Chapter 9

Analysis

Overview

This chapter presents a further analysis of the results from the longitudinal tests. The data points are plotted in three dimensions in order to see at a glance the effect of practice time and test complexity on users' performances. The major findings of the tests are summarised and a distilled version of the users' comments is presented.

9.1 3D presentation of results

The scores gathered in the longitudinal tests can be viewed with respect to the following dimensions:

- Human Test Subject [Subjects 9, 5 & 4]
- Time (session no.) [1-10]
- Test complexity [1-9]
- Interface type [mouse, sliders, multiparametric]

In the previous chapter we saw how the test complexity affected the scores on each interface (section 8.4.1), but this did not take into account any progress over time. The effect of time was shown for the simplest test and the most complex (sections 8.4.2 and 8.4.3), but not as a continuum of test complexity. Finally we showed the overall effect of time on the three interfaces, but without any reference to the complexity (section 8.4.4).

It was desirable to view the effects of time and test complexity in a single graph and so the skills of a mathematician colleague were employed to plot the data in three dimensions (see Acknowledgements and Author's Declaration). This work was carried out using MATLAB and the details are presented in Appendix E.

9.1.1 General form of the 3D plots

Figure 9.1 shows the generic form of all the subsequent three-dimensional plots.

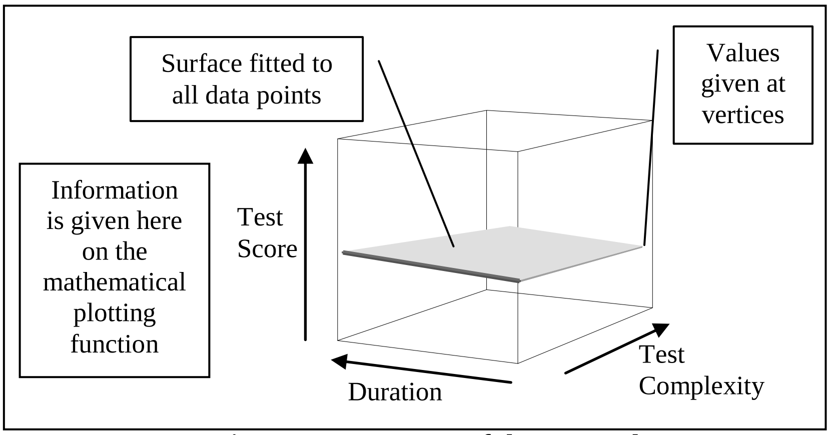

Figure 9.1: Structure of the 3D graphs

The test scores are plotted on the vertical axis, with a ‘perfect’ score of 100% being at the top of the cube. The axis labelled *Test Complexity* is simply the test number (1-9), as the tests are numbered numerically in increasing order of parameter complexity. The labelled *Duration* is the session number and thus represents increasing user contact time with the interface.

Once the data points have been plotted, the MATLAB code fits the points to a curved surface using a Quadratic Surface function. The values for this function are given on the left of the graph. The factor labelled ‘Squared Residuals’ is a measure of how far the data points lie outside the surface. In other words it is a measure of statistical ‘fit’, with a value of 0 meaning that every point lies precisely on the surface. This is discussed in more detail in section 9.1.5.

Finally the data values at the corners of the surface are printed. This is useful for numerically comparing more than one graph in order to see how the surface differs, for example, from one interface to another.

The 3D graphs are now presented. We begin with an overall comparison of the three interfaces, before giving more detailed information about the performance of each of the test subjects.

9.1.2 Results for each interface

The results across all the human test subjects are encapsulated into the following three graphs. The data from the three subjects was averaged before plotting each graph. Various techniques were explored for normalising the data from each subject before plotting, but this has been shown to make no difference to the final plot and so a straight average gives a good representation of the 'joint' performance of several subjects.

Figure 9.2 shows the first of the 3D plots. This graph encapsulates *all* the data collected for the *mouse* interface during the longitudinal tests (as marked by the human judge).

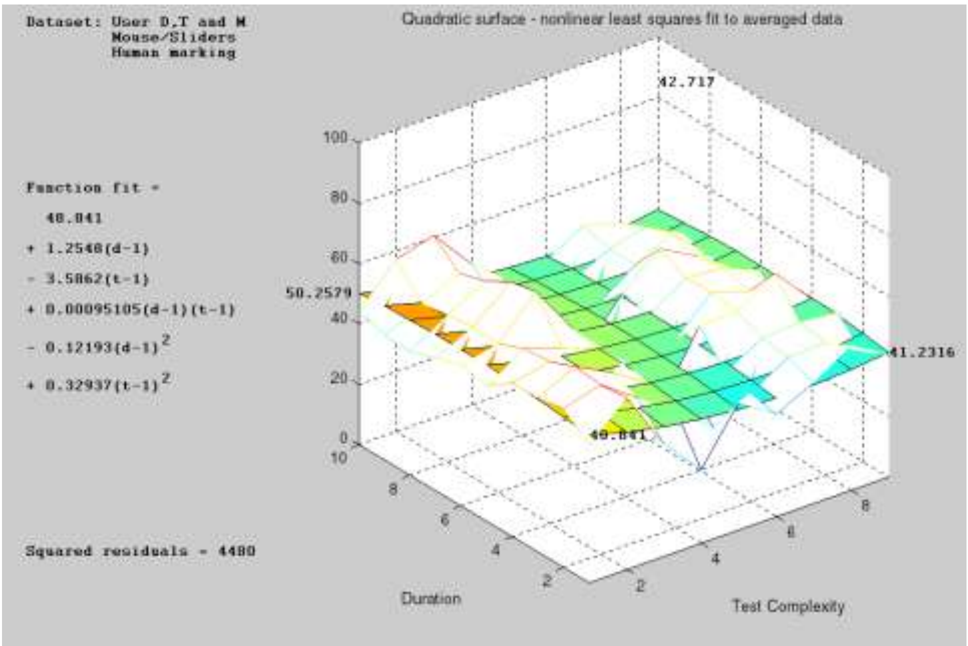

Figure 9.2: 'Mouse' interface data from longitudinal tests

This graph shows that the mouse does indeed give a reasonably 'flat' response over all the tests. There are signs of a very small improvement over time (e.g. an average of 48.84 in session 1, rising to 50.25 after ten sessions). Note the upward bend in the plane that indicates that the best scores are for the simplest few tests.

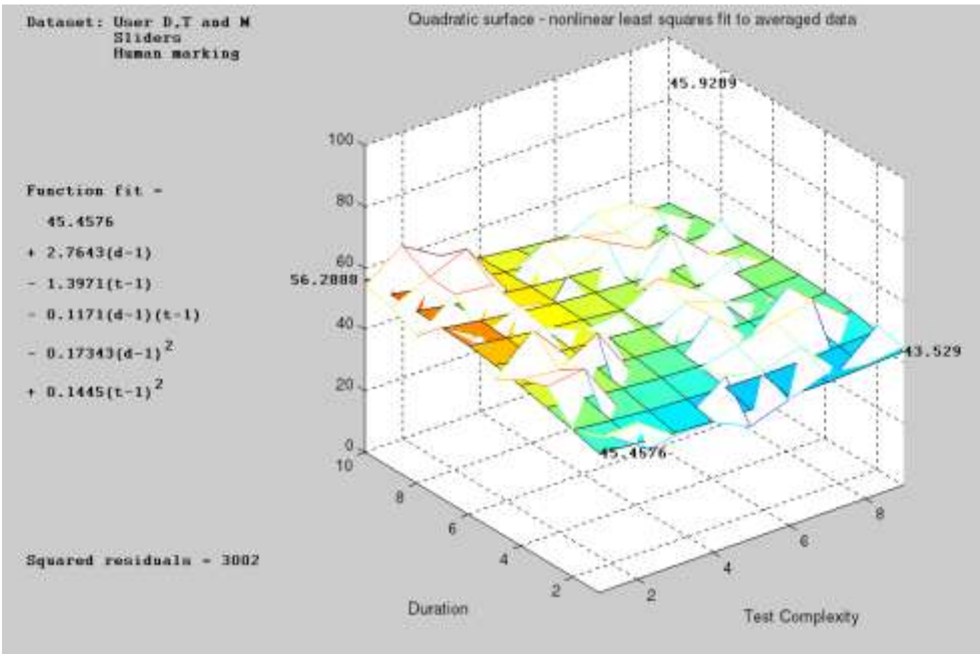

Figure 9.3: 'Sliders' interface data from longitudinal tests

Figure 9.3 shows a similar plot but for the sliders interface. The angle of the surface allows some immediate comparisons to be made with the mouse interface plot above.

- For the simplest tests the initial scores are slightly lower than the mouse.
- There is a rapid learning curve for the simpler tests which means that at the final session the score is much higher than the mouse .
- The sliders performs slightly better than the mouse for the more complex tests.
- The learning curve is only slight for the more complex tests.

This indicates that the sliders feel a bit more difficult than the mouse to start with, but a significant amount of learning can take place for the simpler tests.

Figure 9.4 shows the same type of plot for the multiparametric interface.

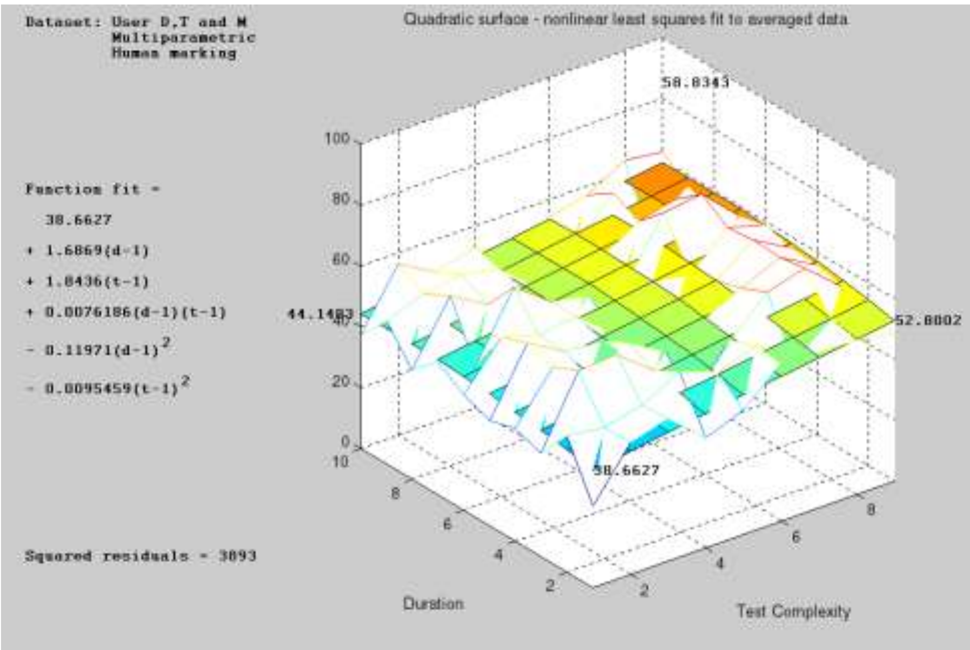

Figure 9.4: 'Multiparametric' interface data from longitudinal tests

This shows a dramatically different picture. The angle of the surface shows clearly that something very different occurred with the multiparametric interface. The following points of comparison with the previous two graphs are noted:

- For the simplest test the scores are always lower than those for the mouse or sliders, but they improve over time.
- The scores **get better** for more complex tests and are much higher than the other two interfaces.
- There is a good improvement over time across **all** test complexities.

The upward tilt of the plane towards the far-right corner is the most notable feature of the graph. It demonstrates that on average, the multiparametric interface performs better on the complex tests and yet allows a general all-round improvement on tests of all complexities. Neither of the other interfaces had this characteristic. The limitation of this interface appears to be that the simplest tests are difficult to achieve and especially so on the first few sessions.

9.1.3 Results for individual test subjects

The above three graphs averaged the results from the three test subjects before plotting. In this section we examine the data gathered from each individual subject in order to compare personal performances.

9.1.3.1 Subject 9's Results

The following three graphs illustrate Subject 9's individual results on the mouse, sliders and multiparametric interfaces respectively.

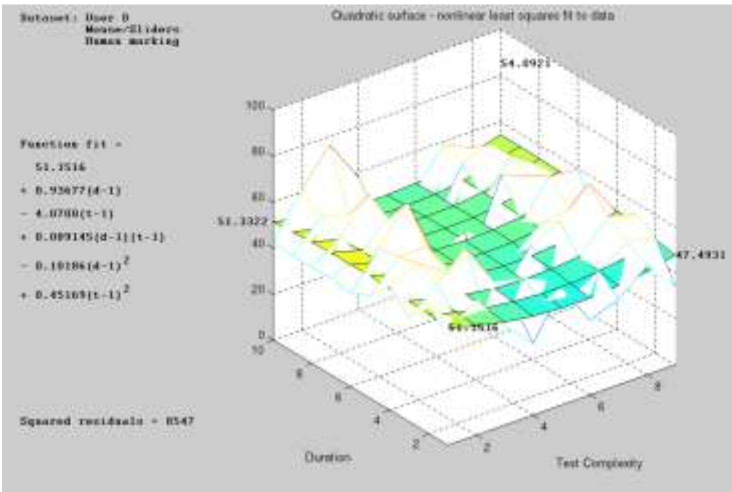

Figure 9.5: Subject 9's results on the mouse interface

His performance on the mouse was somewhat unusual in that the more complex tests did get better over time, and in fact overtook his scores for the simpler ones. His simplest tests were good, but showed no improvement as time progressed.

Figure 9.6 shows that his sliders performance was good, but it deteriorated over time! By the end of the ten sessions the mouse and multiparametric interfaces had overtaken it, apart from for the simplest tests.

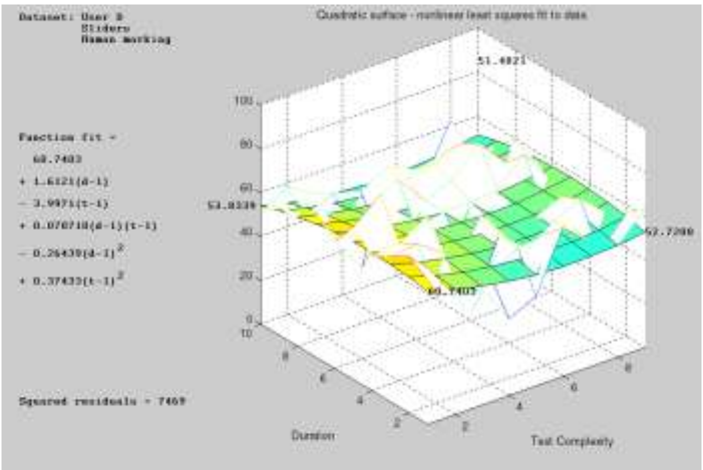

Figure 9.6: Subject 9's results on the sliders interface

Subject 9's performance on the multiparametric interface (see Figure 9.7) shows the upward tilt towards the most complex tests which was shown in the previous section (for all subjects). He clearly gets better over time on all tests, but gets remarkably better on the more complex ones.

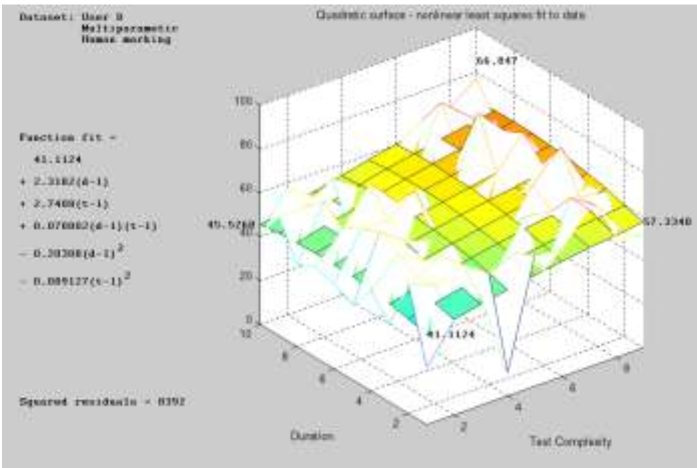

Figure 9.7: Subject 9's results on the multiparametric interface

Overall Subject 9 excelled with the multiparametric interface and improved throughout the sessions.

The sliders interface gave him the best results for the simplest tests.

The mouse gave a fairly flat response, but unusually showed an improvement on the most complex sounds.

9.1.3.2 Subject 5's Results

The following three graphs illustrate Subject 5's individual results on the mouse, sliders and multiparametric interfaces in the same format.

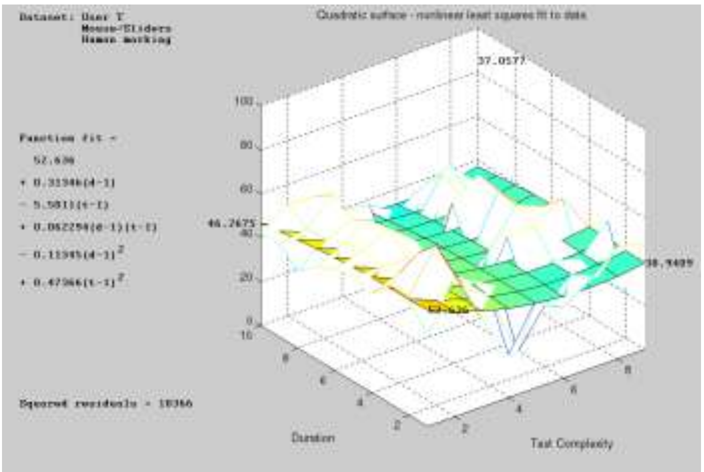

Figure 9.8: Subject 5's results on the mouse interface

Subject 5's graph of his mouse performance (Figure 9.8) is fairly 'flat', but all the scores gradually worsen over time. Note that for Subject 9 this happened with the sliders interface.

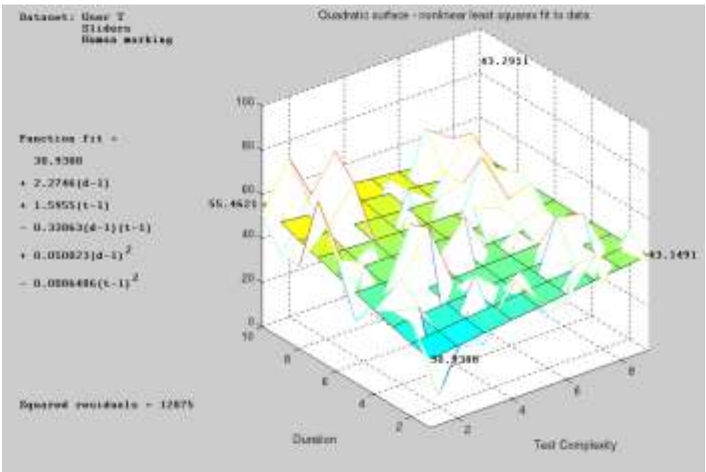

Figure 9.9: Subject 5's results on the sliders interface

Subject 5's sliders graph (Figure 9.9) demonstrates how he started off badly on the simple tests but got better consistently over time. Equally it can be seen that he was performing reasonably on the complex tests but showed no improvement.

Subject 5's multiparametric graph (Figure 9.10) is very similar to Subject 9's and shows overall improvement, especially with the more complex tests.

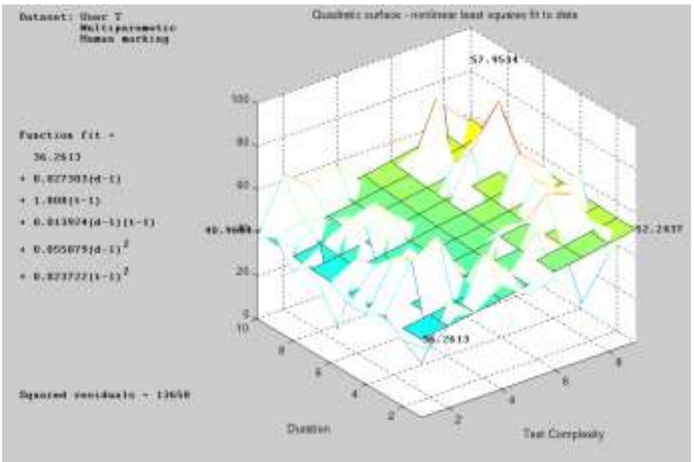

Figure 9.10: Subject 5's results on the multiparametric interface

9.1.3.3 Subject 4's Results

The following three graphs illustrate Subject 4's individual results on the mouse, sliders and multiparametric interfaces in the same format as above.

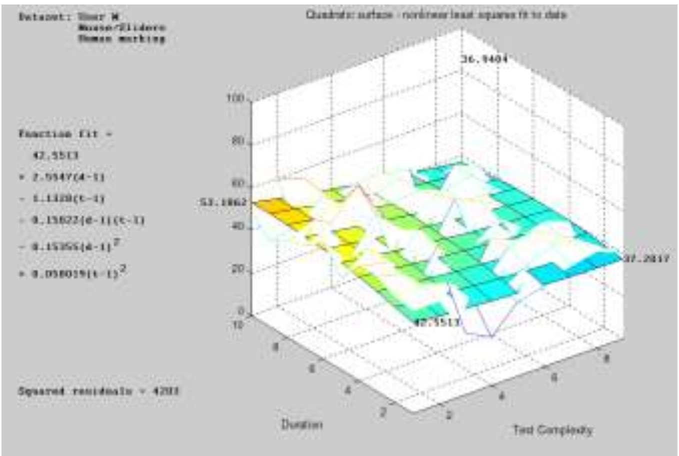

Figure 9.11: Subject 4's results on the mouse interface

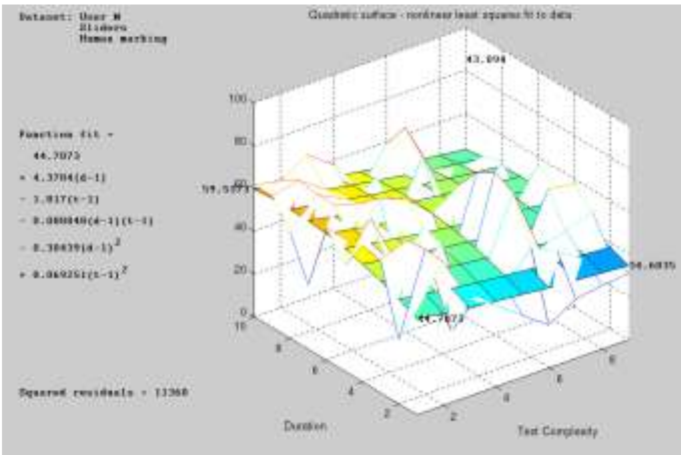

Figure 9.12: Subject 4's results on the sliders interface

Subject 4's graphs for the mouse and sliders interfaces (Figures 9.11 and 9.12) are remarkably similar. The best-fit surfaces are both tilted down towards the right (indicating lower scores for increasingly complex tests) and upwards towards the back (indicating an improvement over time). Closer inspection reveals that the mouse scores only improve for the simpler tests.

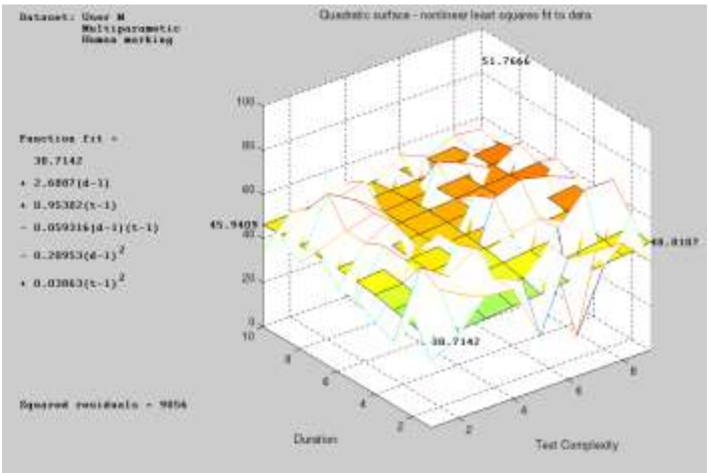

Figure 9.13: Subject 4's results on the multiparametric interface

The multiparametric graph (Figure 9.13) stands out visually because it is tilted the other way, indicating (as did those of subjects 5 & 9) that the scores improve not only with time but with increasing test complexity.

9.1.4 Interpretation of the data

In many studies the main goal is to draw a precise numerical conclusion based on the number of subjects and one or two variables (e.g. "we are 99% confident that adult male feet are between 21cm and 30cm in length"). In contrast, this thesis is mainly concerned with the qualitative claim that a certain type of multiparametric interface will yield relatively better results after a certain amount of practice. There is no single meaningful number that can be extracted from the study to encapsulate all the experimental data. Rather the study has gathered large amounts of information, then presented that data so that the general trends can be seen.

The cross-sectional tests utilised 16 subjects, but for the information we are looking for it is not necessarily the number of subjects that is important. In the cross-sectional tests there were 3456 tests (1152 for each interface) - each of which contained a series of changes to four parameter values. In the longitudinal tests there were only three subjects, but there were 810 tests (270 for each interface).

On each of the 3D graphs 90 points are plotted, but on the graphs which are the average of the three subjects, we have effectively encapsulated 270 points onto a single graph. The 'Squared residuals' value that appears on each graph indicates how well the quadratic surface function actually 'fits' the data set. The higher the value, the more the points lie *off* the surface and the less representative the surface becomes of the data set. In the graphs shown in Figures 9.2 to 9.4 the residual values are 4480, 3002 and 3893 which are quite respectable values for a 270-point plot. These values are less than for any of the corresponding graphs for individuals (Figures 9.5 to 9.13). This means that the average performance of the three subjects is more predictable than that of any individual performance. In other words the most reliable graphs are those shown in Figures 9.2 to 9.4 and these are most likely to indicate the average results of a larger population of users.

Incidentally, the 3D plots of the automated computer marking reveals a different story. Here the residuals are much higher, ranging from 21053 to 44065 for the individual performances. These values show a significant error in the fit of the quadratic surface for a 90-point plot, and so the orientation of the surface can be considered to be less reliable. In conjunction with the aforementioned problem in coping with common human performance traits (section 7.5), this indicates that the computer results should *not* be used as the basis of any major conclusions.

It is important to consider the relevance that these results will have when considering the population in general. For example, here are three characteristics that were implicitly made during the choice of test subject for the cross-sectional tests.

- All are University personnel
- All are within the age-range 24-45
- All are familiar with computers and mouse operation

This is clearly not representative of the population at large, but in the context of this study *this does not matter*. What is being measured is the *relative* effect that the different interfaces have on a person over time. The person's *progress* is recorded over time with all other factors being repeated identically each session. There are many data points which record this progress so we can regard the plots as indicative of how a person responds to the different interfaces. Those features or responses which are found to be common across the subject range are those which are most likely to indicate how another person might respond to the interfaces. It is these common features that are described in the next two sections.

9.2 Major findings

This section summarises the main findings of both the cross-sectional tests and the longitudinal tests.

- For tests where more than one parameter changes simultaneously the multiparametric interface gives the best overall results.
- The multiparametric interface nearly always shows improvement over time (independent of test complexity).
- The performance of the multiparametric interface nearly always *increases* with test complexity.
- For most people the mouse interface gives the best results for the simplest tests, and the multiparametric the worst.
- The scores on the mouse interface do not generally increase at all over time. In fact they may decrease!
- The sliders interface allows increased scores over time for the simpler tests, otherwise it is similar to the performance on the mouse and the more complex tests can get worse as time progresses.

9.3 Summary of comments from taped interviews

The following comments represent an encapsulation of all the comments received by the subjects after each interface trial in the cross-sectional tests.

- The mouse is the easiest interface to use at first, but is clearly limited to controlling one parameter at a time.
- A small proportion of people (about one quarter) favour the sliders interface, and these were mostly the people who did not like having to 'wobble' the mouse to get volume.
- The majority of people found the sliders interface confusing, frustrating or at odds with their way of thinking. This was often focused on the requirement to break down the sound into separate parameters.
- The multiparametric interface allowed people to think spatially, or to mentally rehearse sounds as shapes.

- Users often reported that they felt "out of control" with the multiparametric interface, but this was not always a negative comment - rather a statement that they had achieved a task but couldn't analyse how they had done it. This seems to be linked to the comment that this interface put the least mental load on the subjects. They were free to concentrate on completing the task, or on other things, rather than consciously controlling the interface.
- Several users reported that the multiparametric interface was *fun*.
- Various people reported that the multiparametric interface felt like a musical instrument rather than a tool or computer program. It was described as "flowing", having "more freedom" and a greater dimensionality.
- The majority of users felt that the multiparametric interface had the most long-term potential. Several people commented that they would quite like to continue to use it outside the context of the tests!
- Some people may say that they prefer one interface, but actually perform better with another.
- People's preferences for one interface over another can change from session to session.

9.4 Summary

In this chapter we have seen plots of the progression of users' scores with respect to time and test complexity. The 'average' response has been presented followed by some discussion about the ways in which the data is interpreted. A series of major findings has been listed along with a summary of the users' subjective comments. The three-dimensional graphs make it clear to see that the multiparametric interface elicits a very different response from the users. These findings form the basis for much discussion and further work, which is the subject of the final chapter.