RJ10392 (A0611-003) November 1, 2006
Computer Science

IBM Research Report

Modeling Human Performance of Pen Stroke Gestures

Xiang Cao*, Shumin Zhai
IBM Research Division
Almaden Research Center
650 Harry Road
San Jose, CA 95120-6099

*also with University of Toronto, Toronto, Ontario, Canada

Research Division
Almaden - Austin - Beijing - Haifa - India - T. J. Watson - Tokyo - Zurich

Modeling Human Performance of Pen Stroke Gestures

Xiang Cao[1,2] **Shumin Zhai**[1]
[1]IBM Almaden Research Center
San Jose, CA, USA
zhai@almaden.ibm.com
[2]University of Toronto
Toronto, ON, Canada
caox@cs.toronto.edu

ABSTRACT

Our investigation aims to develop a quantitative human performance model of making single-stroke pen gestures within certain error constraints in terms of production time. Such a model will be useful as a foundation for the design and evaluation of existing and future gesture-based user interfaces at the basic motor control efficiency level, similar to the role of previous "laws of action" played to pointing, crossing or steering-based user interfaces. We report and discuss our experimental results on establishing and validating the model, together with other basic empirical findings in stroke gesture production.

Author Keywords

Pen stroke gestures, experimental study, pen input.

ACM Classification Keywords

H5.2. Information interfaces and presentation (e.g., HCI): User Interfaces, Theory and methods.

INTRODUCTION

Single-stroke pen gestures have been widely used on many pen-based computing devices such as PDAs, tablet PCs or electronic whiteboards. These pen gestures can be used to input text on a letter-level such as Unistrokes [7] and Graffiti, or on a word-level such as in ShapeWriter (also known as SHARK shorthand) [11]. Pen gestures can also be used to trigger system commands and operate applications [8, 14]. With the increasing availability and popularity of pen-based devices, especially mobile devices, it is likely that we will see even more applications of pen gestures in various environments.

In order to enhance the user experience of gesture-based interaction systems, numerous research works have been conducted to improve the performance of gesture recognizers [10, 21] and to design gestures that are easy to learn and remember for users [12]. However, there has been relatively little research on quantitative models of human performance in producing pen gestures that can characterize the efficiency of a given gesture or a gesture set. Such models would help the design and evaluation of existing or future gesture interfaces by quantitatively predicting their efficiency before running extensive user studies.

The existing theoretical tools in user interface design at a motor control level are a set of so-called "laws of action"(See [18] for a brief review). They model human performance on tasks including pointing (Fitts' law [6]), crossing (essentially Fitts' law [2]) and path steering [1, 3, 4]. However, the existing laws only apply to visually guided performance and therefore may not be appropriate to model freehand open-loop stroke gestures.

A highly desirable goal is to extend the family of laws of action to pen stroke gestures. However, freehand gesturing is an inherently complicated behavior that involves planning, chunking, and the variability of behavior across different people and different types of gestures. As we move up the ladder of motor control skills, simple regularities that can be called a law may not exist. The goal of our current work is to build a "computational" model that can predict the production time of single pen-strokes as a function of the stroke's composition, which is still valuable as an approximation tool in analyzing and designing pen-gesture interfaces.

PREVIOUS WORK

Isokoski [9] proposed a model for stroke gestures that used the number of approximating straight line segments in a gesture as a predictor of complexity correlating to production time. The underlying assumption is that drawing a straight line segment takes constant time, regardless of the length of the segment. The model's best correlation result was $R^2 = 0.85$ on Unistroke characters [7], and it achieved R^2 between 0.5 and 0.8 for other gesture sets. This model provides a useful quick prediction, with the attractive merits of simplicity and ease of application. However, defining the number of straight line segments needed to approximate a curved gesture is ambiguous. Furthermore, it does not provide an estimation of the magnitude of the actual production time. While we will maintain the same first order approximation spirit as this model, our goal is to more closely reflect the complexity in the actual strokes.

At a more fundamental motor control theory level, Viviani and colleagues [16, 17] investigated human handwriting and drawing behavior in terms of instant movement velocity as a function of curvature, and proposed a power-formed model. A simple version of their formula is that for a given point on the written/drawn trajectory:

$$V = KR^{\beta} \qquad (1)$$

where V is the instant (tangential) velocity of movement; R is the radius of curvature C (R=1/C), and K and β are constants of the model.

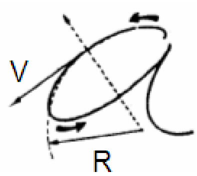


Figure 1. Viviani's power law of curvature.

The model, known as the power law of curvature, indicates that the larger curvature the trajectory has at a given point, the slower the pen motion will be at that point. This model has been tested in different settings, including drawing trajectories with or without visual guidance [17]. We will derive some basic assumptions in our model based on this movement law.

APPROACH

The approach we took was to first find common "building blocks" of pen strokes at an appropriate level. Our model is based on the assumption that any gesture stroke can be decomposed into several "elements", each of which can be modeled by a lower-level model. The total model is represented by integrating the elemental models.

While making this "reductionism" assumption, we recognize the possibility that the interaction between elements and the user's planning as a whole will result in a shorter or longer gesture production time than the simple sum of all elements. However, we hypothesize that the sum of elements may still give a first order approximation or a baseline prediction useful for many applications.

In what follows we first build a set of lower-level models of the common building blocks of stroke gestures, based on information from the existing literature and intuition, to be verified by later experiments.

Smooth Curve

The trajectory of a given gesture G is represented by the parameter equation:

$$G : (x = x(s), y = y(s)) \qquad (2)$$

where

(x,y): coordinates of the current point on the gesture.
s: curve length between the starting point and the current point.
$t = t(s)$: time passed since the starting point until the current point.
$V = V(s) = ds/dt$: instant velocity.
$R = R(s)$: radius of curvature.
S: total curve length of G $(0 \le s \le S)$.
T: total production time of G $(0 \le t \le T)$

Assuming $0<R(s)<\infty$, then according to Viviani's power law of curvature the production time of G is:

$$T = \int_0^T dt = \int_0^S (\frac{dt}{ds})ds = \int_0^S (\frac{1}{V(s)})ds = \int_0^S (\frac{1}{KR(s)^{\beta}})ds = \frac{1}{K}\int_0^S R(s)^{-\beta}\, ds \qquad (3)$$

which gives the theoretical time prediction of a smooth curve gesture articulation $T(curve)$.

A special case of smooth curves is a circular arc with a radius r and a sweep angle α, which has a constant radius of curvature $R(s) = r$ (Figure 2). The total length of the arc is S=αr. The total production time is then:

$$T = \frac{1}{K}\int_0^S R(s)^{-\beta} ds = \frac{1}{K}\int_0^{\alpha r} r^{-\beta} ds = \frac{1}{K}(\alpha r) r^{-\beta} = \frac{\alpha}{K} r^{1-\beta} \qquad (4)$$

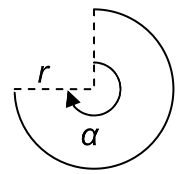


Figure 2. Circular arc.

To analyze the same gesture drawn in different scales (spatial sizes), consider gesture G', which is G spatially scaled to the factor of δ. Then the total curve length of G' is δS, and the radius of curvature for G' becomes $R'(s') = \delta R(s'/\delta)$ $(0 \le s' \le \delta S)$. The total production time of G' is

$$T' = \frac{1}{K}\int_0^{\delta S} R'(s')^{-\beta} ds' = \frac{1}{K}\int_0^{\delta S} \left(\delta R(s'/\delta)\right)^{-\beta} ds' \xrightarrow{let (s = s'/\delta)}$$
$$\frac{1}{K}\int_0^{S} \left(\delta R(s)\right)^{-\beta} (\delta ds) = \delta^{1-\beta}\frac{1}{K}\int_0^{S} R(s)^{-\beta} ds = \delta^{1-\beta} T \qquad (5)$$

Straight Line

For a straight line segment, the curvature is zero anywhere; therefore $R(s) = \infty$. However, V cannot be ∞ in practice.

From previous knowledge and intuition, three plausible candidate models may be proposed for the production time of a straight line. Let L be the length of the line:

(a) *Constant time model*: $T(line) = c$
which is the underlying assumption of Isokoski's model.

(b) *Linear model*: $T(line) = aL + b$
which implies a constant movement velocity $1/a$, plus a constant overhead time b (covering starting and stopping etc.).

(c) *Power model*: $T(line) = mL^n$
this model suggests people tend to move faster with longer lines (but not to the extent that results in constant drawing time), and result in a power-like relationship between production time and length. Later this *power model* proved most valid according to our experimental data.

Although the orientation of a line segment may also affect its production time given the asymmetry of hand and arm anatomy, we hypothesize that such an impact will be relatively small compared to that of the length.

Corner

Pastel [15] studied human performance of steering through paths with corners, and found that people spent more time steering through paths with 90º corners than with paths with 45º and 135º corners. However, since a steering task is fundamentally different from the open-loop movement in making gestures, the same trend may not exist in gesture performance on corners.

Since a corner only exists with the arms that form it, it is difficult to define the operational boundaries of a corner. We therefore define *T*(*corner*) as "the net contribution of the abrupt direction change to the total production time". This value may not necessarily be positive (it is possible that drawing a corner takes less time than drawing 2 separate lines). For generality, we tentatively represent the corner production time as:

$$T(corner) = f(\theta) \qquad (6)$$

f is an arbitrary function of the corner angle θ, to be found through experiments.

We postulate that any given single stroke gesture can be decomposed into these three types of elements: (approximate) straight line segments, corners, and smooth curves, each of which can be modeled by their respective elemental models to be established by experiments. We then put these elements together to form a general summative model of total stroke production time:

$$T = \sum T(line) + \sum T(corner) + \sum T(curve) \qquad (7)$$

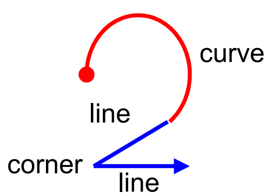


Figure 3. Decomposition of a gesture.

We conducted two experiments to establish the models and test their validity, and to look for other interesting phenomena relevant to stroke gesture design and applications.

APPARUTUS AND EXPERIMENTAL METHOD

All experiments were executed on an Acer TravelMate C110 tablet PC using stylus input. A program was developed for pen gesture study. In each trial, the program displays a sample (target) gesture in the top window, with a red dot and an arrowhead indicating the starting and ending points respectively. The user then draws the prompted gesture with the stylus in the bottom window. Determined by an error metric (described later), the system decides whether the user's input is acceptable. If it is, the original sample is displayed in green, overlaying the user's input to indicate success (Figure 4). Otherwise, it is displayed in red, and the user must repeat the input until it is accepted. This overlaying feedback helps the user to determine how good the input is, and which part needs to be improved. The feedback disappears before the user can make a second attempt, in order to prevent the user from visually tracing it.

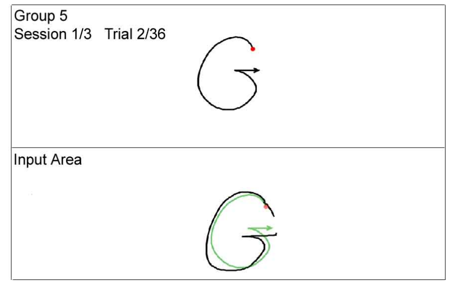


Figure 4. Experiment interface.

The gesture samples presented in the experiments were generated using a semi-automatic authoring program, with which the decomposition of the gestures can be interactively specified.

Unlike previous modeling work on visually guided actions, such as target pointing, goal crossing or path steering with explicitly defined error behavior (on or off target, in or out of path, but see [20] for the complications even in the case of Fitts' law error calculation), the error of a stroke gesture is more ambiguous, subjective (in the case of a human observer), or system dependent (in the case of machine recognition). There are potentially infinite numbers of error metrics that can measure the difference between two instances (the input and the sample) of a stroke gesture. Hence it is difficult, if not impossible, to select the "best" error metric that reflects human's (and recognition system's) sensitivity and habits in all situations. Our tactic in this regard is to define an operational procedure that can capture the essence of holding all gesture strokes to the same error tolerance criterion. We aim to select an error metric that is intuitive, relatively simple (hence more general), and insensitive to scale. In our procedure the input gesture U and the sample gesture V are first normalized by scaling and translating so that both of them have a bounding box with the larger side measuring 1 length unit, and centered at (0,0). Then both U and V are re-sampled into M (=25) equidistant sample points. Let u_i and v_i ($1 \le i \le M$) denote the sample points for U and V respectively, then the error metric E is defined as

$$E = \frac{1}{M}\sum_{i=1}^{M}\left\| u_i - v_i \right\|_2 \qquad (8)$$

which is the average spatial distance of all corresponding pairs of sample points in two gestures. Since all gestures are normalized in size, scaling the gestures does not affect E. E effectively measures the "relative" or "percentage" shape distance between the input and the sample. This error metric has been previously used as the foundation of gesture recognition in ShapeWriter (also known as SHARK shorthand) [11], which is one type of application this current work attempts to address.

In our experiments, the acceptance threshold for E was set at 13.5% of the sample stroke length. Since spatial scaling may also affect the production time of a gesture, we also required the size of the input gesture within 1/1.3 ~ 1.3 times of the sample gesture to be accepted. These thresholds were selected based on pilot trials.

For each successful trial, we recorded the production time T (from the time that the stylus touches the tablet to the time the stylus leaves it), percentage error E, and the number of attempts (denoted A hereafter) made until accepted by the system.

EXPERIMENT 1: BUILDING & VERIFYING THE MODEL

Goal

In this first experiment, we sought to establish the forms and estimated parameters for the elemental models, as well as to verify the validity of the summative model.

Design

The experiment consisted of 5 categories of pen gestures, each tested in a separate phase of the experiment, as shown in Figure 5. The first three phases of the experiment used elemental gestures to determine the elemental models, and the last two used composite gestures to test the validity of the summative model (derived from the first three phases).

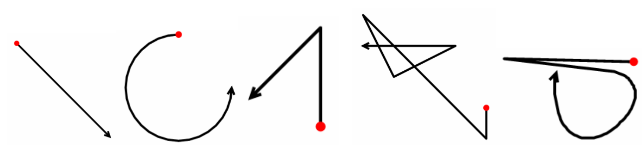


Figure 5. Gesture categories (phases).

Straight Line

To establish the $T(line)$ model, we tested straight lines of 6 lengths (L): 15, 30, 45, 60, 75, 90mm, and 8 orientations: 0º, 45º, 90º, 135º, 180º, 225º, 270º, 315º CCW.

Arc

To establish the $T(curve)$ model, we tested circular arcs of 3 radiuses (r): 12, 24, 36mm, 4 different sweep angles (α): 90º, 180º, 270º, 360º, 2 different start angles (the position angle at which the stroke starts from): 90º, 180º CCW, and 2 directions: CW, CCW. We sought to find the parameters β and K in equation (3) by regression on equation (4) with the experimental data on arcs.

Corner

To establish the $T(corner)$ model, we tested corners with equal arm lengths of 30mm, and 8 corner angles (θ): 0º, 22.5º, 45º, 67.5º, 90º, 112.5º, 135º, 157.5º, 4 start angles (the position angle of the starting arm): 0º, 90º, 180º, 270º CCW, and 2 directions: CW, CCW.

Polyline

Polylines are gestures that consist of only connected straight line segments. This is the type of gesture used in the ShapeWriter text input system [11]. Its gesture prototypes consist of straight lines connecting letters on a soft keyboard. We selected 36 representative gestures from the most commonly used words in the ShapeWriter system as the samples in this category. Out of these 36 samples, 6 of them consisted of 1 line, 6 consisted of 2 lines, …, 6 consisted of 6 lines. The sizes of (the bounding boxes of) the gestures varied between 40 ~ 80mm.

Arbitrary Gesture

This category contains gestures without particular constraints (i.e. the most general). 36 samples were selected from the Graffiti gesture set, including representatives of letters, numbers, and functional gestures. However, in order to reduce the familiarity to the user that may bias the result, all gestures were rotated by 90º. The sizes of (the bounding box of) the gestures were set to be roughly equal (~50 mm).

A fully crossed within-participant factorial design was used for each phase. Each experiment phase was executed in 3 repeated blocks. Within each block the participant conducted one trial for each sample (of length, orientation etc). The presentation order of the samples was randomized in each block. Ten practice trials were performed before each phase started.

Ten right-handed volunteers, aged 20~59, participated in the experiment.

Results

Straight Line

Both length ($F_{5,45}$=29.7, p<.001) and orientation ($F_{7,63}$=7.13, p<.001) had significant effects on the production time T of straight lines. No significant interaction between length and orientation was found ($F_{35,315}$= 1.37, p=.085). Therefore the *Constant time model* for $T(line)$ is clearly invalidated.

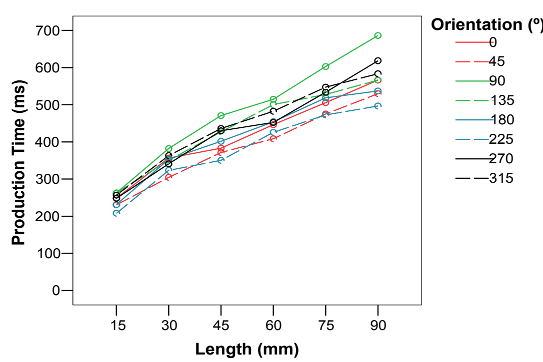


Figure 6. Straight line production time.

The effect of orientation on T is more clearly illustrated in Figure 7. 45º and 225º lines took less time than lines in other orientations. For the right-handed participants 45º and 225º lines can be achieved by wrist rotations. 90º (upward) lines took the most time.

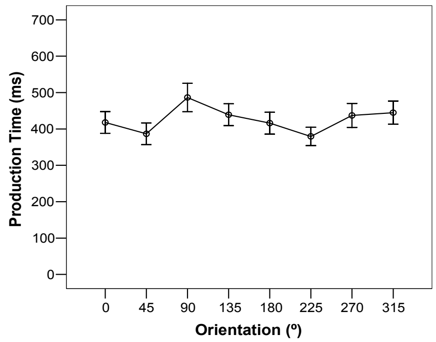


Figure 7. Orientation effect on production time.

Although orientation did affect *T*, the size of the effect was small compared to that of length. Therefore, we felt it safe to only account for length in our *T*(*line*) model.

To compare the *Linear model* and the *Power model*, we performed regression for both models with the experimental data.

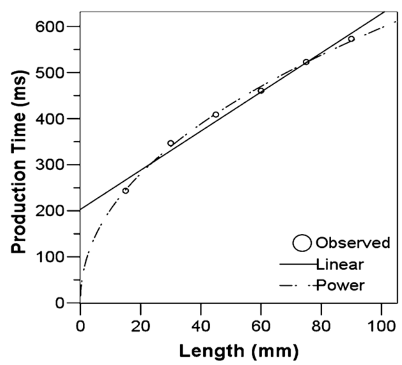


Figure 8. *T*(*line*) model: Linear vs. Power.

Linear Model:

$$T(line) = 4.24\,L + 203 \qquad (R^2 = 0.984) \qquad (9)$$

Power Model:

$$T(line) = 68.8\,L^{0.469} \qquad (R^2 = 0.998) \qquad (10)$$

(L in mm, T in ms)

Note that both models have 2 degrees of freedom; therefore any difference between their fitness with experimental data should not be due to model complexity difference.

The *Power model* best describes the relationship between length and production time. Therefore

$$T(line) = mL^n \qquad (11)$$

is the best validated elemental model for straight lines although the specific parameters *m* and *n* may be subject to individual differences.

Both the number of attempts *A* and percentage error *E* (only for successfully passed trials) are informative measures of input accuracy. Length showed significant effects on both *E* ($F_{5,45}$=25.2, p<.001) and *A* ($F_{5,45}$=22.4, p<.001). Both *E* and *A* decreased as the length increased, which suggests it is harder to maintain the same relative accuracy with smaller-sized gesture (illustrated in Figure 9).

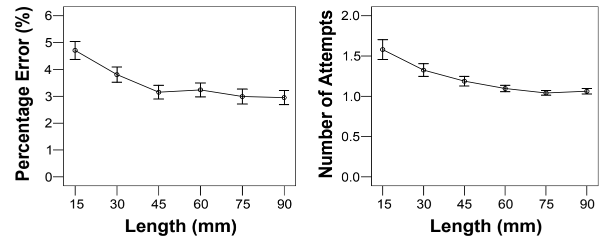


Figure 9. Length effect on input errors.

Orientation also showed significant effects on both *E* ($F_{7,63}$=30.9, p<.001) and *A* ($F_{7,63}$=2.77, p=.014). Movements on the primary axes (horizontal and vertical) tended to be more accurate. Note the different trends of production time and input error as related to orientation. Although the 45º and 225º lines were faster to produce, they were also among the most error-prone compared with other orientations. No significant interaction between length and orientation were observed for either *E* or *A*.

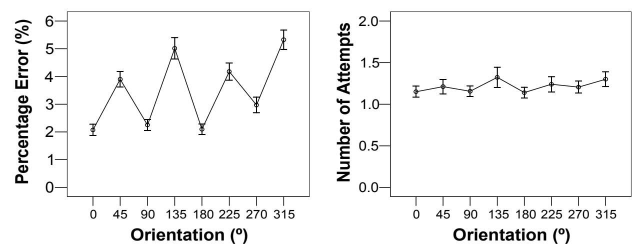


Figure 10. Orientation effect on input errors.

In summary, the length of a straight line was the predominant determinant of the line's production time and accuracy. Within a set accuracy threshold, a power model best described the production time as a function of the length. The orientation of a line also affected time and accuracy, although secondarily. A diagonal line along the wrist rotation direction was somewhat faster to produce, but horizontal and vertical lines were more accurate.

Arc

As we expected based on the power law of curvature derivation and analysis, both radius (r) ($F_{2,18}$=23.5, p<.001) and sweep angle (α) ($F_{3,27}$=23.6, p<.001) had significant effects on production time *T*. The interaction between them was also significant ($F_{6,54}$=13.6, p<.001). Neither start angle nor direction showed significant effects on *T*.

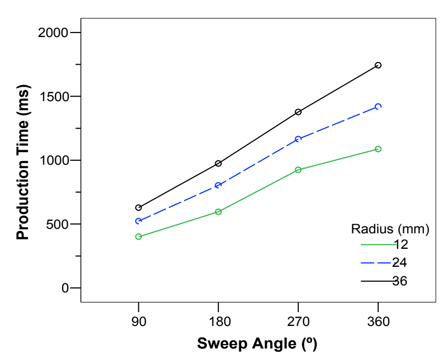


Figure 11. Arc production time.

The regression according to equation (4) gave:

$$T(arc) = \alpha\, r^{1-0.586} / 0.0153 \qquad (R^2 = 0.948) \qquad (12)$$

(α in radians, r in mm, T in ms)

The high regression coefficient validated our derivation of equation (4). Consequently, we obtained $K = 0.0153$, $\beta = 0.586$ for equation (3).

Both radius and sweep angle also had significant effects (p<.001) on E and A. Similar to the straight line cases, both error measures decrease as the radius increases. Arcs with a sweep angle of 180º (semicircle) introduced the least error.

Corner

To compute a corner's "net time contribution" as we proposed previously, we subtracted the time spent on drawing two corner arms (30mm straight lines) from the total production time of the corner gesture. Drawing from the experimental data of the straight lines, the participants spent an average of 370ms on straight lines of 30mm. Therefore T(*corner*) = *sample production time* – 740ms.

The corner angle θ showed a significant effect ($F_{7,63}$= 2.74, p=.015) on T(*corner*) (Figure 12). T(*corner*) fluctuates around zero, with corners of 0º, 22.5º and 90º contributing negative mean time, and all other corners contributing positive mean time. However, the absolute values of these contributions were less than 40ms for all angles ($T(corner)_{mean}$=12.7ms). Given the small value of T(*corner*) compared with other elements, we chose to omit this element in our first-order approximation model.

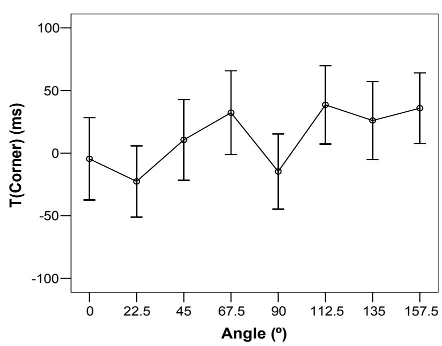


Figure 12. Net corner time contribution.

However, note that omitting T(*corner*) does not mean that the existence of corners has no influence on the total gesture production time. In fact they are critical because of their role in breaking up lines and curves, and causing acceleration and deceleration in movement. For example, drawing two straight lines with length L_1, L_2 connected by a corner do not take the same time as drawing one straight line with length L_1+L_2. The impact of corners is clearly reflected by the power form of the T(*line*) model. Even with the linear model $T(line) = aL + b$, each corner means that an additional constant b (~200ms) is added to the total production time.

The corner angle also had significant effects on both E ($F_{7,63}$= 42.5, p<.001) and A ($F_{7,63}$= 10.6, p<.001). Angles in the middle of the range (around 67.5º) were more error prone than the more extreme angles. Neither start angle nor direction showed significant effects on T(*corner*), E or A.

Summarizing the results from the first 3 phases, we have

$$T = \sum T(line) + \sum T(curve) \qquad (13)$$

$$T(line) = 68.8\, L^{0.469} \text{ (ms)} \qquad (14)$$

$$T(curve) = \frac{1}{0.0153} \int_0^S R(s)^{-0.586} ds \text{ (ms)} \qquad (15)$$

Again the impact of corners is still present in equation (13), although the minor time variance caused by corner angle θ is explicitly omitted.

We will then verify this model with the experimental data in the next two phases with polyline and arbitrary gestures.

Polyline

We verified the proposed model first with the simpler case of polyline gestures, which only involve connected straight line segments. In this case, the model becomes:

$$T = \sum_{i=i}^{N} T(line_i) \qquad (16)$$

where N is the number of line segment in the gesture. Using equation (16) as the summative model and the power equation and parameters obtained in the individual straight line tests (equation (14)) to calculate each element (line segment) of the sample polylines, we computed predictions T_p for the total production time of each gesture sample. This prediction was then compared with the mean value of the actual polyline gesture input time T_d collected in the polyline tests under the same error tolerance criterion. Figure 13 plots the correlation between T_p and T_d. The correlation coefficient $R^2 = 0.960$.

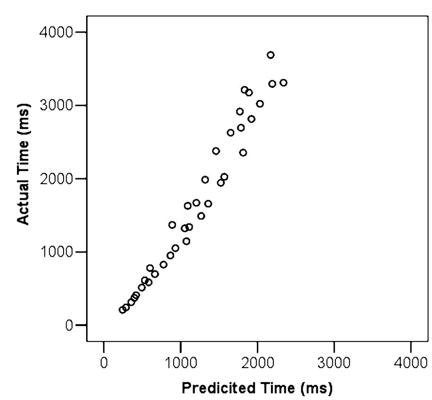


Figure 13. Model prediction for polylines.

Recall that both the number of line segments and the length of each segment influence the total gesture time prediction in our model. In comparison, if we take the total length of the gestures alone as a predictor, the result does not correlate well with T_d ($R^2 = 0.878$).

R^2 value is the validity metric for modeling normally used in testing laws of action – to test how tightly empirical data regress to the model predictions. Such an approach tests the form of a model, but not its specific parameters. The parameters are left floating, and are used as performance indicators of individuals, different muscle groups, devices, or interface methods. By such a metric the validity of the proposed model is quite strong. Next we challenge the model to the next level, to examine the specific value predictions against actual instances of gesture production.

The average value of T_p across all polyline gestures is 1208ms. Compared with the average value of T_d =1684ms, $T_{p\text{-}avg}$ = 0.717 $T_{d\text{-}avg}$. In other words, although on average the magnitude of the prediction is similar to the actual data, it underestimates the participants' production time. Plausibly the longer average actual production time could be caused by cognitive overhead (e.g. pausing or slowing down to plan for the following segments when a polyline gesture has many segments) that are yet to be completely removed by practice. Recall that the experiment only had three blocks of tests and within a block each sample gesture was tested only once. For complex and unfamiliar polyline gestures this might not be enough practice to reach the fluid gesturing behavior that equation (13) intends to model. The more complex a gesture is, the more likely the model would underestimate the actual time when the gesture is still unfamiliar. This could have caused the slight upward bend of the data in Figure 13. We predict that the model's underestimation could be reduced or eliminated (or even become overestimation) if the participants had enough practice with the gesture set. In the following experiment we had the opportunity to test this prediction.

Arbitrary Gesture

Similarly, we verified the proposed model with the arbitrary gestures that consisted of both straight and curved segments. The T_p predictions were computed by equation (13), and verified against T_d, the mean value of the experimental data for each gesture. The integrals for each curve segments were calculated numerically. Figure 14 plots the correlation between T_p and T_d. The correlation coefficient R^2 = 0.914.

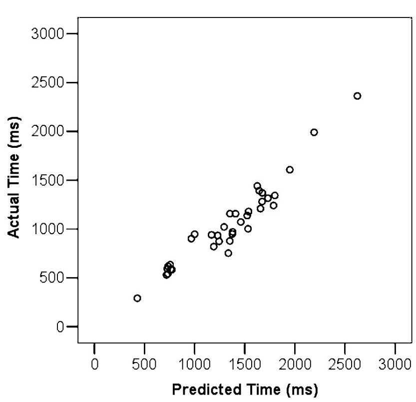


Figure 14. Model prediction for arbitrary gestures.

In contrast to the polyline predictions, the model predictions of the tested arbitrary gestures were longer than actual production time. The average value of T_p across all arbitrary gestures was 1343ms, and the average value of T_d was 1045ms. $T_{p\text{-}avg}$ = 1.285 $T_{d\text{-}avg}$. To clarify, we divided the arbitrary gesture set into those which only consist of straight lines (i.e. "polylines") and those which involve curved segments. Incidentally (and surprisingly) $T_{p\text{-}avg}$ = 1.285 $T_{d\text{-}avg}$ held exactly the same for both subsets. This eliminated the concern that this overestimation of time was caused by the model's bias towards either the straight elements or curved elements. Instead, since the samples used in the arbitrary category were from the Graffiti gesture set, we suspect that the participants' familiarity with Roman letter-like symbols (although rotated) may have helped the participants spend less time producing them. It is conceivable that for a highly familiar gesture, rather than slowing down or pausing to plan for the next move (as in the previous phase), one may chunk the entire gesture (by, for example, cutting corners) and therefore achieve faster speed than what our model predicts.

To summarize, the gesture production time predicted by the proposed model were consistent with the experimental data with the magnitude within ±30% in average, and correlation R^2 > 0.9. Overall, given its high correlation with and the similar magnitude to the experimental data, the proposed model is validated as a good first-order approximation.

EXPERIMENT 2: FURTHER TESTING THE MODEL

Goal

Again, going beyond the correlation between the model and actual human performance, one potential application of the model is to numerically evaluate and compare the efficiency of existing and future gesture set design in terms of expected gesture production time. As an exercise we conducted a small-scale experiment comparing the performance of two pairs of existing gesture sets: Unistrokes vs. Graffiti, and ShapeWriter [11] gestures on two different soft keyboard layouts. We calculated model predictions and collected empirical data for each set of the gestures.

Design

Phase 1: Unistrokes & Graffiti

Unistrokes [7] and Graffiti are two single-stroke character sets. Only the gestures that correspond to letters were tested in this experiment. The sizes of (the bounding box of) the gestures from both sets were set at ~50 mm. There were 52 gesture samples in total, with 26 from each set.

Phase 2: ShapeWriter gestures

We compared two ShapeWriter gesture sets for English words defined on two different keyboard layouts: the traditional QWERTY layout vs. the optimized ATOMIK layout [19] as illustrated in [11]. We selected gestures for the 24 most frequently used words in spoken English (excluding one-letter words, which result in a single-click in ShapeWriter). The gestures for both layouts were generated on soft keyboards whose individual key size was set at 15mm x 15mm.

A within-participant design was used. In each phase gesture samples from the two contrasting sets were mixed and presented in random order. To reach “expert” behavior, each sample was repeated 8 consecutive times, which was in contrast with experiment 1 and gave us a chance to test the prediction that the under-estimation of production time by the model can be reduced by practice. The order of the two phases was counterbalanced among participants. The same error tolerance criterion as in experiment 1 was enforced.

5 volunteers, who were all participants of experiment 1 and still available, participated in experiment 2. This was to ensure that the data from the two experiments were comparable, given the possible large individual differences in the elemental model parameters. We emphasize that running the experiment at such a small-scale was only to empirically test the model’s prediction power for a given group.

Results

We first performed regression on the 5 participants’ data in experiment 1 to derive parameters specifically for them:

$$T(line) = 88.0\,L^{0.394}\ \text{(ms)} \tag{17}$$

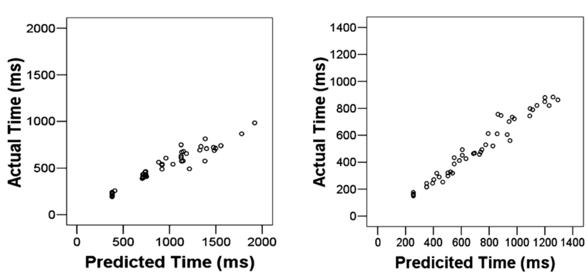
$$T(curve) = \frac{1}{0.0193}\int_0^S R(s)^{-0.566}\,ds\ \ \text{(ms)} \tag{18}$$

Phase 1: Unistrokes & Graffiti

Using the above parameters, we computed time predictions T_p for each gesture sample from Unistrokes and Graffiti. Correlation between T_p and experimental data T_d is plotted in Figure 15(a), with $R^2 = 0.920$.

According to the letter frequencies reported in [13], we calculated the expected gesture production time for both gesture sets:

Model prediction:
Unistrokes : Graffiti = 622 ms : 1125 ms = 0.553 : 1
Experimental data:
Unistrokes : Graffiti = 365 ms : 591 ms = 0.618 : 1



(a) Unistrokes & Graffiti (b) ShapeWriter gestures

Figure 15. Experiment 2 results.

Phase 2: ShapeWriter gestures

Using the same parameters, we computed time predictions T_p for each gesture sample for the ShapeWriter gestures on both keyboard layouts. The correlation between T_p and experimental data T_d is shown in Figure 15(b), with $R^2 = 0.960$.

Weighted by word frequencies estimated from the ANC (http://americannationalcorpus.org/) we calculated the expected gesture production time for the top 24 words on both keyboard layouts.

Model prediction:
ATOMIK : QWERTY = 600 ms : 768 ms = 0.781 : 1
Experimental data:
ATOMIK : QWERTY = 438 ms : 506 ms = 0.865 : 1

This confirmed that ATOMIK is a more efficient layout than QWERTY for ShapeWriter. In addition to the 24 words tested for ShapeWriter in Experiment 2, we can also theoretically estimate an expected production time for the gestures in a complete dictionary, which is difficult to do empirically. Based on 27628 common spoken English words and their frequencies, extracted from the ANC, our model predicts that the expected gesture production time is 903 ms for the ATOMIK layout and 1139 ms for the QWERTY layout. This type of analysis, even theoretically and approximately, had not been possible previously due to the lack of quantitative gesture models.

To compare model predictions with empirical data, we can look at three increasingly stringent indicators. First, the model predictions and empirical data correlated well, with similar R^2 as in experiment 1. Second, between two contrasting designs, the performance order as predicted by the model was always correct and the ratio predicted was similar to that of the empirical data. Third, contrasting with experiment 1, we can see that the difference in production time magnitude between model prediction and empirical data was indeed influenced by practice and previous familiarity. Recall that much more repetition of each gesture was given in this experiment. Consequently, as predicted, the underestimation of time for ShapeWriter gestures in the experiment was indeed eliminated. In fact, the model overestimated the time that the well-practiced ShapeWriter gestures actually took, presumably due to chunking. For the more familiar or simpler Graffiti or Unistrokes characters, more intense practice in this experiment further enlarged the difference between prediction and actual data. We will discuss chunking behavior further in next section.

In summary, the proposed model predicted the comparative performances of different gesture sets reasonably well for the same set of users.

EXPERIMENT 3: SCALING EFFECT

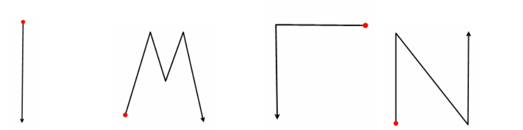
Goal

Equation (5) shows that according to the power law of curvature, the production time of smooth curves is continuously related to the scale the curved gestures are drawn in a nonlinear (power) fashion. The larger a curved stroke is, the longer it takes to articulate such a curve within a set error threshold. The rate of time increase is less than linear. Given that scale is a fundamental aspect of movement with theoretical and practical significance, we conducted an experiment to empirically investigate the effect of spatial scaling, both for smooth curves and other types of stroke gestures.

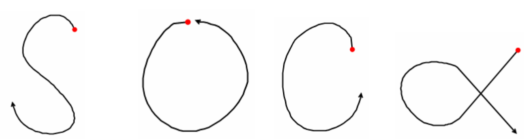
Design

12 gestures were selected from the Graffiti sample set used in Experiment 2. The 12 gestures could be categorized into 3 classes: 4 gestures were connected straight lines with corners (polylines), 4 gestures consist of smooth curves only, and 4 gestures consist of both curves and corners/straight lines (Figure 16).

Polyline (I, M, F, N)



Smooth curve: (S, O, C, K)



Combination: (G, P, Y, E)

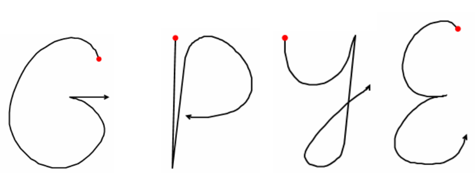


Figure 16. Gesture samples for Experiment 3.

Each gesture were presented in 8 different scales: $1/4\sqrt{2}$, $1/4$, $1/2\sqrt{2}$, $1/2$, $1/\sqrt{2}$, 1, $\sqrt{2}$, and 2 times of the original size (around 50 mm) respectively. We intentionally pushed the scale conditions to a very wide range from large to small. Although one consistent type of scaling relationship may exist in the medium range of scales, different types of scaling effects might take place at the extreme ends, due to the involvement of different muscle groups. Whereas in the medium scales of our experimental design a combination of wrist and finger movements was required to articulate the stroke gestures, arm movement was required to draw at the largest scale (2) and only finger movements were involved in drawing at the smallest scale ($1/4\sqrt{2}$).

A fully crossed within-subject factorial design was used. The experiment was executed in 3 repeated blocks. Within each block the participant conducted one trial for each gesture & scale combination. The presentation order of the samples was randomized in each block. 10 practice trials were performed before the experiment started.

10 right-handed volunteers, aged between 20~49, participated in the experiment.

Results

We performed regression on the average production time as related to scale, using both a power model (motivated by equation 5) and a linear model. In order to decide the valid range of any consistent scaling effect that may exist, we applied regression both including and excluding the extreme scales, and found the range between $1/4$ and $\sqrt{2}$ showed the most consistent behavior. It appeared that the two extreme cases that involved different muscle groups indeed had different or additional scaling effects.

The regression result for the range between $1/4$ and $\sqrt{2}$ is illustrated in Figure 17.

Linear Model:

$T = 230\ Scale + 566$ (ms) ($R^2 = .992$)

Power Model:

$T = 806\ Scale^{0.210}$ (ms) ($R^2 = .984$)

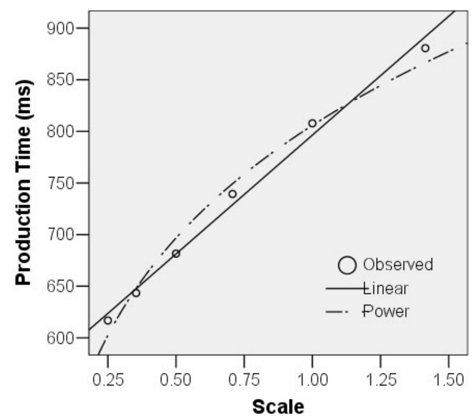


Figure 17. Scaling effect on production time.

Overall the two models matched the aggregated empirical data with similar correlation accuracy. To observe scale effect as related to different stroke gesture characteristics, we further performed regression for each class of gestures. The results are summarized below.

Gesture Group \ Model	Linear	Power
Polylines	$R^2 = .986$	$R^2 = .940$
Smooth Curves	$R^2 = .958$	$R^2 = .995$
Combination	$R^2 = .991$	$R^2 = .959$
Overall	$R^2 = .992$	$R^2 = .984$

Table 1. Scaling effect on production time, by gesture classes.

Not surprisingly, the power model fitted better with smooth curves. This is consistent with the fact that equation (5) was derived from the power law of curvature. For gestures with corners (polyline and combination), a linear model better described the relationship between the scale and the production time. Plausibly the existence of corners contributed a more or less constant overhead to the production time, thus washed out any subtle non-linear relationship that may exists between scale and time. Similar to results in experiment 1, the scale also had significant effects on both the number of attempts *A* ($F_{7,63}$=35.7, *p*<.001) made to pass the set error threshold and the remaining percentage error *E* ($F_{7,63}$=26.2, *p*<.001)Both measures decrease as the scale increases.

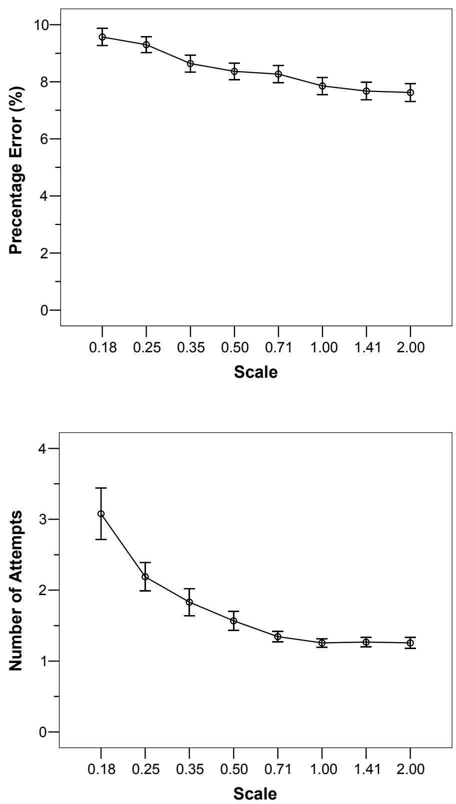


Figure 18. Scaling effect on errors.

In summary, the production time of a stroke gesture increases with the scale the gesture is articulated at. Such an increase is nearly linear, but it is in addition to a rather large constant overhead independent of scale. The overall scale effect can also be modeled as a power function, with greater accuracy when the gesture is a continuous curve. It is important to note the speed accuracy trade-off. While larger stroke gestures take longer to articulate, they tend to be more accurate.

The scale effects revealed in this experiment based on open-loop gestures is clearly different from the scale effects found in visually guided trajectory tasks. For example Accot and Zhai [4] systematically examined the impact of scale on steering law tasks and found that 1. the impact of scale was relatively small in comparison to the impact of the steering law's index of difficulty, and 2. the steering time followed a U shaped function of the scale factor with a quite flat bottom over 16 folds of scale change. Paths with the same index difficulty at the very large and the very small scales both took longer than paths at the medium scales. These effects are clearly different from the monotonic time increase in stroke gesture production. However, the errors in the steering law task followed a similar trend to the current open-loop stroke gesture study: smaller scales tend to cause more errors (Note that in [4] the scale factor was defined as the inverse of the scale factor in the current paper).

DISCUSSION AND CONCLUSION

Conceptual and mathematical analyses as well as knowledge in the motor control literature, particularly Viviani's power law of curvature, led us to a set of candidate elemental models and the summative model for gestures. We tested these models in two multi-phase experiments. The results show that the model forms are quite strong, as reflected by the high correlation with empirical data from a variety of gesture sets, with greater than 0.9 R^2 value in all cases. To our knowledge such a level of precision has never been previously achieved for such a wide range of gestures in varying complexity. In the more demanding tests (beyond what is typically done with much simpler laws of action) that not only required correlation between prediction and empirical data, but also a specific *a priori* time value prediction for each gesture, the proposed model may either underestimate or overestimate the time needed, but nonetheless gave a similar magnitude prediction which can be still useful as a baseline prediction for many design purposes. The order and ratio of predicted performances between comparative pairs were always consistent with the empirical data. We reason that two factors are key to the over or under estimation of time value. First, individual difference may influence the parameters of the model. (In experiment 1, the production time difference between the fastest and the slowest participant was in the factor of ~2.5.) More accurate parameter estimations are needed in future work from a very large pool of participants. Second and more importantly, familiarity and the amount of practice may drive the actual empirical data away from the model's prediction. For unfamiliar and little practiced gestures, such as the polyline gestures in experiment 1, actual articulation of the gesture may be slower than the model prediction due to online visual perception, planning and decision making. For familiar or well practiced gestures, as in the rest of the gestures sets tested in the two experiments, actual production tended to be more rapid than the model prediction. This was likely caused by "chunking" behavior – linking two or multiple elements of a gesture into one action, which can also be explained by our models.

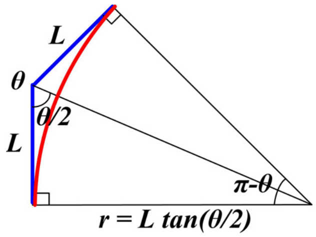


Figure 19. Corner-cutting behavior

For example, for a well-practiced gesture people tend to draw a short arc instead of an abrupt corner between two gesture elements. As illustrated in Figure 16, the time needed to draw the two straight corner arms (blue) is

$$T_1 = 2mL^n \qquad (\textit{according to equation } (11)) \qquad (19)$$

the time needed to draw the "corner-cutting" arc (red) is

$$T_2 = \frac{\pi - \theta}{K}(L \tan(\theta/2))^{1-\beta} \quad (\textit{according to equation } (4)) \qquad (20)$$

Taking the parameters acquired in experiment 1, and approximate $n = 1\text{-}\beta$, we have:

$$T_2 / T_1 = \frac{(\pi - \theta)(\tan(\theta/2))^{1-\beta}}{2mK} = 0.476(\pi - \theta)(\tan(\theta/2))^{0.414} \qquad (21)$$

This function is plotted in Figure 20 (note the ratio becomes zero at θ = 0, π , suggesting that cutting a corner of 0º or 180º is meaningless, since either the radius or the sweep angle of the arc becomes zero), with a maximum ≈ 0.8 at θ ≈ 1 (57.3º). The fact that T_2/T_1 is less than 1 indicates that cutting the corner will always save time. However, this time gain is less noticeable at moderate angles around 60º when compared with the more extreme angles. Note that this is a simplified analysis. A more complete and strict analysis should concern the error tolerance criterion and the context around the corner, etc. Nonetheless, it provides us with a way to more "correctly" apply the summative model (equation (13)) than simply adding elements together. Instead, one may consider the chunking behavior and calculate the elemental and summative models accordingly. Alternatively, we may use the simple summation as a baseline and add a compensatory term to equation (13) to reflect the degree of chunking impact. However, both approaches require further research to be practical.

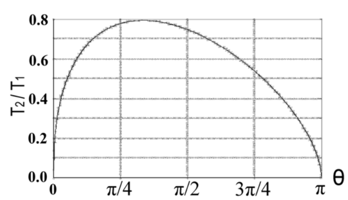


Figure 20. Time for corner-cutting.

In addition to the modeling contribution toward enlarging the theoretical tool box of UI research, design and evaluation, the elemental empirical findings from this work can also be relevant to gesture interface design and interesting to HCI researchers. Due to the space limitation we will only point out two examples. One is that (right-handed) users are faster at drawing straight lines in the 45º and 225º orientations. However, the diagonal directions were also found to be more error-prone, probably due to humans' lesser perceptual sensitivity to these angles than primary axes directions. Therefore gesture interface designers can exploit these directions for faster interaction (such as in [5]), but only if the precise orientation is not essential to the interaction. The second example is that our data showed that it was harder to maintain the same relative accuracy for small gestures, suggesting special challenges for UI design for very small devices.

The current work also revealed many differences between visually guided movement and open-loop gestures. For example, for visually guided motor movement the impact on time performance from scale change in the moderate range is relatively low, as reflected by, for example, the flat bottom of a U shaped function over several orders of magnitude scale change [4]. But in open-loop stroke gesturing, production time is strongly and monotonically influenced by the length of the gesture, both for curved and straight (equation (5) and (11)) gesture elements. Another difference is that while previous study on visually guided steering performance by Pastel [15] found that the degree of a corner significantly influences movement time, the current study found that the presence of a corner is a major determinant of gesture production time; however, the difference between different degrees of corner is negligible. In sum, the current work shows that findings based on visually guided motor control tasks cannot be taken for granted when applied to gesture stroke analysis.

The current work on stroke gesture modeling is probably the most comprehensive to date. Previous work in this domain is rare and has different goals or makes different levels of prediction. The most successful previous model is that of Isokoski's "line counting" model, which has the merits of simplicity and ease of calculation. However, it makes no specific quantitative time prediction for a given gesture. Our proposed model goes beyond the prior work, enabling us to, for example, predict the performance difference between two layouts for ShapeWriter (in contrast, line counting would predict no difference between two layouts as long as the number of lines remains the same).

Our model focused purely on the motor control aspect of gesture strokes; thus it does not model the mental complexity in perceiving and planning the gestures or the transition from novice to expert behavior. In practice these are all important factors that influence the overall user experience. Furthermore, given the limited data we collected, although we have confidence in the form of the model, we do not claim to have found "universal" parameters in the elemental models.

Modeling stroke gestures is an important, complex and challenging task. The current work should certainly not be viewed as an definitive investigation on the topic, but rather as one of the first systematic attempts toward the ultimate modeling and understanding of pen stroke gestures as a human-computer interaction medium.

Acknowledgement

The work presented in this paper was conducted during Xiang Cao's graduate internship at the IBM Almaden Research Center. We thank colleagues at IBM, particularly John Barton, for offering helpful critique and assistance to this project. We also thank Poika Isokoski of University of Tampere for generously sharing his insight and commenting on an early progress report.

REFERENCES

1. Accot, J. and Zhai, S., Beyond Fitts' Law: Models for trajectory-based HCI tasks. *Proc. ACM CHI Conference on Human Factors in Computing Systems* (*1997*), ACM, 295-302.
2. Accot, J. and Zhai, S., More than dotting the i's - foundations for crossing-based interfaces. *Proc. ACM CHI Conference on Human Factors in Computing Systems* (*2002*), ACM, 73-80.
3. Accot, J. and Zhai, S. Performance evaluation of input devices in trajectory-based tasks: an application of the steering law. *Proc. ACM CHI conference on Human factors in computing systems,* , Pittsburgh, Pennsylvania, United States (1999), 466-472.
4. Accot, J. and Zhai, S. Scale effects in steering law tasks. in(ed). *Proc ACM CHI conference on Human factors in computing systems*, ACM Press, Seattle, Washington, United States (2001), 1-8.
5. Baudisch, P., Xie, X., Wang, C. and Ma, W.-Y., Collapse-to-zoom: viewing web pages on small screen devices by interactively removing irrelevant content. *Proc. Proceedings of the 17th annual ACM symposium on User interface software and technology* (*2004*), ACM Press, 91-94.
6. Fitts, P.M. The information capacity of the human motor system in controlling the amplitude of movement. *Journal of Experimental Psychology*, *47*, (1954). 381-391.
7. Goldberg, D. and Richardson, C., Touch-typing with a stylus. *Proc. ACM CHI conference on Human factors in computing systems* (*1993*), ACM Press, 80-87.
8. Hinckley, K., Ramos, G., Guimbretiere, F., Baudisch, P. and Smith, M., Stitching: pen gestures that span multiple displays. *Proc. the working conference on Advanced visual interfaces* (*2004*), ACM Press, 23-31.
9. Isokoski, P., Model for unistroke writing time. *Proc. ACM CHI conference on Human factors in computing systems* (*2001*), ACM Press, 357-364.
10. Kristensson, P.-O. and Zhai, S., Relaxing Stylus Typing Precision by Geometric Pattern Matching. *Proc. ACM Conference on Intelligent User Interfaces* (*2005*), ACM Press, 151–158.
11. Kristensson, P.-O. and Zhai, S., SHARK2: a large vocabulary shorthand writing system for pen-based computers. *Proc. the 17th annual ACM symposium on User interface software and technology* (*2004*), ACM Press, 43-52.
12. Long, C., Landay, J.A., Rowe, L.A. and Michiels, J. Visual similarity of pen gestures. *Proc. ACM CHI conference on Human factors in computing systems*, ACM Press, The Hague, The Netherlands (2000), 360-367.
13. Mayzner, M.S. and Tresselt, M.E. Tables of single-letter and digram frequency counts for various word-length and letter-position combinations. *Psychonomic Monograph Supplements*, *1*, 2 (1965). 13-32.
14. Moran, T.P., Chiu, P., Melle, W.v. and Kurtenbach, G., Pen-based interaction techniques for organizing material on an electronic whiteboard. *Proc. ACM UIST Symposium on User Interface Software and Technology* (*1997*), 127-136.
15. Pastel, R., Measuring the difficulty of steering through corners. *Proc. ACM CHI conference on Human Factors in computing systems* (*2006*), ACM Press, 1087-1096.
16. Viviani, P. and Flash, T. Minimum-jerk, two-thirds power law, and isochrony: converging approaches to movement planning. *Journal of Experimental Psychology: Human Perception and Performance*, *21*, 1 (1995). 32-53.
17. Viviani, P. and Terzuolo, C. Trajectory determines movement dynamics. *Neuroscience*, *7*, 2 (1982). 431-437.
18. Zhai, S., Accot, J. and Woltjer, R. Human Action Laws in Electronic Virtual Worlds - an empirical study pf path steering performance in VR. *Presence*, *13*, 2 (2004). 113-127.
19. Zhai, S., Hunter, M. and Smith, B.A. Performance optimization of virtual keyboards. *Human-Computer Interaction*, *17*, 2,3 (2002). 89-129.
20. Zhai, S., Kong, J. and Ren, X. Speed-accuracy trade-off in Fitts' law tasks - on the equivalency of actual and nominal pointing precision. *International Journal of Human-Computer Studies*Special Issue of Fitts (1954) 50th Anniversary (2004).
21. Zhao, R., Incremental recognition in gesture-based and syntax-directed diagram editors. *Proc. ACM CHI conference on Human factors in computing systems* (*1993*), ACM Press, 95-100.