robotic arm using only non-verbal voice as a control mechanism. Our results provide strong evidence that the further development of non-verbal voice-controlled robotics and prosthetic limbs will be successful.

Author Keywords
Voice-based interface, speech recognition, motor impairment, robotics

ACM Classification Keywords
H.5.2 Information Interfaces and Presentation: User interfaces —*Voice I/O*; K.4.2 Computer and Society: Social Issues—*Assistive technologies for persons with disabilities*

INTRODUCTION
Individuals with motor impairments such as those with paraplegia, spinal cord injuries, war-time injuries, or amputations rely on others to assist them their in daily activities. Advances in assistive technologies have begun to provide an increase in independence for these individuals, but there is great potential for further technological developments to significantly improve their abilities, independence, and overall quality of life. One of the greatest challenges faced in assistive technology, however, is that control options are extremely limited when the target users have little or no use of their limbs. For example, a mouse is useless to someone without arms. Spoken language and automatic speech recognition (ASR) systems are often considered the natural solution to such problems. Unfortunately, natural speech is limited to discrete commands, falling short especially on steering tasks which require smooth continuous control.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.
CHI 2009, April 4-9, 2009, Boston, Massachusetts, USA.
Copyright 2009 ACM 978-1-60558-246-7/09/04...$5.00.

The Vocal Joystick (VJ) [1] seeks to use the human voice without the constraints of a natural language. Instead, the user relies on continuous sounds that can vary in pitch, vowel quality, or amplitude to provide control of computer applications and ultimately electro-mechanical devices. The VJ system has been designed as reusable infrastructure available to any application. Existing work on a VJ-driven mouse has demonstrated the suitability of the system to 2D control [3]. In particular, the VJ mouse application enables the user to navigate a windows/icons/mouse/pointer (WIMP) interface using only the voice. Beyond mouse control, VoiceDraw [4] is a drawing program which mixes speech recognition and Vocal Joystick control in a voice-controlled drawing environment. The success of that work, especially the case study with a motor impaired individual, shows the flexibility of the Vocal Joystick engine and its promise for moving beyond simple WIMP-based interaction. The use of non-verbal vocalization for interaction is a growing field, as evidenced by the recent publication of a humming or whistling interface to control a robotic car [15] or for hand-held devices [19].

In this work, we take a step towards vocal controlled robotics by considering the potential of a VJ system to control a computer simulated two-dimensional robotic arm, and then also a real 5 degrees-of-freedom (DOF) three-dimensional hobbyist robotic arm. Our goal is to leverage the human vocal tract's multiple degrees of freedom to provide an inexpensive control method for a robotic arm. As our later results show, users can indeed complete simple tasks with our simulated robotic arm and the real robotic arm using continuous vocal control. These results are particularly promising given the underlying low cost of a VJ-based system.

BACKGROUND
There exist several assistive technologies using novel control methods for manipulating robotic arms. These include physical joysticks, speech-based robotic controllers, and brain-computer interfaces (BCI). Each seeks to increase independence of individuals with disabilities, and each involves a trade-off between expense, invasiveness, and accuracy.

The most widely accepted method for control of robotic limbs is currently a physical joystick. Commercially available robotic arms often include a joystick feature that allows the user to position the gripper or tool at the end of the arm[1], called the

[1]An example is the Lynxmotion Lynx 6 arm at `http://www.lynxmotion.com/Category.aspx?CategoryID=25`

end effector. These controls tend to be inexpensive, non-invasive, and fairly accurate. For individuals with motor impairments, however, these options are inaccessible. An alternative approach combines speech recognition with semi-autonomous robotics. For example, a system called FRIEND uses a robotic arm attached to an electric wheelchair [11]. The entire system is controlled by a speech interface using simple commands. Sensors and control logic incorporated into the arm allow some complex movements with little user input. Unfortunately, control systems of this type often require a structured physical environment or a large number of voice commands to achieve high accuracy as non-autonomous commands seem to control only 1 dimension at a time. Moreover, such a system is not able to specify continuous control. Recent work by Choi [2], however, uses either a head-mounted laser or a touch screen to help give high-level commands to an autonomous robot.

Looking at non-verbal vocal control methods, such as those of Igarashi [6] or Mihara [12], one finds approaches designed primarily for 1-D control, perhaps including some sort of intensity control; the latter relies on images displayed on a screen for positioning, making it unsuitable for real-world tasks such as robotic arm control. Similarly, the humming or whistling approaches mentioned earlier ([19, 15]) do not currently allow simultaneous control of multiple independent degrees of freedom.

The BCI approach is extremely popular in the mainstream media and is widely accepted to be the great promise of hands-free robotic or prosthetic control. Indeed, BCI has shown considerable progress in the past decade. Non-invasive methods use signals from electrodes on the scalp surface [14] and have recently shown improved 2D control [9]. Invasive BCI devices produce the highest quality signals from measurements of motor neurons, but also require placing electrodes inside the scalp or at nerve endings. The BrainGate implant for humans can deliver 2D cursor control of both direction and speed, although robustness can still be a challenge [7]. The most pressing concern for invasive BCI technology is the need for brain surgery, as well as its ability to function as a long-term solution due to scar tissue formation, and of course the high cost.

The Vocal Joystick, by contrast, relies on the enormous flexibility of the human vocal tract to produce a range of discrete and continuous sounds, providing a unique voice-controlled interface. No specialized or invasive hardware is required, only a standard microphone, sound card, and computer are used. The core engine is a portable library that can be used by any application to provide control parameters. Therefore, possible applications for the VJ are virtually unlimited. The VJ system extracts several vocalic parameters (currently loudness, pitch and vowel quality) which are then mapped into control signals and ultimately motion. In previous work, it has been demonstrated that novices using the VJ for mouse control can provide results at least as good, measured by time to acquire a target, as existing word-based cursor control models [3]. In addition, it has been shown that the Vocal Joystick is also suitable for steering tasks, situations where

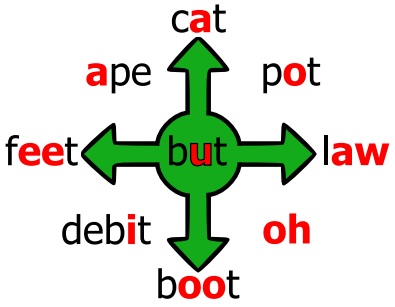


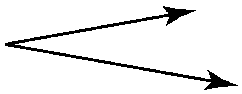
Figure 1. Vowel mapping to cursor movement for mouse control. Words include the nearest vowel in American English.

word-based control models are sorely lacking [8]. While previous work exists for *speech* controlled robotic arms (e.g., in the medical and surgical robotics communities [13]), to our knowledge, our paper demonstrates the first instance of a robot arm being controlled by non-verbal vocalizations.

For mouse control tasks, vowels are mapped into a two-dimensional space, shown in Figure 1, where each vowel determines movement in one direction. Loudness controls cursor velocity with louder sounds corresponding to faster movement. This two-dimensional mapping has been successfully used in drawing applications [4]. Pitch is currently unused in VJ mouse control but is used in the robotic system below. VJ mouse control also uses discrete [ck] and [ch] sounds for buttons — these discrete commands were also used in our robot arm studies, although the underlying VJ software can employ any discrete sounds for this purpose.

INITIAL STUDIES: SIMULATED 2D ARM CONTROL

There are a number of possible control models for a VJ-driven robotic arm. For this first study, we developed a VJ system to simultaneously control three joint angles in two dimensions. We tested three ways of determining joint angles, known as *kinematic models*. In this section, we present the control models, leaving details of the simulated environment for the experimental design section. Note that this section deals with paradigms for control of the arm. A complete exploration of arm control models would also need to examine issues such as what we call "intentional loudness" [10] — estimates of a speaker's intent in an utterance, allowing for discrepancies between intent and the sounds actually produced due to, for instance, physical properties of the vocal tract. Such considerations are beyond the scope of this work, although they may strongly influence results.

Forward Kinematic (FK) Model

Forward kinematics requires the user to control each joint angle explicitly. Such a model is computationally quite simple since the system directly applies the user-supplied input. At the same time, this model may require more cognitive user effort since the user's attention is split between accomplishing the task and the mechanics of realizing that goal. We suspect that an expert user might find this approach appealing while a novice would find it more difficult — lacking true "expert" users (i.e., users who are fluent due to relying on the system as part of their daily routines) means that an answer to this question awaits future research.