NASA-CR-203850

021395

Final Tech

Technical Report

Project No.:   **534336 / 534337**

Title: **Virtual Keyboard for Hands-Free Operations**

Authors:

A. L. AbouAli

Date: **W. A. Porter**

UAH

Submitted to:

August 30, 1996

Mr. Jonnathan Kim
MSFC / NASA

Virtual Keyboard for Hands-Free Operations

Abdel-Latief Abou-Ali[+] and William A. Porter[+]

Introduction

The measurement of direction of gaze (d.o.g.) has been used for clinical purposes to detect illness, such as nystagmus, unusual fixation movements and many others [2-5]. It also is used to determine the points of interest in objects [1]. In this study we employ a measurement of d.o.g. as a computer interface. The interface provides a full keyboard as well as a mouse function. Such an interface is important to computer users with paralysis or in environments where hand-free machine interface is required. The study utilizes the commercially available (ISCAN Model RK426TC) headset which consists of an infrared (IR) source and an IR camera to sense deflection of the illuminating beam. It also incorporates image processing package that provides the position of the pupil as well as the pupil size.

The study shows the ability of implementing a full keyboard, together with some control functions, imaged on a head mounted monitor screen. This document is composed of four sections:

1. The Nature of the Equipment

2. The Calibration Process

3. Running Process

4. Conclusions

1. Nature of the Equipment

The data produced by the vendor equipment is the pupil position and the pupil size. The pupil position is calculated in two dimension coordinates. We call them x_1-pupil horizontal and the x_2-pupil vertical. The data is sampled at a 60 Hz rate. The data is processed through

_____

[+] Laboratory for Advanced Computer Studies, University of Alabama in Huntsville, Huntsville, AL 35899.

proprietary software coming from an image processing chip. The image is subject to noise which in turn affects the output data. Figures 1-3 are typical examples of noise affecting the output of the equipment. The figures shows x_1, x_2 and pupil size as the eye is dwelling on a fixed target. The data of importance to our study is the pupil position which we will concentrate on as we proceed. The distribution of the typical dwelling (fixations) intervals is depicted in Figures 4 and 5. The distributions resemble more or less a Gaussian Character. The mean value of the distribution is the value of interest assuming we have a large enough sample size.

To make our notation clear we will use the x^i to denote vectors in the equipment coordinates (i.e., signal processor outputs) we use y^i to denote the vectors in the screen coordinates. The capital X,Y will be used to refer to a set of vectors in equipment and screen coordinates, respectively. The components of the vectors x,y will be referred to as (x_1, x_2) and (y_1, y_2), respectively.

During the running of an experiment the eye makes transitions from point to point. However, wandering and blinking will also be a part of the process. It is beneficial to see these aspects in terms of the equipment output. Figure 5 shows the blinking of the eye. From the figure we can see that the blinking period is "about 14 samples," and there will be an overshoot before the eye comes to a stability point. To illustrate the eye transition we cycle the position of a highlighted point on the screen between 5 different positions in a random way. Figure 7 shows the transitions between these positions. From the figure we can see that time lag is of about 30 samples. Also there will be some overshoot around the stability points. The eye wandering is depicted in Figures 8 and 9.

2. The Calibration Process

The utility of equipment hinges on the ability to calibrate it for each specific user. The calibration process in effect models the mapping $f : X \rightarrow Y$ from the equipment coordinate to the screen coordinates. It is also used to model the inverse mapping $g : Y \rightarrow X$ from screen coordinates to equipment coordinates. Moreover, calibration eliminates user variations and

2

temporal variations in these maps. The sets X,Y are ordered sets of corresponding points in equipment coordinates and screen coordinates, respectively. The choice of the calibration set cardinality should be moderate. A high cardinality gives a good calibration, however it will also be tedious for the users. There is an uncertainty about the nature of mapping (i.e., linear, nonlinear with low order or highly nonlinear) that must be resolved.

2.1 Choice of Calibration Points

It is obvious that the calibration points should be well distributed on the screen, and that the user should look at each point during the data collection. The process also should include some triggering. This process includes many aspects like user delays, eye wandering, and maybe eye blinking. That besides the noise of a fixation recall Figures 1 - 5. As it is clear from the figures a good choice for extracting a fixation from the noise is the mean of the Gaussian, i.e., the maximum likelihood MXL point. That point x^i is such that the probability $P(x^i) \geq P(x^j)$ for all x^j in the sampling interval for the fixation at that point. But the sampling interval includes other types of noise as mentioned above. To clarify this issue recall Figure 7 for transition between file points on the screen. From the figures we can see that the delay and overshoot interval is about 35 samples and assume one blinking and overshoots of 16 samples then we need at least 4 times that to get a reliable MXL value. Figure 10 and 11 show the distribution for interval of 225 for transition fixation

2.2 Mapping Schemes

Assuming we have two sets of points X and Y, what is the possible nature of the mappings f,g. We studied several choices of approximate mapping of different natures, namely

(1) Affine mapping $\quad f : (x_1, x_2, 1) \rightarrow (y_1, y_2)$

$$g : (y_1, y_2, 1) \rightarrow (x_1, x_2)$$

(2) Nonlinear of order 2 $\quad f : (x_1, x_2, x_1 x_2, x_1^2, x_2^2,) \rightarrow (y_1, y_2)$

3

$$g : (y_1, y_2, y_1 y_2, y_1^2, y_2^2,) \rightarrow (x_1, x_2)$$

(3) Higher nonlinear orders

The study of the nature of the mapping was based upon data taken over a full set keyboard in both the screen and the equipment coordinate. The study involved different personnel with different eye characteristics. The mapping is constructed by optimal regression using the full data sets. The images created by the mapping are compared with the actuals. Figures 12-15 show examples of these mappings for the linear and nonlinear of order 2 consequently. The highly nonlinear cases studied did not give a significant improvement over the linear case.

The experiments pointed to the linear mapping as an optimal choice. The nonlinear maps fit slightly better, but exhibit more ripple and require more points in the calibration set.

2.3 Constructing the Interpolation Functions

For convenience we use $x = (x_1, \cdots, x_n) \in R^n$ and $y = (y_1, \cdots, y_m) \in R^m$ to denote independent (input) and response (output) variables, respectively. In many applications $y = F(x)$ where F is an unmodeled function $R^n \rightarrow R^m$.

We assume that a collection of experimental data

$$\{(x^i, y^i) : i = 1, \cdots, N\}$$

is available as the basis for our design. an analytically determined map, $\hat{F} : R^n \rightarrow R^m$ is said to be an approximation to F provided

$$m^2 = \sum_i \left\| y^i - \hat{F}(x^i) \right\|^2$$

is acceptable. When $\hat{F}$ is *linear* it is synonymous with an $m \times n$ matrix, W, which computes it. When $\hat{F}$ takes the form

$$\hat{F}(x^i) = Wx^i + \xi$$

where $\xi \leftarrow R^m$ is a fixed vector, is said to be *affine*. Linear maps may also have interpreted as generating a subspace (namely the span of the columns of W) in R^m. Affine maps generate translated subspace (by ξ), i.e., hyperplanes.

to facilitate our discussion of optimal affine interpolators (regression, hyperplanes) we adopt a notational convention. Each $x^i \in R^n$ has an additional component added and its value is fixed at 1. Similarly ξ is added as the right hand column of W. It is easy to show that

$$Wx + \xi = [W : \xi]\begin{pmatrix} x \\ 1 \end{pmatrix}$$

hence affine maps are linear maps in an expanded space.

Another convention is also useful. For this let X denote the matrix constructed by using $\{(x^i, 1) : 1 = 1, \cdots, N\}$ as columns. Let Y denote the matrix constructed using $\{y^i : i = 1, \cdots, N\}$ as columns. Then

$$E = Y - WX$$

is the error matrix in which the ith column is the ith error vector. In fact $m^2 = trace(E^*E)$.

Using standard techniques [ ] we may minimize m^2 with respect to W. Indeed the optimal matrix, W_o, giving the minimum rms error, is

$$f^* = W_o = YX^*(XX^*)^{-1}$$

In some applications it is necessary to model an inverse map. That is given $\{y^i\}$ approximate $\{x^i\}$. The process of constructing affine approximations is the same. We augment each y^i with a component 1. The optimal matrix T, becomes

$$g^* = T_o = XY^*(YY^*)^{-1}$$

Since neither f^* nor g^* are necessarily square. Their relationship is more complicated than simple inverses of each other.

2.4 Calibration Set Cardinality

The calibration for linear mapping requires at least cardinality three. In this case an exact fit will be the case, but the calibration normally occurs at the beginning of a typing session which

makes some possibility for noisy choices of points due to the noise and quantization in of the equipment and some psychological human factors. So, a 3-point calibration will not provide a means of implicit diagnostics. Also, the plane will be sensitive to any variation . A choice of 4 points provides a means of diagnostics, but does not provide implicit means of isolating the error point. A choice of 5 points or more provides a mean of diagnostics and isolation. Clearly as the cardinality increases the number of noisy points that could be detected goes up. The choice of the point locations should feature a homogeneous distribution over the keyboard.

2.5 Calibration Check Procedure

From the study of a large number of test cases the linear mapping proved to fit well the mapped keyboard. Also it interpolates with good accuracy the position of the eye gaze, giving a noise free calibration point. Also the calibration check could be done to one of the mapping functions, and the other one is just a mirror.

Giving the calibration sets

$$X = \{x^1, x^2, \cdots x^6\} \qquad \textit{Equipment Coordinates}$$
$$Y = \{y^1, y^2, \cdots y^6\} \qquad \textit{Screen Coordinates}$$

Considering $g^* : Y \to X$ where g^* is the optimal mapping as described before. Applying the mapping g^* to the set Y we will get the estimation for the set X say X'.

$$\text{rms} = \textit{root mean square error estimation in that} = \frac{1}{6} \sum_{i=1}^{6} \left\| x^i - x'^i \right\|$$

The rms represents how much deviation is in the two sets X,Y from the linear relation. The deviation is normally due to noisy choice in one or more points. Assuming $\text{rms} < \varepsilon$ a typical value for $\varepsilon = 0.95$ is pointing to a good 6 points calibration. Otherwise we take two steps in sequence.

(a) Error Isolation

1. i = 1

2. Pick point x_i. Assume it is noisy $X^i = X - x^i$, $Y^i = Y - y^i$.

3. Construct $g_i^* : Y^i \to X^i$

4. $\mathrm{rms}_i = \frac{1}{5} \sum\limits_{k=1}^{5} \left\| x^k - x'^k \right\|$ *for all* $x^k \in X^i$

5. $i = i + 1$

6. $i > 6$ exit, or else go to 2

7. $\mathrm{rms}_L = \min\{\mathrm{rms}_i\} < \varepsilon$. Accept g_L^* as the mapping function, otherwise step (b).

(b) Error Correction

At this step we have at least two points are subject to errors. We have 15 different choices to come up with 2 out of the six points. The 4 points may not feature a homogeneous distribution over the keyboard, and that may imply unreliable mapping and rms values. The error isolation procedure using two points is employed to determine if the error is in two points or more. For two points in error, a new value for these points is captured and the check procedure is repeated all over again. For more than two points in error, the process of isolation at this step become lengthy since each time the mapping function computation requires matrix inversions. So we have decided to give notice to the user to repeat the whole calibration process again, since it is an indication for gross instability.

3. Running Process

After the calibration stage of the system is ready to estimate the instantaneous location for the line of gaze of the user using the mapping function f^*. In keyboard interface of a fixed position the process could be handled in another domain using the function g^*. Since the keyboard keys center positions in the Y domain are known and fixed, then pre-running stage on skewed image of that keyboard could be constructed in the equipment domain. Giving that the entire process could be handled without the need to the mapping function. However, predicting

the line of gaze position on the screen is beneficial for applications such as mouse function. For such a dual function we need the two way mapping.

We talk about the two function in following sections.

3.1 Typing process

The typing process includes (1) recognition of user dwelling, (2) deciding the character on which the dwelling happened, (3) giving feedback to the user, and (4) performing the associated function with the character / function key.

3.1.1. Recognition of User Dwelling and Deciding Character

One way of recognizing dwelling is to watch for eye speed, i.e. giving consecutive samples (x_1^i, x_2^i), (x_1^{i+1}, x_2^{i+1}) if $\left| x_1^i - x_1^{i+1} \right| + \left| x_2^i - x_2^{i+1} \right| < \varepsilon$ is valid for a sequence $(x^1, x^2, \cdots, x^n)$, then we say that the user is dwelling. The indices values in this sequence will have at most ε from each other. We call the integer n the dwelling number. The corresponding dwell time $T_d = n * S$ where S is the sampling rate of the equipment. An estimation for the intended line of gaze could be the mean value

$$x_1^\ell = \frac{1}{n} \sum_{i=1}^{n} x_1^i$$

$$x_2^\ell = \frac{1}{n} \sum_{i=1}^{n} x_2^i$$

A similar approach is used in [ ] and they called the algorithm "Running Mean." The "Running Mean" provides a good typing mechanism. However, it has some drawbacks in our application such as

(1) A single noisy value could collapse an entire sequence of good values.

(2) A sample for a blink at the end of the dwelling period will also make the entire process useless.

(3) A treatment for such problems is not an easy task in the context of Running Mean.

(4) A few accepted noisy samples could shift the mean of a long sequence pointing to a character to another.

8

To build an algorithm which is more beneficial in our application and environment, we employed a new technique based on the maximum likelihood (MXL). The MXL technique uses the distribution of the sequence in deciding the character. Giving a sequence $(x_1^i, x_2^i), (x_1^2, x_2^2), \cdots, (x_1^n, x_2^n)$, the sequence is pointing to the key set $C_1, C_2, \cdots, C_k$ not necessarily a sequence $k \le n$ with distribution $S_1, S_2, \cdots, S_k$ where

$$\sum_{i=1}^{k} S_i = n$$

We define the maximum likelihood character (MXLC) C_L where $S_L \ge S_i \quad \forall\, i = 1, \cdots, k$. Also we define the quality number (QN)

$$QN = S_L / n, \qquad 0 < QN \le 1$$

We accept a sequence with $MXLC = C_L$ as a representation to gazing at C_L if

$$QN(C_L) \ge \xi, \qquad 0 < \xi \le 1$$

ξ represents the efficiency of the decision which could be used to control the number of noisy samples that could be accepted in a sequence. Clearly this technique is more global in view than the previous one, and does not have the drifting effects from biased noise. Appendix A shows an overview for the typing process.

3.1.2. Feedback and Function Execution

Once the character is decided a feedback should be issued to the user to move to the next character / function. Two types of feedback could be suitable for such an application visual and audio. We have employed the two by highlighting part of the key or changing colors of prints for some time, as well as a tone. It is well known that a user will not recognize the changes until 100 milliseconds (ms) after or more, also it takes some thinking and motion delay to get positioned to the next character. So, collected data in that period is useless. So a delay of at least 700 rms should be imposed at that point of time to avoid meaningless repeating characters/functions.

3.1.3. Mouse and Shift Function

The mouse function in using the mapping function is enabled and disabled by a keyboard key "mouse." The instantaneous mouse is a direct use of f^* per sample. That does not provide

a stable mouse due to the noise effect. An approach of MXL or the Running Mean could be used to take out the noise effect. , and then use the mapping f^* will provide a stable mouse function.

A full keyboard requires small and capital letters, numerals, special character sets, and function keys. The line of gaze pointed at the center of the screen key provides the means to corrected recognition. So we implemented the shift function to change the character layout of the keys from small to capital, numbers to special characters and vice versa. The keyboard layout is shown in Figures 16, 17. Appendix B shows a flowchart of the typing process based on the MXL and Running Mean.

4. Conclusions

The implementation of a full keyboard in half screen is feasible and implemented by tracking the pupil position. The MXL technique is used to get rid of the noise imposed by the digitization effect as well as human factors. A calibration of 6 points or higher provides a mean for diagnostic, error isolation, and good performance. The linear mapping provides a good estimation and generalization. The typing process looks stable enough to be suitable for implementation. Experiments indicate that a typing rate of 55 characters per minute is possible with the existing equipment. The mouse function is also added to the design that could be used in activating processes and easy selections. The equipment we used proved to be sufficient for this purpose. However, more stability could be added to the equipment to sense and eliminate the noise. Also, the support points of the equipment should be studied to be independent of the eye motion.

References

1. Clynes, M., and J. H. Milsum, *Biomedical Engineering Systems*, McGraw Hill, 1970.

2. Nodine, C. F., and others, "Recording and Analyzing Eye-Position Data Using a Microcomputer Workstation," *Behavior Research Methods, Instruments, and Computers*, 1992.

3. Cawthorne, J., M. R. Dix, C. S. Hallpike, and J. D. Hood, "The Investigation of Vestibular Function," *Britt. Med. Bull.*, 12, pp. 131-142, 1956.

4. Van Noorden, G. K., H. S. Thompson, and M. W. Van Allen, "Eye Movements in Myotonic Distrophy, and Investigative Study," *Invest. Opthalmol.*, 3, pp. 314-324, June 1964.

5. Shapiro, A., "Instrumental Techniques for the Study of Human Behavior," *Ann. NY Acad. Sci*, 118, pp. 65-70, Sept. 1964.

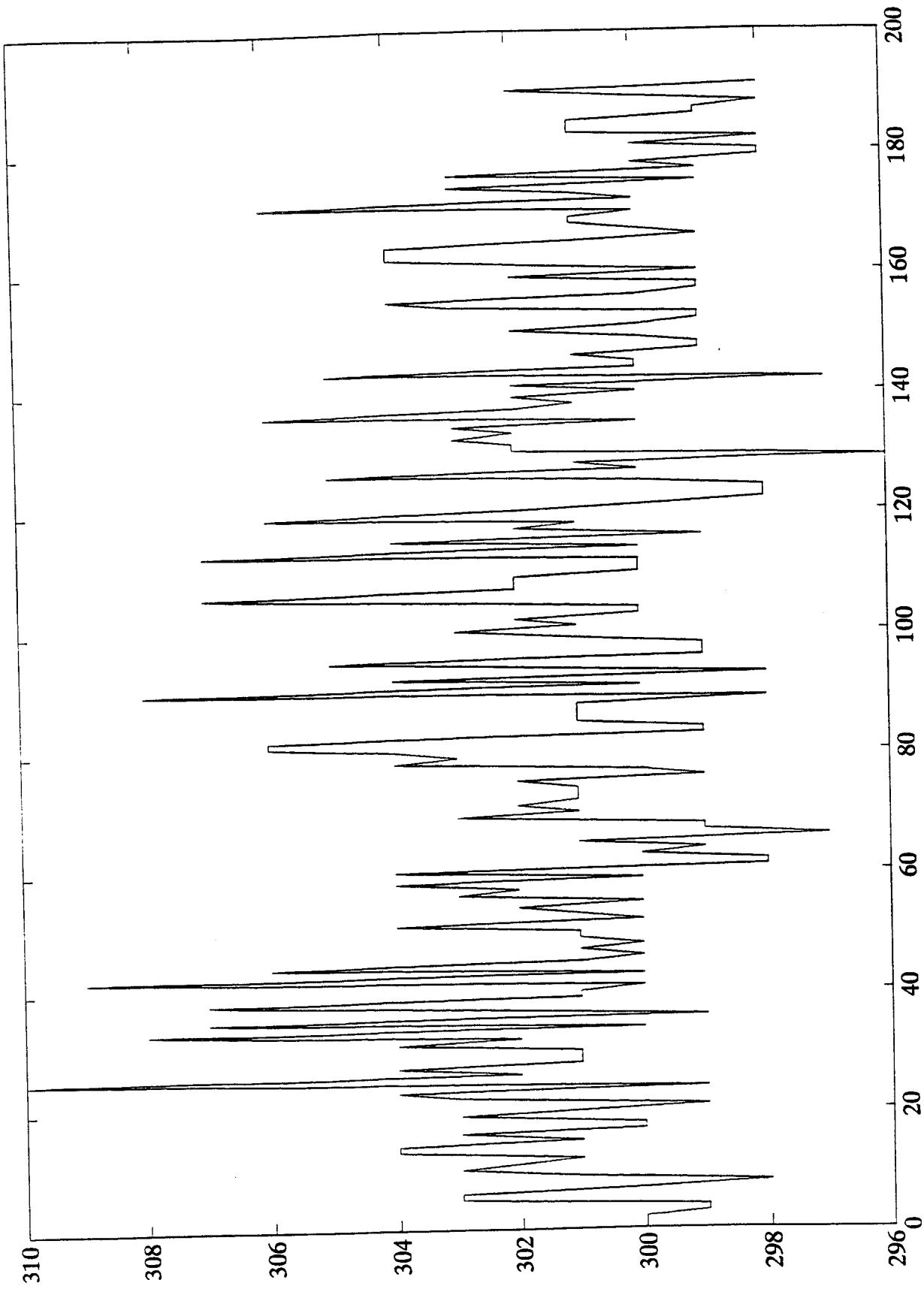Figure 1. Pupil horizontal in fixation.

Figure 2. Pupil vertical in fixation.
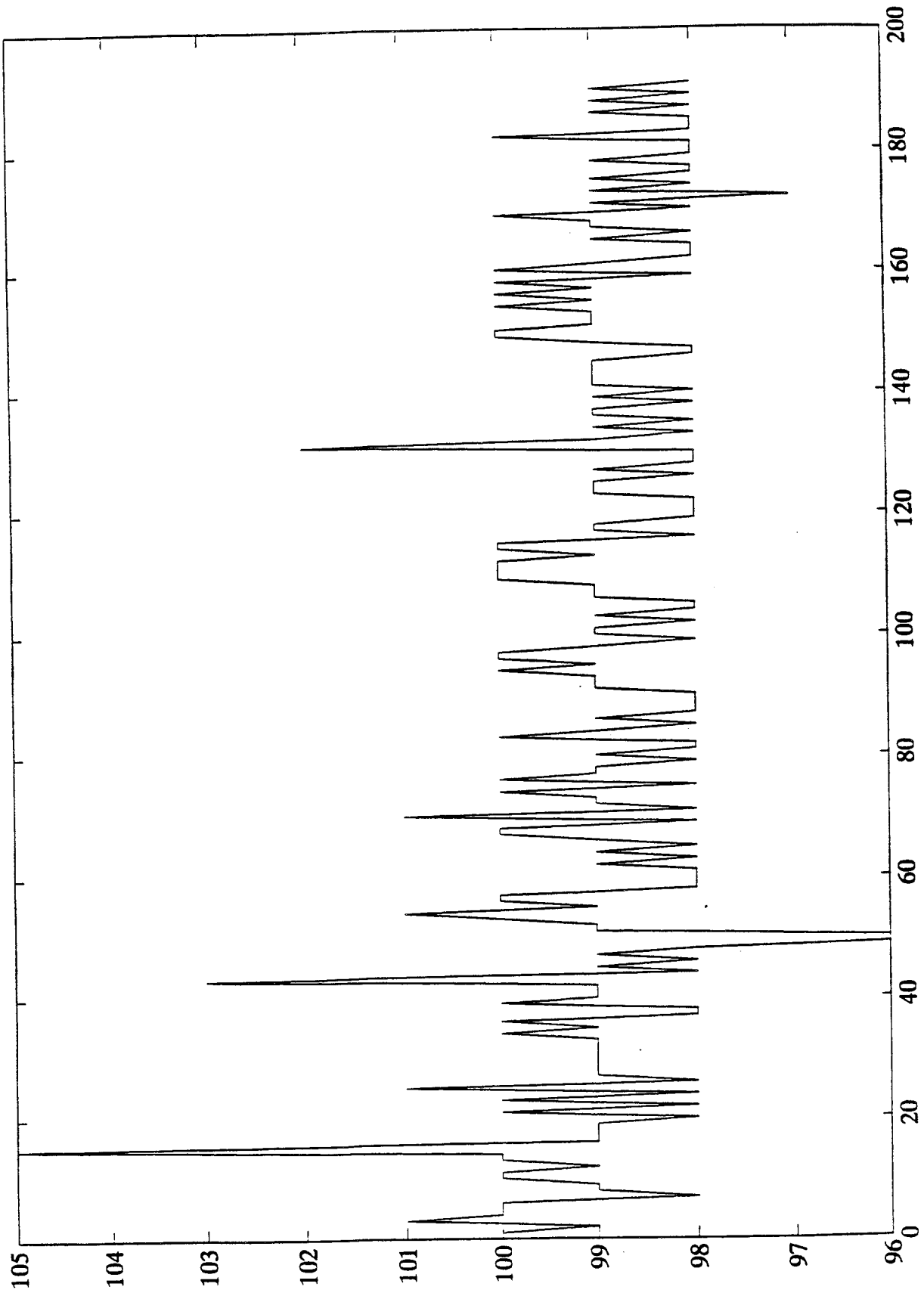
Figure 3. Pupil size in fixation.
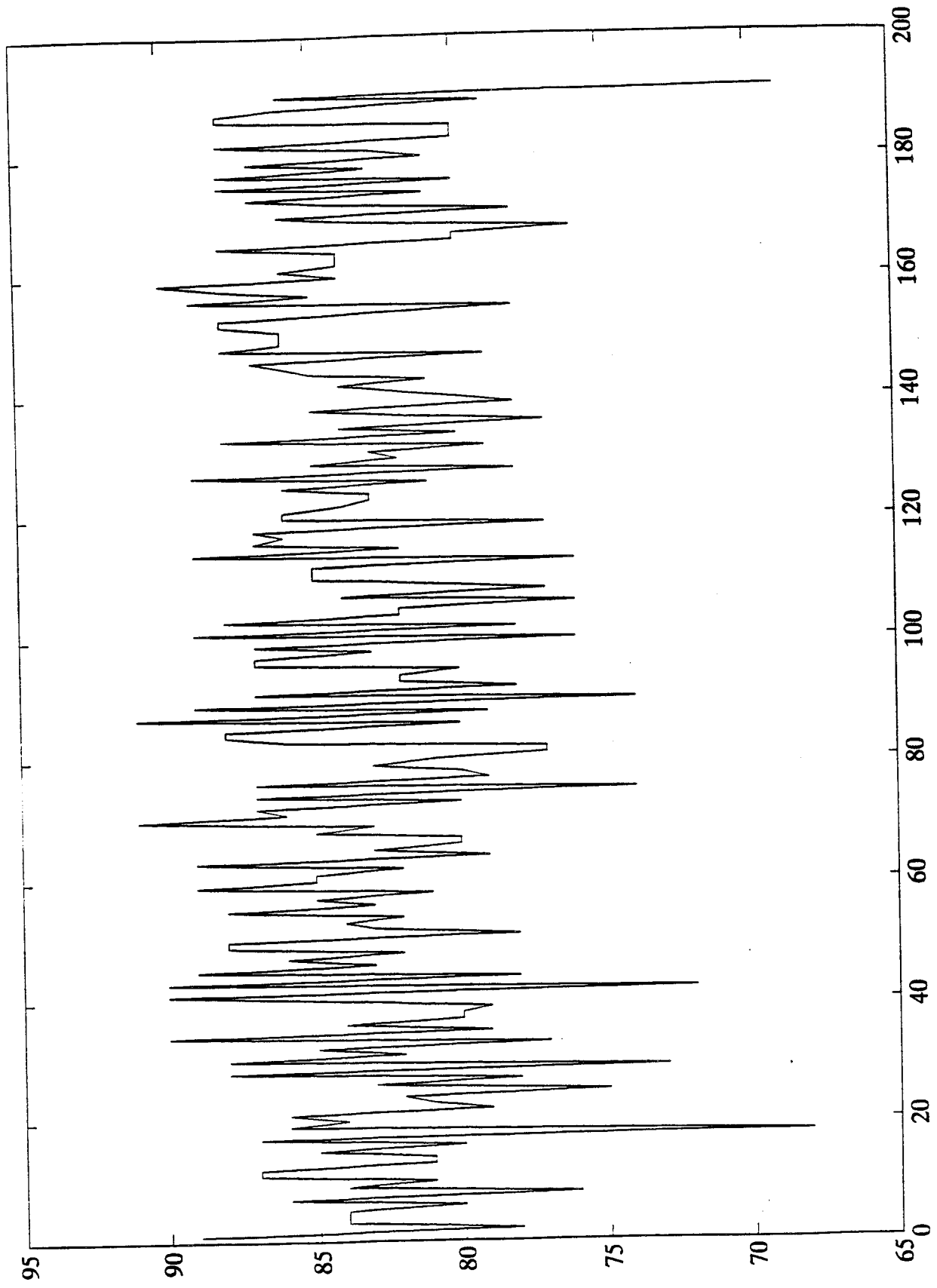
Figure 4. Distribution of x_1 in fixation.
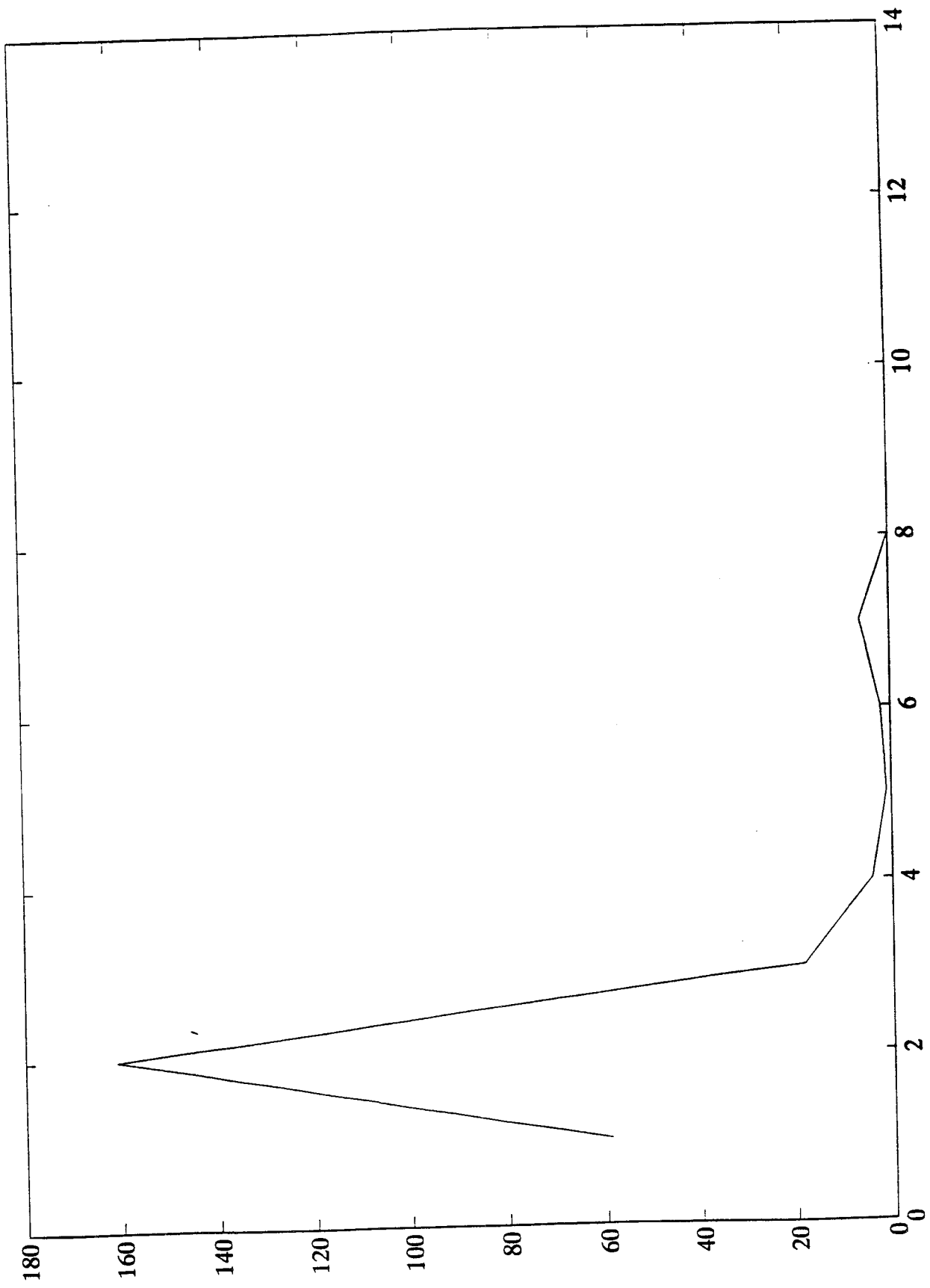
Figure 5. Distribution of x_2 in fixation.
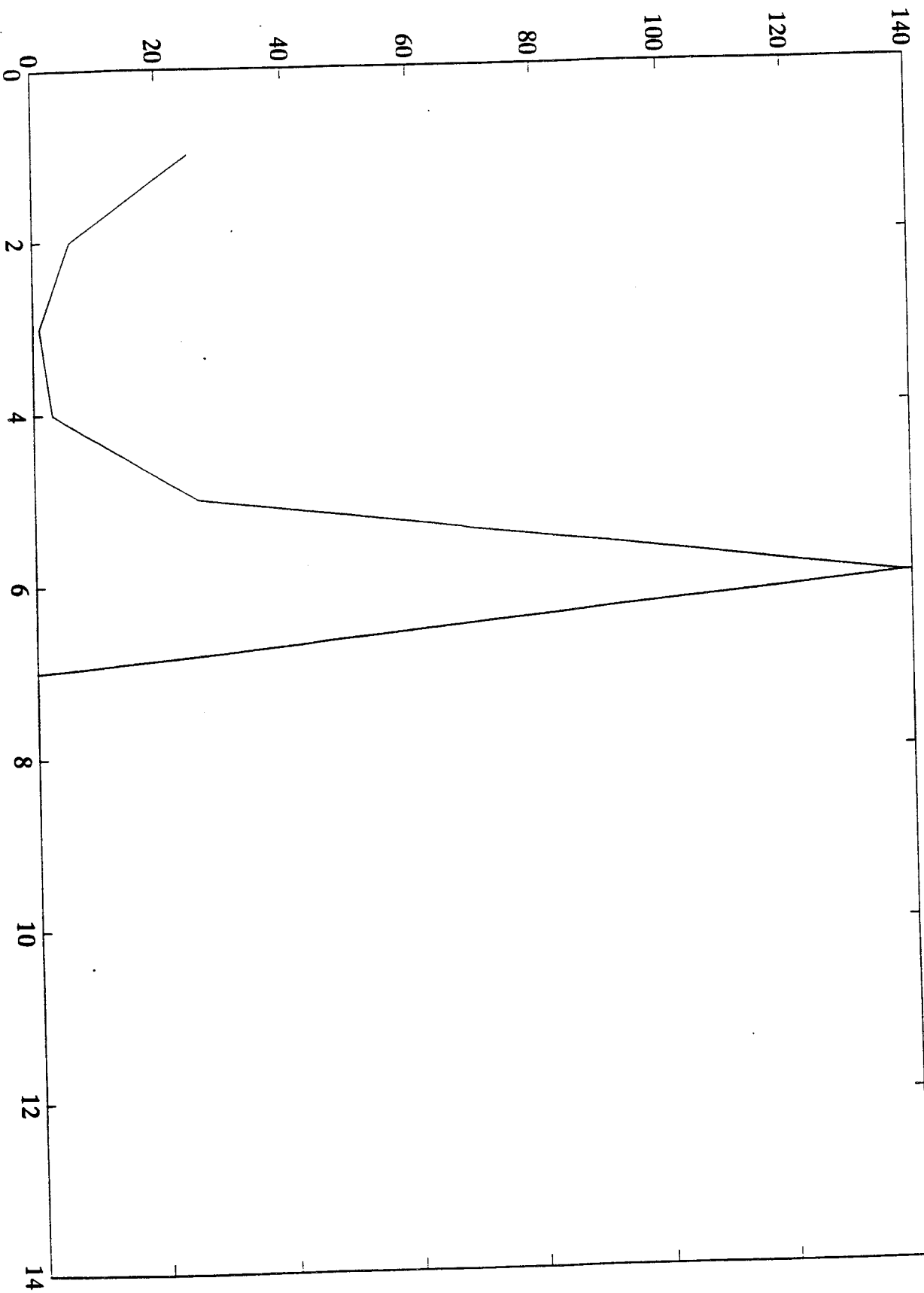
Figure 6. Eye blinking.
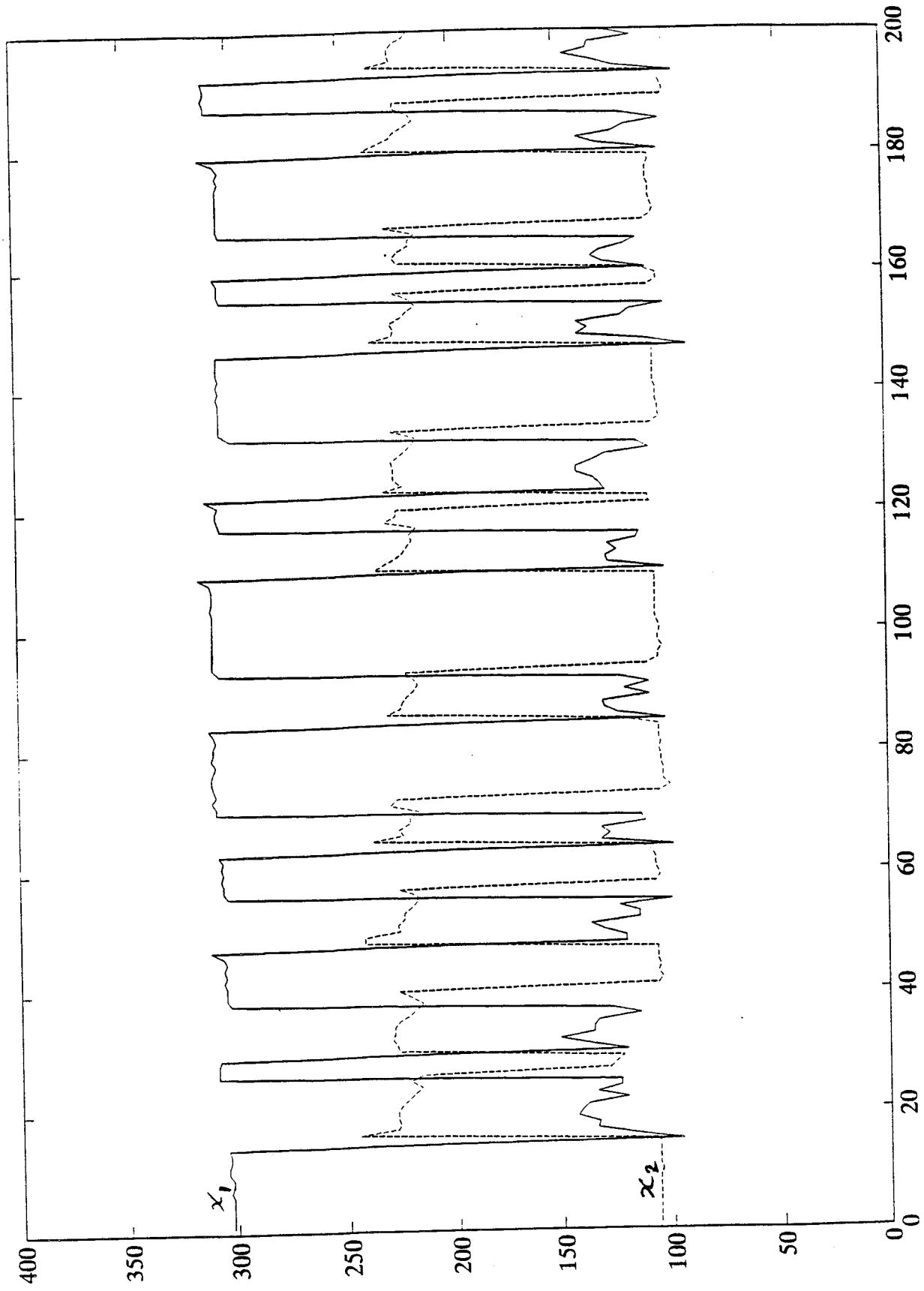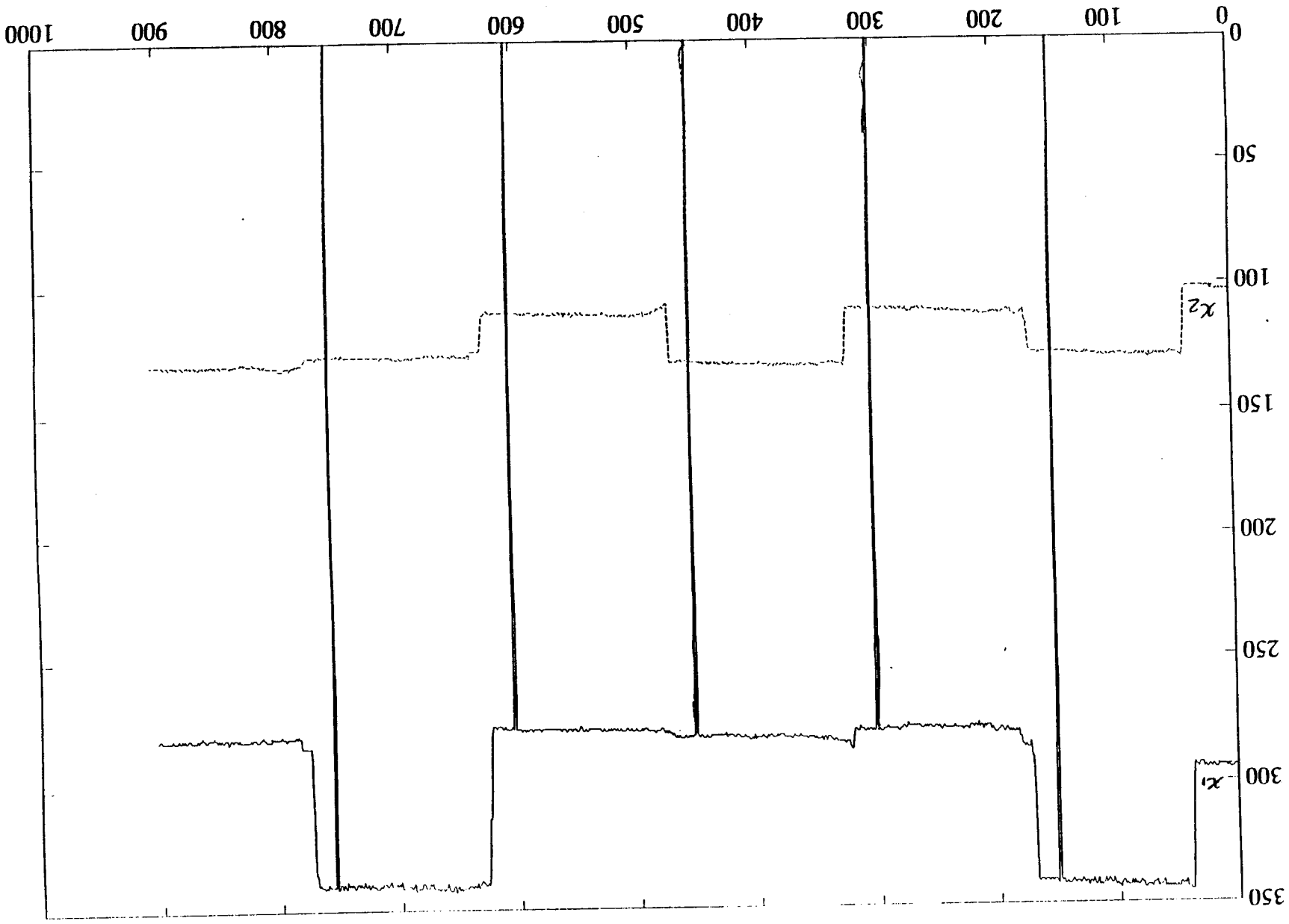
Figure 7. Eye transition.

Figure 8. Pupil horizontal in eye wandering.

Figure 9. Pupil size in eye wandering.
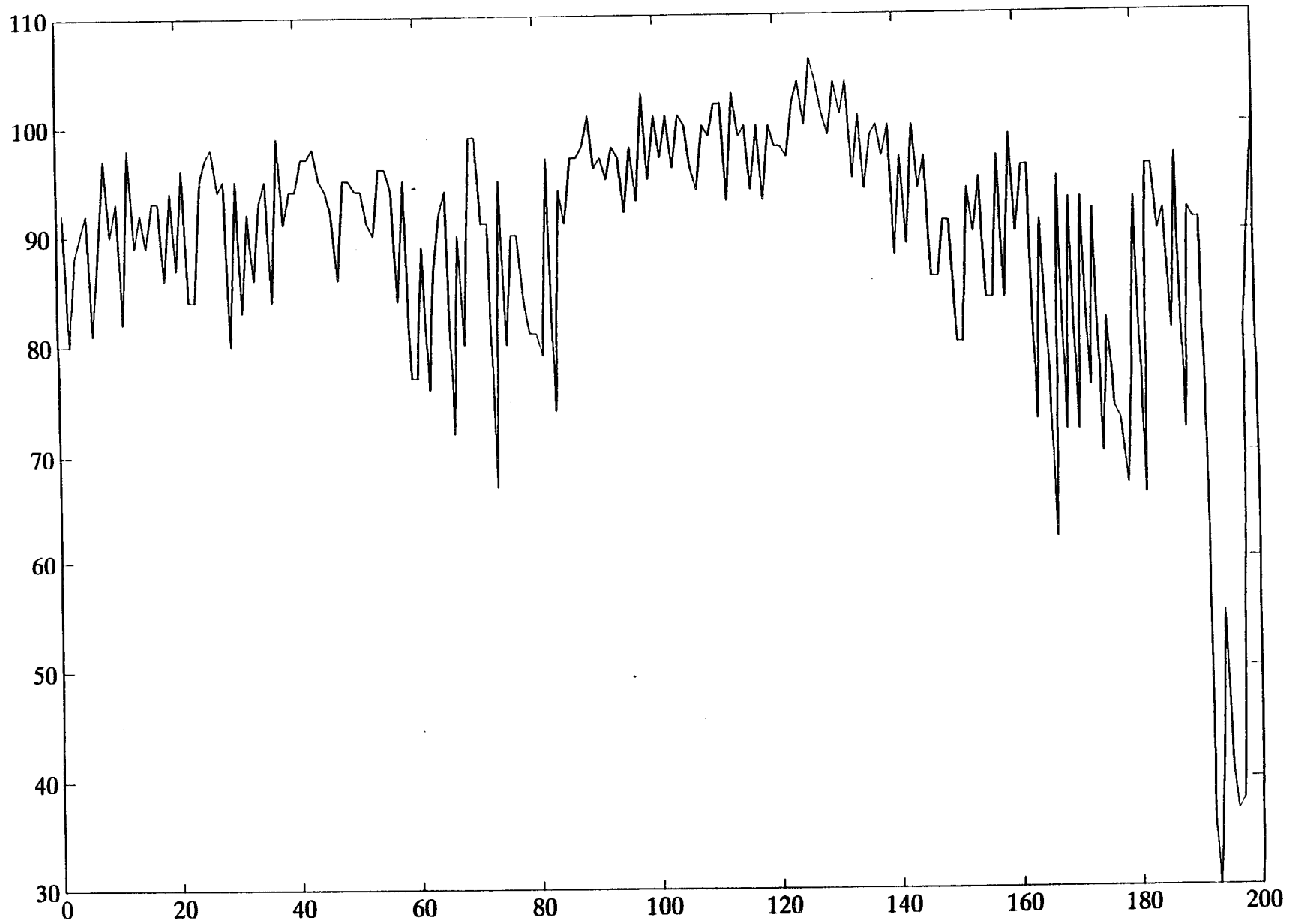
Figure 10. Distribution of pupil horizontal transition fixation.
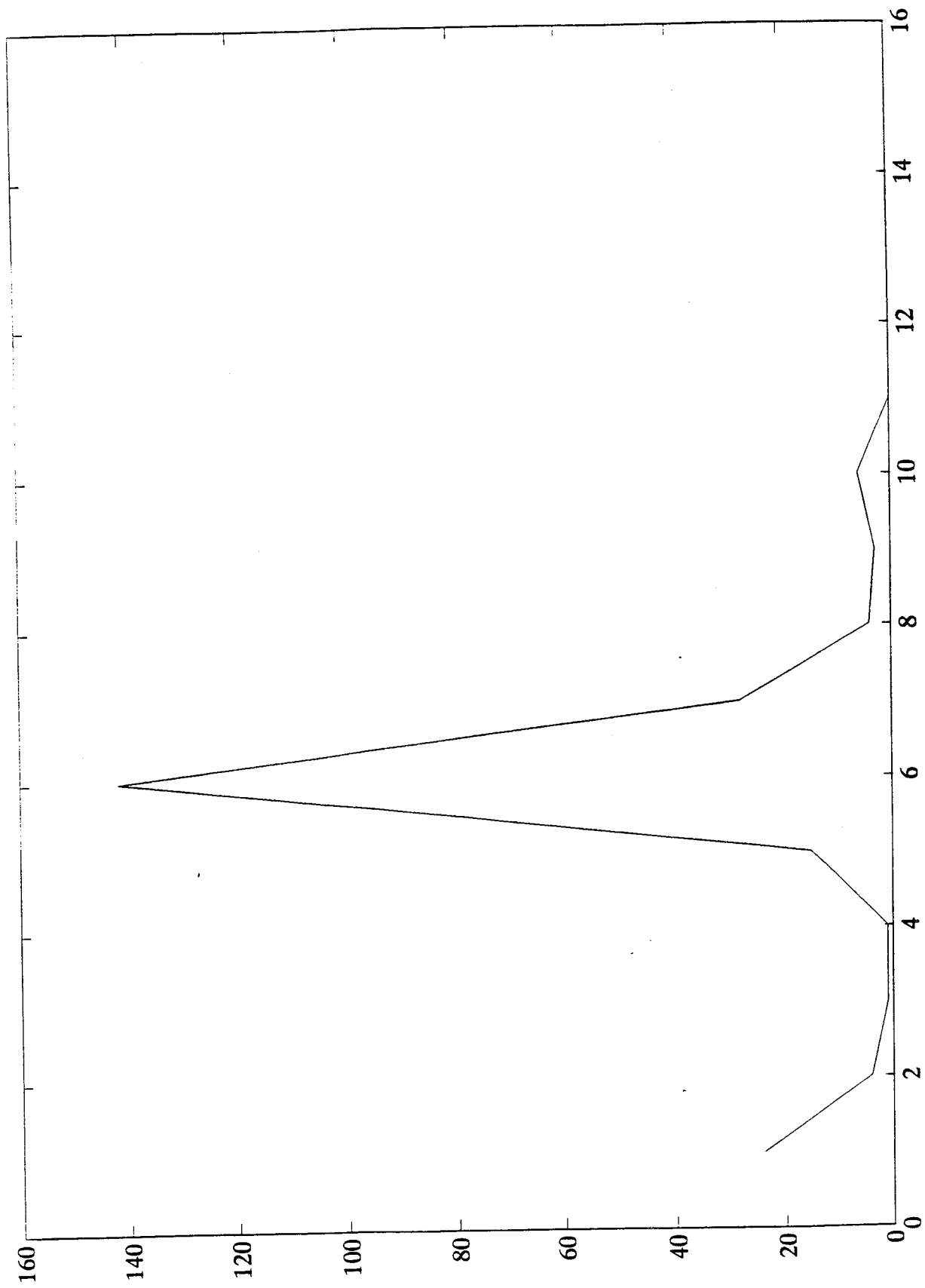
Figure 11. Distribution of pupil vertical transition fixation.
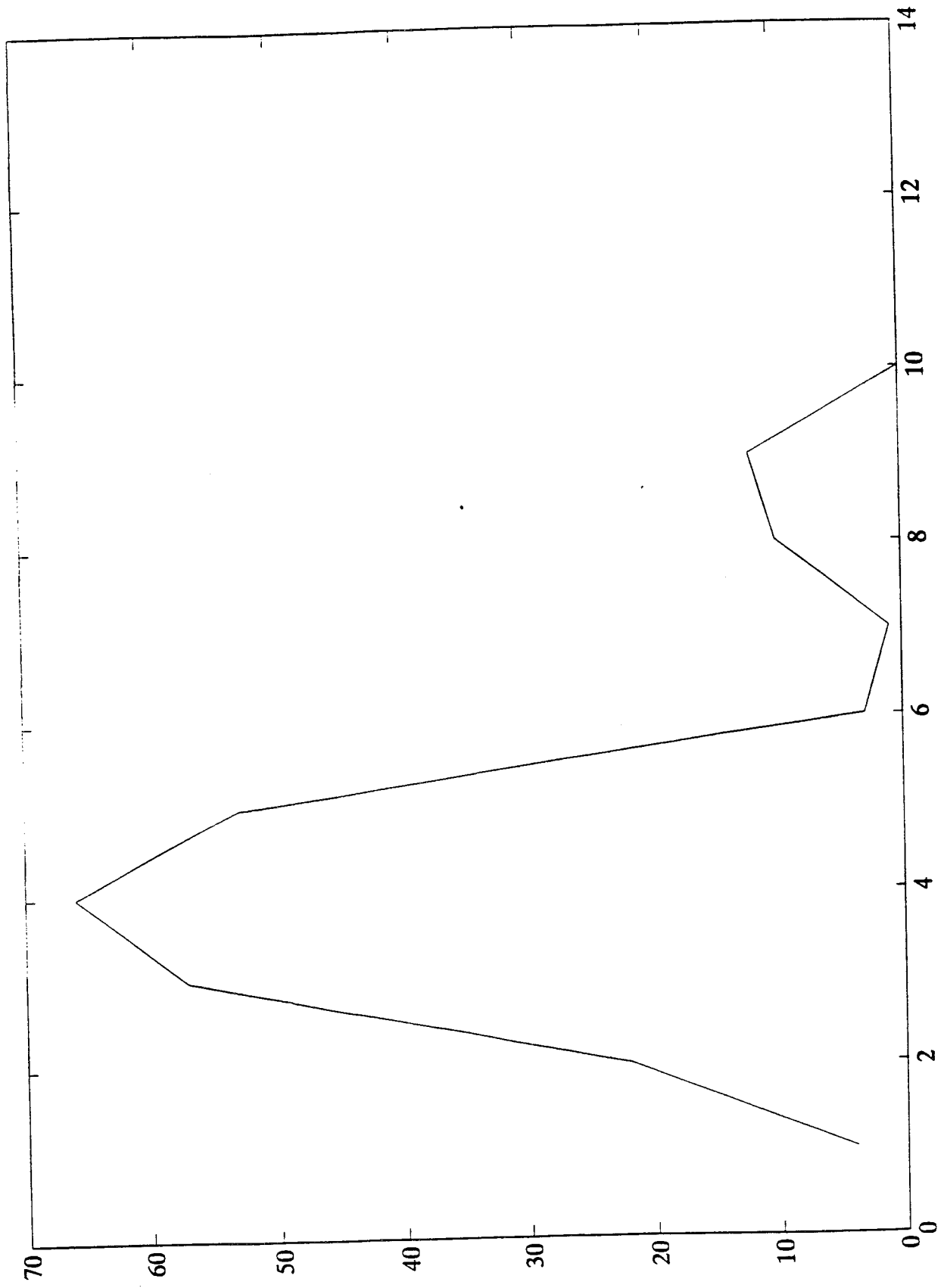
Figure 12. Linear mapping of pupil horizontal vs. actual.
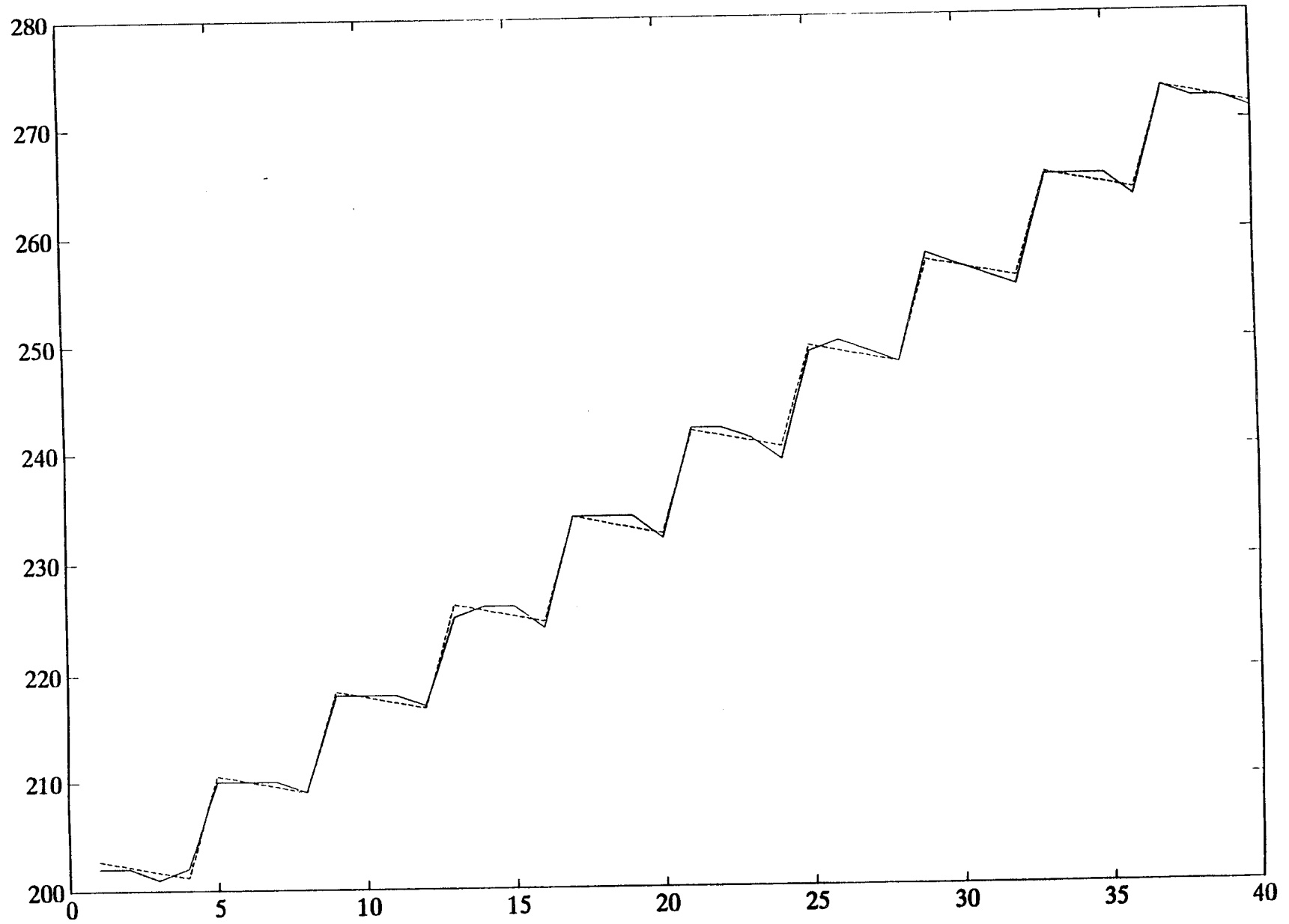
Figure 13. Linear mapping of pupil vertical vs. actual.
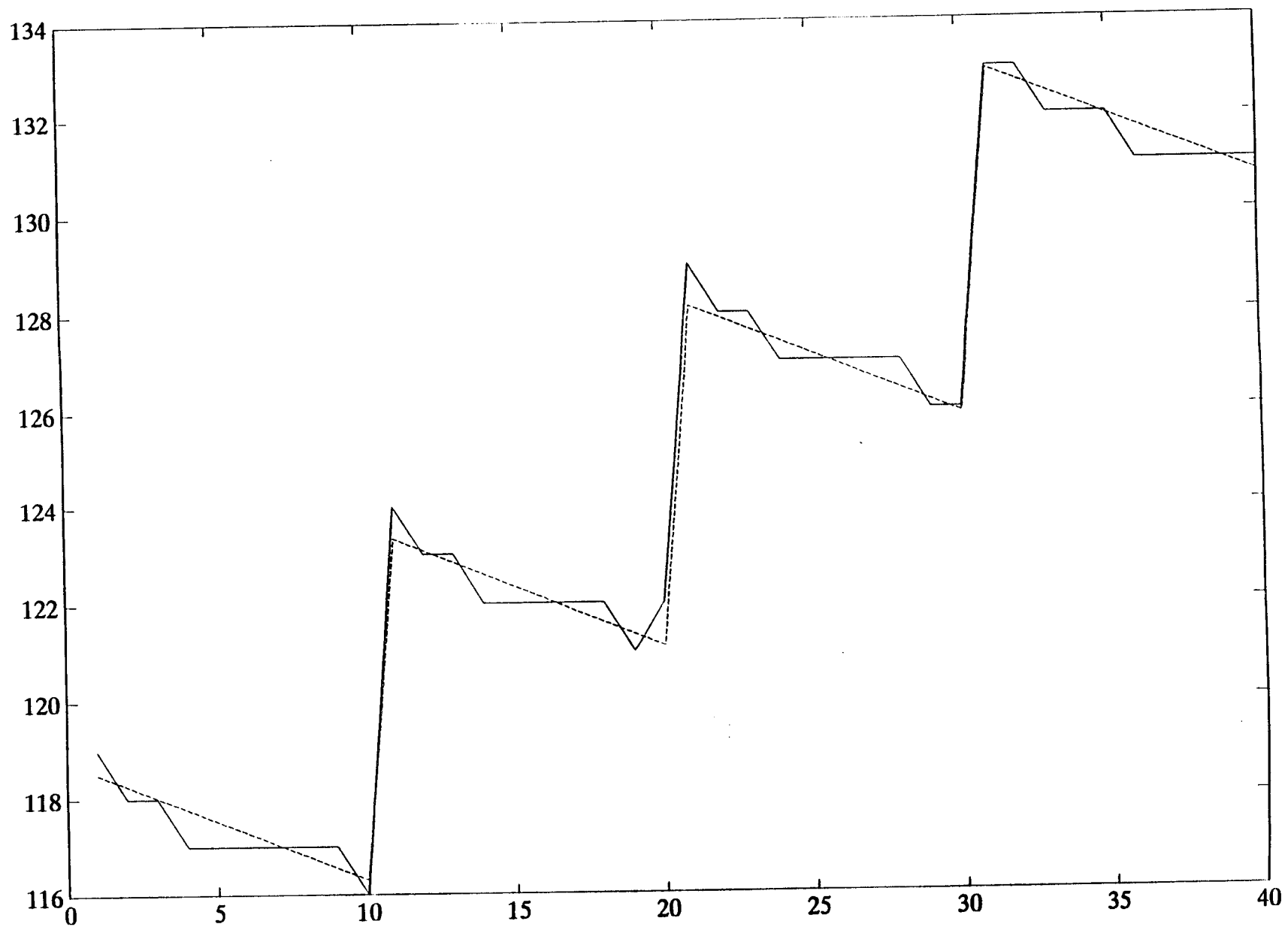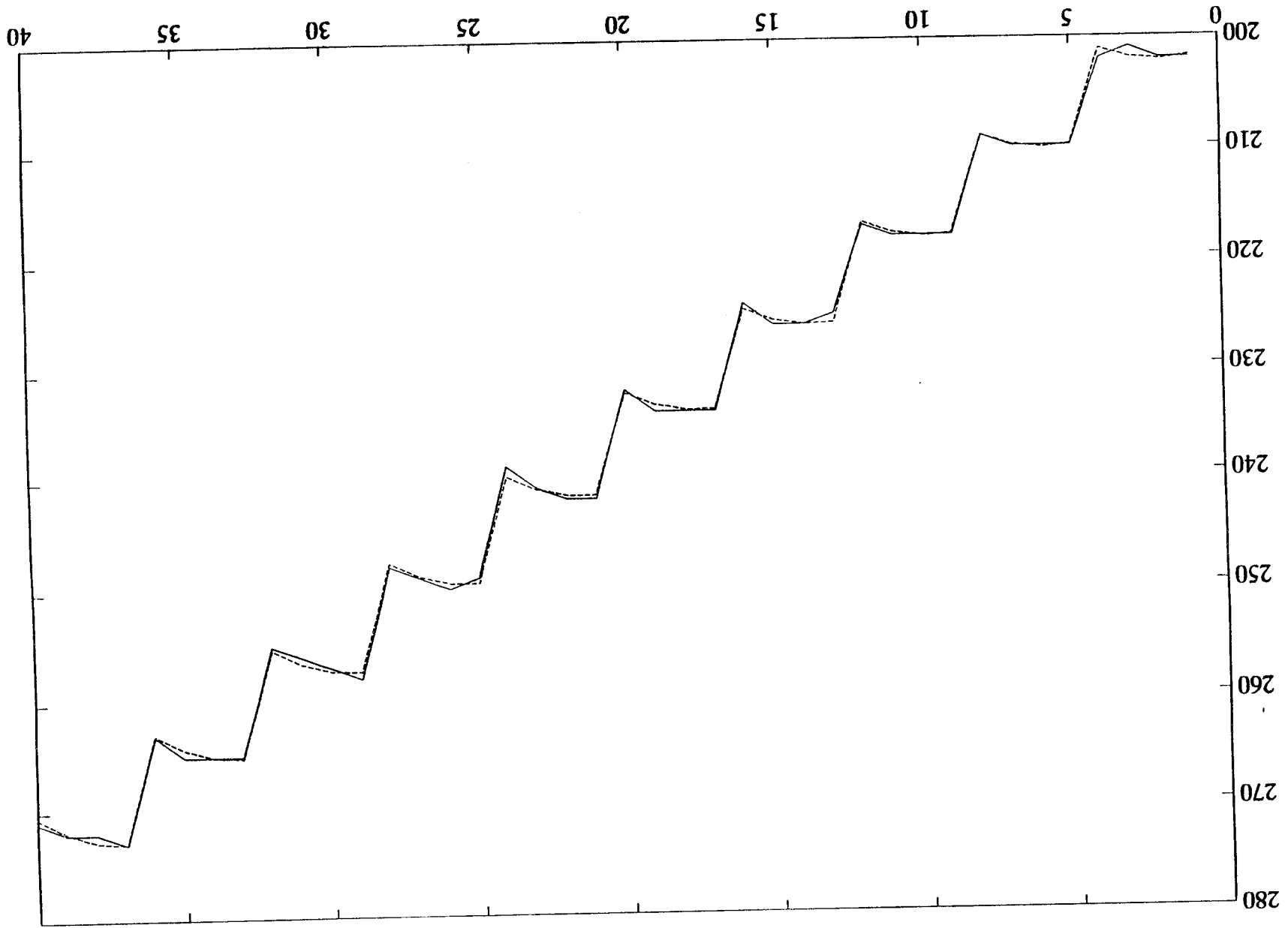
Figure 14. Non-linear mapping order-2 pupil horizontal vs. actual.
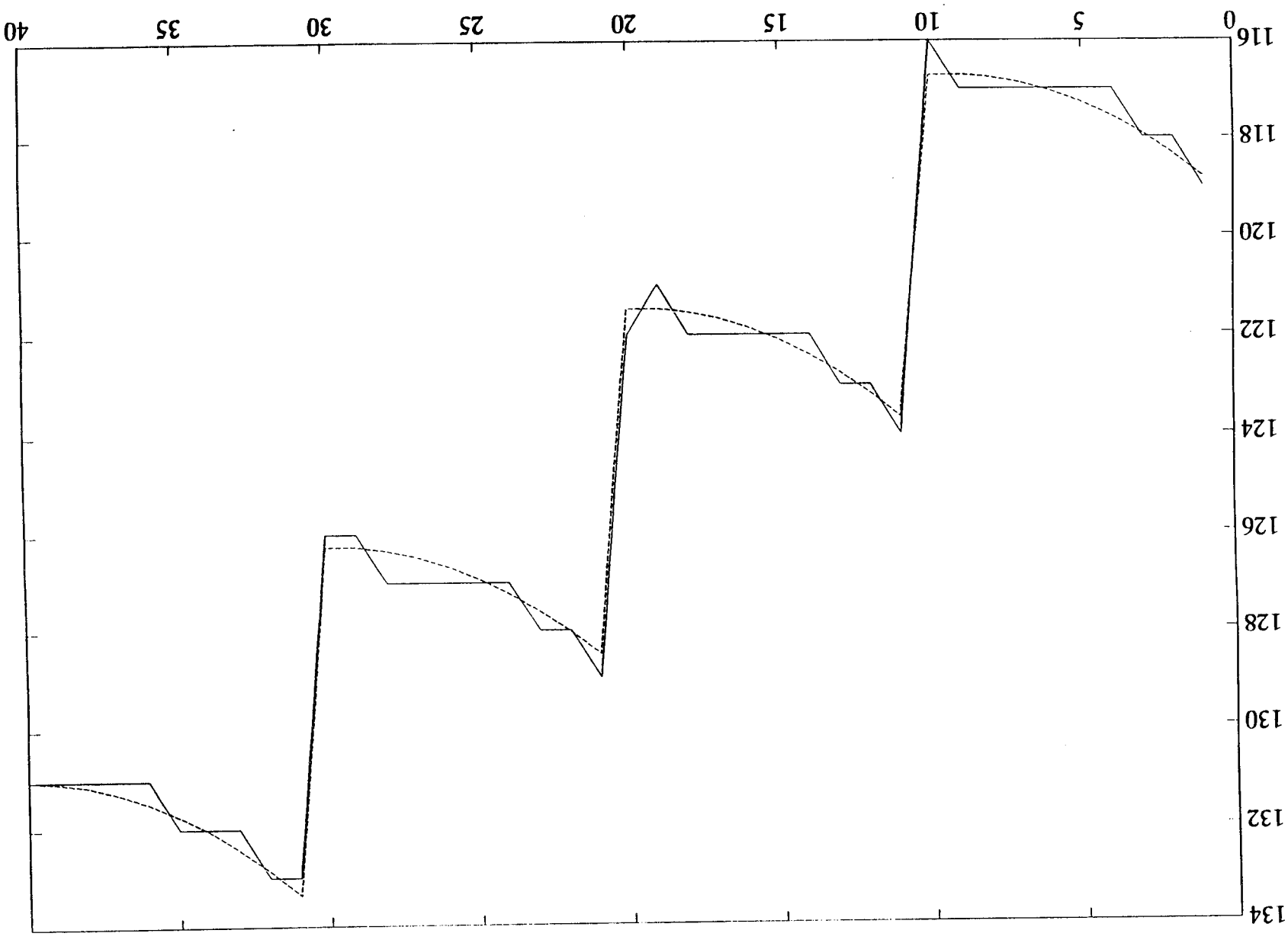
Figure 15. Non-linear mapping order-2 pupil vertical vs. actual.

Typing Models

User   **Computer**

```
┌──────────────┐              ┌──────────────┐
│   Dwelling   │              │    Sample    │
└──────────────┘              └──────────────┘
       │                              │
       ▼                              ▼
┌──────────────┐              ┌──────────────┐   N    ┌──────────────┐
│   Typing     │              │  Character   │───────▶│    Sample    │
│ Recognition  │              └──────────────┘        │    Delay     │
└──────────────┘                     │ Y             └──────────────┘
       │                              ▼
┌──────────────┐              ┌──────────────┐
│  Thinking of │              │  Highlight   │
│     Next     │              └──────────────┘
│  Character   │                     │
└──────────────┘                     ▼
       │                      ┌──────────────┐
       ▼                      │ Recognition  │
┌──────────────┐              │  Thinking/   │
│              │              │  Searching   │
│  Eye Motion  │              │    Delay     │
│              │              └──────────────┘
└──────────────┘
```
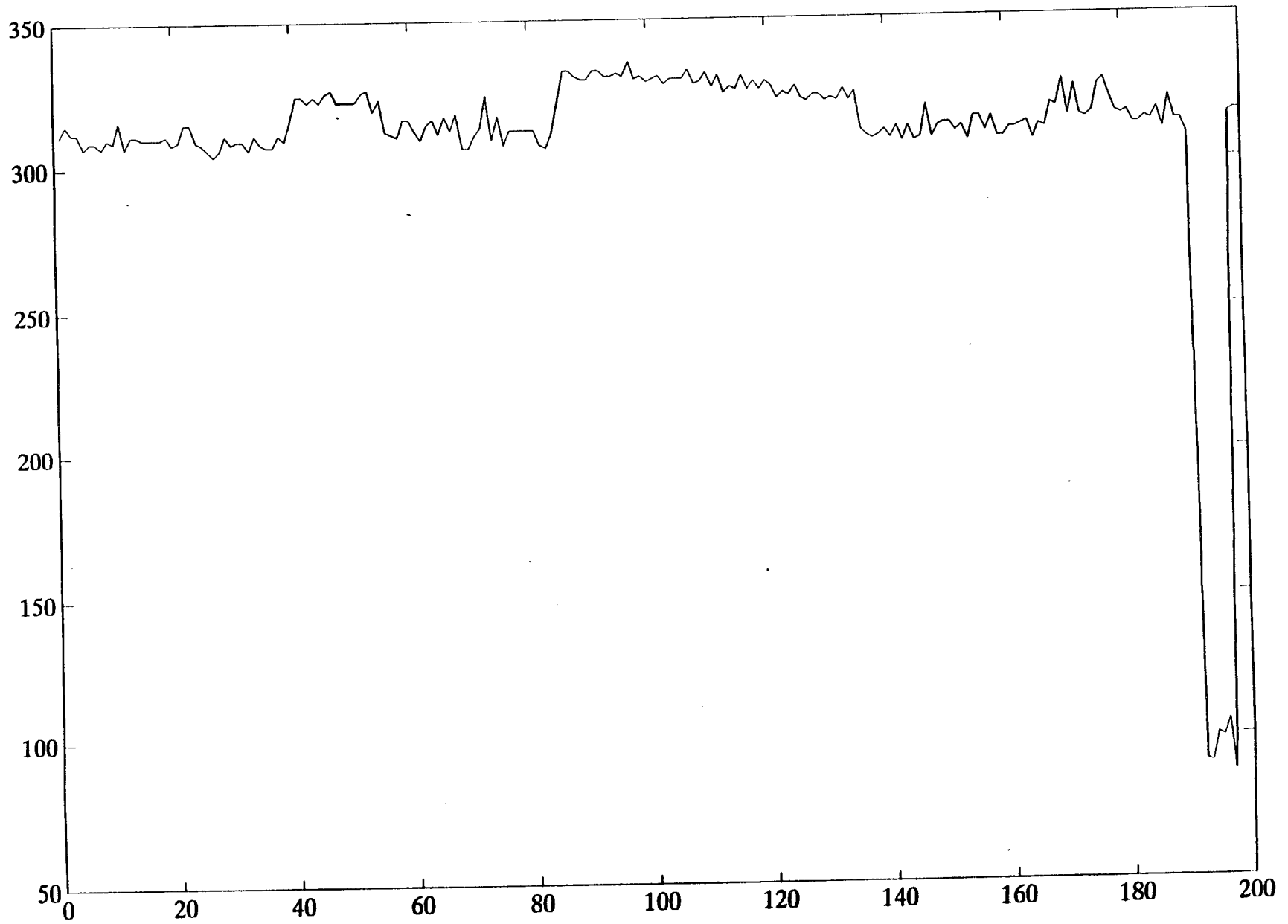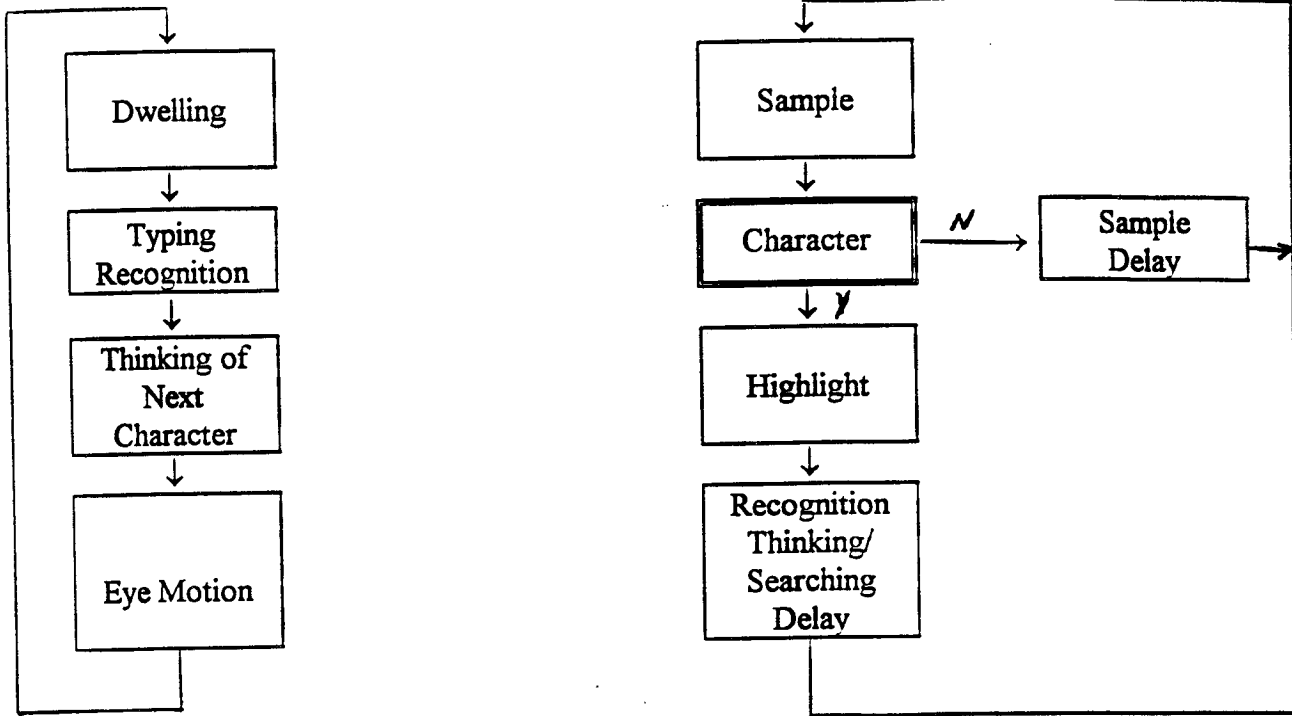
Recognition / Thinking / Motion

Average = 55 Samples ⇒ Delay of 900 ms

Appendix B

Typing Routine Chart

CNT = count
RNG = range
RN = running

```
                                    Setup work space
                                          ↓
                                  Initialize pointers &
                                       counters
                                          ↓
                                     Time Stamp
                                          ↓
                                  Get ISCAN data
                                          ↓
                              n
              ┌──────────────────────  In Range
              ↓                           ↓ Y
        Increment                       Reset
      OUT-Range-CNT                   OUT-RNG-CNT
              ↓                           ↓
  Clear - All ──  OUT-Range           Get nearest neighbor
              │   CNT > limit              ↓
              │       ↓                Update distribution list
              │      ③                     ↓
                                    Increment MXL-CNT
                                          ↓
        Reset              Y
   ① ← RN-MEAN-CNT ← ───────────  speed > limit
        RN-MEAN                            ↓ n
                                      Increment
                                     RN-MEAN-CNT
                                          ↓
                                    Compute new
                                     RN-MEAN
                                          ↓
                                         ①
```
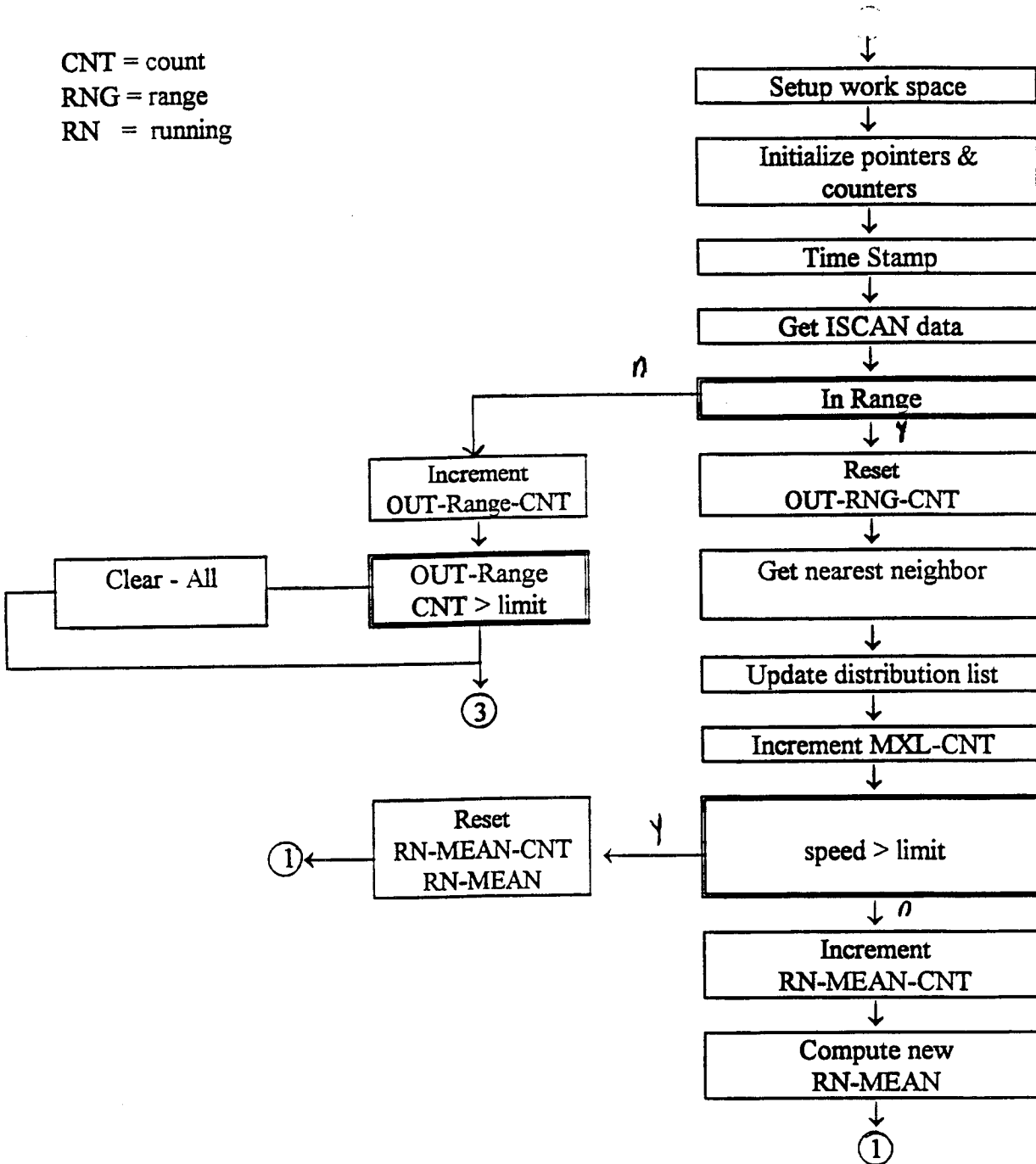
Typing Routine Chart

QLT = Quality Number
MX = Maximum
DWN = Dwelling Number
MXL = Maximum Likelihood
UNBL = Unable to Make Decision

① → Add Sample to Buffer/Increment Buffer-MX- Pointer

→ Run-Mean-CNT > DWN

Get Character ← Run-Mean-CNT > DWN

Get Character → Inititlize MXL counters

Inititlize MXL counters → Feedback Type

Feedback Type → MXL → MXL-CNT > DWN

Feedback Type → Sound + Colored Character

MXL-CNT > DWN — n → ③

MXL-CNT > DWN — Y → Get MXL-Char Compute MXL-QLTY

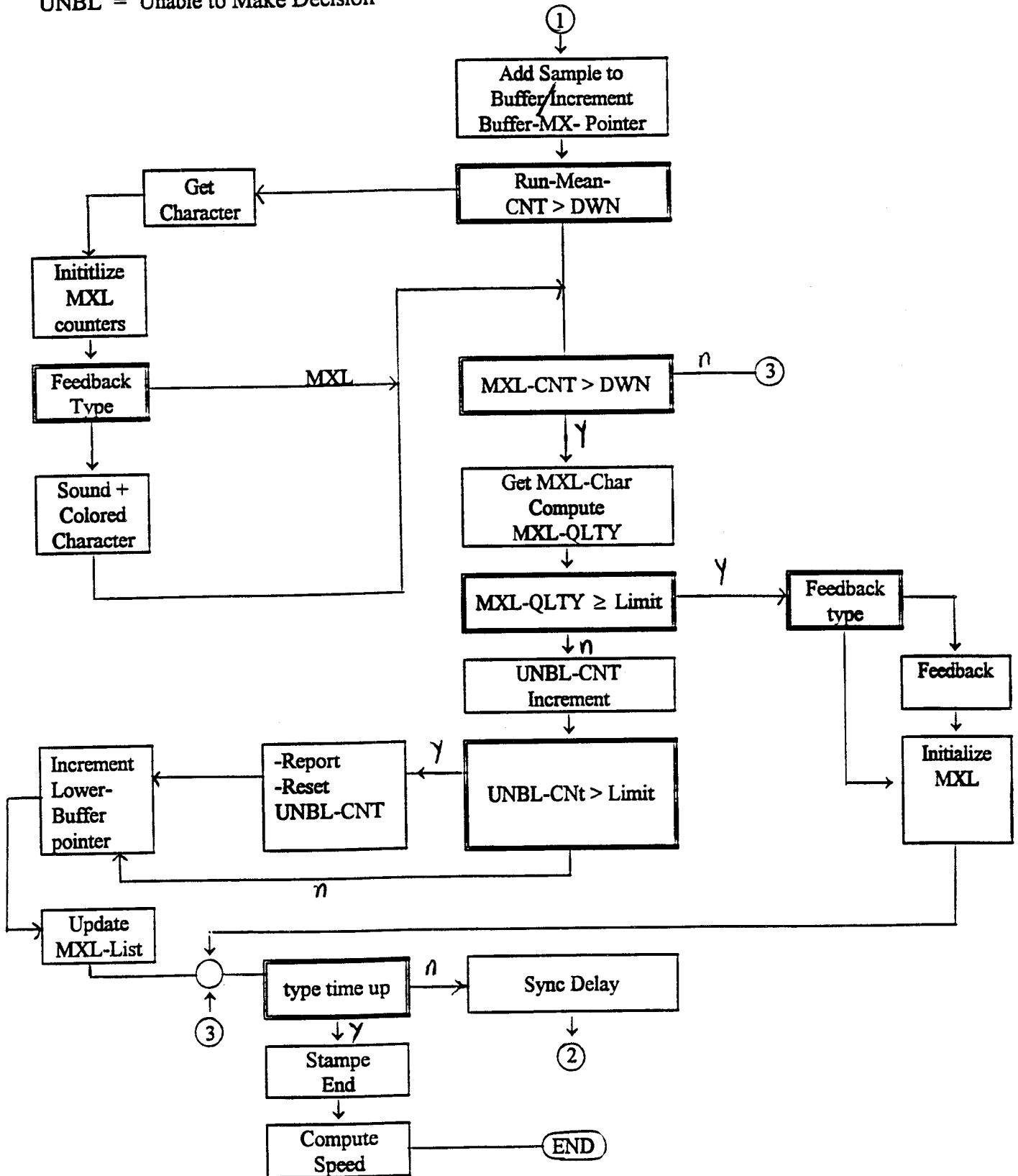Get MXL-Char Compute MXL-QLTY → MXL-QLTY ≥ Limit

MXL-QLTY ≥ Limit — Y → Feedback type

MXL-QLTY ≥ Limit — n → UNBL-CNT Increment

Feedback type → Feedback → Initialize MXL

UNBL-CNT Increment → UNBL-CNt > Limit

UNBL-CNt > Limit — Y → -Report -Reset UNBL-CNT

UNBL-CNt > Limit — n →

-Report -Reset UNBL-CNT → Increment Lower-Buffer pointer

Increment Lower-Buffer pointer → Update MXL-List

Update MXL-List → ○ → type time up

type time up — n → Sync Delay

type time up — Y → Stampe End

Sync Delay → ②

③ → ○

Stampe End → Compute Speed → END

Report Document Page

NASA
National Aeronautical and
Space Agency

1. Report No.	2. Government Accession No.	3. Recipient's Catalog No.

4. Title and Subtitle

F/NAS/Virtual Keyboard for Hands Free Operations

5. Report Due

9/5/96   FINAL REP

6. Performing Organization Code

University of Alabama in Huntsville

7. Author(s)

W. A. PORTER
A H Abouali

8. Performing Organization Report No.

10. Work Unit No.

9. Performing Organization Name and Address

University of Alabama in Huntsville
Huntsville, Alabama 35899

11. Contract or Grant No.

NAS8-38609 D.O. 158

13. Type of report and Period covered

FINAL

12. Sponsoring Agency Name and Address

National Aeronautics and Space Administration
Washington, D.C. 20546-001
Marshall Space Flight Center, AL 35812

14. Sponsoring Agency Code

NAS8-38609 DO 158

15. Supplementary Notes

SEE ATTACHED REPORT

16. Abstract

SEE ATTACHED (Paragraph #1)

17. Key Words (Suggested by Author(s))

Eye Tracking
Calibration

18. Distribution Statement

19. Security Class. (of this report)	20. Security Class. (of this page)	21. No. of pages	22. Price
Unclassified	Unclassified		

NASA FORM 1626 OCT 86

Enclosure #2