A Hybrid Feedback Control Model
for a Gesture-based Pointing Interface System

Kazuaki Kondo, Genki Mizuno and Yuichi Nakamura

Academic Center for Computing and Media Studies, Kyoto University, Yoshida Honmachi, Kyoto, Japan

Keywords: Human-in-the-loop System, Pointing Behavior, Feedback Control Theory.

Abstract: This study proposes a mathematical model of a pointing interface system that includes indicator behaviors and pointer visualization based on the measurement of pointing postures. The key idea for simulating pointing behaviors under measurement noise involves constructing a hybrid feedback control model. It switches a target value follow-up phase to a noise compensate phase at a specific elapsed time. We evaluated its performance in terms of a simulation error with given step inputs as target values and random-walk noise sequences. The results indicate that the proposed hybrid model simulates the actual pointing trajectories within several centimeter errors.

1 INTRODUCTION

In daily communication, "pointing pose" is typically used as an intuitive way to indicate target objects, locations, directions, and areas. However, the pointing pose could be potentially ambiguous for audience such as when a questioner at a distance from a wide screen points out a specific location on it. In these type of cases, a pointing stick or a laser pointer is generally used to clearly indicate the target. The purpose of our study is constructing a pointing support interface as opposed to the additional equipments. Figure 1 illustrates an example of a construction that uses visual sensing. It recognizes an indicator's pointing posture, estimates a target location on the screen, and shows a visual pointer at the estimated location.

This study proposes a mathematical model of the pointing interface system based on feedback control theory including sensors, computers, visual devices, and an indicator. It allow us simulation-based performance evaluation and interface design. Once a mathematical model of a target pointing interface system is established, its behaviors under various pointing situations can be simulated and then its usability is also evaluated. It aids in designing the interface by employing a trial-and-error strategy without any experimental evaluations in the real world. This advantage is extremely important for human inclusive system because many real evaluations with participants require immense effort and involve difficulties in configuring the same experimental conditions. Fur-

thermore, repeatability of human behavior is not very high, and thus each participant must constantly repeat pointing under the same conditions to collect a sufficient number of samples from which general and essential analysis/evaluations are obtained.

A key idea for modeling involves assuming that an indicator switches a pointing strategy from approaching to the target location to attempting to maintain a pointer in its neighborhood. These behaviors are formulated with two feedback control phases, namely a target value follow-up phase and a noise compensate phase. The characteristics of the two phases are represented with different parameter sets in the same formulation with an assumption that the essential body control schemes in those phases are same. This type of hybrid framework is a first trial in modeling pointing interface systems that were investigated. Thus in this paper, we evaluate the performance of the proposed model in accuracy of pointing trajectory simulation.

2 RELATED WORKS

Easy-to-use interfaces should suit human perception and behaviors. Specifically, transient behaviors from the beginning to end of pointing were analyzed by several extant studies. Woodworth et. al. proposed a pointing action model by using a combination of feed-forward motions for a rapid approach to a target position, and subsequent feedback adjustments
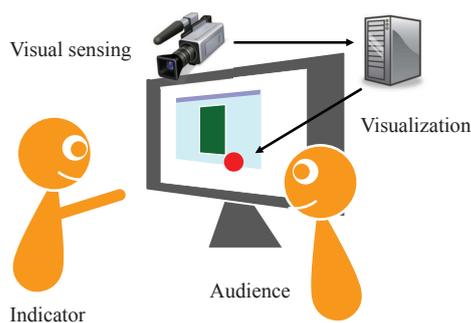
Figure 1: An overview of a vision-based pointing interface. The vision sensor measures indicator's pointing posture and estimates the indicated location to show a pointer.

(R.S.Woodworth, 1899). Fitts et. al. reported that pointing duration depends on a distance and a target size(Fitts, 1954). This fact is known as "Fitts's law", which has been used in many studies because of its accurate approximation in various conditions. McGuffin and Balakrishnan proposed zooming the region around a pointer (McGuffin and Balakrishnan, 2005). This gives a perceptual illusion that the target size and its distance from a pointer appear larger than the actual amounts in neighborhood of a destination. In order to obtain a similar effect, several previous studies also controlled pointer size and/or its speed (Worden et al., 1997; Grossman and Balakrishnan, 2005; Blanch et al., 2004). However, those conventional methods were designed for mouse interfaces and visualization on laptop monitors. Hence, it is necessary to apply those scheme with respect to the design of gesture-based pointing interfaces for a distant and/or large screen.

A particular problem of the present gesture-based pointing interface system is difficulty in estimating the location that an indicator wants to point. One reason is accuracy of pointing pose measurement. In the pointing interface system, significant measurement accuracy is required especially for a target at a distance from an indicator because even extremely small errors on body coordinates are amplified on a screen. Although various type of motion capture techniques developed with acceleration sensors (Slyper and Hodgins, 2008), magnetic sensors (O'Brien et al., 2000), or visual markers (Loper et al., 2014) could satisfy the accuracy requirement, they are not appropriate in daily pointing situations. Several previous studies proposed markerless methods based on visual sensing that do not interfere with an indicator's behavior (Shotton et al., 2011; Yoshimoto and Nakamura, 2015; Nickel and Stiefelhagen, 2003). However the performance of these methods does not satisfy the requirement because visual sensing basically

corresponds to semi-2D measurement, and thus it is weak in terms of self-occlusion and non-rigid deformation by clothing.

Another problem in estimating pointed position is body pose ambiguity. It corresponds to inconsistency between 3D pose structure and a pointed location in indicator's intent. Although previous studies investigated the relationship between a pointing posture and an indicated position, it still be uncertain and influenced substantially by pointing conditions. Additionally each person possesses individual characteristics in a pointing pose. Initially, Fukumoto et. al. reported that a target position is placed on an indicating vector defined by a fingertip and a reference point inside an indicator's body (Fukumoto et al., 1994). The reference point moves according to the pointing pose. It is placed on an eye position when an indicator's arm is straight while resting on an elbow when it is bent. Kondo et. al. reported on selecting a suitable model for intermediate poses with slightly bent elbows (Kondo et al., 2016). Ueno et. al. indicated that despite a straight arm, an intended location does not stay on the indication vector. Although calibrating the amount of the disparity as horizontal and vertical offsets for targets at various locations (Ueno et al., 2014), they did not propose its general model.

Extant studies do not discuss the mathematical model that assumes estimation errors on pointed locations caused by the above problems. Although Kondo et. al. proposed a feedback control model for pointing interface system (Kondo et al., 2015), it focused on pure step responses under no or sufficiently small noises.

3 CONTROL MODEL OF POINTING INTERFACE SYSTEM

3.1 Fundamental Formulation

With a vision-based pointing interface system as shwon in Fig. 1, a pointing procedure is modeled as a feedback loop as shown in Fig. 2. Based on classical control theory, Kondo et. al. formulated it with four transfer functions H_g, H_p, H_s and H_v to denote indicator's body dynamics, visual perception catching a pointer location, computer estimating pointed position, and visualization filter, respectively (Kondo et al., 2015). However they did not consider the estimation error caused by the visual sensing accuracy and the pointing posture ambiguity. Our study assumes that H_s correctly estimates the pointed position
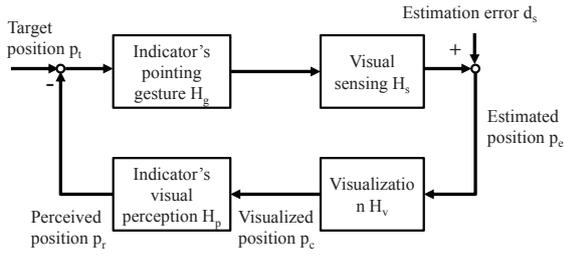
Figure 2: The feedback loop scheme of the vision-based pointing interface system. The modules H_g, H_p, H_s, and H_v describe the indicator's body dynamics, visual perception, estimation of the pointing location, and pointer visualization. Additionally d_s refers to disturbance given to the pointed position estimation.

and its estimation error is inserted as unexpected incoming noise d_s. Additionally no visualization filter is assumed to model a simple pointing interface system. Therefore the transfer functions H_s and H_v consisting of the pointing interface side are formulated as follows :

$$
\begin{aligned}
H_s(s) &= e^{-\tau_s s} \\
H_v(s) &= e^{-\tau_v s}
\end{aligned}
\tag{1}
$$

where τ_s and τ_v denote the latencies for the pointed position estimation, and the pointer visualization, respectively.

The pointing dynamics H_g is assumed as a second-order lag element

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

based on the preliminary analysis of pure step responses, where τ_g denotes the dead time to begin a pointing action. The human visual perception H_p is simply formulated as a first-order lag element

$$
H_p(s) = \frac{K_p}{T_p s + 1}
\tag{3}
$$

to simulate a cognitive delay.

3.2 Hybrid Feedback Model

Let the output of H_s be a control value denoted as p_s, because an indicator can not know amount of incoming noise and thus tries to make p_s closer to p_t.

Given four transfer functions and input signals $p_t(t), d_s(t)$, the control value p_s consists of two terms of a target value response and a noise response:

$$
P_s(s) = P_t(s)G_{st}(s) + D_s(s)G_{sd}(s)
\tag{4}
$$

with an assumption of their linear independency. $P_s(s)$, $P_t(s)$, and $D_s(s)$ correspond frequency domain descriptions of $p_s(t)$, $p_t(t)$, $d_s(t)$, respectively, with

Laplace transform. $G_{st}(s)$ and $G_{sd}(s)$ are described as

$$
\begin{aligned}
G_{st}(s) &= \frac{H_g(s)H_s(s)}{1 + H_g(s)H_s(s)H_v(s)H_p(s)} \\
G_{sd}(s) &= -\frac{H_g(s)H_s(s)H_v(s)H_p(s)}{1 + H_g(s)H_s(s)H_v(s)H_p(s)}.
\end{aligned}
\tag{5}
$$

based on a closed loop theorem.

Pointing behaviors is also influenced by values of the parameters $K_g, T_g, T_p, \zeta, \tau_s, \tau_v$ and τ_g. These should be configured so that Eq. (4) well simulate actual pointing behaviors. We have an important constraint for the parameter configuration. With enough small or ignorable noise, a trajectory of pointed positions generally converges to a target position after sufficient time goes. Additionally p_t can be given as a step signal because an indicator usually configures a target position at a distant from the initial position and it remains for a certain duration. Thus the convergence behavior corresponds to a mathematical constraint that stationary error $e(t) = p_t(t) - p_s(t)$ at $t = \inf$ must be 0 under the condition of $p_t(t)$ being a step signal and $d_s = 0$. This can be formulated as

$$
\frac{p_s(t)}{p_t(t)}\Big|_{t=\infty} = \lim_{s \to 0} s \cdot \frac{1}{s} G_{st}(s) = 1
\tag{6}
$$

based on the final-value theorem. It results in

$$
K_p = \frac{K_g - 1}{K_g}
\tag{7}
$$

where K_g corresponds to the gain of H_g. Here K_g should not be extremely high when considering human body dynamics. Specifically, the previous study reported that estimated values of K_g are slightly exceed 1 (Kondo et al., 2015). This fact and Eq. (7) lead to a significantly small feedback gain K_p. Thus a feedback effect to compensate noise influence is also small. Essentially, it is difficult for a simple feedback model to simultaneously satisfy precise tracking to a target value and compensating noise influence. Thus a more advanced scheme is necessary to construct a compatible model.

In order to solve this issue, this study focuses on a transition of control strategy as reported in (R.S.Woodworth, 1899) ; namely human pointing behavior switches from a rapid approach at the beginning to a subsequent adjustment. Similarly the proposed model switches the control model from a target value follow-up phase to a noise compensation phase at a specific time T. Given this hybrid framework, the control value $p_s(t)$ is formulated as

$$
\begin{cases}
\mathcal{L}^{-1}(P_t(s)G_{st}(s)) & t < T \\
\mathcal{L}^{-1}(P_t(s)G_{st}(s))|_{t=T} + \mathcal{L}^{-1}(D_s(s)G_{sd}(s)) & t \geq T
\end{cases}
\tag{8}
$$

and the control strategy $G_{st}(s)$ for a target value input and $G_{sd}(s)$ for an incoming noise signal are obtained from the same feedback scheme shown in Fig. 2 but with individual parameter sets $\phi^{\mathbf{t}} = [K_g^t, T_g^t, T_p^t, \zeta^t, \tau_s, \tau_v, \tau_g^t]$, $\phi^{\mathbf{d}} = [K_g^d, K_p^d, T_g^d, T_p^d, \zeta^d, \tau_s, \tau_v, \tau_g^d]$, respectively. Because even the different pointing strategy appears to follow the same body control scheme and visual perception rule. K_p^t is automatically determined from the values of K_g^t by using Eq. (7) while it is not applied to $\phi^{\mathbf{d}}$. The reason is that the incoming noise is a probabilistic sequence and thus it is less meaningful to discuss convergence to that. τ_s and τ_v are assumed as common because the parameters of artificial components do not depend on the human pointing strategy.

4 EXPERIMENTS

4.1 System Implementation

A pointing interface system was implemented as shown in Fig. 3 for experimental evaluation of the proposed method. The $1280\,pixel \times 800\,pixel$ visual contents are projected on a screen (white wall in the figure) approximately corresponding to $2m \times 1m$ by using a short focal length projector RICOH PJ WX4141. Subjects stand at a distance of approximately $2.5m$ from the screen. For natural pointing, a contactless visual sensing should be used. A Kinect v2 sensor is placed close to the screen at the left of the subjects. However in the following experiments, the magnetic field-based 3D pose sensor POLHEMUS Liberty is used to accurately measure pointing postures. The magnetic sensors are attached on finger tips of subject's right hands and temples. Although this implementation is not allowed in practical use, it can be accepted for analyzing human pointing behaviors.

The subjects are requested to straighten their arms during pointing behaviors. Therefore the indicated location is estimated as where the indicating vector that connects the center of the temples coordinates and the finger tip across to the screen, based on a previous report (Fukumoto et al., 1994). This corresponds to the processing of H_s. Then an approximately $1cm$ circular pointer that is assumed as a "point" for the subjects is drawn at the estimated position on the screen.

4.2 Model Calibration

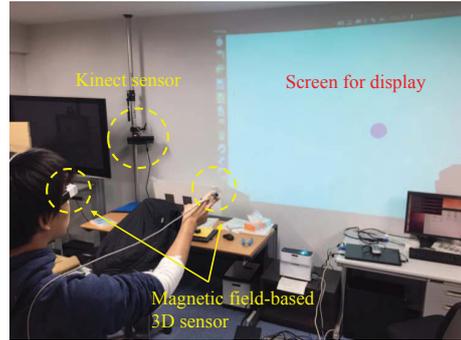It is necessary to determine the parameters used in each pointing phase prior to evaluating performance

Figure 3: The experimental environment.

of the proposed model. In order to simplify a calibration, we assumed that both of the parameter sets $\phi^{\mathbf{t}}$ and $\phi^{\mathbf{d}}$ are independent for each other and do not change. It allows to individually calibrate them beforehand.

The latency parameters τ_s and τ_v included in H_s, H_v, respectively, are assumed as inherent and stable features. Thus they are estimated in advance from directly measured durations of sensing, calculation and display. The remaining parameters are estimated via a non-linear optimization that minimizes residuals between $\hat{p_s}(t)$ from the model and the actually measured $p_s(t)$.

4.2.1 Target Value Follow-up Phase

A step input signal and its response are used to calibrate G_{st} that describes the target follow-up phase. They correspond to the pointing target that suddenly arises at a distance from the initial location and the transient trajectory till an indicator complete pointing the target, respectively. Given this consideration, the experimental procedures for step response measurement are configured as follows.

1. The measurement starts when a subject indicates an initial target visualized on the screen and the pointer remains in that location.

2. The initial target suddenly disappears. Simultaneously, a new target pops up at a distance of approximately 70 cm 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.

A sufficiently small noise $d_s(t) = 0$ was expected because of the measurement accuracy of the magnetic sensor and the straight arm condition. Thus it was assumed the estimated values of pointed positions corresponded to $p_s(t)$. The subjects include three univer-
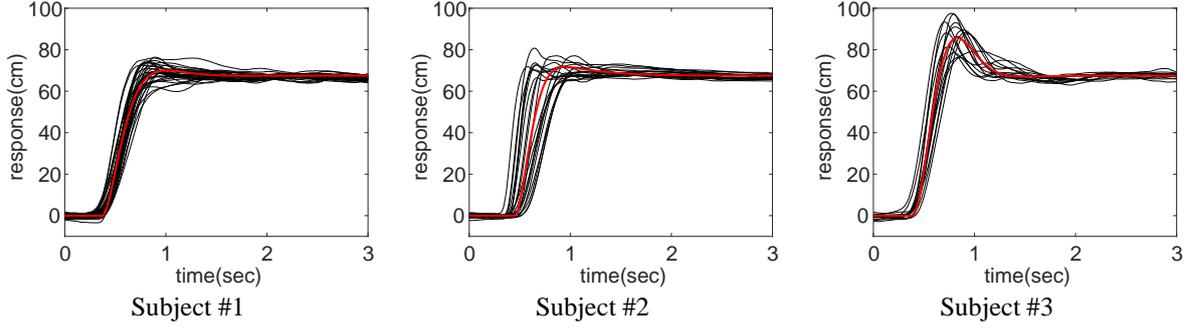
Figure 4: The calibration results of G_{st}. The horizontal and vertical axes denote the elapsed time from when the new pointing target appears and the distance from the initial location, respectively. (black) the measured step response trajectories. (red) the trajectories simulated by G_{st} with the optimized parameter sets $\phi^{\mathbf{t}}$ shown in Table 1.

Table 1: The optimized parameter set $\phi^{\mathbf{t}}$ for each subject.

Subject	K_g^t	τ_g^t	T_p^t	K_p^t	T_g^t	ζ^t
#1	1.02	0.37	0.12	0.02	0.12	0.81
#2	1.04	0.45	0.36	0.03	0.10	0.79
#3	1.08	0.39	0.26	0.07	0.12	0.49

Table 2: The optimized parameter set $\phi^{\mathbf{d}}$ for each subject.

Subject	K_g^d	τ_g^d	T_p^d	K_p^d	T_g^d	ζ^d
#1	1.73	400	14.5	1.69	317	3.22
#2	1.73	403	18.3	1.63	296	14.2
#3	1.73	403	14.2	1.69	288	2.32

sity students. Each of them conducted $N = 20$ trials under the same conditions.

The parameter optimization for the calibration is formulated as

$$\hat{\phi}^{\mathbf{t}} = argmin \sum_t \left(p_s(t) - \mathcal{L}^{-1} \left(\frac{1}{s} G_{st}(s, \phi^{\mathbf{t}}) \right) \right)^2 \quad (9)$$

that means minimizing sum of squared errors with regard to the response trajectory. A trust region nonlinear optimization method is applied to solve Eq. (9). The upper and lower limits of the parameters were configured as $1 < K_g \le 2$, $1 \le T_g \le 10^4$, $0 < \zeta \le 50$, $300 \le \tau_g \le 700$, $0 \le T_p \le 10^4$ by considering of human body dynamics and visual perception. The optimal parameter set that minimizes Eq. (9) is selected from the results of multiple optimizations beginning with various initial values so that at least few of them reached the global minimum. With this procedure, an optimal parameter set $\phi^{\mathbf{t}}(i, j)$ for each pointing trial i of each subject j is determined. After that, the final parameter set for each subject is estimated as their average values $\frac{1}{N} \sum_i \phi^{\mathbf{t}}(i, j)$. A direct optimization of the common parameter set for all trials is simpler than the above two step calibration. However it estimates a parameter set that averages dispersed pointing trajectories in a real domain but not in a frequency domain. The two step calibration averages the parameter sets that reflect the system behaviors in both the real and frequency domains, and thus it appears to maintain essential characteristics.

The black and red curves in each figures in Fig. 4 show the trajectories of each subject's 20 pointing trials and their simulated trajectories with the calibrated

parameter sets as shown in Table 1, respectively. Even the 20 samples under the same experimental condition had certain diversity because of unconscious pointing posture perturbation. The calibrated models approximated almost the center of the diversities. The amount of the diversity depends on the subject. The pointing trajectories of subjects #1 and #3 appear to be adequately simulated while those of subject #2 are not so much.

4.2.2 Noise Compensation Phase

Similarly the parameter set $\phi^{\mathbf{d}}$ of G_{sd} for the noise compensation phase is estimated by using the pointer location stabilizing behaviors against to given noises. The experimental procedures for the measurement are configured as follows :

1. The measurement begins when a subject indicates a stable target visualized on the screen and the pointer stays at that location.

2. A noise sequence is added to the measured pointing location $p_s(t)$. The subject changes its own pointing posture to maintain the pointer on the stable target.

3. The measurement stops after a certain experimental duration $T_d = 4.0$ seconds.

The pointing target does not move, and thus only the responses to given noises are appeared. A vision-based pose measurement usually searches a neighborhood of that in the previous frame. It corresponds to that the estimation error is accumulated. Thus in this experiment a random walk sequence is assumed as the given noise. It is formulated as
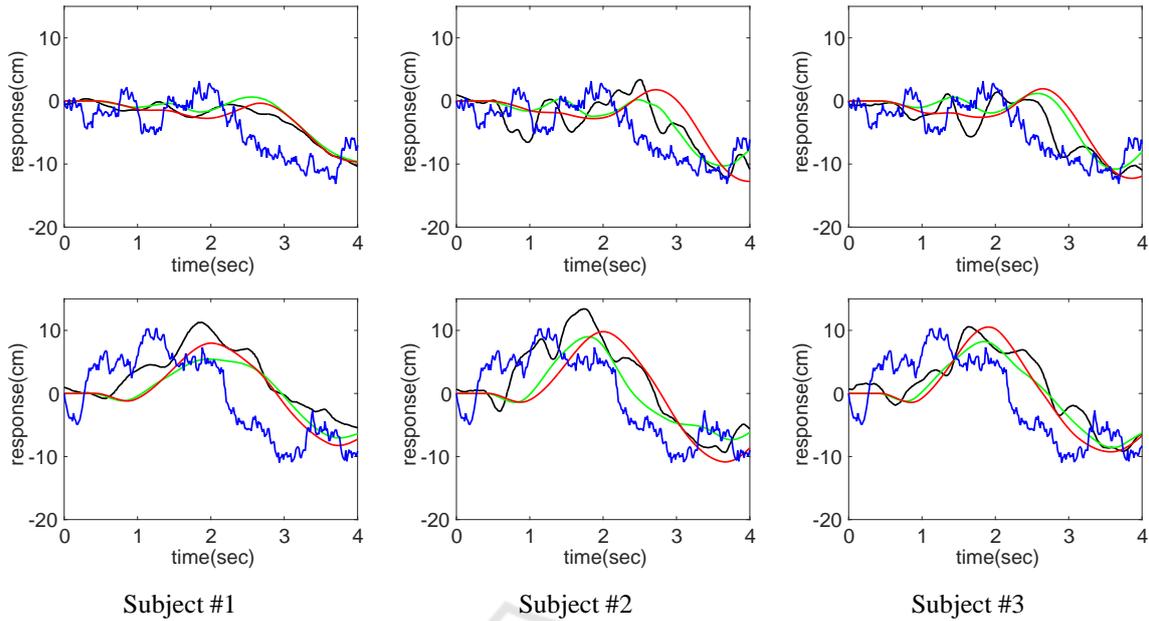
Figure 5: The calibration results of G_{sd}. The figures in each column correspond to results with respect to two representative trials of each subject. (blue) given random-walk noise sequences. (black) the measured noise compensation trajectories. (green) the trajectories simulated by G_{sd} with the optimized parameter sets $\phi^{\mathbf{d}}(\mathbf{i,j})$ for each pointing trial i and subject j. (red) those with the final parameter sets $\phi^{\mathbf{d}}(\mathbf{j}) = \frac{1}{N}\sum_i \phi^{\mathbf{t}}(i,j)$ shown in Table 2.

$$n_r(t+1) = n_r(t) + sx, \quad x \sim \mathcal{N}(0,1) \qquad (10)$$

with an update factor x that follows an independent and identical standard normal distribution where $n_r(0) = 0, t < T_d$, and $s = 2.0$ cm were configured. In the experiment, 20 different random walk sequences $\{n_r(t)\}$ were given to each subject. The manner to acquire an optimal parameter set is the same as that in the previous calibration experiment.

Figure 5 shows the calibration results of each subject. The blue, black, green, and red curves in the figures denote the given random-walk noises, response trajectories, simulated trajectories with the individually optimized parameter sets for each trial, and those with the final parameter sets as the results of the two step estimation, respectively. A high frequency component filtering with particular latency appears in a translation from the given noise trajectories to their responses. Simultaneously we can see that small disparities remain between the given noise sequences and their compensation response. Our hypothesis is that human pointing dynamics is not so fast to follow high frequent changes. Thus too much trial to do so may result in worse. In stead of that, the subjects appeared to allow the certain amount of remaining gap between the shown pointer and the target position. This non-linear characteristic is a possible reason of not so good approximation performance shown as the inconsistency in the black and green (or red) curves in Fig. 5.

While the proposed G_{sd} roughly simulate the noise responses, it does not adequately explain relatively high frequent components. The proposed model consists of a tamdemly connected 2nd-order and 1st-order lag elements, and thus it is difficult to explain the non-linear characteristics. A visual perception component that is insensible to small position disparities be necessary.

4.3 Model Evaluation

The performance of the proposed hybrid model was evaluated in the situations such as when an indicator attempts to move a pointer under a certain noise. A method to evaluate is analyzing how it approximates actual pointing trajectories.

In a manner similar to the calibration experiments, $p_t(t)$ and $d_s(t)$ given to the subjects were configured as step inputs and random-walk sequences. The way to measure the response trajectories $p_s^{gt}(t)$ as ground truth was almost same as that shown in section 4.2.1 with the exception of random-walk noises being added during pointing. Their simulated trajectories $p_s^{prop}(t)$ were generated by the proposed hybrid feedback model Eq. (8) by using G_{st}, G_{sd} calibrated in the previous two experiments and the same $p_t(t)$ and $d_s(t)$ given to the subjects. The model switching time T was manually configured to 1.2 seconds (just after the overshoot in most cases). The reference
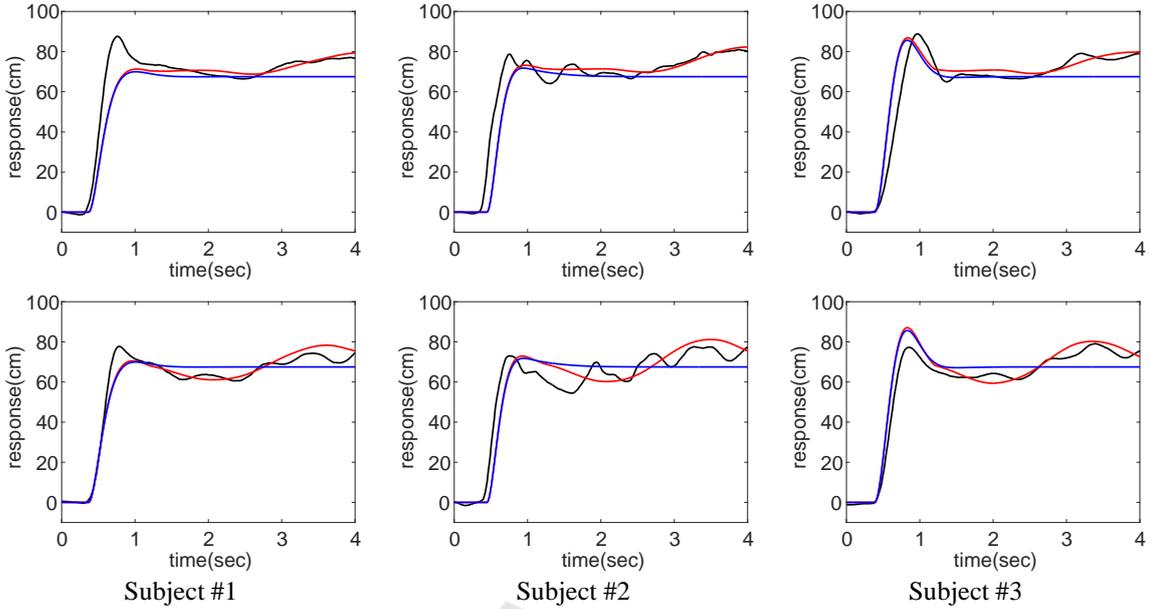
Figure 6: The predicted pointing trajectory by simulation. The figures in each column correspond to results with respect to two representative trials of each subject. The given random-walk noise sequences are different from those used in the calibration phase. (black) the ground-truth trajectories $p_s^{gt}(t)$. (red) the simulated trajectories $p_s^{prop}(t)$ by using the hybrid model. (blue) those by using reference model.

trajectories $p_s^{ref}(t)$ for comparison were generated by using the common parameter set ϕ^t for both of G_{st} and G_{sd}. This assumed the conventional non-hybrid model that does not consider the influence of the estimation error.

The predction results by simulation of each subject are shown in Fig. 6. The proposed hybrid model begins to compensate for the noise influence after $t = T$, and thus the trajectories $p_s^{prop}(t)$ more adequately approximate the ground truth $p_s^{gt}(t)$. However the fact that the noise compensation behaviors can not be well approximated as shown in the calibration results appeared also in this evaluation situation. Another reason for the remaining errors corresponds to the influence of the phase switching time T. In this moment, a method to determine T is a quite naive way and it should be automatically configured based on model parameters and input signals. Thus the influence of T appears to be approximated to the stationary bias. This indicates that the trajectories simply go up or down in the time-location coordinate according to T. A total comparison in statistic performance of the simulation errors $e^{prop}(t) = p_s^{prop}(t) - p_s^{gt}(t)$, $e^{ref}(t) = p_s^{ref}(t) - p_s^{gt}(t)$ are shown in Table 3. The simulation accuracies of the proposed method are within several centimeters. It is approximately half of those with respect to the reference method.

The proposed hybrid model mainly focuses on compensating noise influence at a stationary state be-

Table 3: The statistic values of the simulation accuracy $|e^{prop}(t)|$ and $|e^{ref}(t)|$ (cm).

Subject	$|e^{prop}(t)|$		$|e^{ref}(t)|$	
	ave.	s. d.	ave.	s. d.
#1	3.87	4.05	8.12	5.74
#2	5.15	5.23	8.87	6.63
#3	4.29	3.99	8.01	5.92

cause it is expected that a little perturbation does not exhibit a significant influence during a transient period. However the simulation accuracy at $t < T$ is still not high. This confirmed that it is also necessary to compensate for a relatively low amount of disturbance at the beginning of pointing. In actual pointing situations, pointing often starts by holding indicator's hand hanged down. In this case, large amount of noise may be caused even in a transient period. An indicator does not attempt to cancel such type of large noise and instead configure a new target value from the current pointer location to the original target location.

5 CONCLUSION

This study involved proposing a mathematical model of a pointing interface system to simulate its behaviors with respect to step signal targets and pointed position estimation errors. The proposed hybrid feedback control model switches its characteristics from a

target value follow-up phase to a noise compensation phase. The simulation accuracy becames fairly higher than that of the conventional non-hybrid model. Nevertheless a large amount of simulation error at a transient period continues to constitute a considerable problem.

In the experiments, a very simple pointing situation was assumed in which the subjects stand at the same location, point with their arm being straight, and start pointing from previous pointing postures. The travel distance to the new target is also limited to only 70 centimeters. The performance in more various conditions must be evaluated. A further essential issue of human behavior diversity requires a more advanced framework. Additionally in order to really contribute interface design, it is necessary to apply and model various visualization methods such as an area pointer, a blurred pointer, and showing it at a smoothed location.

REFERENCES

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, pages 519–526.

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*, pages 281–290. ACM.

Kondo, K., Mizuno, M., and Nakamura, Y. (2016). Analysis of human pointing behavior in vision-based pointing interface system - difference of two typical pointing styles-. In *In Proc. on The 13th IFAC/IFIP/IFORS/IEA Symposium on Analysis, Design, and Evaluation of Human-Machine Systems 2016*.

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*.

Loper, M., Mahmood, N., and Black, M. J. (2014). Mosh: Motion and shape capture from sparse markers. *ACM Transactions on Graphics*, 33(6):220:1–220:13.

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.

Nickel, K. and Stiefelhagen, R. (2003). Pointing gesture recognition based on 3d-tracking of face, hands and head orientation. In *In Proc. of The IEEE Int. Conf. on Multimodel Interfaces*, pages 140–146.

O'Brien, J. F., Bodenheimer, R., Brostow, G., and Hodgins, J. K. (2000). Automatic joint parameter estimation from magnetic motion capture data. In *Graphics Interface*, pages 53–60.

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, pages 1297–1304.

Slyper, R. and Hodgins, J. K. (2008). Action capture with accelerometers. In *In Proc. of the 2008 ACM SIGGRAPH/Eurographics Symposium on Computer Animation*, SCA '08, pages 193–199.

Ueno, S., Naito, S., and Chen, T. (2014). An efficient method for human pointing estimation for robot interaction. In *In Proc. of IEEE Int. Conf. on Image Processing 2014*, pages 1545–1549.

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*, pages 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 Int. Conf. on Pattern Recognition Applications and Methods*, pages 210–218.