



A 3D head pointer: a manipulation method that enables the spatial position and posture of supernumerary robotic limbs

Joi Oh¹, Fumihiko Kato², Yukiko Iwasaki¹, Hiroyasu Iwata³

¹ Waseda University, Graduate School of Creative Science and Engineering, Tokyo, Japan

² Waseda University, Global Robot Academic Institute, Tokyo, Japan

³ Waseda University, Faculty of Science and Engineering, Tokyo, Japan

ABSTRACT

This paper introduces a novel interface '3D head pointer' for the operation of a wearable robotic arm in 3D space. The developed system is intended to assist its user in the execution of routine tasks while operating a robotic arm. Previous studies have demonstrated the difficulty a user faces in simultaneously controlling a robotic arm and their own hands. The proposed method combines a head-based pointing device and voice recognition to manipulate the position and orientation as well as to switch between these two modes. In a virtual reality environment, the position instructions of the proposed system and its usefulness were evaluated by measuring the accuracy of the instructions and the time required using a fully immersive head-mounted display (HMD). In addition, the entire system, including posture instructions with two switching methods (voice recognition and head gestures), was evaluated using an optical transparent HMD. The obtained results displayed an accuracy of 1.25 cm and 3.56° with the 20-s time span necessary for communicating an instruction. These results demonstrate that voice recognition is a more effective switching method than head gestures.

Section: RESEARCH PAPER

Keywords: VR/AR; hands-free interface; polar coordinate system; teleoperation; SRL

Citation: Joi Oh, Fumihiko Kato, Iwasaki Yukiko, Hiroyasu Iwata, A 3D head pointer: a manipulation method that enables the spatial position and posture for supernumerary robotic limbs, Acta IMEKO, vol. 10, no. 3, article 13, September 2021, identifier: IMEKO-ACTA-10 (2021)-03-13

Editor: Bálint Kiss, Budapest University of Technology and Economics, Hungary

Received March 31, 2021; **In final form** September 6, 2021; **Published** September 2021

Copyright: This is an open-access article distributed under the terms of the Creative Commons Attribution 3.0 License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Corresponding author: Joi Oh, e-mail: joi-oh0924@akane.waseda.jp

1. INTRODUCTION

In recent years, there has been a considerable amount of research and development on the use of supernumerary robotic limbs (SRLs) for 'body augmentation'. In previous studies, robotic technology, especially wearable robots, has been developed for use as prostheses for rehabilitation purposes. An SRL aims to provide its users with additional capabilities, enabling them to accomplish tasks that they would otherwise be incapable of performing. In this respect, an SRL is different from other types of existing wearable robots; a lightweight, sufficient torque and a highly manoeuvrable SRL developed by Vernonia et al. [1] is a classic example. These robots can be used in any context, from helping individuals to perform household chores to improving industrial productivity.

To effectively assist in routine tasks (e.g., opening an umbrella or stirring a pot), users require an interface that indicates the target point location to the end effector of the SRL without requiring them to interrupt their actions. However, such a

method has not yet been established. Parietti et al. [2],[3] developed a manipulation technique in which the operator's movements were monitored by a robot, following which the robotic arm performed the corresponding movements. Iwasaki et al. [4] proposed an interface that allowed the operator to actively control the SRL by using the orientation of the face, while Sasaki et al. [5] developed a manipulation method that enabled more complicated operations of the robotic arm with the user's feet as the controllers. Previous studies have overlooked the balance between ensuring the operator's limbs move freely and providing detailed instructions to the SRL, and there are further challenges with respect to multitasking in the context of daily life. Therefore, in this study, a method for manipulating SRLs so that two parallel tasks do not interfere with each other is proposed and then evaluated for its usefulness.

In the present study, a two-stage experiment was conducted. This section describes the hypothesis of the method, and Section 2 presents the method for position instruction along with the experimental results. In Section 3, a manipulation method that includes posture instructions is proposed and the experimental

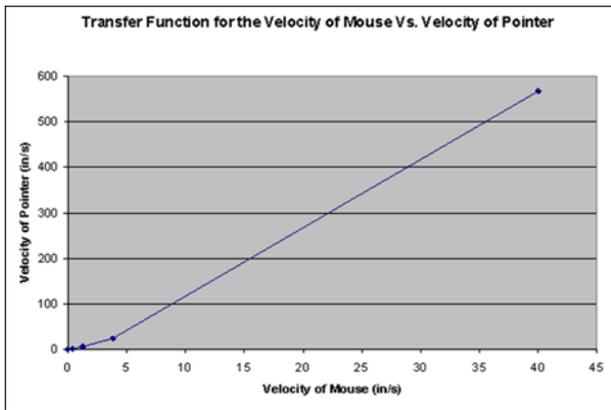


Figure 1. Microsoft's mouse-cursor speed-change settings [15].

results are presented, and the two experiments are then discussed. Section 4 presents comparisons with other similar methods and discusses the limitations, and finally, Section 5 presents the conclusions.

The following two elements are considered essential for achieving daily support for parallel tasks:

- 1) undisturbed movement of the operator's limbs,
- 2) an indication of spatial position and posture.

To date, several hands-free interfaces have been proposed to satisfy requirement 1, with some operated by the tongue [6], eye movement [7] or voice [8] and used for either screen control or robot manipulation (or both). Methods to control robotic limbs with brain waves [9] are also being investigated.

However, this study focuses on requirement 2 and the construction of a more intuitive instructional method. When the operator provides directions related to a 3D-space location, they must accurately indicate the target point. The field of view, within which a person can perceive the shape and position of an object, is as narrow as 15° from the gazing point [10]; hence, to compensate, it is necessary to direct the face and gaze in the instructional space to provide spatial position instructions. The interface proposed in this study takes advantage of this compensatory action and uses it as an instruction method.

Methods for using the head as a joystick have already been proposed. One method involves the manipulation of the head for instruction in a 2D plane, such as on-screen operations [11]. Another method involves switching between the vertical and horizontal planes by nodding towards the plane to be manipulated, supplementing the plane manipulation by the head so that only the head is used to manage the 3D space [12]. However, these methods do not use the compensatory head motion as a manipulation technique.

2. PROPOSAL FOR A POSITIONING METHOD USING HEAD BOBBING

Turning one's head can be used to instruct the radial direction of the target point in polar coordinates. In this section, we propose a pointing interface that combines head bobbing with head orientation in a polar coordinate system. Head bobbing is a small back and forth motion of the head that does not interfere with the operator's movements.

This research was performed using the standard morphology of a Japanese man, as recorded by Kouchi et al. [13]. According to these data, the head-bobbing range was determined as approximately 9.29 cm, which allows the operator to keep the zero-moment point in the torso of the body and operate a



Figure 2. External view of the robotic arm proposed by Nakabayashi et al. [14] and Amano et al. [16].

robotic arm without losing balance. A doughnut-shaped area with an innermost and outermost radius of 30 and 100 cm, respectively, around the operator was defined as an example of an SRL operating range [14]. The head-bobbing depth-change factor was $70 / 9.29 = 7.53$ or more. The range of motion that can be performed using head bobbing is considerably lower than that of arms.

The preliminary experiments demonstrated that at high magnification, the instructional accuracy of head bobbing was lower than that of other comparable methods. Additionally, the required instructions were shown to be longer, and an increase/decrease factor (IDF) that gradually changes the depth of the head-bobbing task based on head velocity was therefore introduced. The IDF allows precise instructions while maintaining a high magnification. In this study, the IDF was constructed using the mouse-cursor change factor shown in Figure 1, set by Microsoft Windows [15].

2.1. Evaluation test with a fully immersive head-mounted display

This section examines the usefulness of the IDF and 3D head pointer as a whole. This study was conducted based on the previously developed robotic arm proposed by Nakabayashi et al. [14] and Amano et al. [16], as shown in Figure 2. The arm has a reach of up to 1 m, and its jamming hand, shown in Figure 3, can be used as an end effector to grasp an object, with an error of up to 3 cm [16]. Therefore, the allowable indication error at the interface in this experiment was set to 3 cm.

In this study, the validation was performed in a virtual reality (VR) environment. The indication of radial direction by head orientation was measured from the front of the head-mounted display (HMD). The depth indicator was implemented by setting up a sphere with the operator at the centre, as shown in Figure



Figure 3. External view of the jamming hand.



Figure 4. 3D image of the head pointer operation.

4, and by changing the radius of the sphere controlled by head bobbing.

An HMD is used in the proposed method (HTC VIVE [17]). The experimental procedure is as follows:

- 1) The participant wears the VIVE headset and grasps a VIVE controller in each hand, holding them up in front of their chest, as shown on the right in Figure 5. This is defined as the 'rest position'. The subject's avatar is displayed in the VR space, as shown on the left in Figure 5.
- 2) The 3D head pointer's control cursor (the red ball in the centre of Figure 6) appears 65 cm in front of the eyes. Simultaneously, the target sphere with a 10-cm diameter (the blue transparent sphere in the upper-right corner of Figure 6) appears at any of the eight locations at a ± 30 -cm height, ± 20 -cm width and ± 20 -cm depth, and it is positioned ± 20 cm from the cursor.
- 3) The participant aligns the cursor with the centre of the target sphere by using the 3D head pointer.
- 4) When the participant perceives that they have reached the centre of the target sphere, they verbalise the completion of the instruction. As shown in Figure 7, the target sphere has a reference frame with its origin at the centre of the sphere. The participant adjusts the position of the cursor accordingly.
- 5) Steps (1)–(4) are performed for all eight target sphere positions.

In the present study, the above-mentioned procedure was performed by two groups of six participants each. The experiments were performed once under different conditions for each group. Table 1 shows the experimental conditions and group distribution. Group 1 was asked to perform the tasks described above, but with a predefined time limit for instruction execution, while group 2 was asked to perform the experiment either with or without an IDF.

Figure 8 shows the relationship between head-bobbing speed and magnification. 'IDF not available' is a condition in which the

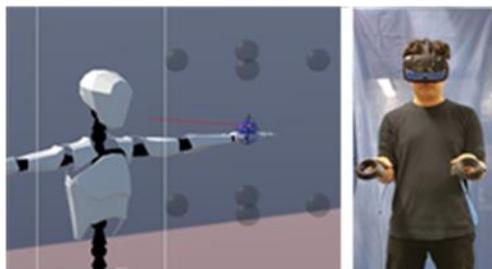


Figure 5. The experimental interface operation. Left: instructional target spheres and participants within the VR; right: participant wearing the HMD and holding the controllers.

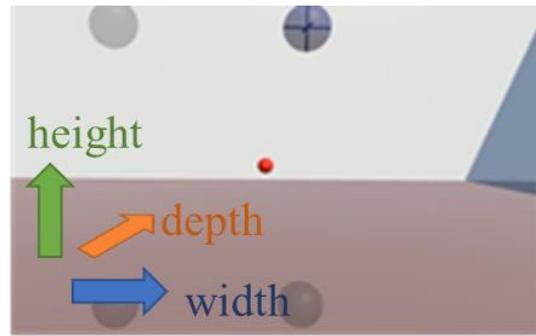


Figure 6. Subjective view of the user's experience.



Figure 7. Target sphere and cursor visibility.

rate of change in depth due to head bobbing is fixed at 10 times without using the IDF.

Based on the aforementioned experiments, the usefulness of the 3D head pointer was evaluated using the average indication error condition (a) shown in Table 1, the relationship between the indication accuracy error and operation time in conditions (a)–(f) and the maximum arm sway of the subject measured by the VIVE controller according to condition (a). At the same time, the usefulness of the IDF was tested by comparing the instructional error between conditions (a) and (g).

2.2. Results and discussion on the fully immersive HMD

In this study, the Wilcoxon signed-rank test was used to verify the significant differences between two conditions. This is a nonparametric test used when the population does not follow a normal distribution. The difference in $Z_i = Y_i - X_i$ between the experimental values of two conditions X_i and Y_i performed on the i -th participant was obtained. Next, Z_i was arranged in order of decreasing absolute value, and rank R_i was assigned to the smaller value. The Wilcoxon signed-rank test quantity of W was then calculated as follows:

$$W = \sum_{i=1}^n \phi_i R_i. \quad (1)$$

However, in this case, ϕ_i was calculated as

Table 1. The experimental conditions and group distribution.

condition	Requirement	Group
(a)	No requirements	1, 2
(b)	2-s time limit for instruction	1
(c)	3-s time limit for instruction	1
(d)	4-s time limit for instruction	1
(e)	6-s time limit for instruction	1
(f)	8-s time limit for instruction	1
(g)	the rate of change in depth due to head bobbing is fixed at 10 times	2

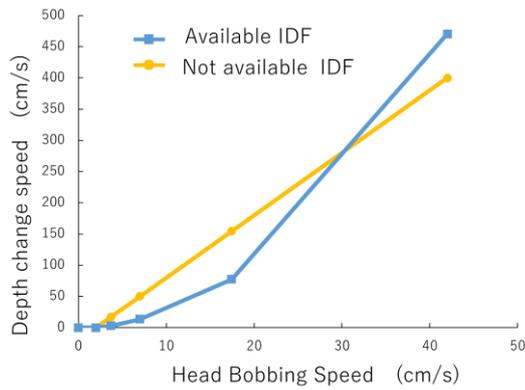


Figure 8. Change in head-bobbing magnification with and without IDF.

$$\emptyset_i = \begin{cases} 1(Z_i > 0) \\ 0(Z_i < 0) \end{cases} \quad (2)$$

Significant differences were calculated by comparing test quantity W to the Wilcoxon signed-rank table [18]. In this experiment, instead of the table, the Excel statistics function (Microsoft Inc.) was used to calculate significant differences.

2.2.1. Indication Error

The instructional error of the distance from the centre of the target sphere to the control cursor was measured upon completion of the instruction.

This was done in VR by using an IDF-based 3D head pointer for 12 people, divided equally into two groups (1 and 2). The results are presented in Table 2.

In this study, a jamming hand [16] capable of grasping an object with an error of up to 3 cm in target point indication was used as a reference-index end effector. The average error of the instructions in this experiment was approximately 1.32 cm, with the highest instructional error of 2.5 cm. These results suggest that the indication error of the 3D head pointer is within the range of absorbable error in the case of grasping and manipulating an object with the specific end effector.

The standard deviation of the indication error was 0.65 cm, and the error varied widely from person to person. This result may be related to the familiarity level of each individual in the use of a VR space. The results were validated by considering VR experience.

2.2.2. Change in indication error at each indication time

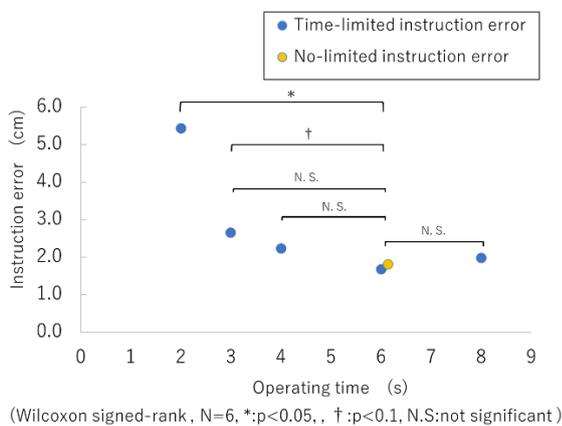


Figure 9. Instruction error per operating time in the evaluation test.

Table 2. Average instruction error.

Subject	Instructional error (cm)
1	1.20
2	2.50
3	1.54
4	2.19
5	2.41
6	1.06
7	1.06
8	0.882
9	0.757
10	0.905
11	0.695
12	0.668
Average	1.32

The experiment was conducted under conditions (a)–(f) for six members of group 1. The relationship between the instruction error and instruction time is shown in Figure 9.

The average operation time under condition (a), with no time limit, was 6.2 s. When the operation time was limited, the indication error decreased rapidly with the increase in time limit from 2 to 3 s. When the time was greater than 4 s, this error remained almost constant regardless of the time taken. This suggests that the operation with the 3D head pointer itself had already been completed by 4 s.

2.2.3. Maximum arm sway

The maximum arm sway of the six participants in group 1 was measured from the movement of the VIVE controller while standing upright and compared to the maximum arm sway when the 3D head pointer was manipulated in condition (a). The results are presented in Figure 10.

The comparison results demonstrated that the maximum arm sway was greater with a 3D head pointer. However, the Wilcoxon signed-rank test did not show any significant difference between the two conditions ($N = 6, p < 0.1$), suggesting that the proposed method allows a user to continue performing regular arm movements while following the instructions. Because the proposed method requires visibility of the target space for performing tasks with SRL, multitasking is sometimes impossible, and interruption of the task being performed by the user is unavoidable. However, if the operator's hand position can be maintained while using the 3D head pointer, the interrupted

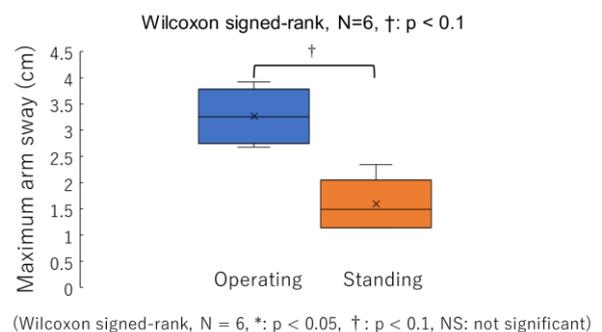


Figure 10. Maximum arm sway when standing upright and operating the 3D head pointer.

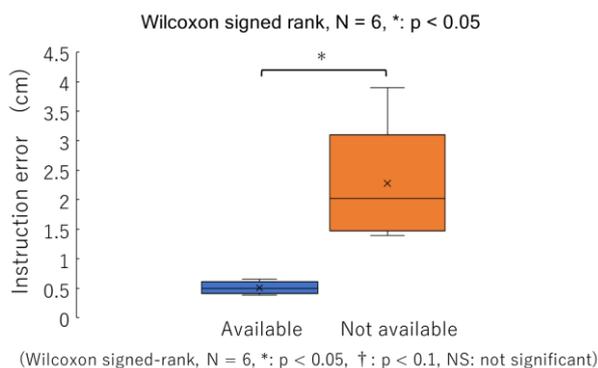


Figure 11. Depth error based on head bobbing with and without IDF.

task can be resumed quickly following instructions to the SRL; this is significantly more efficient than performing the two tasks separately.

2.2.4. Differences in indication error with and without IDF

We conducted the experiment under conditions (a) and (g) for the six members of group 2 and measured the instruction errors of the 3D head pointer and the depth-only instruction errors for head bobbing. The results are shown in Figure 11 and Figure 12, respectively. The use of IDF reduced the average instruction error by approximately 77.6 % for the depth instruction by head bobbing and approximately 67.0 % for total error in the three axes (x, y, z). Additionally, a significant difference was observed between the two conditions with and without IDF in the case of the Wilcoxon signed-rank test ($n = 6, p < 0.05$). It was therefore confirmed that the introduction of the IDF greatly improved the accuracy and demonstrated its usefulness. Nevertheless, it is still necessary to verify whether the accuracy can be further improved with the additional fine-tuning of the parameters related to the magnification change ratio.

3. PROPOSAL FOR COMBINING THE POSITION AND POSTURE INDICATION METHOD

The previous section showed the effectiveness of the position indications for SRL. However, without posture instructions at the interface, the SRL cannot perform complex routine tasks (e.g., holding an umbrella at an angle to strong winds, pouring the contents of a bottle into a cup). Some objects can only be grabbed from certain directions. In this study, a method is proposed that uses the head for SRL to provide posture indications. Because it is difficult to provide stereotactic and

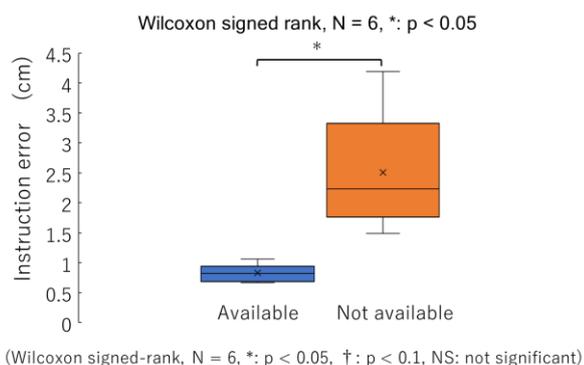


Figure 12. Total error in the three axes due to the 3D head pointer with and without IDF.



Figure 13. The three different rotation axes of the head.

posture instructions simultaneously with the head, a 'switching indication' function was also proposed, which switches between position and posture indications.

3.1. Proposal for a posture-indication method using isometric input

Figure 13 shows that the human head can rotate in three axes using Unity-chan as the model (a humanoid model created by Unity Technologies Japan [19]). The use of head-rotation axes for SRL posture indication (yaw, pitch and roll) facilitates intuitive instructions. However, the head has limited angles of yaw, pitch and roll ranging from -60° to $+60^\circ$, -50° to $+60^\circ$ and -50° to $+50^\circ$, respectively [20]. If the displacement of the head is used as an input device, the SRL cannot be instructed to posture at an angle beyond the limits of the angle of the head. In addition, according to requirement (2) in Section 1, if the head moves more than 15° , the operation target will be out of the operator's effective field of view.

In this study, the three-axis rotation of the head was used as an isometric input-device parameter that determines the rotational velocity of the pointer according to the rotational angle of the head [21]. The maximum input angle of the head was set to 15° , which is the maximum angle limit of the effective field of view. To avoid incorrect input, head rotations of $\leq 3^\circ$ were not detected as inputs. The changes in the rotational velocity were spherically interpolated using trigonometric functions. Figure 14 shows the relationship between the amount of rotation of the head and the rotation speed of the posture indicator. The

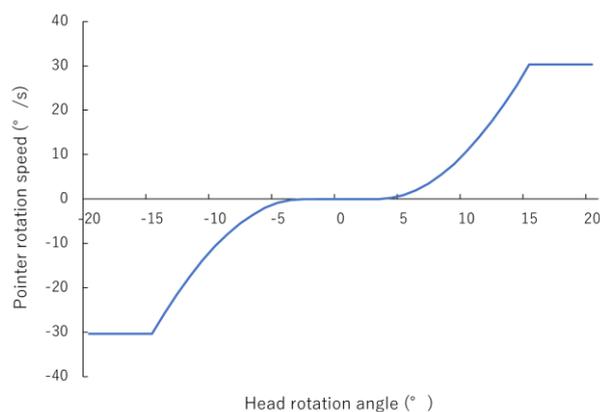


Figure 14. Relationship between head rotation angle and posture rotation speed.

Table 3. Voice command list.

Voice command	Function
Indicate position	Switch from posture indication to position indication
Indicate posture	Switch from position instructions to posture instructions
Finish	Signals that the indication has been completed. (Used for evaluation tests)

reference angle for head rotation is the direction that the user is facing when switching to the posture indication.

3.2. Proposal for a mode-switching method using voice recognition

An increase in the number of body parts used for manipulation is undesirable because it leads to an increase in the body load. The switching method was constructed using the head or voice. In this study, two types of switching instruction methods were proposed and then compared in an evaluation test.

3.2.1. Voice-recognition-based switching indication method

A switching method based on voice recognition is less physically demanding and has less impact on the operator's limbs than physical operations. Table 3 lists the commands used for voice indications.

3.2.2. Head-gesture-based switching indication method

A method for switching between posture and position instructions using head gestures was also proposed. In this method, a 'head tilt' motion was performed to switch from position to posture instructions (top of Figure 15), while the 'head bobbing' motion was performed to switch from posture instruction to position instruction (bottom of Figure 15). Because the user only has to indicate the operation mode required, the head-gesture-based switching method requires little cognitive load, and switching can be done intuitively.

3.3. Evaluation test with optical transparent HMD

This section presents an evaluation of the usefulness of the posture and switching instructions in the 3D head pointer as well



Figure 15. Top: Switch to posture instruction; Bottom: Switch to position instruction.



Figure 16. Pointer cursor corresponding to posture indication.

as an evaluation of the usefulness of the 3D head pointer in real space. To operate the SRL on a real machine, the tip of the SRL and target object must be visible. There are two ways to see the tip of the SRL on a real machine: by using a video transparent HMD or an optical transparent HMD [22]. The video transparent system may not be able to cope when the SRL malfunctions because of the delay in viewing the actual device. In this experiment, the proposed method was constructed using an optical transparent HMD (Hololens2 [23]) to evaluate the usefulness of the entire 3D head pointer. To provide posture instructions, the pointing cursor was changed from a red sphere to a blue-green bipyramid, as shown in Figure 16.

The indication of the radial direction based on head orientation was measured from the front of the HMD. The depth indicator was implemented by changing the radius of the sphere through head bobbing, as described in Section 2.1. The amount of head rotation in the posture indication was determined by measuring the posture of the HMD. Compared to position indication, it is difficult to evaluate the amount of operator input required for posture indication. To visually display the user's head rotation, the user interface (UI) is displayed during posture instruction, as shown in Figure 17. The white point on the UI is aligned with the centre and moves up, down, left and right according to the amount of yaw and pitch fed as the input. The roll-angle input is displayed as a white circle in the UI, and the circle rotates according to the amount of roll input. This UI allows the operator to visually understand how much operator input is. For speech recognition, Microsoft's Mixed Reality Toolkit was used [24].

In this experiment, a pointing task was set up as the target appearing in the air. The experimental procedure is described as follows:

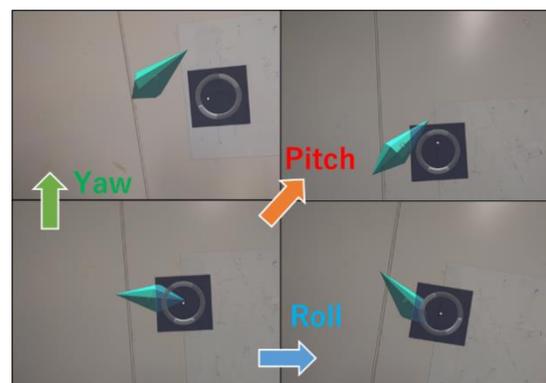


Figure 17. Auxiliary user interface for posture instruction.



Figure 18. Cursor and target in the experiment.

- 1) The subjects stood upright while wearing the HMD and Bluetooth headset in a room with white walls.
- 2) The 3D head pointer cursor (blue–green bipyramid in Figure 18) and the target (purple bipyramid in Figure 18) were displayed in front of the participant. The target appeared at a random position within 15° to the left and right of the subject’s direction of gaze and at a depth of between 30 and 100 cm, as shown in Figure 19. The direction of the target was determined randomly from six possible directions: up, down, left, right, front and back.
- 3) The participant moved the cursor to the same position and posture as the target using a 3D head pointer. When the subject perceived that the operation had been completed, they verbalised ‘instruction complete’ into the Bluetooth headset. Markers were displayed at the centre of the cursor and at the target position and rotation, as shown in Figure 18. These markers were always visible to the participant regardless of the position and posture of the cursor and target, and the operator relied on these markers for position and posture indications.
- 4) Steps 1)–3) were performed 12 times in succession in one experiment. The evaluation experiment was conducted under the following two conditions:
 - A) switching indications by voice recognition,
 - B) switching indications by head gesture.
 A verbal questionnaire was administered after the operation was complete.

The experiment was conducted using a total of six men and six women in their 20s and 30s, with the order of conditions A) and B) randomised. Procedures 1)–4) were performed at least once as a practice run before conducting the experiment, and additional practice was conducted until the subject judged that they were proficient. Based on the above experiments, the

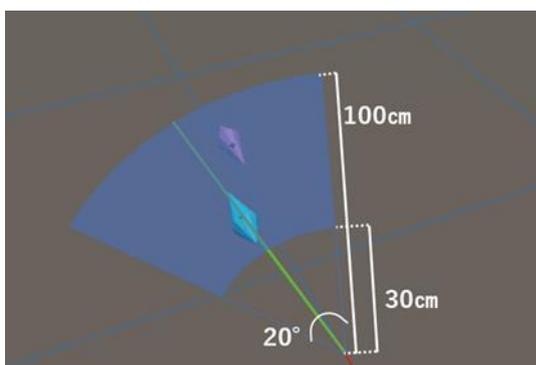


Figure 19. Area where the target appears (blue area in the Figure).

usefulness of posture indication was verified according to the posture error and operation time. The usefulness of the switching instruction was verified by comparing the position error, posture error and operation time in each condition. Finally, the usefulness of the 3D head pointer as a whole was verified based on the position error, posture error and operation time. Section 3.4 describes these results.

3.4. Results and discussion on the optical transparent HMD

3.4.1. Position error and posture error

The average values of the position and angle errors for each condition for the six subjects are shown in Figure 20. In this experiment, the tolerance was set assuming the same use of SRL as in the experiment discussed in Section 2.2.1, and the tolerance of the position indication was 3 cm. In the jamming hand of the SRL, when reaching vertically to a cylindrical or spherical object, the success rate for grasping did not decrease if the angular error was within 30° [16].

The average position error of the instructions in this experiment was approximately 1.25 cm for the voice switching method and approximately 2.82 cm for the head-gesture switching method, and a significant difference was observed between the two conditions in the Wilcoxon signed-rank test. This result demonstrates that the voice recognition method is more accurate in terms of indicating the position. Since the instruction error of the position instruction alone in Section 2.2.1 was 1.32 cm, this result shows that the head-gesture switching method has a negative effect on the accuracy of the location instruction. The increased error in the head-gesture result can be attributed to the shift in the position indication; when the head is tilted to switch from position to posture instructions, the direction of the face moves accordingly. In addition, in the

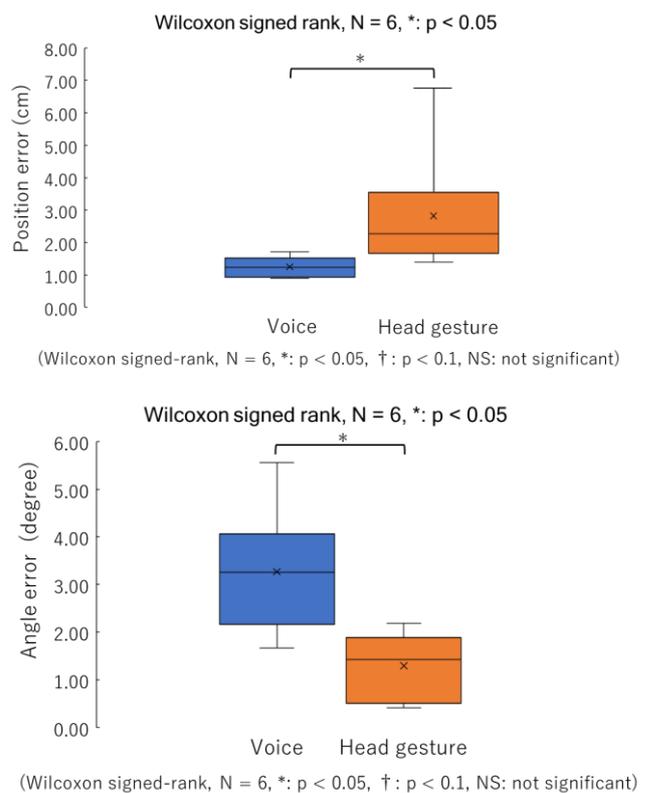


Figure 20. Top: Error in position indication; Bottom: Error in posture indication.

questionnaire, there were several comments noting that it was difficult to tilt the head without changing the direction of the face while indicating using the head gesture.

The average error for the posture instruction was approximately 3.56° for the voice switching method and approximately 1.78° for the head-gesture switching method, and a significant difference was observed between the two conditions in the Wilcoxon signed-rank test. This result shows that the accuracy of posture indication is higher when using head gestures. This can be attributed to the fact that posture instruction is an isometric input; as long as the head is rotated from the origin, the posture of the cursor will continue to rotate. If the operator uses head gestures, the instruction can be rapidly switched to stereotactic instructions, and consequently, the cursor posture can be fixed at the moment the continuously rotating cursor reaches the target posture. In the voice-based switching method, there is a delay between the time the voice command is uttered and the time the uttered voice is recognised as a command by voice recognition.

The voice-based switching method might cause the cursor to rotate during the time the user wants to switch; however, a time delay occurs when the operation actually switches to the position instruction, resulting in a posture error. These results show that voice-based switching is effective in terms of position indication, and head-gesture-based switching is effective in terms of posture indication. Furthermore, when switching using voice, the posture error increases; but even for the subject with the largest error, the average error was 5.56° , which is within the acceptable range of 30° . However, the subject with the largest error in the case of head-gesture-based switching had an average position error of 6.74 cm, which is far beyond the acceptable error of position instruction. Thus, it can be concluded that the voice-based switching method is more useful in terms of instructional accuracy, as all the values are within the acceptable error range for the SRL assumed in this experiment.

3.4.2. Operation time

The mean values of the operation time for each condition for the six participants are shown in Figure 21. The average operating time was approximately 20.3 s for the voice switching method and approximately 20.8 s for the head-gesture switching method. There was no significant difference between the two conditions in the Wilcoxon signed-rank test. This indicates that there is no significant difference between the two switching methods in terms of operation time. When combined with the

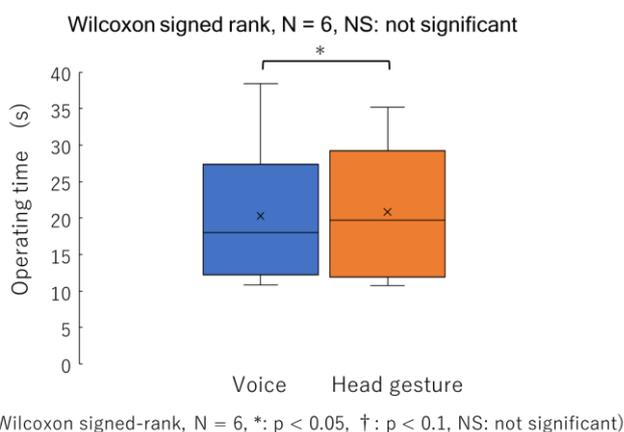


Figure 21. The mean values of the operation time.

results of instructional accuracy, the results suggest that voice switching is more practical.

Moreover, the average operation time for position instructions alone, as discussed in Section 2.2.2, was 6.2 s. In this experiment, the operation time was three times higher than when only position instructions were used owing to the addition of posture and switching indications. In addition, compared to the participant with the shortest average operation time, the participant with the longest average operation time had an operation time that was three times longer. When the subjects were asked about the cause of the increase in operation time in the questionnaire, some of them explained that the operation took longer when the posture indication did not perform well. The causes of the delay for posture indication were as follows:

- 1) when giving posture instructions, incorrect rotation was mistakenly fed as input,
- 2) compared to position instructions, it was difficult to correct errors when they occurred,
- 3) it was difficult to understand the posture of the cursor or target during rotation instructions.

Since posture manipulation by intentionally moving the neck along the three axes is not performed in daily life, the reason for cause 1) was verified. The reason for cause 2) was the length of time it took to correct the error because the error had to be corrected by indicating the amount of displacement in the posture indication. This is in contrast to position indication, which can directly specify the correct position when an error occurs. The reason for cause 3) was related to depth perception and size perception in the peripheral vision. The permissible eccentricity for recognising the position and shape of an object in the peripheral vision is 15° [10], but the perceptible eccentricity for depth is less than 12.5° , and the perceptible eccentricity for size is less than 5° [25]. In addition, the accuracy of both depth perception and size perception decreased with eccentricity from the gazing point. Because posture indication recognises the posture of an object from changes in the size and depth of each side of the cursor or target, it required more visual information than position indication. These reasons made it difficult to recognise the posture of the object when the face was turned away by up to 15° during posture manipulation.

3.4.3. Evaluation of the usefulness of the 3D head pointer as a whole

In the case of the voice switching method, the error in both position and posture indications was within the acceptable range, suggesting that the accuracy of the 3D head pointer is also effective for indications in real space through an optical transparent HMD. In terms of operation time, there was a large variation, and the indication time was not stable, indicating room for improvement. The improvement in posture instruction, which is the most significant factor for the increase in operation time, is considered to be effective, and from the results of the questionnaire, the improvements to be made are as follows:

- 1) construct the manipulation method using routine head movements,
- 2) use isotonic input,
- 3) do not leave the operator's gaze point.

Of these, 1) and 2) can be solved by using face orientation for posture indication, but there is a potential problem in how to provide posture instructions by rotating the head beyond its movable angle limit. In terms of finding a solution for 3), when the operator removes their gaze point from the cursor and target object in the posture indication state, the target object and cursor

can be improved by continuing to display them in front of the operator in augmented reality (AR). However, displaying real objects in AR in real time is a demanding process for AR devices.

In order to display AR in real time, it is necessary to devise a way to reduce using the processing power, such as detecting the mesh of objects in real space and displaying them.

4. DISCUSSION ON THE PRACTICAL APPLICATION OF A 3D HEAD POINTER

In this section, the practical application of the proposed method presented in this study is discussed. The advantages of the 3D head pointer can be clarified by comparing this method with other manipulation methods. Following the comparison, concerns about using this interface in real life are discussed.

4.1. Comparison with other similar methods

Based on the results of the previous section, the proposed method was compared with other similar methods.

a. *Physical controller*

Some SRLs, such as those made by Vernonia [1], use a physical controller that is similar to a gamepad, with an analogue stick and buttons as the method of operation. The advantage of the 3D head pointer is that its operation is more intuitive and easier to understand than that of a physical controller, and it can be operated hands free.

b. *SRL manipulation method using the feet*

The proposed method can operate the SRL in any standing or seated position unlike methods operated by the feet [5]. However, manipulation with the feet can simultaneously indicate the position and attitude of the SRL. A short operation time is the main advantage of foot operation.

c. *Head joystick and nodding to switch between the vertical and horizontal planes*

Because the 3D head pointer uses the compensatory motion of the head, it has a lower operational burden than methods that use the head as a joystick [11],[12]. In contrast, the nodding method [12] allows for digital input from the head alone and may be used in conjunction with the 3D head pointer.

4.2. Limitations

In this study, voice recognition was used to give instructions, such as for switching, but voice recognition has the disadvantage of not being able to operate in a noisy environment or while the operator is having a conversation. Some prior examples of command-type instructions use gaze to provide instructions [26],[27]. The combination of pointing instructions with the head and gaze-based instructions could provide a more flexible environment for SRL indications.

If there is a need to use SRL for complex or long movements in daily life, the movements must be registered and played back. Registering and replaying behaviours require many commands, but the number of command-type instructions that can be intuitively memorised and selected is as few as six [28]. When building a system with seven or more commands, it is necessary to devise a way to remember commands, such as displaying a menu screen in the HMD.

5. CONCLUSIONS

In this study, a spatial position and posture indication interface for SRLs was proposed to improve functional efficiency in the execution of routine tasks. The required functions for indicating spatial position and posture have been described, and a position indication method, the 3D head pointer, has been proposed, which combines head-bobbing-type depth indication for spatial position and polar direction indication by face orientation. In a VR environment, evaluation tests of the 3D head pointer and IDF were conducted. The results showed that the 3D head pointer had sufficient accuracy without requiring the operator to interrupt their actions.

In addition, to provide not only position but also posture guidance by using a 3D head pointer, a posture guidance method using head rotation as isometric input and two types of switching guidance methods using voice recognition and head gestures were proposed. In addition, a comparative study of two switching instruction methods using an optical transparent HMD and a test to evaluate the usefulness of the 3D head pointer as a whole was conducted. The results showed that the switching method based on voice recognition was effective for using the assumed SRL, and it was confirmed that the 3D head pointer was sufficiently accurate to be useful for operating robotic arms using an optical transparent HMD. These results provide useful knowledge for improving the SRL interface.

In the future, an intuitive posture instruction method will be developed that is not affected by compensatory head movements and that will incorporate a command instruction method that replaces voice recognition. In addition, an SRL will be considered as an interface for use as a third arm in situations, such as banquets and construction sites, where an individual's hands are not sufficient.

ACKNOWLEDGEMENT

This research is supported by Waseda University Global Robot Academic Institute, Waseda University Green Computing Systems Research Organization and by JST ERATO Grant Number JPMJER1701, Japan.

REFERENCES

- [1] C. Veronneau, J. Denis, L. Lebel, M. Denninger, V. Blanchard, A. Girard, J. Plante, Multifunctional 3-DOF wearable supernumerary robotic arm based on magnetorheological clutches, IEEE Robotics and Automation Letters 5 (2020) pp. 2546-2553. DOI: [10.1109/LRA.2020.2967327](https://doi.org/10.1109/LRA.2020.2967327)
- [2] C. Davenport, F. Parietti, H. H. Asada, Design and biomechanical analysis of supernumerary robotic limbs, Proc. of the IEEE/ASME International Conference on Advanced Intelligent Mechatronics, Fort Lauderdale, Florida, United States, 17-19 October 2012, pp. 787-793. DOI: [10.1115/DSCC2012-MOVIC2012-8790](https://doi.org/10.1115/DSCC2012-MOVIC2012-8790)
- [3] H. H. Asada, F. Parietti, Supernumerary robotic limbs for aircraft fuselage assembly: body stabilization and guidance by bracing, Proc. of IEEE International Conference on Robotics and Automation, Hong Kong, China, 2014, pp. 119-125. DOI: [10.1109/ICRA.2014.6907002](https://doi.org/10.1109/ICRA.2014.6907002)
- [4] Y. Iwasaki, H. Iwata, A face vector - the point instruction-type interface for manipulation of an extended body in dual-task situations, IEEE International Conference on Cyborg and Bionic Systems, Shenzhen, China, 25-27 Oct. 2018, pp. 662-666. DOI: [10.1109/CBS.2018.8612275](https://doi.org/10.1109/CBS.2018.8612275)
- [5] T. Sasaki, M. Saraji, K. Minamizawa, M. Inami, MetaArms: body remapping using feet-controlled artificial arms, Proc. of the 31st

- Annual ACM Symposium on User Interface Software and Technology, New York, United States, 14 October 2018, pp. 65-74.
DOI: [10.1145/3242587.3242665](https://doi.org/10.1145/3242587.3242665)
- [6] S. G. Terashima, J. Sakai, T. Ohira, H. Murakami, E. Satho, C. Matsuzawa, S. Sasaki, K. Ueki, Development of a tongue operative joystick - for proposal of development of an integrated tongue operation assistive system (I-to-AS) for seriously disabled people, *The Society of Life Support Engineering* 24 (2012), pp. 201-207.
DOI: [10.5136/lifesupport.24.201](https://doi.org/10.5136/lifesupport.24.201)
- [7] R. Barea, L. Boquete, M. Mazo, E. Lopez, System for assisted mobility using eye movements based on electrooculography, *IEEE Transactions on Neural Systems and Rehabilitation Engineering* 10 (2002), pp. 209-218.
DOI: [10.1109/TNSRE.2002.806829](https://doi.org/10.1109/TNSRE.2002.806829)
- [8] R. C. Simpson, S. P. Levine, Voice control of a powered wheelchair, *IEEE Transactions on Neural Systems and Rehabilitation Engineering* 10 (2002) pp. 122-125.
DOI: [10.1109/TNSRE.2002.1031981](https://doi.org/10.1109/TNSRE.2002.1031981)
- [9] S. Nishio, C. I. Penalzoza, BMI control of a third arm for multitasking, *Science Robotics* 3 (2018) 20.
DOI: [10.1126/scirobotics.aat1228](https://doi.org/10.1126/scirobotics.aat1228)
- [10] T. Miura, Behavioral and Visual Attention, Kazama Shobo, Chiyoda, Japan, 1996, ISBN 978-4-7599-1936-3.
- [11] R. Hasegawa, Device For Input Via Head Motions, Patents WO 2010/110411 A1, Japan, 30 September 2010.
- [12] A. Jackowski, M. Gebhard, A. Gräser, A novel head gesture based interface for hands-free control of a robot, Proc. of the IEEE International Symposium on Medical Measurements and Applications, Benevento, Italy, 15-18 May 2016, pp. 1-6.
DOI: [10.1109/MeMeA.2016.7533744](https://doi.org/10.1109/MeMeA.2016.7533744)
- [13] M. Kouchi, M. Mochimaru, AIST Anthropometric Database, Pub. National Institute of Advanced Industrial Science and Technology, Japan, January 2005. Online [Accessed 4 September 2021]
https://www.airc.aist.go.jp/dhrt/91-92/fig/91-92_anthrop_manual.pdf
- [14] L. Drohne, K. Nakabayashi, Y. Iwasaki, H. Iwata, Design consideration for arm mechanics and attachment positions of a wearable robot arm, Proc. of the IEEE/SICE International Symposium on System Integration, Paris, France, 14-16 January 2019, pp. 645-650.
DOI: [10.1109/SII.2019.8700355](https://doi.org/10.1109/SII.2019.8700355)
- [15] Windows Dev Center - Hardware, Pointer Ballistics for Windows XP, 2002. Online [Accessed 4 September 2021]
<http://archive.is/20120907165307/msdn.microsoft.com/en-us/windows/hardware/gg463319.aspx#selection-165.0-165.33>
- [16] K. Amano, Y. Iwasaki, K. Nakabayashi, H. Iwata, Development of a three-fingered jamming gripper for corresponding to the position error and shape difference, *RoboSoft* (2019), IEEE International Conference on Soft Robotics, Seoul, Korea (South), 14-18 April 2019, pp. 137-142.
DOI: [10.1109/ROBOSOFT.2019.8722768](https://doi.org/10.1109/ROBOSOFT.2019.8722768)
- [17] HTC VIVE, 2011. Online [Accessed 4 September 2021]
<https://www.vive.com/eu/product/vive/>
- [18] C. Zaiontz, Wilcoxon Signed-Ranks Table, 2020. Online [Accessed 4 September 2021]
<http://www.real-statistics.com/statistics-tables/wilcoxon-signed-ranks-table/>
- [19] Unity Technologies Japan/UCL, Unity-chan!, 2014. Online [Accessed 4 September 2021]
<https://unity-chan.com/>
- [20] Committee on Physical Disability, Japanese Orthopaedic Association, Joint range of motion display and measurement methods, *Japanese Journal of Rehabilitation Medicine* 11 (1974) pp. 127-132.
DOI: [10.2490/jjrm1963.11.127](https://doi.org/10.2490/jjrm1963.11.127)
- [21] S. A. Douglas, A. K. Mithal, *The Ergonomics of Computer Pointing Devices*, Springer, London, 1997.
- [22] J. P. Rolland, R. L. Holloway, H. Fuchs, Comparison of optical and video see-through, head-mounted displays, Proc. of The International Society for Optical Engineering, 21 December 1995, pp. 292-307.
DOI: [10.1117/12.197322](https://doi.org/10.1117/12.197322)
- [23] Hololens2. Online [Accessed 4 September 2021]
<https://www.microsoft.com/en-us/hololens/buy>
- [24] Mixed Reality Toolkit. Online [Accessed 4 September 2021]
<https://hololabinc.github.io/MixedRealityToolkit-Unity/README.html>
- [25] A. Yasuoka, M. Okura, Binocular depth and size perception in the peripheral field, *Journal of the Vision Society of Japan* 23 (2011) pp. 103-114.
DOI: [10.24636/vision.23.2_103](https://doi.org/10.24636/vision.23.2_103)
- [26] M. Yamato, A. Monden, Y. Takada, K. Matsumoto, K. Tori, Scrolling the text windows by looking, *Transactions of the Information Processing Society of Japan* 40 (1999), pp. 613-622. Online [Accessed 4 September 2021]
https://ipsj.ixsq.nii.ac.jp/ej/?action=pages_view_main&active_action=repository_view_main_item_detail&item_id=12841&item_no=1&page_id=13&block_id=8
- [27] T. Ohno, Quick menu selection task with eye mark, *Transactions of the Information Processing Society of Japan* 40 (1999), pp. 602-612. Online [Accessed 4 September 2021]
https://ipsj.ixsq.nii.ac.jp/ej/?action=pages_view_main&active_action=repository_view_main_item_detail&item_id=12840&item_no=1&page_id=13&block_id=8
- [28] Y. Iwasaki, H. Iwata, Research on a third arm: analysis of the cognitive load required to match the on-board movement functions, Poster presented at: The Japanese Society for Wellbeing Science and Assistive Technology, 6-8 September 2018, Tokyo, Japan, Session No. 2-4-1-2.