Visual Similarity of Pen Gestures

A. Chris Long, Jr., James A. Landay, Lawrence A. Rowe, and Joseph Michiels
Department of Electrical Engineering and Computer Science
University of California at Berkeley
Berkeley, CA 94720-1776
{allanl,landay,rowe,cujoe}@cs.berkeley.edu
+1-510-643-7106
http://www.cs.berkeley.edu/~{allanl,landay,rowe}

ABSTRACT

Pen-based user interfaces are becoming ever more popular. Gestures (i.e., marks made with a pen to invoke a command) are a valuable aspect of pen-based UIs, but they also have drawbacks. The challenge in designing good gestures is to make them easy for people to learn and remember. With the goal of better gesture design, we performed a pair of experiments to determine why users find gestures similar. From these experiments, we have derived a computational model for predicting perceived gesture similarity that correlates 0.56 with observation. We will incorporate the results of these experiments into a gesture design tool, which will aid the pen-based UI designer in creating gesture sets that are easier to learn and more memorable.

Keywords

Pen-based user interfaces, pen gestures, multi-dimensional scaling, similarity, perception

INTRODUCTION

Pen and paper is a versatile, powerful, and ubiquitous technology [17]. Pen-based user interfaces are becoming more widespread [9] and have great promise in power and versatility [4, 8, 17, 25]. Gestures, or commands issued with a pen, are one desirable feature of pen-based UIs. Because command and operand can be specified in one stroke, they are fast [5]. They also are commonly used and iconic,[1] which makes them easier to remember than textual commands [19]. Gestures are useful on displays ranging from the very small, where screen space is at a premium, to the very large, where controls can be more than arm's reach away [20].

A survey of PDA users [14] showed that users think gestures are powerful, efficient, and convenient. They want more gestures in applications and the ability to define their own gestures. However, the survey also revealed problems with gestures. Specifically, users often find gestures difficult to remember, and they become frustrated when the computer misrecognizes gestures. Users of other systems have also found gestures to be awkward [7].

1. By "iconic", we mean that the gesture shape suits or suggests its meaning.

We believe gestures can be difficult to use because they are difficult to design. We are developing a tool to assist pen-based UI designers in creating and evaluating gestures for pen-based UIs [15]. The primary benefit of the tool will be to advise designers about how to improve their gesture set. For example, it will notify the designer of gestures that: 1) are likely to be perceived as similar by users, 2) may be difficult for users to learn and remember, and 3) may be misrecognized by the computer. This advice will enable the designer to improve their gestures early in the design process before investing in expensive user studies.

The current work is an investigation into gesture similarity. The goal of this work is to develop a computable, quantitative model of gesture similarity that can be incorporated into the gesture design tool. Perceived similarity is useful for designers to know because it affects how easily users can learn and remember the gestures. We contend that similar operations with a clear spatial mapping, such as scroll up and scroll down, should be assigned similar gestures. Conversely, gestures for more abstract operations that are similar, such as cut and paste, may be easily confused if they are visually similar.

To determine what features affect perceived gesture similarity, we ran a pair of experiments to measure the similarity of a variety of gestures. The data collected in these experiments enabled us to derive an algorithm for computing how similar novel gestures are. In conjunction with information about the impact of gesture similarity on their ease of learning and recall (from another experiment whose results are being analyzed), this will allow our gesture design tool to provide better advice.

The remainder of the paper is organized as follows. The first section discusses related work. The next section describes the gesture similarity experiments. A discussion of the results of both experiments follows. Finally, we present future work and conclusions.

RELATED WORK

This section discusses relevant prior work. The first subsection gives some background on pen-based user interfaces, which is the context for this work. The second discusses prior work on perceptual similarity of gesture-like shapes. The last section introduces multi-dimensional scaling, a data analysis technique used in our experiments.

Pen-based interfaces

The device that popularized pen input was the Apple Newton MessagePad. It was designed primarily for pen input. It minimized the use of overlapping windows and encouraged the user to focus on one document at a time. Its core applications were a notepad, to-do list, calendar/scheduler, and address book. By default it recognized text as it was entered, but the user could elect to enter ink and

To appear in *Proceedings of Human Factors in Computing Systems* (SIGCHI '00).

recognize it later. It also supported a small number of gestures for editing text and drawings. Our work deals with the same type of gesture used on the Newton: single-stroke and iconic. The Newton's handwriting recognition was widely criticized when it was first introduced, but by the last model the recognition had greatly improved.

More recently, the 3Com PalmPilot has become a popular pen-based platform. Its display is even smaller than the Newton's. Its applications have very few on-screen controls, and the pulldown menu at the top of the screen is normally hidden. Its core applications are the same as the Newton's. The Pilot does not recognize normal English letters, but uses the Graffiti character set [12], which must be entered in a dedicated area of the screen. It does not use gestures of the kind we focus on, but instead uses a keyboard accelerator facility. The user may write a special "command" stroke to indicate that the next stroke is a command. For example, in many applications "command" followed by the "d" stroke invokes the delete operation.

Pens have been used in many different computer applications, on desktop- and on wall-sized displays. These applications have included spreadsheets, word processors, a disk manager, music editors, an equation editor, a GUI design tool, an air-traffic control UI, and note-taking applications [4, 6, 7, 10, 18, 25, 27].

Pens performed well for many of these applications. In their standard office applications, Briggs and associates found that users liked using a pen for navigation and positional control, although not for text entry [4]. This is in spite of a lack of pen-specific interaction techniques.

Wolf and colleagues added pen input and gestures for editing operations to a drawing application, a spreadsheet program, a music editor, and an equation editor [25]. Users reported that gestures were easier to recall than keyboard commands, and edited spreadsheet documents 30% faster with the pen than with keyboard.

Zhao used a combination of gestures and menus to facilitate object creation in a structured drawing program [27]. It allowed users to draw the object creation gesture and select from the object type menu in either order. This technique could allow rapid access to many commands while limiting the number of gestures that the user is required to learn.

Landay developed an interface design tool based on sketching that was well-suited for pen input and gestures [10]. It used iconic gestures of the same type as the ones in our study for creating and editing the UI elements.

Chatty and Lecoanet discussed how pen input and gestures are useful for air-traffic control [6]. Their interface allowed controllers to navigate the airspace and change aircraft attributes such as speed and heading. An evaluation of their system showed that although a few gestures had to be modified because of confusion, they were still useful.

Lopresti and Tomkins advocated treating electronic ink as a first class datatype [16]. They developed a prototype system that supported ink input and searching based on matching feature vectors. This technique was used to search a library of gestures using the gesture as the key.

We concentrate on gestures in the spirit of copy editing [13, 22] rather than marking menus [23], because we believe that traditional marks are more useful in some circumstances. For example, they can specify operands at the same time as the operation, and they can be iconic. For simplicity of analysis and because of restrictions in our gesture recognizer, we use only single-stroke gestures.

Perceptual similarity

Psychologists have investigated similarity of simple geometric shapes, which are less complex than our gestures. Attneave studied how changes in geometric and perceptual properties of different kinds of simple figures influenced their perceived similarity [2]. Participants in one experiment reported how similar they perceived parallelograms of differing sizes and tilts to be. Attneave found that similarity was correlated with the log of the area of the parallelograms and with their tilt. Also, parallelograms that were mirror images of one another were perceived as similar.

Another study by Attneave indirectly measured perceived similarity by measuring how easily names of triangles and squares were remembered. The assumption was that similar shapes will be misremembered. The squares varied in reflectance and area; the triangles varied in tilt and area. The result of these experiments indicated that similarity of form caused more confusion than similarity of area.

Based on these experiments, Attneave concluded the following. In general, the logarithm of quantitative metrics was found to correlate with similarity. Also, if the range of differences in stimuli is small, these differences are linearly related to perceived similarity, and when multiple dimensions of the stimuli change, their dimensions combine nearly linearly to give the change in perceived similarity. When the range of differences is large, the relationship between stimuli value and similarity is not linear, and the stimuli do not combine linearly.

Lazarte and colleagues studied how rectangle height and width affected perceived similarity [11]. They found that reported similarity was related to rectangle width and height and they derived a model to fit the reported similarity data. Also, they found that not only did different people use different similarity metrics, but that the same participant may have used different metrics for different stimuli.

Multi-dimensional scaling

Multi-dimensional scaling (MDS) is a technique for reducing the number of dimensions of a data set so that patterns can be more easily seen by viewing a plot of the data, usually in two or three dimensions. It takes as input one or more sets of pairwise proximity measurements of the stimuli. It outputs coordinates and/or a plot of the stimuli in a predetermined number of dimensions (typically 2–3) such that the pairwise inter-stimuli distances in the new space correlate with the input proximities of the stimuli.

There are several decisions to make in using MDS. One is how to use data from multiple participants. A simple method is to average the pairwise proximities and analyze the resulting proximity matrix as if it came from a single participant. However, there is evidence that this method does not give good results [1], and it also prohibits analyzing the differences among participants. Fortunately, we were able to use a version of MDS, INDSCAL [26], that takes as input a proximity matrix for each participant and takes individual differences into account.

Another decision when using MDS is how many dimensions to use in the analysis. Like other MDS methods, INDSCAL outputs how well the distances it produces correlate with the input proximities. A graph of this correlation vs. dimension ideally has an obvious "knee" in the graph, which indicates the number of dimensions to use.

Also, a rule of thumb for standard MDS is to use no more than a quarter as many dimensions as there are stimuli.[2]

How to measure distance is another issue in doing MDS analysis. The most often used metric is Euclidean distance, which is a special case of the Minkowski distance metric:

$$d_{ij}^{p} = \sum_{a}^{r} \left| x_{ia} - x_{ja} \right|^{p}$$

where d is the distance, there are r dimensions, and x_{ia} is the coordinate of point i on dimension a [26]. When p is 2, this is Euclidean distance. Another common p value is 1, in which case it is called city-block or Manhattan distance. Infinity is also sometimes used for p, which makes the sum equal to the distance along the dimension that differs most. There are sometimes psychological reasons to prefer city-block or Euclidean [24], but generally researchers use the metric that fits their data best.

The final step in MDS analysis is assigning meaning to the axes. Sometimes the experimenter may know the axes in advance. In this work that was not the case, and so two methods were used to determine the axes: 1) inspecting plots of the stimuli and 2) correlation with measurable quantities. More details of MDS analysis are available in [26].

GESTURE SIMILARITY EXPERIMENTS

To better understand the principles people use to judge gesture similarity, we performed an experiment to measure how people perceived similarity among gestures in a predefined set. We hoped that the experiment would enable us to derive metrics for predicting the human-perceived similarity of gestures.

We ran this experiment twice with different data sets and different subjects in order to confirm our results. The two trials are described below.

Similarity Trial 1

In the first experiment, we attempted to make a gesture set consisting of gestures that varied widely in terms of how people would perceive them. The gesture set is shown in Figure 1. It was designed by one of the authors (Long) based on personal intuition with two criteria in mind: 1) to span a wide range of possible gesture types, and 2) to have differences in orientation (e.g., 14 and 15, 20 and 22).

We considered whether the participants should draw the gestures or not. Drawing them would mimic actual usage more closely, but it would have lengthened the experiment. In order to accommodate more participants and more gestures, we decided not to have participants draw the gestures. Instead, the test program animated the gestures to show participants the dynamic nature of the gestures.

Participants

We recruited twenty-one people from the general student population to participate in the experiment. We only required that they be able to operate the computer and tablet. Each participant was paid $10 (US).

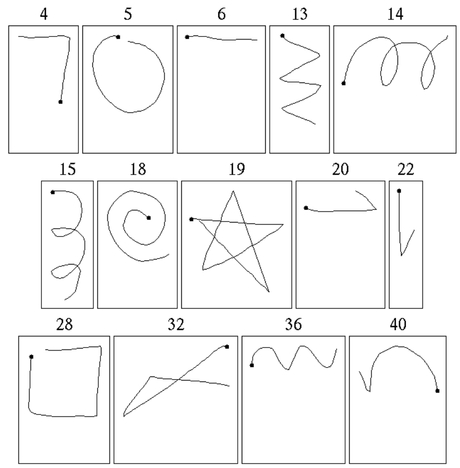

Figure 1 Gesture set for first gesture similarity experiment. A dot indicates where each gesture starts. (Gestures are not numbered 1, 2, 3... because they were chosen from a larger set.)

Equipment

For the main part of the experiment participants used a display tablet (a Mutoh MVT-12P) attached to an IBM PC compatible computer running Windows NT. The experimental application was written in Java.

Procedure

Participants were first shown an overview of the experiment, which outlined the tasks they would perform. Participants were then shown the tablet and the task was explained to them. The program displayed all possible combinations of three gestures, called triads. The order of the triads was randomized independently for each participant, as was the on-screen ordering of gestures within each triad. Figure 2 is a representative screen shot of the triad program. For each triad, participants selected the gesture that seemed most different from the others by tapping on it with a pen. The program recorded the selections of the users and computed the dissimilarity matrix.

The program was run using a practice gesture set of five gestures, shown in Figure 3, so that participants could become familiar with the program and the tablet. Participants were asked to select the gesture in each triad that seemed most different to them. After the practice, they performed the task again using the experimental gesture set of fourteen gestures (Figure 1). All participants saw all

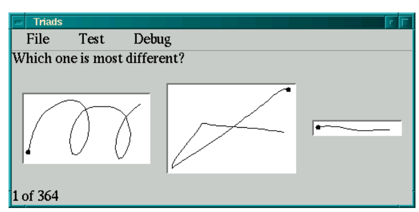

Figure 2 Triad program.

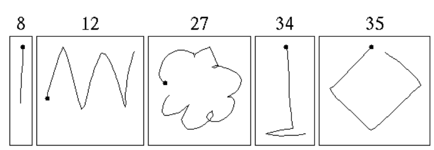

Figure 3 Practice gesture set.

2. ALSCAL uses more information than standard MDS, so it is reasonable to think that more dimensions would be valuable. Unfortunately, we were unable to find an analysis of how many.

To appear in *Proceedings of Human Factors in Computing Systems* (SIGCHI '00).

possible triads of gestures exactly once, for a total of $\binom{14}{3} = 364$ triads.

When the experimental task was completed, participants filled out a questionnaire which asked: 1) their impressions of the task and program and 2) demographic information about themselves. This questionnaire was a web form that was filled out on a different computer than the one used for the experimental task [21].

Analysis

The goals of the analysis were: 1) to determine what measurable geometric properties of the gestures influenced their perceived similarity and 2) to produce a model of gesture similarity that, given two gestures, could predict how similar people would perceive those two gestures to be.

The first goal was addressed using plots of gestures generated by MDS. In these plots, the Euclidean inter-gesture distances corresponded to the inter-gesture dissimilarities reported by the participants. By examining these plots, we were able to determine some geometric features that contributed to similarity. To determine the number of dimensions to use, we did MDS in two through six dimensions and examined plots of stress and goodness-of-fit (r^2) versus dimension to find the "knee" in the curve[3]. Similarity data was analyzed with MDS as interval/ratio and as ordinal. The ordinal model gave a better fit, so it was used for all subsequent analysis. We used Euclidean distances since it provided a better fit to our data than the city-block metric did.

The second goal was achieved by running regression analysis to determine which of many measurable geometric features of a gesture correlated with the reported similarity. Regression also produced weights indicating how much each feature contributed to the similarity. To compute the similarity of two gestures, their feature values are computed. The feature values and weights together give the positions of the gestures in feature space. The similarity of the gestures is given by the Euclidean distance between the two gestures in the feature space, where smaller distance means greater similarity.

The features used in our regression analysis came from a few sources. Some features were taken from Rubine's gesture recognizer [22]. Others were inspired by plots from the MDS analysis. The list of features that we thought might predict similarity is given in Table 1. We wanted our model to be computable, so we did not include in the final regression analysis features whose values were only obtainable by subjective judgement.

Results

We were able to derive a model of gesture similarity that correlated 0.74 ($p < 0.003$, 2-tailed t-test) with the reported gesture similarities.

The multi-dimensional scaling indicated that the optimal number of dimensions was five (S-stress = .13, $r^2 = 0.76$). For ease of comprehension, we plotted the gesture positions two dimensions at a time. Examination of the plot of dimensions 1 and 2 (shown in Figure 4) quickly showed that dimension 1 was strongly correlated with how "curvy" the gestures were — for example, g5 and g40 are curvy and g32 and g28 are straight. The curviness metric was derived in an

1. Cosine of initial angle	**12. Aspect [abs(45° – #4)]**
2. Sine of initial angle	**13. Curviness**
3. Size of bounding box	14. Total angle traversed / total length
4. Angle of bounding box	**15. Density metric 1 [#8 / #5]**
5. Distance between first and last points	**16. Density metric 2 [#8 / #3]**
6. Cosine of angle between first and last points	**17. Non-subjective "openness" [#5 / #3]**
7. Sine of angle between first and last points	18. Area of bounding box
8. Total length	19. Log(area)
9. Total angle	20. Total angle / total absolute angle
10. Total absolute angle	**21. Log(total length)**
11. Sharpness	**22. Log(aspect)**

Table 1 Possible predictors for similarity. Features 1-11 are taken from Rubine's recognizer. Bold features were found to be significant and so are used in the model.

attempt to capture our intuitive notion of curviness and to match this dimension from the MDS plot.[4]

It was observed in the MDS plots that short, wide gestures were perceived as being very similar to narrow, tall ones and that both types were perceived as different from square gestures. Angle of bounding box represented the difference between thin and square gestures, but not the similarity of tall vertical and short horizontal ones. We created the aspect feature to represent this relationship.

Table 2 shows which features strongly correlate with each dimension, based on a regression analysis. Although the most important (i.e., lower numbered) dimensions are predicted by relatively few features, the other dimensions require many features.

A separate regression analysis was done for each dimension, using the computed feature values as the independent variables. From these regressions we derived a set of equations to predict the position of an arbitrary gesture in the feature space. Given the predicted positions of two gestures, the degree of similarity humans would

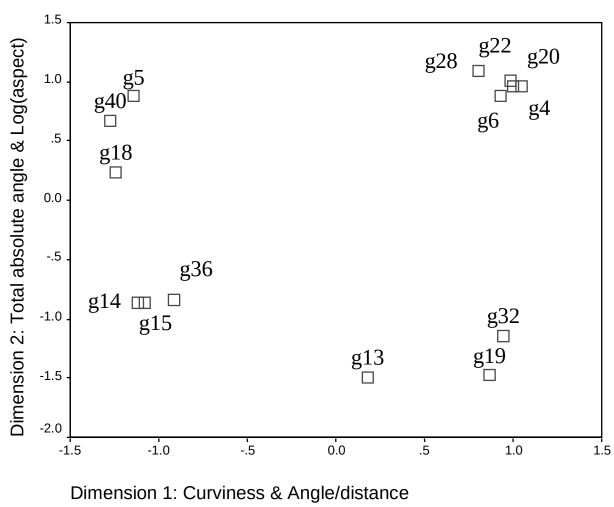

Figure 4 MDS plot of dimensions 1 and 2 for first similarity experiment.

3. Examination of stress vs. dimension and r^2 vs. dimension is a standard MDS technique for determining dimensionality.

4. Curviness of a gesture was computed by adding up all inter-segment angles within the gesture whose absolute value was below a threshold (19°). The threshold was chosen so that the metric would agree with the author's curviness judgements of gestures in trial 1.

Dimension	Correlated features (in order of decreasing importance)
1	Curviness, Angle / distance
2	Total absolute angle, Log(aspect)
3	Density 1, Cosine of initial angle
4	Cosine of angle between first and last points, Cosine of initial angle, Sine of initial angle, Distance between first and last points, Angle of bounding box
5	Aspect, Sharpness, Cosine of initial angle, Total angle

Table 2 Predictor features for similarity trial 1, listed in decreasing order of importance for each dimension.

perceive is approximated by the Euclidean distance between the gestures in the feature space. The derived model predicts the reported gesture similarities with correlation 0.74 ($p <$ 0.003, 2-tailed t-test). The MDS model upon which it is based fits the data only slightly better, so this is a good fit.

Another interesting aspect of trial 1 was the differences among participants. As expected, the degree that different features affected similarity judgements varied across participants. This disparity is consistent with the finding in other perception experiments that different people judge similarity using different features [11]. What was surprising was that the participants seemed to be clumped into two distinct groups. We separated the data for the two groups of participants and analyzed them separately. However, the resulting MDS models were not as good as the original, combined model, so we did not pursue the analysis further. It would be interesting to see if more participants reinforced this trend and illuminated a pattern.

Participants took an average of 26 minutes to complete the experimental task. The total time for each participant was approximately 40 minutes.

Similarity Trial 2

The results of the first similarity trial were encouraging, but we wanted to test the predictive power of our model for new people and gestures. We also were interested in exploring how systematically varying different types of features would affect perceived similarity.

To investigate how varying particular features would affect perceived similarity, three new gesture sets of nine gestures each were created. The first was designed to explore the effect of total absolute angle and aspect (Figure 5). The second was designed to explore length and area (Figure 6). The third was designed to explore rotation-related features such as cosine and sine of initial angle (Figure 7).

In addition to examining the effects of particular features, we wanted to determine the relative importance of the features. The most straightforward way to perform this test is to combine all gestures into one big set and have participants look at all possible triads from the combined set. Unfortunately, combining all of these gesture sets into one results in far too many triads, based on the time per triad taken for the first experiment. To allow us to compare the three sets against one another without a prohibitively large gesture set, two gestures from each of the three gesture sets were chosen and added to a fourth set of gestures (see Figure 8). All participants were shown all possible triads from all four gesture sets.

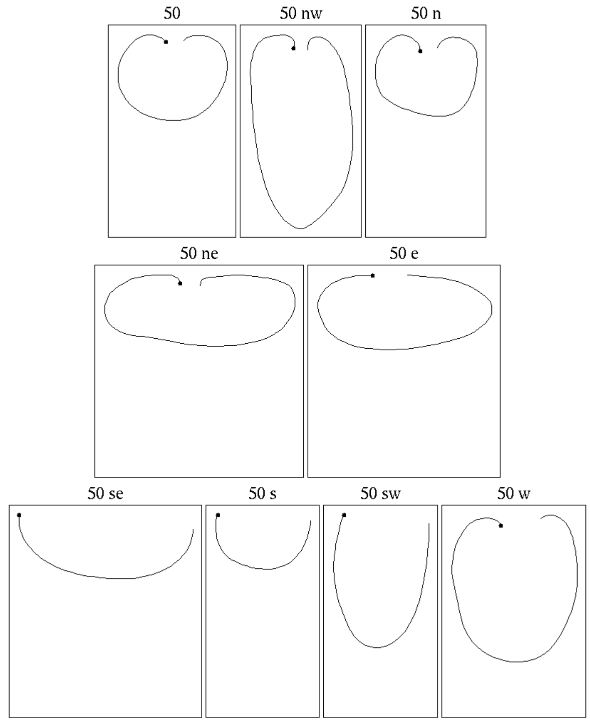

Figure 5 Similarity trial 2, gesture set 1. It was used to explore absolute angle and aspect.

Participants

Twenty new participants were recruited from the general student population. As in trial one, the only requirement was that they be physically able to use the computer and stylus. Each participant was paid $15.

Equipment

The same equipment was used as in the first trial.

Procedure

The procedure was the same as in trial one, except that participants saw triads from four gesture sets rather than one. Each participant saw all possible triads of gestures from each set, for a total of $3\binom{9}{3}+\binom{13}{3} = 538$ triads.

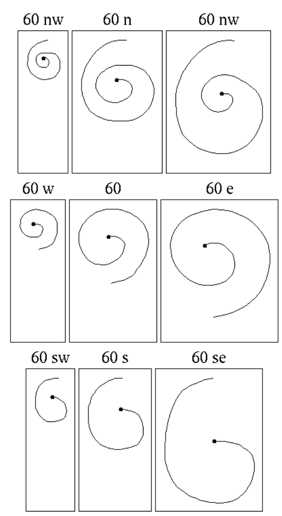

Figure 6 Similarity trial 2, gesture set 2. It was used to explore length and area.

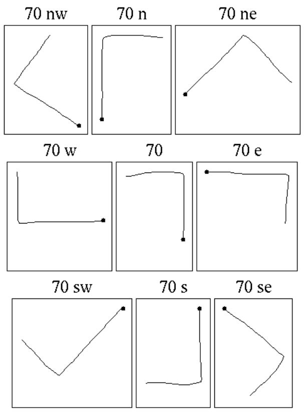

Figure 7 Similarity trial 2, gesture set 3.It was used to explore rotation.

Analysis

This trial was analyzed using the same techniques as the first trial, MDS and regression. First, a combined analysis was done, using the data from all four gesture sets. The goal of the combined analysis was the same as trial one: to determine what features were used for similarity judgements and to derive a model for predicting similarity. Many pairwise dissimilarity measures were missing from the data, because not all possible triads of all gestures were presented to the participants. However, this was not a problem because MDS can accommodate missing data.

In addition to the combined analysis, data from each of the first three sets was analyzed independently. The focus of the independent analyses was to determine how the targeted features affected similarity judgements.

Lastly, the results of trial 1 were used to predict the perceived similarity of the gestures in trial 2. These

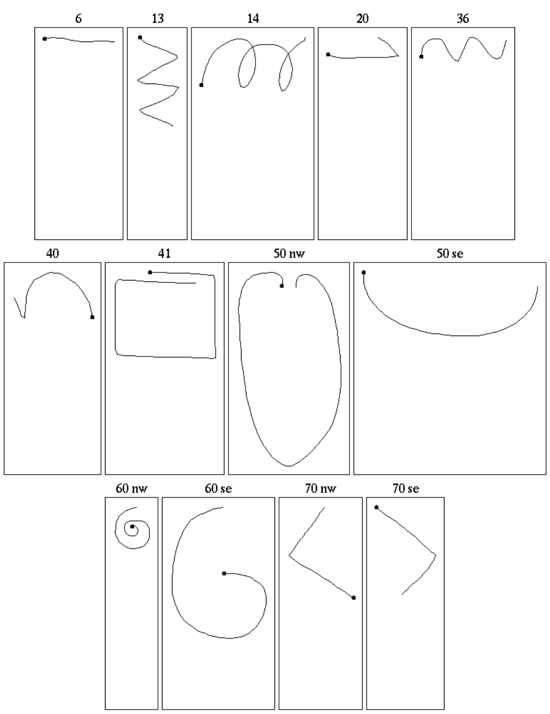

Figure 8 Similarity trial 2, gesture set 4. It includes gestures from Figure 5, Figure 6, and Figure 7.

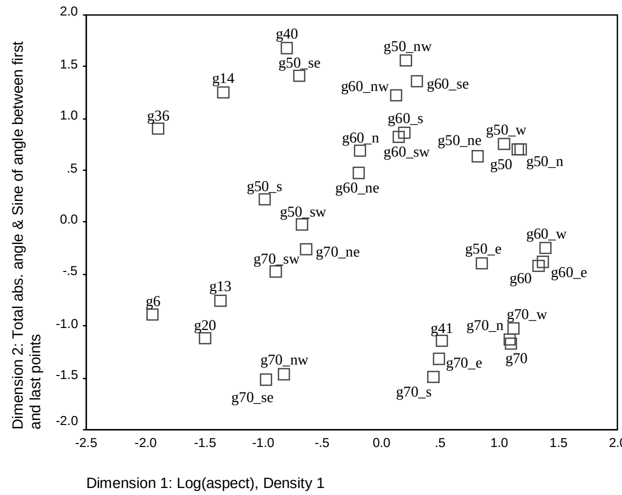

Figure 9 MDS plot of dimensions 1 and 2 for second trial in similarity experiment (combined data).

predictions were compared with the similarities reported by participants in trial 2.

Results

The best number of dimensions for MDS for the trial 2 data was 3 (S-stress = 0.08663, r^2 = 0.89539). Unfortunately, when the data was plotted, the meaning of the dimensions was not as obvious as in trial 1. (A plot of dimensions 1 and 2 is shown in Figure 9).

Table 3 shows which features correlate with each dimension, based on a regression analysis. The derived model predicts the reported gesture similarities with correlation 0.71 (p < 0.000003, 2-tailed t-test). Separate analyses of individual gesture sets (shown in Figures 5, 6, and 7) revealed that: 1) bounding box angle is an important feature and 2) alignment or non-alignment with the normal coordinate axes is significant for similarity.

Analysis of the first gesture set ("50" series, shown in Figure 5) gave a three dimensional MDS plot. This gesture set was intended to show the effects of absolute angle and aspect. We found that absolute angle is highly correlated with the first dimension (–0.81) and aspect is highly correlated with the second (–0.77). Unfortunately, the absolute angles of gestures in this set covaried greatly with the values of several other features, so it was not possible to determine whether absolute angle was significant. Strong covariance with other features was not a problem for aspect. However, bounding box angle correlated even more strongly with dimension two (–0.92) than aspect did.

Data from the second gesture set ("60" series, shown in Figure 6) were surprising. Its analysis was done in four

Dimension	**Correlated features (in order of decreasing importance)**
1	Log(aspect), density 1
2	Total absolute angle, Sine of angle between first & last points
3	Density 2, Non-subjective openness

Table 3 Predictor features for similarity trial 2 (using data from *all* trial 2 gesture sets).

dimensions. It was intended to discover the effect of length and area, but although length and area correlate well with dimension four (–0.83 and –0.92, respectively), they are both only weakly correlated with the first three dimensions (highest value of 0.46 for length and dimension 3). Since INDSCAL dimensions are ranked in order of importance, it appears that neither length nor area are very significant contributors to similarity judgement.

The third gesture set ("70" series, shown in Figure 7) also provided interesting results. One might expect similarity among gestures that are rotations of one another to be proportional to the amount of rotation, but this was not the case. Instead, the gestures whose lines were horizontal and vertical were perceived as more similar to one another than to those gestures whose components were diagonal. The perceived similarity of gestures whose components are aligned in the same directions is consistent with findings on texture in the vision community [3]. This set was analyzed in five dimensions.

As in trial one, there were differences among participants in their similarity judgements. Unlike trial one, the participants did not separate easily into two groups, but consisted instead of one clump with outliers trailing off. We experimented with removing outliers from consideration, but they did not appreciably improve the MDS model.

Similarity analysis

To validate the models produced by the two trials, each model was used to predict the similarities between all pairs of gestures used in the other trial. These predictions were compared with the reported similarities from the other trial. The correlation between the prediction of trial 1 and the data from trial 2 was 0.56 ($p < 0.0005$, 2-tailed t-test). The correlation between the prediction of trial 2 and the data from trial 1 was 0.51 ($p < 0.058$, 2-tailed t-test). Based on these correlations, the model derived from experiment 1 is a slightly better predictor of gesture similarity than the model from experiment 2.

DISCUSSION

This section discusses the results of the experiments described above and the challenges involved in designing the experiments and analyzing their results.

Results

Human perception of similarity is very complicated, even for simple shapes [11]. Shapes like pen gestures can be viewed as similar or dissimilar based on many different perceptual cues. In the face of this difficulty, we are pleased at how well our model predicts similarity. It correlates with its own data 0.74 ($p < 0.003$, 2-tailed t-test) and with novel (i.e., trial two) data 0.56 ($p < 0.0005$, 2-tailed t-test).

We were pleased to find that a small number of features explain the three most salient dimensions. In experiment 1, we saw that dimensions one through three can be predicted based on only two features each. Several possible explanations exist for the larger number of features needed for dimensions four and five. One is that the underlying perceptual model is complex. Another is that the gesture set used in the experiment was not complex enough or did not vary in the right way to illuminate those dimensions.

It was surprising to us that neither length nor area were significant factors in experiment 1, so the "60" series (Figure 6) in trial 2 was designed to investigate the effect of these two features. Trial 2 confirmed the results of trial 1; neither length nor area was a significant feature for similarity.

Our results are consistent with Attneave's [2]. We found that the logarithm of aspect had more influence on similarity than aspect itself. Also, the range of distances among feature values of our gestures was large and did not combine linearly (as shown by the better fit of the Euclidean distance metric over the city-block metric).

Design and Analysis

The primary challenge in designing both the similarity and memorability experiments was creating good stimuli (i.e., gesture sets). For the first similarity experiment, we wanted the stimuli to span the perceptual feature space. However, this was difficult because we did not know the structure of the perceptual feature space in advance. We culled gestures from an informal survey of colleagues and from another experiment [15] in an attempt to create a "well-rounded" gesture set.

For the second similarity trial, we wanted gestures that varied with respect to particular features. Our gesture design tool [15] was modified to display values for these features, but the process was still difficult. In particular, some of the features we wanted to investigate covaried with other features, which made the results difficult to interpret.

We were concerned at the outset that developing a model for similarity would be complicated by differences among the participants. However, in spite of the individual differences the model does have predictive power. Although analyzing different groups of participants separately was not useful for our data, more data might make it feasible to create multiple models, each of which models a subset of users well. In that case, a gesture design tool could use multiple similarity metrics and notify the designer about similar gestures along any metric. The designer may want two gestures to be similar or dissimilar, depending on the semantics of the operations they are used for.

The two similarity experiments each resulted in a model for similarity, and they are different. It is difficult to say which is better, but we think the model from trial one is slightly preferable. It predicts the data from the other trial slightly better than trial two predicts its data. Also, it uses more features, and thus may capture more about the underlying psychological model.

We found MDS to be very useful, but also limited. It was extremely helpful in the early stages of analyzing trial one, when we had little idea of what features might affect similarity. It inspired us to invent several significant features, including curviness, aspect, and density. Another potential benefit of MDS is the ability to analyze differences in participants, which are discussed above. Although it was useful for discovering candidate predictors for similarity, our use of MDS was qualitative. For the quantitative analysis, when we needed to create a predictive model, we used standard linear regression.

FUTURE WORK

We have run an experiment to investigate geometric properties of gestures that influence learning time and memorability, and to explore how similarity relates to learnability and memorability. The analysis is not complete, but preliminary results indicate that when similar gestures are used for similar operations, they are easier to remember.

We plan to incorporate these discoveries about gesture similarity into a tool for gesture set design that we are developing [15]. Gestures are an important part of pen-based UIs, and we believe that designers of pen-based UIs could greatly benefit from a gesture design tool that informed them of gestures that may be difficult for the computer to recognize or for people to learn and remember.

These experiments also suggest avenues for more psychological experiments. For example, it would be interesting to measure memorability and similarity of other gestures to validate and/or refine our current models. Also, in the two experiments described here, participants saw animated gestures but did not draw them. It is possible that different similarity criteria would emerge if participants drew the candidate gestures before judging their similarity.

CONCLUSIONS

Gesture set designers may want their gestures to be similar or dissimilar depending on the semantics of the operations. We have shown that perceptual similarity of gestures is correlated with well-defined computable features such as curviness. With these features, we have derived a computable, quantitative model for perceptual similarity of gestures that correlates 0.56 with reported similarity. Using our model, we can predict how similar people will perceive gestures to be. We expect similarity predictions to be a useful addition to our gesture design tool.

Our model and our experiences of experimental design and analysis should provide an excellent starting point for further investigation into gesture similarity, memorability, and learnability. When integrated with our gesture design tool, our model will allow designers to create gestures that are more memorable and learnable by users.

ACKNOWLEDGMENTS

The authors would like to thank Richard Ivry for his invaluable advice on experimental design and analysis, and the experimental participants for their time.

REFERENCES

[1] Ashby, F. G., Maddox, W. T., and Lee, W. W. On the dangers of averaging across subjects when using multidimensional scaling or the similarity-choice model. *Psychological Science*, May 1994. 5, 3, 144–151.

[2] Attneave, F. Dimensions of similarity. *American Journal of Psychology*, 1950. 63, 516–556.

[3] Beck, J., Prazdny, K., and Rosenfeld, A. A theory of textural segmentation. In *Human and Machine Vision* (Beck, J., Hope, B., and Rosenfeld, A., eds.), vol. 8 of *Notes and Reports in Computer Science and Applied Mathematics*. Academic Press, New York, NY, 1983 1–38. Proceedings of the Conference on Human and Machine Vision, Aug. 1981.

[4] Briggs, R., Dennis, A., Beck, B., and Nunamaker, Jr., J. Whither the pen-based interface? *Journal of Management Information Systems*, 1992-1993. 9, 3, 71–90.

[5] Buxton, W. There's more to interaction than meets the eye: Some issues in manual input. In *User Centered System Design: New Perspectives on Human-Computer Interaction* (Norman, D. A. and Draper, S. W., eds.). Lawrence Erlbaum Associates, Hillsdale, N.J, 1986 319–337.

[6] Chatty, S. and Lecoanet, P. Pen computing for air traffic control. In *Human Factors in Computing Systems (SIGCHI Proceedings)*. ACM, Addison-Wesley, Apr. 1996 87–94.

[7] Forsberg, A., Dieterich, M., and Zeleznik, R. The music notepad. In *Proceedings of the ACM Symposium on User Interface and Software Technology (UIST)*. ACM, ACM Press, New York, NY, Nov. 1998 203–210.

[8] Frankish, C., Hull, R., and Morgan, P. Recognition accuracy and user acceptance of pen interfaces. In *Human Factors in Computing Systems (SIGCHI Proceedings)*. ACM, Addison-Wesley, Apr. 1995 503–510.

[9] International Data Corporation. Smart handheld device market will enjoy robust growth in 1999 and 2000, May 1999. Available at http://www.idc.com/Data/Personal/content/PS052699PR.htm.

[10] Landay, J. and Myers, B. Interactive sketching for the early stages of user interface design. In *Human Factors in Computing Systems (SIGCHI Proceedings)*. ACM, Addison-Wesley, Apr. 1995 43–50.

[11] Lazarte, A. A. and Schonemann, P. H. Saliency metric for subadditive dissimilarity judgments on rectangles. *Perception and Psychophysics*, Feb. 1991. 49, 2, 142–158.

[12] Lee, Y. L. PDA users can express themselves with Graffiti. *InfoWorld*, Oct 3 1994. 16, 40, 30.

[13] Lipscomb, J. A trainable gesture recognizer. *Pattern Recognition*, Sep. 1991. 24, 9, 895–907.

[14] Long, Jr., A. C., Landay, J. A., and Rowe, L. A. PDA and gesture use in practice: Insights for designers of pen-based user interfaces. Tech. Rep. UCB//CSD-97-976, U.C. Berkeley, 1997. Available at http://bmrc.berkeley.edu/papers/1997/142/142.html.

[15] Long, Jr., A. C., Landay, J. A., and Rowe, L. A. Implications for a gesture design tool. In *Human Factors in Computing Systems (SIGCHI Proceedings)*. ACM, ACM Press, May 1999 40–47.

[16] Lopresti, D. and Tomkins, A. Computing in the ink domain. In *International Conference on Human-Computer Interaction* (Anzai, Y., Ogawa, K., and Mori, H., eds.), vol. 1 of *Advances in Human Factors/Ergonomics*. Information Processing Society of Japan and others, Elsevier Science, Jul. 1995 543–548.

[17] Meyer, A. Pen computing. *SIGCHI Bulletin*, Jul. 1995. 27, 3, 46–90.

[18] Moran, T. et al. Implicit structures for pen-based systems within a freeform interaction paradigm. In *Human Factors in Computing Systems (SIGCHI Proceedings)*. ACM, Addison-Wesley, Apr. 1995 487–494.

[19] Morrel-Samuels, P. Clarifying the distinction between lexical and gestural commands. *International Journal of Man-Machine Studies*, 1990. 32, 581–590.

[20] Pier, K. and Landay, J. A. Issues for location-independent interfaces. Tech. Rep. ISTL92-4, Xerox Palo Alto Research Center, Dec 1992.

[21] Reeves, B. and Nass, C. *The media equation: how people treat computers, television, and new media like real people and places*. Center for the Study of Language and Information; Cambridge University Press, Stanford, Calif.: Cambridge [England]; New York, 1996.

[22] Rubine, D. Specifying gestures by example. In *Computer Graphics*. ACM SIGGRAPH, Addison Wesley, Jul. 1991 329–337.

[23] Tapia, M. and Kurtenbach, G. Some design refinements and principles on the appearance and behavior of marking menus. In *Proceedings of the ACM Symposium on User Interface and Software Technology (UIST)*. ACM, Nov. 1995 189–195.

[24] Thomas, H. Spatial models and multidimensional scaling of random shapes. *American Journal of Psychology*, 1968. 81, 4, 551–558.

[25] Wolf, C., Rhyne, J., and Ellozy, H. The paper-like interface. In *Designing and Using Human-Computer Interfaces and Knowledge Based Systems* (Salvendy, G. and Smith, M., eds.), vol. 12B of *Advances in Human Factors/Ergonomics*. Elsevier, Sep. 1989 494–501.

[26] Young, F. W. *Multidimensional Scaling: History, Theory, and Applications*. Lawrence Erlbaum Associates, Hillsdale, NJ, 1987.

[27] Zhao, R., Kaufmann, H.-J., Kern, T., and Müller, W. Pen-based interfaces in engineering environments. In *International Conference on Human-Computer Interaction* (Anzai, Y., Ogawa, K., and Mori, H., eds.), vol. 20B of *Advances in Human Factors/Ergonomics*. Information Processing Society of Japan and others, Elsevier Science, Jul. 1995 531–536.