Available online at www.sciencedirect.com

ScienceDirect

IFAC Papers Online
CONFERENCE PAPER ARCHIVE

IFAC-PapersOnLine 49-19 (2016) 367–372

# Analysis of Human Pointing Behavior in Vision-based Pointing Interface System - difference of two typical pointing styles -

**Kazuaki Kondo** * **Genki Mizuno** * **Yuichi Nakamura** *

* *Academic Center for Computing and Media Studies, Kyoto University, Yoshida-Honmachi, Sakyo-ku, Kyoto 606-8501, Japan (email: kondo, mizuno, yuichi@ccm.media.kyoto-u.ac.jp)*

**Abstract:** This paper reports human pointing behaviors in vision-based pointing interface system, to make a mathematical model of them for designing easy-to-use interface. In natural pointing situations, we point targets at distant position with various postures, for example straight arm style or bent elbow style. We analyze their difference in pointing behaviors with assuming the pointing interface system as a feedback control model including an indicator. The difference had been confirmed in the step responses and the estimated parameters in the transfer functions, and matches to our actual experiences in those pointing styles. The estimation accuracy of indicated position from indicator's posture in the intermediate styles has been also analyzed. The results said that the reference point of indication smoothly moves from indicator's eye to his or her elbow according to the elbow joint angle.



*Keywords:* Pointing systems, Human-machine interface, Control theory, Closed-loop systems

## 1. INTRODUCTION

Nowadays we can get wide visual display devices inexpensively, which makes us easy to show various contents on large regions. This accelerates distance between display and users. In such environments, easy and intuitive remote interaction scheme is required rather than oscillatory interface like touch panel. For example, in presentation scene with slides on a wide screen, it is not reasonable for audiences to come close to the screen for pointing particular portion on the slide, directory. We usually use pointing gesture in such a case. The pointing interface system shown in Fig. 1 supports the remote pointing using on visual measurement. It shows a pointer at an estimated indication position on a display based on capturing indicator's posture. With this system, an indicated position becomes clear to the indicator and audience, which encourages smooth communication. We often use a raiser pointer for similar purpose, but it can be used only for indicating. The vision-based pointing interface controls displayed contents and materials according to user's posture, and will construct interactive scheme beyond mere indicating. Additionally when robots living with human and/or virtual agents connect to the pointing interface, these can get information about indicators and behave adaptively.

Kondo et al. (2015) tried to assume the vision-based pointing environment as a feedback loop model under classical control theory. It mathematically describes indicator's pointing behaviors and visual perception using transfer functions to simulate them with various configurations for designing easy-to-use interface, and to predict pointing behaviors for adaptive display. However the proposed model assumed an indicating posture with an arm being straight, and did not assume that with an elbow joint being bent that also appears in natural pointing situations. The purpose of this paper is to analyze human pointing behavior in the two pointing styles to expand the proposed pointing interface model to general one that can deal with arbitrary pointing types. We analyze how the pointing styles affect to pointing behaviors and how these can be described in the interface system model through measurement of actual pointing. Additionally what kind of method to estimate indicated positions should we select, in the case of the intermediate pointing style, e. g. slightly bent elbow postures are also investigated.

## 2. RELATED WORKS

The real-time vision-based pointing interface becomes practical with the progress of computational and visual sensing performance. However, we still have two considerable problems for natural pointing in daily environments. One is the difficulty in accurate measurement of pointing pose. Significant accuracy is required, especially for a target at a distance because even tiny errors on coordinates of body parts are amplified on a screen. We have several marker-less motion capture techniques based on visual sensing that do not interfere with an indicator's behavior like Shotton et al. (2011); Yoshimoto and Nakamura (2015), but the performance of those methods does not satisfy the requirement, because visual sensing is unstable to illumination change and occlusion. The latency arising from the sampling time of visual sensing and processing time for estimation also cannot be ignored.


10.1016/j.ifacol.2016.10.593

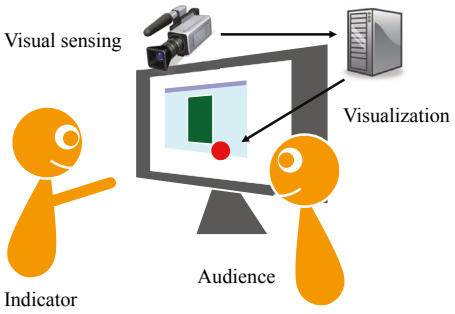


Fig. 1. An overview of a vision-based pointing interface. The vision sensor measures indicator's posture and estimates indicated position to display a pointer at that location.

The other is ambiguity in human pointing behavior. It depends on position of the indicating target, intention in indicator, and also individuality. Pointing pose relative to a screen indicates approximately where a person is pointing. Fukumoto et al. (1994) reported that a target position is on a line defined by a fingertip and a reference point inside of an indicator's body. The reference point moves depending on the pointing pose. As shown in Fig. 2, it is placed at an eye position for a distant target, but at an elbow position for a relatively near target. In addition, a geometric environment perceived by a human may not match an actual one. Knowledge of the relationship between a pointing posture and the indicated position is complicated and influenced substantially by various conditions. This makes it difficult to estimate indicated position from pointing pose accurately.

Characteristics of a transient pointing behavior during an indicator changes a pointing target have been analyzed for a long time. R.S.Woodworth (1899) proposed a pointing action model with a combination of feed-forward motions for rapid approach to a target position followed by feedback adjustments. Fitts (1954) reported that pointing duration increases with a larger moving distance or a smaller target, a notion known as "Fitts's law", which has been used in many studies because of its accurate approximation in various conditions.

The above problems suggest a need for additional schemes that reduce the influence of the ambiguity of pointing behaviors and measurement error to construct an easy-to-use pointing interface. Most of the conventional methods tackling this issue focus on how to visualize a pointer and contents on a screen. McGuffin and Balakrishnan (2005) proposed zooming the region around the pointer. This means that the target size and the distance from it feel larger in its neighborhood. To obtain a similar effect, controlling cursor size or cursor speed has also been proposed by Worden et al. (1997); Grossman and Balakrishnan (2005); Blanch et al. (2004). Retaining the pointer trajectory within the last short duration proposed by Baudisch et al. (2003) helps to recognize and predict the behaviors of the pointing interface intuitively. However, those methods assume a mouse interface and have not been evaluated with a remote vision-based pointing interface, as we assume here. Thus, we need a general framework that enables us to evaluate, compare, combine, and improve such conventional methods under various conditions.

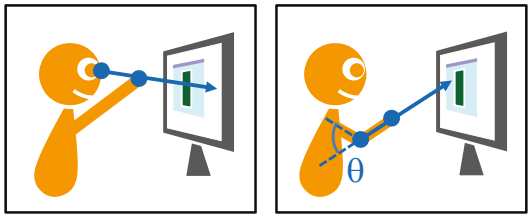


Fig. 2. The indicated position model for two typical pointing styles. (Left) with straight arm (Right) with bent elbow

## 3. MODEL OF POINTING INTERFACE SYSTEM

In this paper, We assume a vision-based pointing system of the sort in Fig. 1. It proceeds as follows.

(1) The indicator has the target he wants to indicate in his or her intention. We define its position as a reference position $p_t$.
(2) The computer estimates its location $p_e$ based on visual sensing through the camera. The pointer is displayed at the location $p_c$ conducted from $p_e$ with some filters for visualization.
(3) The indicator recognizes the pointer location as $p_r$ and adjusts his pointing posture to move it close to the target.
(4) Steps (2)-(4) continue until $p_t - p_r$ becomes 0. [1]

This procedure can be modeled as a feedback control loop, as shown in Fig. 3. The control model is constructed by $H_g$ for the indicator's body kinematics, $H_p$ for his or her visual perception catching a pointer, $H_s$ for a computer estimating indicated position, and $H_v$ for visualization filter. In this model, the indicator works as a controller with $H_g$ and adjusts a feedback gain with $H_p$, simultaneously. The visualization $H_v$ includes the pointer's shape, position, and so on. The two problems in a vision-based pointing interface, the pointing pose ambiguity and the error of pointing pose estimation, are in $H_g$ and the noise $d_s$, respectively.

### 3.1 Pointing interface part

The estimation of indicated position $H_s$ consists of visual sensing via cameras and a pose estimation algorithm based on the measurement. We assume that it correctly estimates indicated position with particular latency $\tau_s$ and the estimation error included in the noise term $d_s$, to formulate $H_s$ as

$$H_s(s) = e^{-\tau_s s}. \tag{1}$$

On the other hand, the formulation of $H_v$ is determined by the visualization methods of a pointer and contents. We assume to display a sufficiently small circle pointer at a smoothed position of the estimated pointing position during the latest short duration $N$. This visualization formulates $H_v$ as

$$H_v(s) = \frac{1}{N}\sum_{n=0}^{N-1} e^{-(n\tau_s+\tau_v)s}, \tag{2}$$

[1] In this paper, we focus on only the indicator's behavior with assumption of audience perception being same as him.

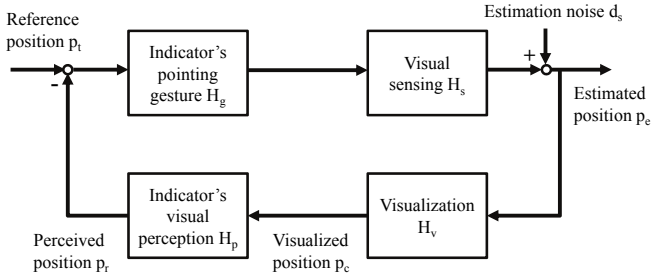


Fig. 3. The control model of the vision-based pointing interface system. The modules $H_g, H_p, H_s$ and $H_v$ describe indicator's body kinematics, his visual perception, estimating pointing position, and cursor visualization. $d_s$ means disturbance into the position estimation.

where $\tau_v$ is the latency for displaying the cursor on a screen. Eq. (2) with $N = 1$ means a typical display without smoothing. Other pointer visualizations can also be implemented using similar formulations. Note that a large circle pointer or displaying trajectory of estimated positions can have multiple factors affecting human visual perception. This requires a multidimensional design for $H_v$ and $H_p$.

### *3.2 Indicator part*

The characteristics of human pointing action can still be uncertain, especially from an analytic viewpoint. We configured a mathematical model of $H_g$ based on the measurements of actual pointing behaviors. In the preliminary experiments, we analyzed indicated position trajectories without cursor visualization, i.e., no feedback information is provided to an indicator. Pointing behavior when a pointing target suddenly moves to a distant location can be considered to be a step response of $H_g$. We confirmed that the trajectories converge with some overshoots and then formulated $H_g$ as a second-order lag element

$$H_g(s) = e^{-\tau_g s} \frac{K_g}{T_g^2 s^2 + 2\zeta T_g s + 1}, \tag{3}$$

where $\tau_g, T_g$ and $\zeta$ mean dead time to begin a pointing action, a parameter determining pointing speed, and a damping coefficient of pointing fluctuation, respectively.

$H_p$ is also difficult to formulate explicitly, because $p_r$ is an internal value of indicator and cannot be measured well. Thus, we consider $H_p$ to be a first-order lag element

$$H_p(s) = \frac{K_p}{T_p s + 1} \tag{4}$$

to reflect a delay in human perception and to avoid overfitting in the identification phase.

## 4. SYSTEM IMPLEMENTATION FOR EXPERIMENTS

The overview of our implemented pointing interface for the experiments is shown in Fig. 4. The *pixel* × *pixel* visual contents are projected on the $m \times m$ screen ( the white wall ) using the short focal length projector. Subjects indicate particular points on the screen at approximately

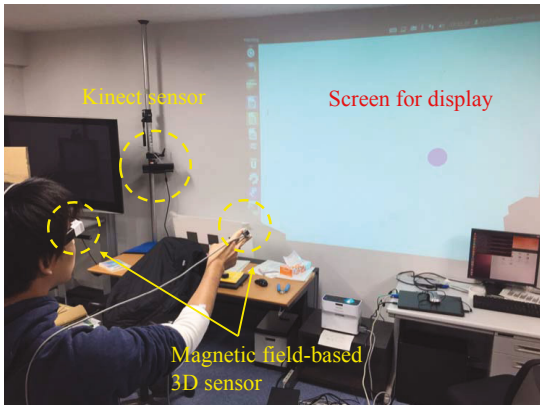


Fig. 4. The experimental environment

$2.5m$ distance from that. For natural pointing, indicating postures are measured by Kinect v2 sensor close to the screen at the left of the subjects. But in the following experiments, we use the magnetic field-based 3D position sensor attached on the subject's body instead of the Kinect sensor to realize the assumption $d_s = 0$ as explained in the section 3.1. While this implementation should not be allowed in practical use, it can be accepted for analysis of human pointing behaviors.

The available measurement space of the magnetic field sensor is almost $1m$ cube. It is difficult to calibrate the geometric relation to the screen, directly. Thus we estimated it via the Kinect sensor coordinate system, e. g. the combination of the rigid transformation between the magnetic field sensor and the Kinect sensor, and that between the Kinect sensor and the screen. We assumed homography transformation between the screen and the contents in the computer.

The indicated position is estimated as the crossing position of the screen and the indicating vector that connects the reference point in the body and the finger tip, based on the Fukumoto et al. (1994)'s report. This corresponds to the actual processing of $H_s$. When the indicator's arm is straight ( named "pointing style A" in this paper ), the magnetic field sensors are attached to his or her temples and finger tip to acquire the indicating vector ( named "eye-reference estimation model" ). In the case of the elbow joint being bent ( named "pointing style B" ), the sensor on the elbow is used for the reference point ( named "elbow-reference estimation model"). The approximately $1cm$ circular pointer is displayed at the estimated position without temporal smoothing with $N = 1$ in $H_v$. This pointer can be assumed as a point for the indicator.

## 5. DIFFERENCE OF POINTING BEHAVIORS IN TWO POINTING STYLES

We focus on step responses of the pointing interface system to compare pointing behavior between the two pointing styles. A step input signal and its response correspond to the pointing target suddenly popping up at a distant from the initial position and the transient trajectory until the indicator feels to finish indicating the target, respectively. The experimental procedures are below.

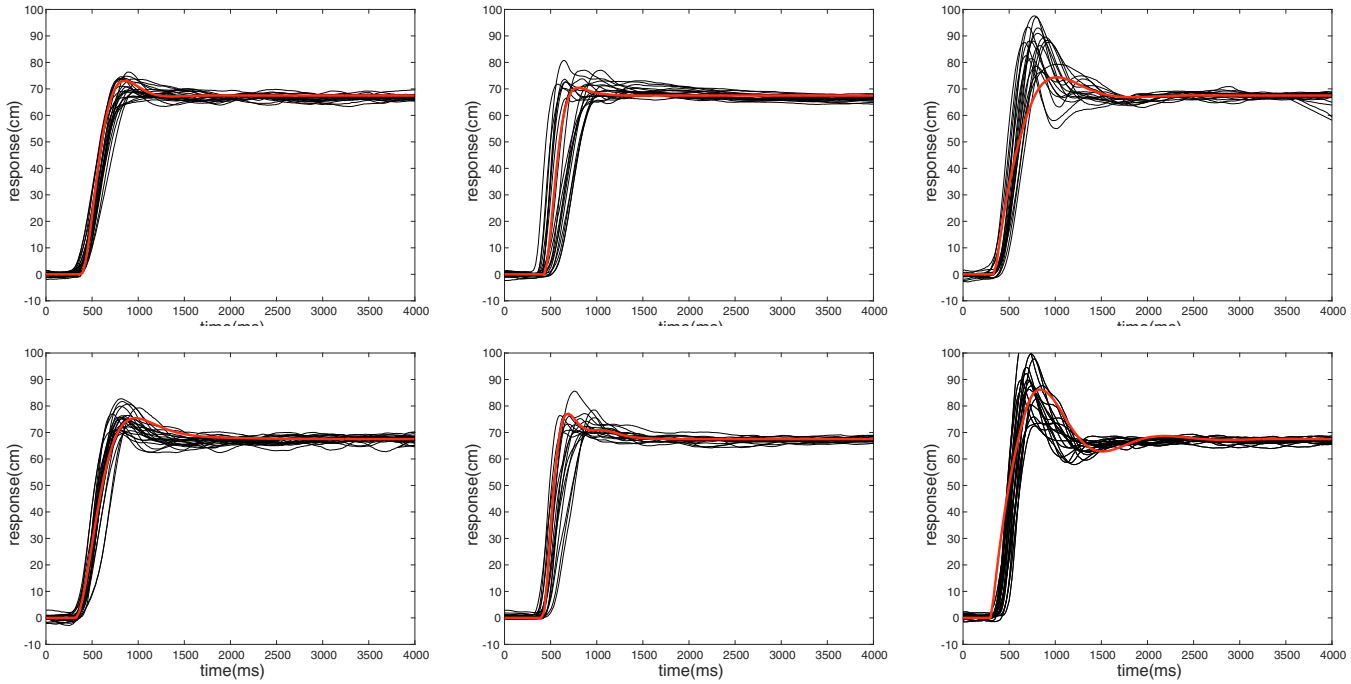


Fig. 5. The step response trajectories for the two pointing styles. The horizontal and vertical axes represent transit time from the step input and distance from the initial location, respectively. The black colored trajectories mean the actually measured samples. The red one are simulated trajectories with the optimized model parameters estimated from the measured samples. (Top-row) the pointing style A (Bottom-row) the pointing style B

(1) The measurement begins when a subject indicates an initial target visualized on the screen and the pointer remains to that position.
(2) The initial target suddenly disappears. Simultaneously, a new target pops up at a $70cm$ distance from the initial target. The subject changes his or her posture to move the pointer onto the new target.
(3) The measurement stops when the subject calls the finish of the pointing action.

The measurement results for three subjects are shown as black colored trajectories in Fig. 5. We can see larger overshoots and fluctuation in the results of the pointing style B than style A. This matches to our daily experiences. The possible reasons are that small motion of a finger tip is reflected to larger motion on the screen in the pointing style B compared to the style A because the distance from the indication reference point and the finger tip is larger. Furthermore it is difficult for indicators to correctly percipient the indication vector because they can not see the elbow joint well and just have to perceive its location with only body sensation. In the pointing style A, rising velocity has larger variance compared to the style B. It may come from sizes of body portions to be moved in pointing ; an indicator needs to move whole arm in the pointing style A while only lower arm and hand in the most cases of the style B.

For quantitatively analyzing the such differences between two pointing styles, we estimated the model parameters from the measured samples shown in Fig. 5 based on Kondo et al. (2015)'s method. This is to search parameters that minimize the difference between simulation and measured values of the pointing trajectories. The unknown parameters in the model are $\tau_g, K_g, T_g, \zeta, K_p, T_P$. The remained latency parameters $\tau_s, \tau_v$ included in $H_s, H_v$ can be determined by their measurement. We applied a convergence constraint to decrease their degree-of-freedom. Considering natural pointing situations, a cursor position $p_e$ converges on a target position $p_t$ after sufficient time passes. This corresponds to a stationary error of $p_t - p_e$ being zero for a step input. The evaluation function $E$ is a residual of step responses on $p_e$, formulated as

$$E(\tau_g, K_g, T_g, \zeta, T_P) = \sum_{i=0}^{N} \sum_{t=0}^{T_e} (\hat{p_e}(t) - p_e(t))^2, \qquad (5)$$

where $T_e, p_e(t), \hat{p_e}(t)$ indicate the measurement duration, measured value of pointing position, and simulated value, respectively. Eq. (6) is minimized about $\tau_g, K_g, T_g, \zeta, T_P$.

The results of the model parameter estimation and the simulated step responses using them are shown in Table 1 and the red colored trajectories in Fig. 5. Unfortunately simulated trajectories do not describe general characteristics because the system model may not have enough DOG to deal with the various measured samples. Thus we roughly analyze the optimized parameters. Focus on the subject 1 and 2, first. $K_g$ in the pointing style B is larger than that in the style A, while the dead time $\tau_g$ is smaller. This explains easiness of moving only the lower arm in the style B. $\zeta$ has relation to fluctuation of step response trajectory. The $\zeta$ values in the case of the style B is smaller than that of the style A. It corresponds to larger fluctuation and matches to the measured samples. $T_p$ becomes a coefficient of a derivative term in the closed-loop transfer function. Thus it means degree of dependency on prediction of the displayed pointer motion when an

Table 1. The estimated model parameters for two pointing styles.

StyleA : with straight arm

| sub. | $K_g$ | $T_g$ | $\zeta$ | $\tau_g$ | $T_p$ |
|---|---|---|---|---|---|
| 1 | 1.010 | 115.0 | 0.6423 | 381.8 | $6.491 \times 10^{-3}$ |
| 2 | 1.010 | 79.23 | 0.7348 | 430.4 | $5.890 \times 10^{-4}$ |
| 3 | 1.400 | 8.884 | 19.27 | 347.4 | 0.1317 |

Style B : with bent elbow

| sub. | $K_g$ | $T_g$ | $\zeta$ | $\tau_g$ | $T_p$ |
|---|---|---|---|---|---|
| 1 | 1.059 | 153.3 | 0.6763 | 318.9 | 531.3 |
| 2 | 1.056 | 68.11 | 0.6232 | 410.8 | 256.7 |
| 3 | 1.780 | 22.36 | 7.918 | 306.0 | 2.344 |

indicator changes own pointing posture. An indicator can not see own elbow well in the case of the pointing style B. Thus he or she seems to strongly depend on prediction.

We did not find clear results as noted above in the case of the subject 3. One of the reasons is that the subject 3 is not accustomed to the pointing interface system. The values of $\zeta$ and $T_p$ being similar in the both pointing styles can explain that novice have not learned sensation of body dynamics and do not depend on own uncertain prediction, respectively.

## 6. HOW TO ESTIMATE INDICATED POSITION FOR INTERMEDIATE POINTING STYLES

In practical pointing situations, people do not use the either pointing style alternatively. The intermediate posture with an elbow joint being slightly bent often appears. Here we conduct fundamental investigation to make it clear the relationship between pointing posture and indicated position in such case. We focus on whether the estimation model of indicated position smoothly changes from the eye-reference to the elbow-reference, or switches them at the particular elbow angle. In the experiments, the positions of the temple, the elbow, and the finger tip are measured by the magnetic field sensors when the indication posture becomes stable for the static pointing target. For various elbow joint angles, estimation errors are calculated as Euclid distances between the estimated position and the displayed target position on the screen, and used for analyzing which estimation model is suitable. The error distribution corresponding to the elbow joint angle $\theta$ for 3 subjects are shown in Fig. 6.

Look at the estimation errors drawn with the black circles that assume the eye-reference model. The errors monotonically and smoothly increase according to the elbow joint angle in the result of all subjects. The estimation error drawn with the gray squares that assume the elbow-reference model also changes smoothly but not monotonically. Because the elbow is relatively close to the vector connecting the eye and the finger tip when the elbow angle is small, e. g. the two estimation model are similar. The smaller errors switch from the eye-reference to the elbow-reference at different angle for the 3 subjects ; approximately 55 degrees for the subject 1, 2, and 70 degree for the subject 3, but the similar error amounts. As noted in section 5, the subject 3 is novice to the pointing interface and does not perceive own body posture well, especially when he bends his elbow. Thus the estimation accuracy with the elbow-reference model for around 60

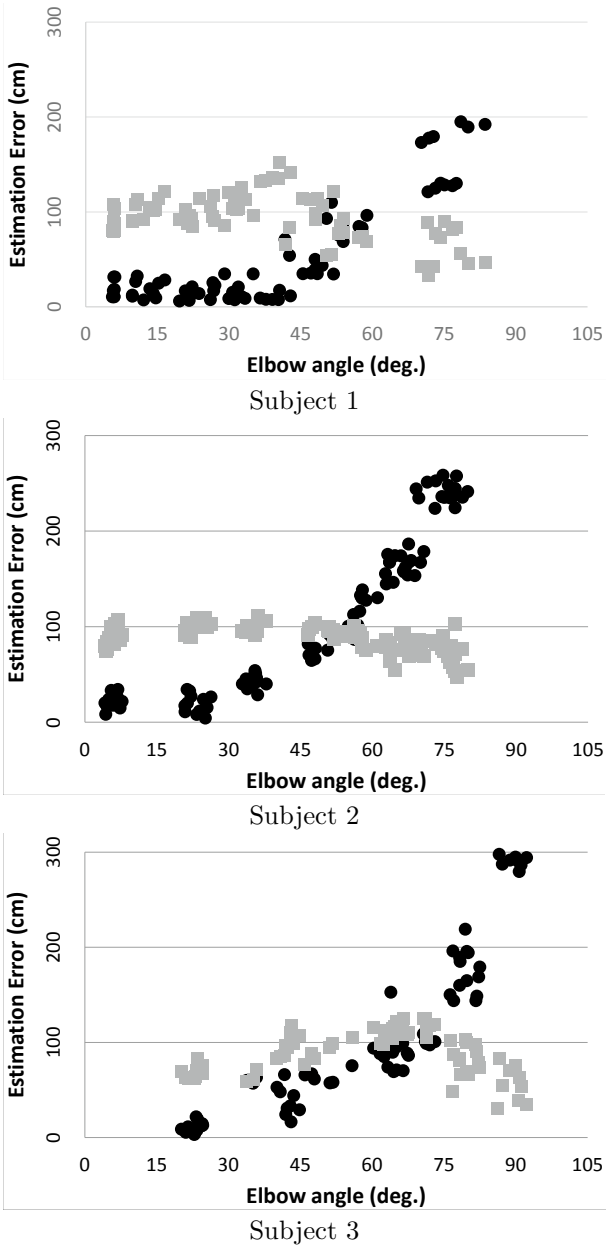


Fig. 6. The estimation errors of the indicated position. The horizontal axis represents elbow joint angle $\theta$ shown in Fig. 2 constructed by a shoulder-elbow vector and a elbow-hand vector. 0 and 90 degrees correspond to the pointing style A and B, respectively. The black circles and gray squares correspond to the estimation errors using the eye-reference model and the elbow-reference model.

degree seems to be still large. From the above analysis, a general estimation model for indicated position in which the reference point smoothly moves from eye to elbow may be suitable than the existing two models.

## 7. CONCLUSION

In this paper, we analyze human pointing behavior with two typical pointing styles in vision-based pointing interface. The difference of pointing behaviors between the two pointing styles can be confirmed as the step responses and the parameters in the transfer functions in the con-

trol model. The mathematical characteristics of each style considered from those results well match to our daily experiences on actual remote pointing situations. From those analysis we confirmed that the parameters in the control model reflect human pointing behaviors in two pointing styles. But we also confirm that the current control model do not well describe the trajectory variations. The methods for estimating the model parameters should be improved because the residual minimization in the actual domain does not always well represent the pointing trajectories as shown in the cases of the subject 3 in Fig. 5. ( over smoothed ). The minimization in the frequency domain may match better.

The estimation error of the indicated position for intermediate pointing styles with the two reference model indicates that the suitable estimation model for arbitrary elbow angle smoothly connects the eye-reference model and the elbow-reference model. Future works are to expand the current control model and the indicated position model to more general models based on the results, and to design the visualization filter $H_v$ to construct easy-to-use interface adapting various pointing styles.

## REFERENCES


Baudisch, P., Cutrell, E., and Robertson, G. (2003). High-density cursor: A visualization technique that helps users keep track of fast-moving mouse cursors. In *In Proc. of Interact2003*, 236–243.

Blanch, R., Guiard, Y., and Beaudouin-Lafon, M. (2004). Semantic pointing: improving target acquisition with control-display ratio adaptation. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '04, 519–526. ACM, New York, NY, USA.

Fitts, P.M. (1954). The information capacity of the human motor system in controlling the amplitude of movement. *Journal of Experimental Psychology*, 47(6), 381–391.

Fukumoto, M., Suenaga, Y., and Mase, K. (1994). Finger-Pointer: Pointing interface by image processing. *Computers & Graphics*, 18(5), 633–642.

Grossman, T. and Balakrishnan, R. (2005). The bubble cursor: enhancing target acquisition by dynamic resizing of the cursor's activation area. In *In Proc. of the SIGCHI conference on Human factors in computing systems*, 281–290. ACM.

Kondo, K., Nakamura, Y., Yasuzawa, K., Yoshimoto, H., and Koizumi, T. (2015). Human pointing modeling for improving visual pointing system design. In *In Proc. of Int. Symp. on Socially and Technically Symbiotic Systems (STSS) 2015*.

McGuffin, M.J. and Balakrishnan, R. (2005). Fitts' law and expanding targets: Experimental studies and designs for user interfaces. *ACM Trans. Comput.-Hum. Interact.*, 12(4), 388–422.

R.S.Woodworth (1899). The accuracy of voluntary movement. *Phychological Review Monograph Sullplement*, 3(13), 1–119.

Shotton, J., Fitzgibbon, A., Cook, M., Sharp, T., Finocchio, M., Moore, R., Kipman, A., and Blake, A. (2011). Real-time human pose recognition in parts from single depth images. In *In Proc. of the 2011 IEEE Conference on Computer Vision and Pattern Recognition*, CVPR '11, 1297–1304. IEEE Computer Society, Washington, DC, USA.

Worden, A., Walker, N., Bharat, K., and Hudson, S. (1997). Making computers easier for older adults to use: area cursors and sticky icons. In *Proceedings of the ACM SIGCHI Conference on Human factors in computing systems*, 266–271. ACM.

Yoshimoto, M. and Nakamura, Y. (2015). Cooperative gesture recognition: Learning characteristics of classifiers and navigating user to ideal situation. In *In Proc. of The 4th IEEE International Conference on Pattern Recognition Applications and Methods*, 210–218.