

Chapter 5

Testing the Biofeedback Pointer

5.1 Introduction

There are two basic questions about the Biofeedback Pointer which can only be answered empirically. The first is: Is the neural network adaptable enough that the Biofeedback Pointer can be used by different people? The second is: How well does the Biofeedback Pointer work compared to other graphic input devices? In order to answer these questions, a series of short tests were administered to 3 people to determine their indices of performance on a standard PC mouse and the Biofeedback Pointer. This would determine if the Biofeedback Pointer can adapt to different users and the range of performance one can expect from it.

In addition to the two main questions, there is the question of the structure of the neural network. Are there any trends in the weights from person to person, or even with the same person who has trained it more than once? Any trends discovered could be used to improve the neural network and thereby the performance of the Biofeedback Pointer. Comparing the networks generated by the subjects in the tests might show some potentially exploitable patterns.

5.2 Materials

5.2.1 The Mouse

The mouse used in this experiment was a two-button MicrosoftTM serial, PS/2 compatible mouse. The mouse was used in the right hand on a mouse pad. All input was performed using the left button.

5.2.2 The Biofeedback Pointer

The Biofeedback Pointer was used as described in Section 3.3, with one exception: The device was designed only for moving a pointer, not for selection. Selection is intended to ultimately be done by using the shift buttons on the Chording Glove, but this was impractical for the performance tests, since it would provide much more hardware and software than would be necessary. Instead of using the Chording Glove, a mouse was held in the subject's left hand and was used to activate the buttons. Since the mouse was held, and not placed on a surface, it had no effect on the pointer's motion. This was found to be sufficient for the tests.

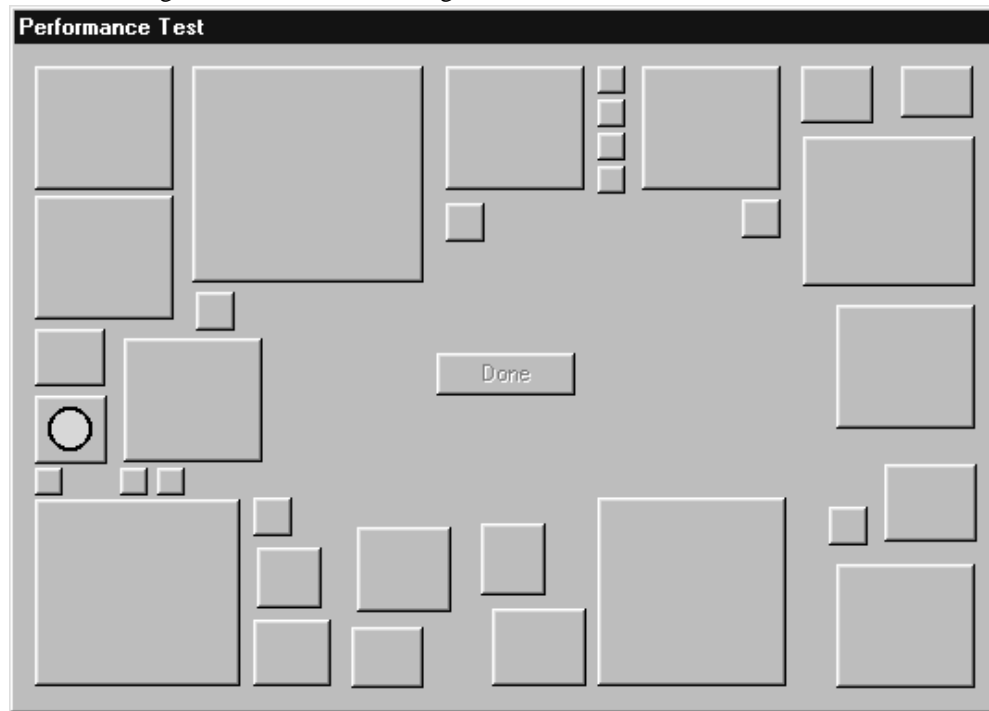
3M Red DotTM pediatric electrodes were used in the tests. In preparation for use, the last $\frac{1}{2}$ cm of the electrode was cut off to fit properly in the amplifier. One of each pair of electrodes was marked with the number of the corresponding amplifier.

5.2.3 The Performance Test Window Setup

The performance test was run using the EMG pointer software (see Appendix H) on a Pentium 133MHz computer in the Windows 95 environment. EMG data was read and filtered in software during both the mouse and Biofeedback Pointer trials. Even though this data was ignored in the mouse trials, calculating it ensured that any software delays brought about due to the computationally intensive calculations of the data filtering would be the same for both input devices.

The test window consisted of 30 near-square buttons of various sizes and positions (Figure 5.1). In addition there was one button labelled "Done" in the middle, which was disabled throughout the trial.

Figure 5.1: The button arrangement in the Performance Test window



When the trial was finished, the “Done” button would enable, requiring the subject to press it to end the trial.

In the Windows 95 environment, a click on a button widget is registered only when the pointer has stopped on the button. If the pointer is moving while clicking on a button, the focus is changed to the button, but it is not activated. There are few problems with this when using a mouse, because, in practice, the user always stops moving before clicking on a button. The mouse trials kept this standard method for button activation.

The Biofeedback Pointer could not activate buttons in the above manner. When using the Biofeedback Pointer, the user merely slows down when clicking on a button. Using the original method, the button would only become the focus, requiring the user to make a second or third pass before slowing down enough to have the click activate the button. As a consequence the behaviour of the environment had to be changed to accommodate the Biofeedback Pointer. The buttons in the performance test window were adjusted so that any click on a button would register as a full press and activate the button.

5.3 Method

The experiment consisted of two parts. In the first part, at least 2 trials were performed using the mouse. From these trials the mouse’s index of performance was calculated. These same trials were repeated using the Biofeedback Pointer to calculate its index of performance.

The mouse trials consisted of just the performance test, as described in Section 5.3.2 below. No training or practice was necessary. Before the Biofeedback Pointer trials, the subject would need to train the pointer’s neural network and then practice using the pointer. Training was done by the method described

in Section 3.3.4. After training, the subject would practice using the pointer until they stated that they felt comfortable with it. During the practice, if the subject was not satisfied with the performance, they could retrain the pointer. However, the subject was not allowed to retrain after the first trial had started. This was to ensure that the same neural network was used for each trial. Each subject performed at least four trials.

5.3.1 The Subjects

There were three subjects used in the tests. Two of the subjects were female and one male. Different sexes were used because there are significant differences in amplitude and frequency distribution of EMGs between males and females (Cioni et al., 1988). All of the subjects were right handed and the right arm was used in all trials with the mouse and Biofeedback Pointer. In addition all the subjects had used the Biofeedback Pointer at least once before these tests. This was to help minimise the effects of the subject being unfamiliar with the device. Subject C was a computer science graduate student who had used the Biofeedback Pointer once before. Subject B had only a passing familiarity with computers and had also used the Biofeedback Pointer once before. Subject A was the author and had used the Biofeedback Pointer many times before.

Daily Preparation

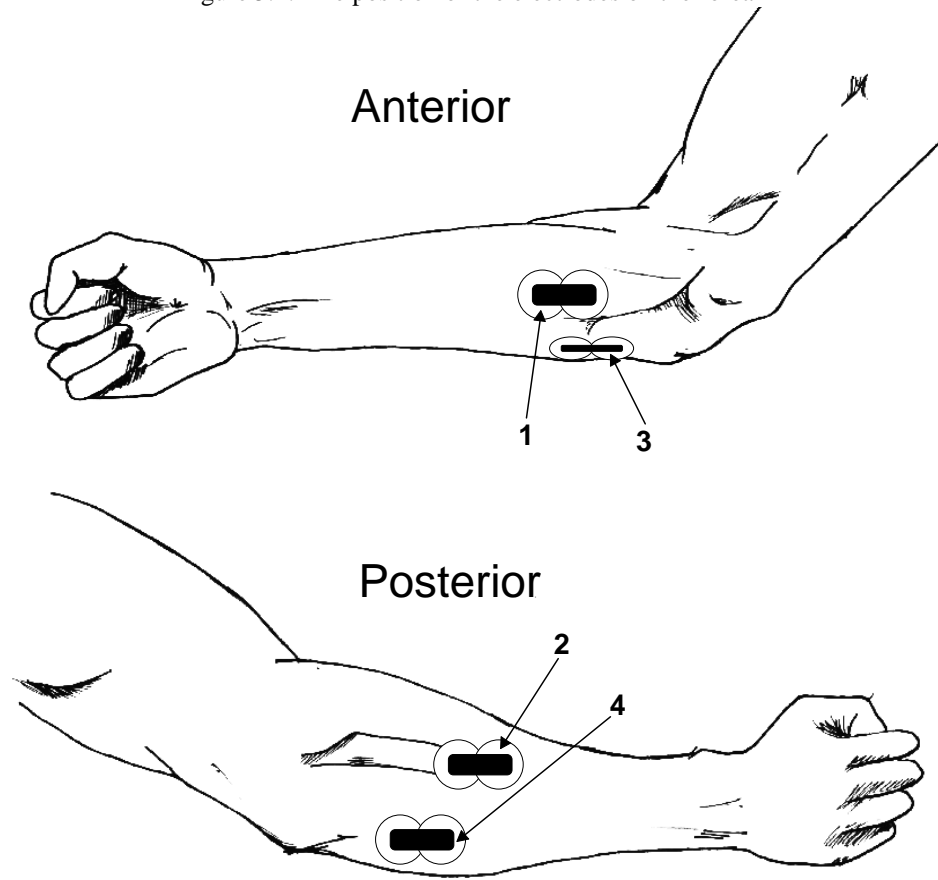
Before each day of testing, the battery was tested. If it was significantly less than 9 volts it was replaced with a fresh one. This was to ensure that the amplifiers were giving their maximum performance. The next step would be to prepare the electrodes by clipping the ends and numbering them. After that the EMG collection hardware would be turned on and attached to the ADC on the parallel port. The last step before attaching the electrodes was to load and run the EMG Pointer software.

The electrodes were attached to the right arm in their standard locations (see Section 3.3.3). The reference electrode was placed over the acromium process in the right shoulder. The other electrodes were placed in the following locations: 1 - flexor carpi radialis, 2 - extensor digitorum, 3 - flexor carpi ulnaris, 4 - extensor carpi ulnaris (Figure 5.2). The electrodes would be attached by the following procedure:

1. The data collection on the EMG Pointer software is turned on. This displays the input from each of sensors on the monitor (see Appendix H for more details).
2. The subject is asked to place the reference electrode over “the bony part” of their right shoulder. The subject was allowed to do this because exact positioning of this electrode is not important.
3. The experimenter attaches each electrode in turn by palpating the muscles in the subject’s arm and then positioning the electrodes to give the best reading of the EMG displayed on the computer monitor.

The component box was usually attached to the subject’s belt to keep it secure and out of the way. If the subject did not have a belt, they usually placed it in their lap while performing the test. This was not a problem as the subject would be sitting for the duration of the test.

Figure 5.2: The position of the electrodes on the forearm



5.3.2 The Performance Test

The performance test is made up of a series of button-pressing tasks of varying levels of difficulty. In each task, one of the blank buttons, selected at random is marked by a large yellow dot accompanied by an audible beep. When this occurs, the subject moves the pointer over the button and presses it as quickly as possible. The test then pauses for a random amount of time between 0.5 to 2 seconds before the next task. This is repeated 40 times. When the last task is completed, the “Done” button enables, accompanied by a beep. When the subject presses it, the window closes and the trial is over.

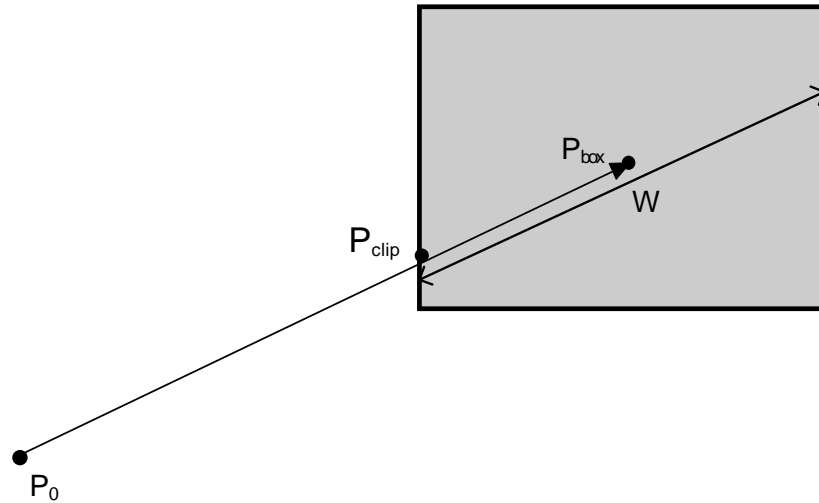
The software saves the beginning and ending time of each task, along with the entire path of the pointer during the test. In addition the neural network used for each subject is saved.

5.4 Results

5.4.1 Training

Each subject started to train the Biofeedback Pointer as soon as they were connected. After it was trained they were told to move around the pointer and see if they were happy with how it worked. If they were unhappy with it, they were allowed to retrain. Subjects B and C retrained twice, and subject A retrained three times before being satisfied with the neural network.

Figure 5.3: Calculating the effective width of the target. P_0 is the starting position, P_{box} is the centre of the target. The vector from P_0 to P_{box} is the approach vector. The point at which the approach vector intersects the target is P_{clip} . W is twice the distance from P_{box} to P_{clip} .



5.4.2 Index of Performance

Calculating W

The width (W) in Fitts' law is a one-dimensional term. The buttons in the performance test were two-dimensional with varying aspect-ratios ranging from 0.90 to 1.39¹. MacKenzie (1992) addresses this problem and describes a few methods for determining the effective width. One common method described in the paper is to take the larger of the height or the width. This method was deemed too simplistic, despite its popularity and was passed over in favour of the method of using the span of the target along the approach vector. The approach vector is defined as the line from the position at the start of the task to the centre of the button. In this method, described in Figure 5.3, the approach vector is clipped against the button's edge. Twice the distance from the clipping point to the centre of the button is W . The specific details of the calculation are described in Appendix E.

Mouse

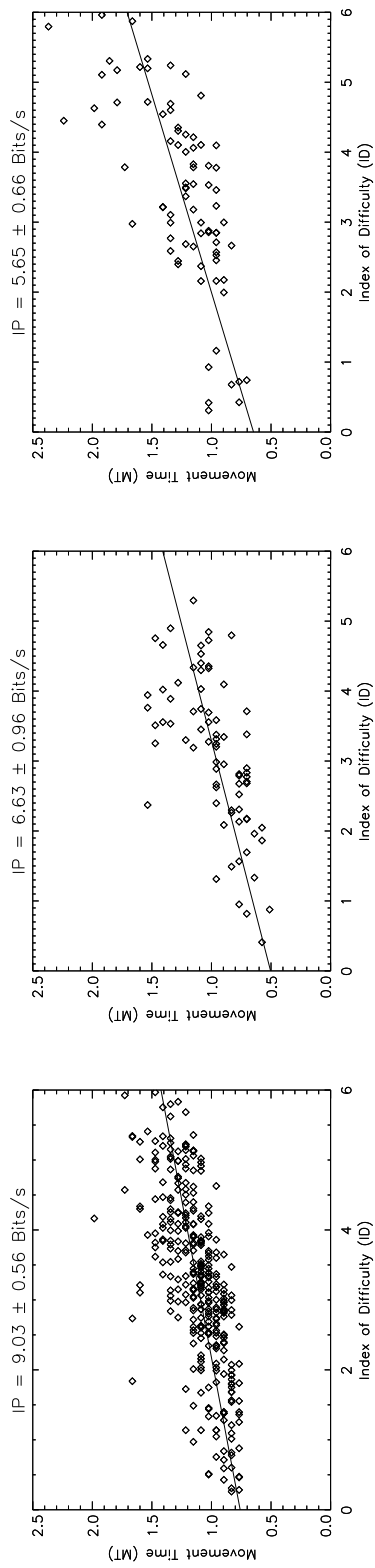
The Index of Performance (IP) for the mouse was calculated by the method described in Section 2.3.1. This consisted of finding the line which best fits the plot described by the movement time (MT) plotted against the index of difficulty (ID). IP is calculated by the inverse of the slope of the line. The y-intercept (a) corresponds to the reaction time.

The values for IP were 9.03 ± 0.56 , 6.63 ± 0.96 , and 5.65 ± 0.66 , which average to 7.10. The reaction times were (in ms): 504 ± 72.0 , 647 ± 76.0 , and 759 ± 24.1 , which average to 637. These can be seen in Figures 5.4(a)–5.4(c).

The values for the mouse described in MacKenzie (1992) range from $a = 1030$ and $IP = 10.4$ to

¹Except for the "Done" button which has an aspect ratio of 3.26

Figure 5.4: Index of performance for the mouse. The vertical grouping is an artifact caused by the fact that MT is measured in multiples of $1/16$ th of a second.

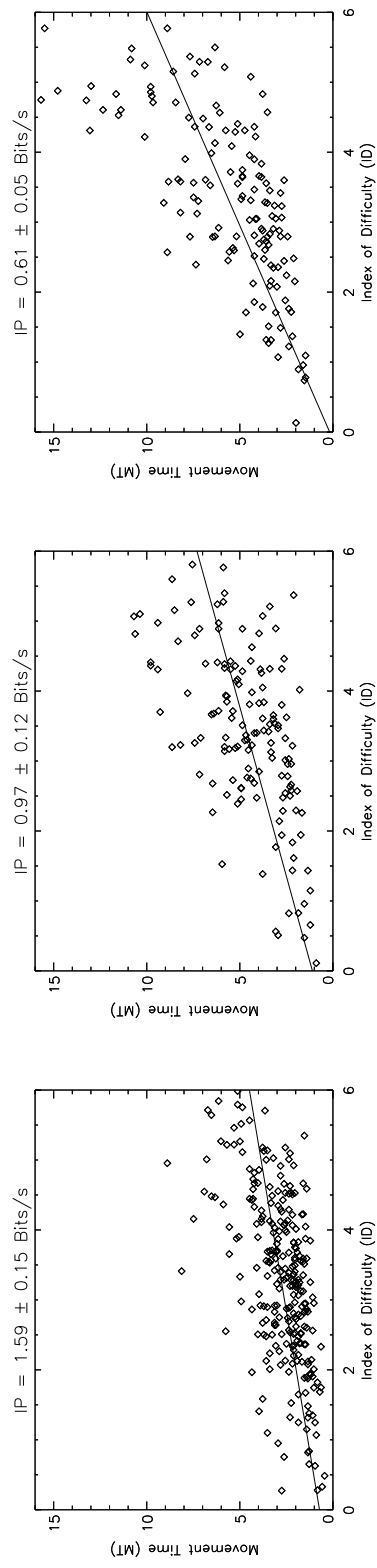


(a) Subject A

(b) Subject B

(c) Subject C

Figure 5.5: Index of performance for the Biofeedback Pointer



(a) Subject A

(b) Subject B

(c) Subject C

$\alpha = 108$ and $IP = 2.6$. The values for IP and α from the experiment fit into this range well enough to be considered valid. The exact value of IP for the mouse is not as important as its relation to the IP for the Biofeedback Pointer.

Biofeedback Pointer

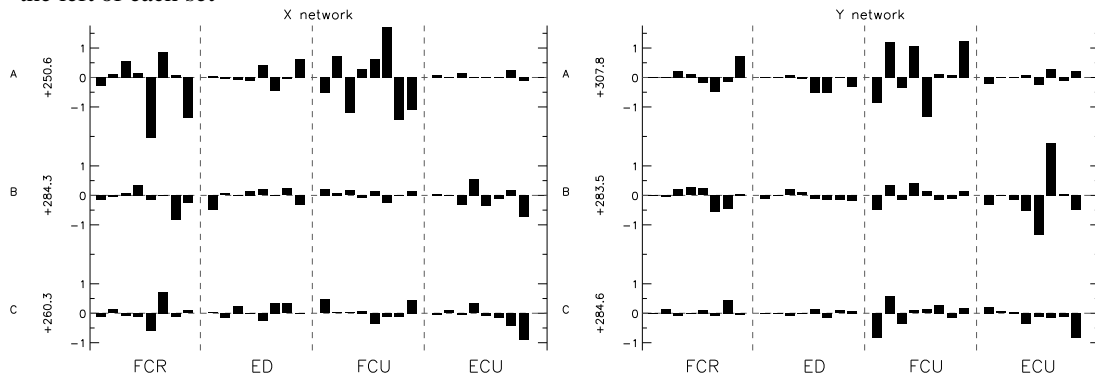
The IP for the Biofeedback Pointer was calculated in the same manner as for the mouse. The plots for each subject can be seen in Figures 5.5(a)–5.5(c). Table 5.1 summarises the values of IP and α for each subject. As the table shows, the IP for the Biofeedback Pointer ranges from 11% to 18% of mouse's IP , averaging at 14%. The y-intercept value tends to vary much more, ranging from 27% to 219% of the mouse's value.

Table 5.1: Comparison of performance results for the Biofeedback Pointer and mouse

Subject	Mouse				Biofeedback Pointer				Ratios			
	IP_M		a_M		IP_B		a_B		IP_B/IP_M		a_B/a_M	
	mean	σ	mean	σ	mean	σ	mean	σ	mean	σ	mean	σ
A	9.03	0.56	759	24	1.59	0.145	720	200	0.177	0.019	0.949	0.265
B	6.63	0.96	504	72	0.967	0.118	1101	454	0.146	0.028	2.186	0.953
C	5.65	0.66	647	76	0.611	0.053	174	503	0.108	0.016	0.270	0.779
Average	7.10	0.75	637	62	1.06	0.112	665	408	0.144	0.012	1.135	0.420

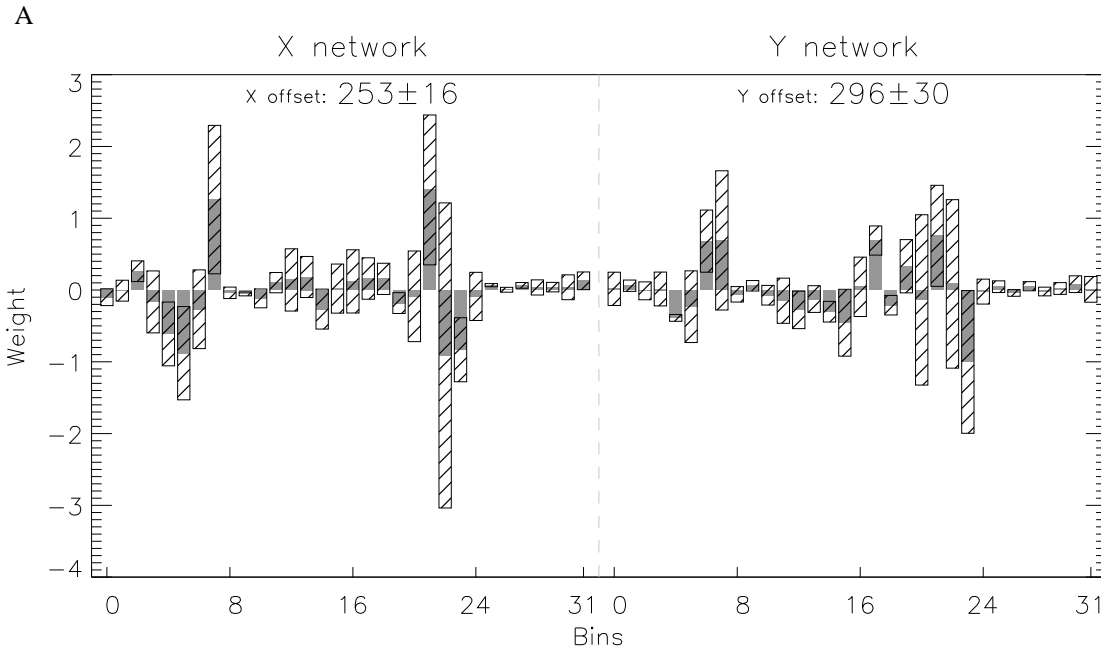
5.4.3 The Neural Networks

Figure 5.6: The neural networks weights used in the experiment for each subject. The offset is shown on the left of each set



Each neural network consists of 66 weights. One set of 33 weights for the x network and the other 33 weights for the y network. The first value in both sets of 33 weights is the offset value, which is added to the sum of the product of each weight and its input. The remaining 32 values consist of 4 groups of 8. Each group corresponds to one sensor, or more specifically, one muscle. The 8 values in each group are the weights for the power of the muscle's EMG spectrum which falls in the corresponding frequency bin. The value of all the weights for each subject are shown in Figure 5.6.

Figure 5.7: The average and standard deviation of the weights of 4 neural networks generated by subject



In examining the weights we are looking for three potentially exploitable trends. The first is a consistency over the entire set of weights from one session to another, or one person to another. If the variation in the value of weights is small, this would indicate that averaging the weights together could improve the accuracy of the network. The second trend comes from determining if specific muscles tend to have larger or smaller weights. This would indicate the importance of the muscle to the overall network, and whether or not it is actually necessary to measure the muscle. The third possibility is a relationship between the weights and frequencies. It could be possible that certain frequencies are more important than others. For example, if the high frequencies turn out to be unimportant, a lower sampling rate can be used.

To start off, we examine the variation in weights for a single subject. Figure 5.7 shows the average and standard deviation of the weights for 4 networks trained by subject A. All of these networks were trained on the same day, without moving any electrodes. The solid bar shows the average value while the striped bar shows the standard deviation.

If the value of a weight is completely random, then the average of different sets would be zero. A significantly non-zero average shows that the weight is consistent between sessions. There are only 5 values out of the 64 weights which *might* be significantly different from zero at the 5% level. The origin of the x and y offsets is at (270,250), indicating that these values are not significant at the 5% level. The only inference one can make from this plot is that the high variability of individual weights means that a linear combination of multiple sets of weights will not yield a more accurate neural net.

Figure 5.8 shows the sum of the absolute values of all the weights for each muscle. This is averaged over the 4 networks generated by subject A. The absolute value is used to determine the importance of each muscle, not the effect. As the graph shows, the variation of each of the weights is much smaller than Figure 5.7, with mean to standard deviation ratios ranging from 2.4 to 9.2. The height of the bar

Figure 5.8: The total of the absolute values of all the weights for each muscle, averaged over the 4 neural networks produced by subject A

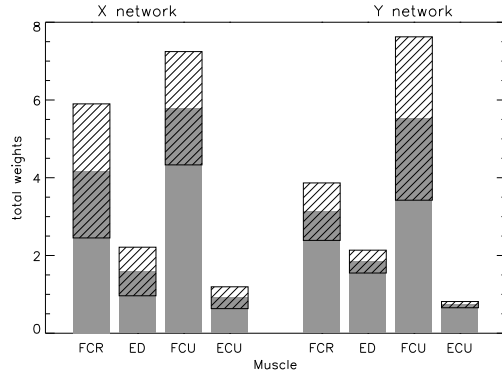
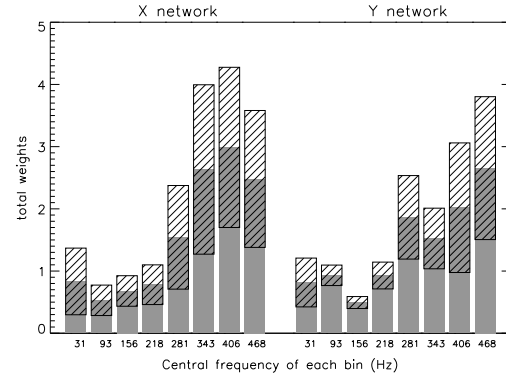


Figure 5.9: The total of the absolute values of each frequency bin, averaged over the 4 neural networks produced by subject A



is related to how much amplification the muscle needs. The ECU and FCU have smaller EMGs. The weights of the ECU are also small, implying that it has a small effect on the pointer. The FCU has a large weight, which means it is amplified by a fair amount. This could mean the FCU has anywhere from a small to a very large influence on the pointer, depending on the values of the EMG and the signs of the individual weights. The FCR has a large EMG which means that it is likely that the FCR has a large effect with regards to the pointer motion, but, like the FCU, it cannot be shown from these data alone. Another interesting point to note is that the muscles have a more-or-less equal relative importance in the x and y networks.

Figure 5.9 shows the weights grouped by frequency range. This is averaged over the 4 networks generated by subject A. The mean to standard deviation ratio of this plot is smaller than Figure 5.8, with a range from 2.1 to 5.7. The shape of the plot may be due to the fact that the spectra have higher amplitudes at lower frequencies, which drop off at high frequencies. This could imply that all frequencies of the muscle are important, but this cannot be known without further data.

Since there are trends in the neural networks which are consistent for one person, it begs the question if similar trends also can be found when comparing the neural networks generated by different people. The above analysis was applied to the neural networks generated by all the subjects in order to see if the same patterns, or any new ones emerged.

Two neural networks were used for each subject, one was used during the performance test, and one from a previous time. Both networks were recorded on different days, thus the electrode positioning may not have been exactly the same. All six networks were averaged together in the same manner as above. Figure 5.10 shows the average and standard deviation for all 3 pairs of networks for each subject. None of the values are significantly different from zero. This indicates that it is unlikely that there can be a “perfect” set of weights for everyone.

Comparing total weights for each muscle for all subjects (Figure 5.11) to that of just one subject (Figure 5.8) shows that the differences from person to person are larger than for an individual, which is what we would expect. An interesting point to note is that the plots for the x and y networks have similar shapes, quite similar to the pattern for a single subject (Figure 5.8).

Figure 5.10: The average of the neural networks of all the subjects

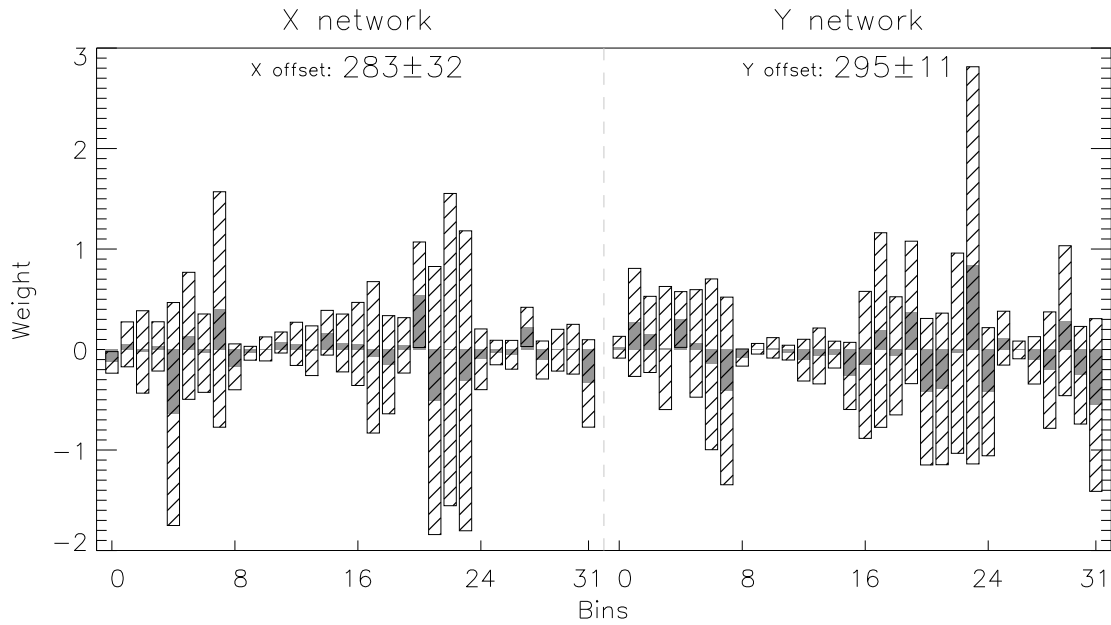


Figure 5.11: The total weights for each muscle, averaged over all subjects

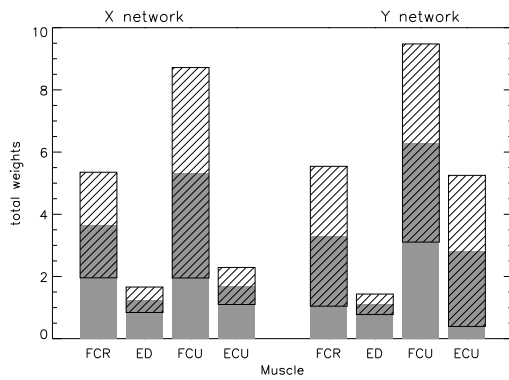


Figure 5.12: The total weights for each frequency bin, averaged over all subjects

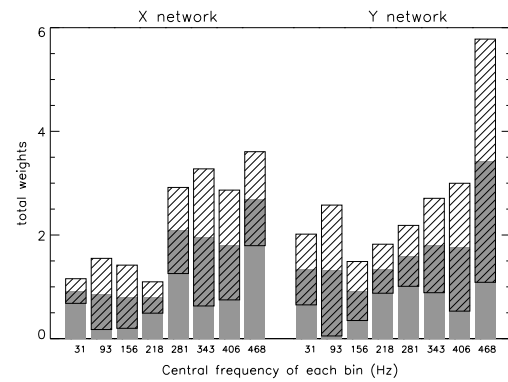


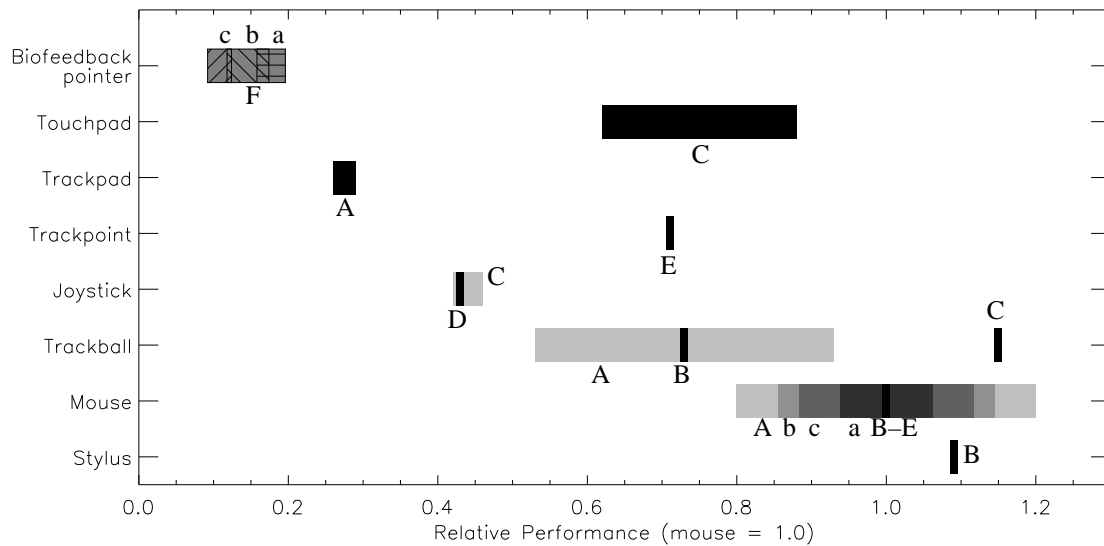
Figure 5.12 groups the weights by frequency range. Higher frequencies tend to have larger weights, and the x and y networks have similar shapes. There is a larger level of variation in the weights as compared to the frequency plot for a single person (Figure 5.9), but the overall shape is fairly similar.

5.5 Discussion

The Biofeedback Pointer was adaptable enough to be used by three different people. The neural network was able to cope with the individual differences between the subjects to provide an interface that each of them could use. While the tests were performed using too few people to generalise that the Biofeedback Pointer can be used by *anyone*, no problems have been identified which could prevent an able bodied person from using it. Further research must be done to observe the effects of differing body types (i.e. various levels of fat and muscle) on the device.

An analysis of the weights in the neural network showed certain trends when looking at the effect of muscles and frequency distribution. It was found in the overall analysis that there was too much variation

Figure 5.13: Relative performance on various graphic input devices. The colours are used to help differentiate between experiments. Notes: A = MacKenzie & Oniszczak (1998); B = I. Scott MacKenzie & Buxton (1991); C = Epps (1986); D = Card et al. (1978); E = Mithal & Douglas (1996); F (a-c) = This experiment (subjects a-c)



in each weight for an average of the weights to have any meaning. This means that multiple neural networks cannot be combined to create a more accurate network. In the muscle analysis it was determined that the ECU has the least influence over the outcome. This supports the theory mentioned in Section 3.3.2 that only 3 muscles are necessary to provide the 2 degrees of freedom. The frequency analysis shows that the weights are larger at higher frequencies. Since the high frequencies are low power, the larger weights are most likely to compensate for this. This implies that all frequencies are important to the final outcome. Had this been otherwise, and the higher frequencies had a low influence, then the EMGs could be sampled at a lower rate, which would be less computationally expensive and could use a slower ADC. As it stands, the full kilohertz sampling is necessary and a fast processor is required to handle all the computations.

The subjects' performance on the Biofeedback Pointer averaged at just over $\frac{1}{7}$ their performance with a mouse, which is fairly low compared to most graphic input devices (Figure 5.13). The implication is that, in its current version, the Biofeedback Pointer is not adequate to be used as a general pointer input. According to the chart, the Biofeedback Pointer's IP is also about half of the trackpad, which is a graphic input device common with notebook computers. This implies that the performance of the Biofeedback Pointer is not far off from an acceptable level. The Biofeedback Pointer does show promise and further research must be done to improve its performance by at least a factor of two.