Assistive Robot Control Using the Vocal Joystick

Mike Chung, Eric Rombokas, Yoky Matsuoka, Maya Cakmak
Computer Science & Engineering
University of Washington
Seattle, WA, USA
{mjyc,eric,yoky,mcakmak}@cs.washington.edu

Jeff Bilmes
Electrical Engineering
University of Washington
Seattle, WA, USA
bilmes@ee.washington.edu

1. INTRODUCTION

In the United States, there are over a quarter million individuals with spinal cord injuries, 47% of which are quadriplegic[1] (i.e., have restricted use of both their lower and upper limbs). To help these individuals with severe motor impairments achieve greater independence, various alternative interfaces have been developed to increase the accessibility of computers and robots; including verbal commands [5], eye-tracking [6] and tongue [4]. Our work focuses on an alternative interface called the *vocal joystick*, which uses varying properties of continuous non-verbal vocalizations as a control input to a robotic system.

The vocal joystick is particularly interesting as it offers continuous, fluid control without the need of expensive equipment. Our previous work demonstrated that the vocal joystick can be successfully used for controlling a cursor as well as a toy robotic arm [2]. This motivates the use of the vocal joystick as a control interface for mobile manipulators that are physically capable of performing tasks to assist people. In this paper we summarize our work, detailed in [1], that explores this possibility. We propose three different control methods that address the challenge of having fewer control dimensions than are required to control the robot's degrees-of-freedom (DoF). We present results from a user study (N=9) that involves users performing two realistic everyday tasks through these interfaces.

2. APPROACH

The signal obtained from human vocalizations[2] involves continuous values for *vowel quality velocities* (2D), *pitch* (1D) and *loudness* (1D). In addition, we obtain *discrete identifiers* from the vocalizations (e.g. discrete phonemes like [k] and [tS]) which are used for discrete commands (e.g. mode switching or hand actions). The control methods proposed in this paper differ in the ways they convert the continuous input signals to control signals for moving a 7-DoF robotic arm (Fig. 1)[3].

(a) (b)

Figure 1: Experimental setup for the two real-world tasks: (a) users move water bottles into a recycling bin (b) users pick up a grocery bag by the handle and move it to a specified region of another table.

Method 1: Direct control. This method involves direct control of a subset of joint angles while being able to select which subset is currently active. At any given time the user controls two co-located joints on different axes, through the two-dimensional continuous values from *vowel quality velocities*. The speed of the joint motions are controlled by the loudness of the user's voice. A monitor next to the robot (Fig. 1) displays a unique color indicator of the currently active joints: blue for shoulder, green for elbow, and orange for wrist. The discrete sound [tS] is used to switch between joints and the discrete sound [k] is used for opening/closing the robot hand.

Method 2: Cartesian control (IK). In this method the user controls the cartesian position of the end effector, while the appropriate joint angles are computed through inverse kinematics (IK). The *vowel quality velocities* and *pitch velocity* are used for controlling the 3D position of the end effector. Users can also switch (via the discrete sound [tS]) to a fine manipulation mode where they get direct control of the wrist joint. The monitor displays a color indicator of the current mode. As before, loudness controls the speed of motion and the discrete sound [k] controls the robot hand.

Method 3: Synergy. The last control method is an approach inspired by the *synergy hypothesis* [3] in the study of neural movement control. It involves computing a control subspace from one demonstration of the task. Specifically, we map the *vowel quality velocities* and the *pitch velocity* to

[1] `http://www.sci-info-pages.com/facts.html`

[2] We use a headset microphone to acquire input signals

[3] A video illustrating the different control methods can be viewed at `http://youtu.be/1d1MBc6UAEk`

the first three dimensions computed from running Principle Component Analysis (PCA) on a single example execution of the task recorded in 7-DoF. The loudness and the discrete sound [k] function the same way as the other methods.

3. EVALUATION

We conducted a within-groups, two-session user study to compare the three control methods developed for the vocal joystick and investigate their learnability. The two sessions were conducted in two consecutive days and participants did two trials in each session (i.e. a total of 4 trials). Each trial involved completing two different tasks (see Fig. 1) with each of the three different control methods. The two tasks were (a) placing as many bottles as possible from the table to the recycling bin in 5 minutes and (b) moving a 3kg paper grocery bag from one table to another table as fast as possible. The order of tasks and control methods were counterbalanced.

Our study was completed by nine naive healthy volunteers (eight male and one female) with ages varying between 20-35. The participants had varying occupations (students, department staff, and manager) and native languages (Chinese, English, Korean, Mongolian). They did not have any experience using the vocal joystick or similar systems.

We evaluated the performance of the different control methods through task efficiency. For the bottle task, we looked at the number of bottles dropped in the recycle bin in 5 minutes and the time taken per bottle. For the bag task, we looked at the time taken to successfully transport the bag to its target location. The metrics were measured through video recordings of each experimental run for each user. In addition we had a questionnaire asking participants to compare the different control methods at the end of each session.

4. FINDINGS

Comparison of control methods. We did not observe a sustained significant difference between the control methods in terms of the task efficiency. However, by the fourth trial, the *synergy* method was the most efficient method in both tasks. This difference was particularly notable in the bag moving task where the completion using synergy method took 39% *less time* than with the (next best) *IK* method with statistical significance ($p<0.05$). The main problem in the *IK* method was that participants got stuck in configurations where the IK solver had difficulties finding solutions to get out of the configuration. This problem came up more frequently in the bag transfer task, causing the *IK* method to be the least efficient method for this task.

Despite its efficiency, the *synergy* method was subjectively considered as the most difficult by the participants (Fig. 2(d)). The major complaints were related to the difficulty of understanding the mapping between vocal joystick control signals and the control dimension of the robotic arm.

Learnability. The performance measures of both tasks (Fig. 2(a)-(c)) indicate that all three control methods are learnable, with statistically significant performance improvement from trial one to trial four ($p<.05$). For the three different methods (*direct*, *IK*, and *synergy* respectively) the completion time decreased by 45%, 50%, and 52% for the bottle task and by 67%, 63%, and 78% for the bag task. Also, the participants' perception of difficulty decreased from the end of trial 2 to the end of trial 4 for all methods (Fig. 2(d)).

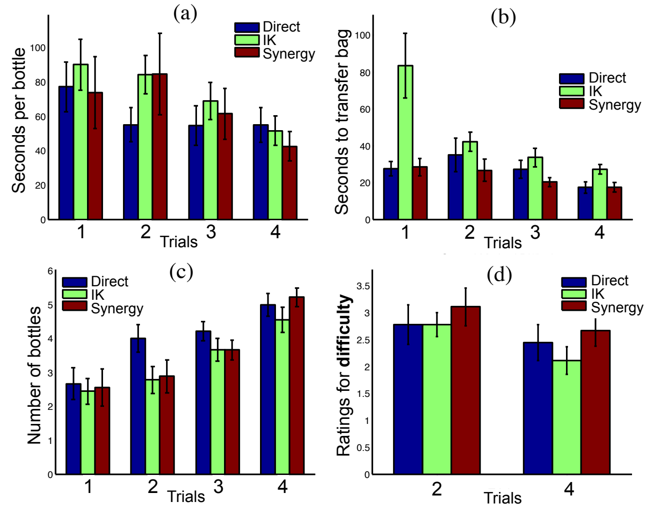

Figure 2: Results of the user study: (a) time taken per bottle, (b) time taken to complete the bag task, (c) total number of bottles recycled in 5 minutes, and (d) Likert scale ratings for the *difficulty* of the control method (1:*very easy*, 5:*very difficult*).

5. CONCLUSION

We present methods for mapping continuous vocalization signals to robot manipulator control signals. We demonstrated the feasibility of the methods through a user study in which participants with no prior exposure to the system were able to accomplish realistic tasks, even in their first interaction. Moreover, we observed a clear learning effect over multiple trials, suggesting that further improvement may occur with more practice. Comparing alternative approaches, we found that the *synergy* method, which moves the arm on a task-specific manifold, was competitive or more efficient than simpler methods. As next steps, we plan to improve the synergy method with visualizations that allow the user to predict the effects of their input and we plan to evaluate the final system with our target population (i.e., persons with mobility impairments).

6. REFERENCES

[1] M. Chung, E. Rombokas, Q. An, Y. Matsuoka, and J. Bilmes. Continuous vocalization control of a full-scale assistive robot. In *Intl. Conf. on Biomedical Robotics and Biomechatronics*, 2012.

[2] B. House, J. Malkin, and J. Bilmes. The VoiceBot: a voice controlled robot arm. *CHI*, 2009.

[3] M. L. Latash. *Neurophysiological basis of movement.* Human Kinetics Urbana Champaign, IL, 1998.

[4] M. Mace, K. Abdullah-Al-Mamun, R. Vaidyanathan, S. Wang, and L. Gupta. Real-time implementation of a non-invasive tongue-based human-robot interface. In *IROS*, 2010.

[5] C. Martens, O. Prenzel, and A. Gräser. The rehabilitation robots friend-I & II: daily life independency through semi-autonomous task-execution. *Rehabilitation Robotics*, 2007.

[6] H. A. Yanco. Wheelesley: A robotic wheelchair system: Indoor navigation and user interface. In *Assistive Technology and Artificial Intelligence*. Springer, 1998.