# Chapter 7

# Cross-sectional Tests

## Overview

This chapter gives the details of the 'Listen and Copy' tests which were undertaken by a range of human subjects. The musical examples are described and the test environment outlined from the point of view of the test subjects. The analysis methods are then discussed and the results given. A summary of users' comments gives insight into what the subjects thought of each interface.

## 7.1 Overview of Cross-sectional tests

The purpose of the cross-sectional tests was to compare the performance of several users on the three interfaces. Each user was asked to spend approximately 15 minutes on each interface, making up a session of 45 minutes duration. Within each 15-minute period they undertook to listen to and reproduce 24 sound examples. These sound examples were the same for all three interfaces. Users came back after several days to repeat the entire process and then again a few days after that. Each human subject therefore experienced (24 sounds x 3 interfaces x 3 sessions) = 216 tests altogether.

There are many variables in these tests, but every example is a comparison over four parameters of what the computer plays and what the human subject manages to perform. The results have been studied to compare how different human subjects respond to each of the interfaces and how their performances varied with time over the three sessions. We will also see how the results are affected by the test complexity, for example the number of parameters that are altered simultaneously.

## 7.2 Musical Test Examples

Each test set consists of 24 sound examples each of which lasts between 2 and 4 seconds (a duration range which initial trials showed to be short enough to remember, but long enough to make a decent sound trajectory). They vary in complexity and can be notionally divided into three groups (although the user is not made aware of these groupings).

Group A (tests 1-8) contains sounds which have simple stepwise changes in one parameter only while the other parameters are held constant (see Figure 7.1).

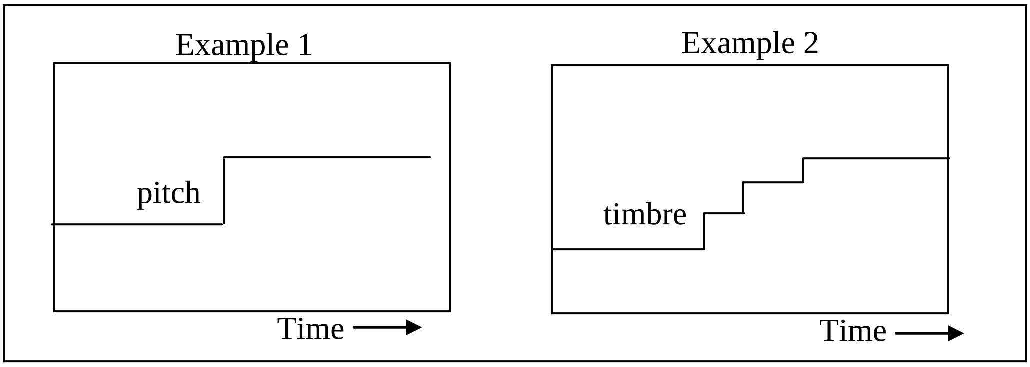


**Figure 7.1: Group A sounds: stepwise uni-parameter changes.**

For example the pitch may step up a tone after 1 second while timbre, volume and panning do not alter (as in Example 1, above). Towards the end of this group, there may be several step changes in a single parameter (as in Example 2 in the Figure).

Group B (tests 9-16) introduces 'trajectories' for a single sound parameter. These are perceived by the user as more of a continuous change. Towards the end of this group more than one parameter changes during the test, but *not at the same time*. Figure 7.2 shows examples from either end of this group.

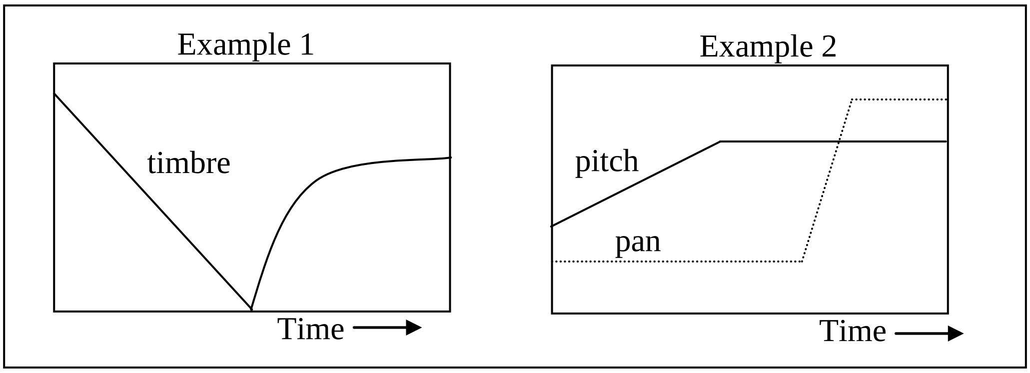


**Figure 7.2: Group B sounds: continuous non-simultaneous changes.**

Group C (tests 17-24) consists of sounds where more than one parameter is changing at once (see Figure 7.3). Towards the start of this group just two parameters change simultaneously, whereas for the last few all four parameters are moving with their own trajectories.

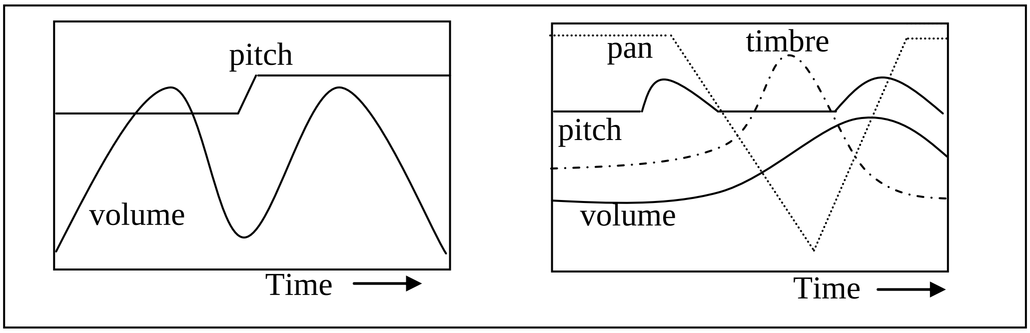


**Figure 7.3: Group C sounds: continuous simultaneous changes.**

Comments have been made that when listening to some of these sounds they take on a specific character e.g. "a wolf-whistle effect that gets brighter while spinning from left to right". This tendency to mentally 'fuse' several concurrent sonic events may have some bearing on the later analysis.

The tests were created in a variety of ways. Some were produced on the interfaces themselves. Others were typed in as breakpoint files. Another set were individually constructed from a combination of those two methods. The key point is that all the tests are the same for *every user* and for *every interface* and so can be used to compare individual differences in performance.

## 7.3 Choice of Test Subjects

Sixteen subjects were selected from University staff and students with a variety of musical and non-musical backgrounds. However it should be noted that all the subjects are in a University environment and are familiar with the use of computers (at the very least for word-processing), are used to a computer mouse and were very willing to take part in the tests. These facts need to be considered before drawing conclusions about the population in general.

The subjects are listed in the following table along with a summary (in their own words) of their previous musical experience.

| Subject | Musical Experience (as described by subjects) |
|---|---|
| 1 | Cello and keyboard (to quite a high standard) |
| 2 | Flute (high standard) and singing |
| 3 | Recorder at school "a long time ago" |
| 4 | Percussion (advanced) |
| 5 | Violin as a child, nothing since |
| 6 | Live performance with synthesisers ("dials & sliders") |
| 7 | None whatsoever |
| 8 | Flute and guitars (quite a high standard) |
| 9 | Piano, organ, singing, synthesisers (advanced) |
| 10 | Keyboard (moderate) |
| 11 | Guitar (moderate) |
| 12 | None whatsoever |
| 13 | Piano & flute (advanced) |
| 14 | Piano, synthesisers (particularly analogue) |
| 15 | No instruments |
| 16 | None whatsoever |

**Table 7.1: Test subjects for the cross-sectional tests**

## 7.4 The Test Environment

This section describes the study from the users' point of view, before outlining how the computer co-ordinated the tests and gathered the data.

### 7.4.1 Instructions to Users

The following sheet of instructions was given out to each of the test subjects *in advance* of their first session, thus allowing them time to assimilate the structure of what they would be doing.

# User Interface Testing

Pre-test information

## Welcome & Overview

Thanks for agreeing to take part in this musical user-interfacing test. In this series of tests we are comparing different types of *interface* for achieving a musical task. We would like to see how people learn to control these interfaces over several weeks and to see which ones are better suited to certain tasks.

## The Task

You will be asked to listen to a short burst of sound. It may sound musical, or it may not. You listen to it *just once* and then you have to try to play it back. You do this by trying to reproduce the sound on an interface (details below) consisting of sliders and the computer mouse. There will be many sounds, each of which you hear once and will try to copy.

## The Interfaces

There are three interfaces to test :

1. You will use the computer *mouse* (and its Left button) to control a set of sliders on the computer screen. Each slider controls one musical dimension (such as the 'pitch' of the sound, or its 'loudness'). As you point to a slider and press the mouse button the slider becomes controlled by you and as you move the mouse you will hear the sound changing.

2. This second interface also consists of sliders, but this time they are *physical* sliders. You do not need the mouse as you can put your fingers directly on the sliders. It is rather like a small 'mixing desk', so you may use two hands if you wish.

3. The final interface involves both the mouse *and* a pair of sliders. One hand operates the sliders, while the other uses the mouse. You will find that this interface responds rather differently to the others. Your job is to find out how it works, and then use it to try to recreate the sounds you will hear as part of the test.

## The Test Situation

You will be asked to take part in several sessions over a number of weeks. This is so we can assess how you accumulate knowledge about the interfaces and skill in using them. Each session should take about half an hour to complete. During that time you will work with each of the three interfaces. On each interface you will be asked to do your best to reproduce 24 sounds. You will be seated facing a computer, with the interface in front of you, and wearing a pear of headphones. Over the headphones you will hear some spoken instructions and, later, the test sounds themselves.

## Information about you

In order to make valid comparisons between people and interfaces we need to compare the following information about you. Please complete the form below and return it to Andy Hunt at your first test.

- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -

Name:

Are you a regular computer user ? yes/no

Do you play (or have you played) an acoustic musical instrument ? yes/no

Do you have practical experience with the use of computers for making sound/music ? yes/no

### 7.4.2 Description of Test Environment

The test subjects sat in front of a Silicon Graphics computer terminal (see Figure 7.4). The standard QWERTY keyboard was moved to one side, but the 'spacebar' was still accessible by the subject's left hand. In front of them was the particular interface being tested. They wore a pair of headphones that allowed them to hear the tests, but also to hear my voice if they needed further instructions.

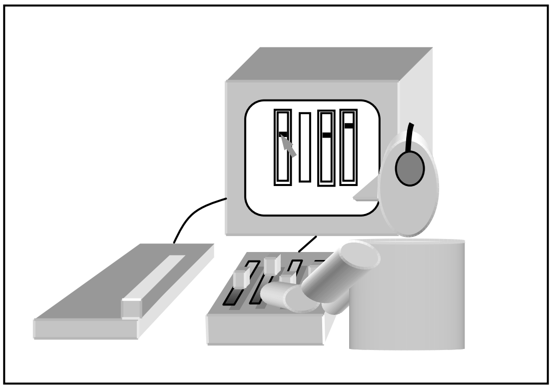

**Figure 7.4: The test environment**

Each test batch (i.e. 24 sounds on one interface) was completely automated. The MIDAS program co-ordinated playback of each sound and the gathering of the user's performance in response to that sound. It achieved this by giving the user recorded instructions such as:

*"Next test. Press the spacebar to hear the sound"*

<user presses spacebar>

*(computer plays sound)*

*"Now it's your turn. Press the spacebar to start"*

<user presses spacebar>

*(four 'beep' count-in)*

<user performs the sound on the interface>

*"Next test. Spacebar "*

etc.

In other words the test subject steps through each stage of the test by tapping the spacebar (on the QWERTY keyboard). This ensures that the user is in control of the overall timing of the procedure. The instructions get shorter for the subsequent tests so that they take up less time as the subject gets used to the procedure.

The *four 'beep' count-in* mentioned above gives the user time to move his or her hand from the spacebar, and set it in position on the interface. The time at which the (non-existent) fifth beep would occur is when the computer begins to record the user's data. In other words users synchronise themselves to the count-in as follows “Beep, beep, beep, beep, play”.

Test subjects were given two minutes to freely explore each interface before the set of tests began.

### 7.4.3 Organisation of Tests

The sixteen test subjects were each asked to attend an individual test session once a week for three weeks. Each session consisted of a series of 24 graded musical examples which had to be performed on each interface.

The order in which the interfaces were approached in each session was rotated. This was done for two reasons:

- People might tire during a session, so it was necessary to give each interface a chance at being the first one used by each subject.
- The first interface encountered in the session might actually influence the mental processes of the test subject, and thus affect the performance on those interfaces that follow. Chapter 2 describes some research which shows that certain interface styles can induce certain learning modes.

The order of the sound examples *within* a batch was changed for each session too. This was done to counteract a potential tendency for a user to "get the hang of" each session, thus giving a consistently better performance for sounds in the middle or towards the end of a batch.

The order of the interfaces and the tests was the same for everyone (see Table 7.2).

| **Session** | **1st interface** | **2nd interface** | **3rd interface** | **Sounds arranged** |
|---|---|---|---|---|
| 1 | Mouse | Sliders | Multiparametric | Increasing complexity |
| 2 | Multiparametric | Sliders | Mouse | Decreasing complexity |
| 3 | Sliders | Mouse | Multiparametric | Randomly |

**Table 7.2: Order of interfaces and tests during the 3 sessions**

## 7.5 Methods of Analysis

Every test result was stored on the computer and later given a score by both a computer algorithm and a human marker.

### 7.5.1 Computer Marking

A computer marking procedure is arguably completely objective in this experiment as it could compare the amount of deviation from the original. A program was written to establish a single 'total error score' for every test (see Appendix D for details). A score of 0 represents a perfectly reproduced musical example. The program performs a root-mean-square (r.m.s.) difference comparison for every parameter (pitch, volume, timbre, and panning) in each test. However, there are substantial difficulties in getting such an algorithm to mark human performances with any degree of confidence (these difficulties are summarised in the following section). Consequently the computer marking algorithm is not yet considered as a reliable method of marking, and will not be used in this thesis.

### 7.5.2 Human Marking

The automated procedure outlined above cannot take into account several very common human performance traits.

- Transposition – everything about the sound is reproduced *relatively* correctly, but the parameter starts off higher or lower than the original (see Figure 7.5). The computer marks this as a constant error and gives the test a bad score, whereas a human might say "you're mostly there, but you just need to start a bit lower next time".

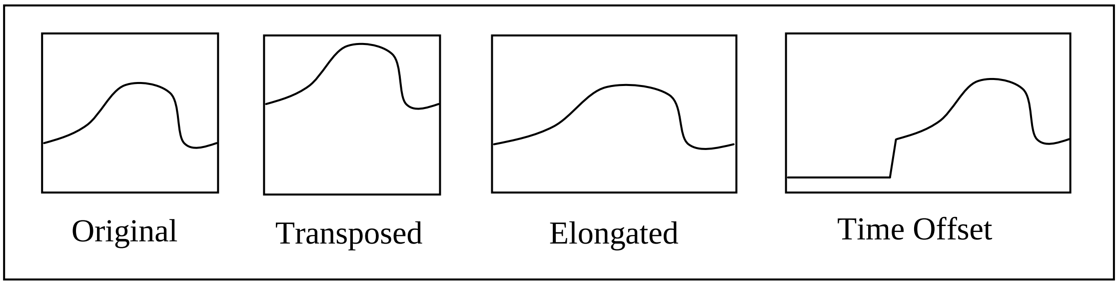


**Figure 7.5: Different types of human performance trait**

- Elongation – the overall shape and sound is fine, but is played back at a slightly slower or faster tempo. Again the moment-by-moment comparison carried out by the computer gives a poor mark whilst a human listener might say "well done, but it was just a little bit slow".

- Time Offset – the sound is perfectly reproduced but is started a short time after the given countdown signal. The computer marks low again whereas a human listener would probably just ignore the late start and give it a good score.

These features of human performance are not rare occurrences; they happen to some extent in every test. Whilst there are further mathematical correlation techniques which can be used to compensate for some of the above characteristics, it was thought better to use a human marker to give a score to every test.

The marker's job was to listen to the original sound, then to the performed sound (as many times as necessary) and give a mark out of 10 for each of three categories:

- Timing Accuracy (how do the *timings* of the sonic events match the original?)
- Parameter Accuracy (how near to the original are the parameter *values*?)
- Trajectory Accuracy (how well do the parameters *move* compared to the original?)

The final outcome for each test is a percentage score (where 100% means that the sound has been perfectly reproduced).

Human marking of audio material is normally considered essential for music exams, competitions, recitals at college, and feedback in music lessons etc. These tasks are never left to a computer, although in some instances (such as our experiment) a computer can provide an 'absolute error count' for the recorded data. Therefore the program which carries out the computer marking also prompts a human judge to give a score out of 10 for each of the categories outlined above. The judge can listen to the original sound and the performed sound as many times as required.

A single human judge (the author) marked the entire range of (216 per subject x 16 subjects) = 3456 tests. This took a large investment in time, but it meant that a consistent marking scheme was established – with no problems of equating marks between separate judges. A moderator was hired to randomly select several test sets and to mark them. The results of this process are shown in Appendix H.

### 7.5.3 Taped interviews

Every subject was interviewed after each set of tests on each interface. They were asked how they felt about their performance and the interface. Towards the end of the session they were asked to sum up how they had done overall and to compare the different interfaces. All of these interviews were recorded on audio cassette and later transcribed. This produced 48 pages of interview notes (representing a page for each of the 3 sessions for 16 users).

## 7.6 Results

This section described how the results of the cross-sectional tests were gathered, stored and made available for analysis. A selection of graphs are presented to summarise the overall findings of the study. Particular emphasis is given to the description of how the different groupings of test complexity are dealt with on each of the three interfaces. Finally some example graphs are given to show how different subjects reacted to the tasks.

### 7.6.1 Storage and Presentation of Results

Section 7.5 describes how a computer program was written to co-ordinate the human marking procedure, and to produce a computer 'error' mark even though this is not used in the final analysis. Each test was therefore given a human score (as a percentage, where 100 is perfection) and a computer error mark which ranged from 0 (perfection) to approximately 150 (very bad). Each of these marks was entered into spreadsheets in Microsoft *Excel* (one for the human marks and one for the computer error scores). From the spreadsheet a number of graphs were drawn to represent the different dimensions of the data. A summary of these graphs is given in the following sections. The full set of graphs is given for reference in Appendix B.

### 7.6.2 Results across all subjects and tests

The graph shown in Figure 7.6 is a summary of the entire cross-sectional study. The horizontal axis shows time in terms of session number. The points on the graph represent the average score for *all subjects* and for *all sound examples* on a particular interface. All scores are from the human marker.

The points on the graph are joined by lines for visual clarity, although it should be made clear that there is no 'continuity of information' between sessions. For the sake of consistency the line styles for each interface are kept constant throughout this whole document.

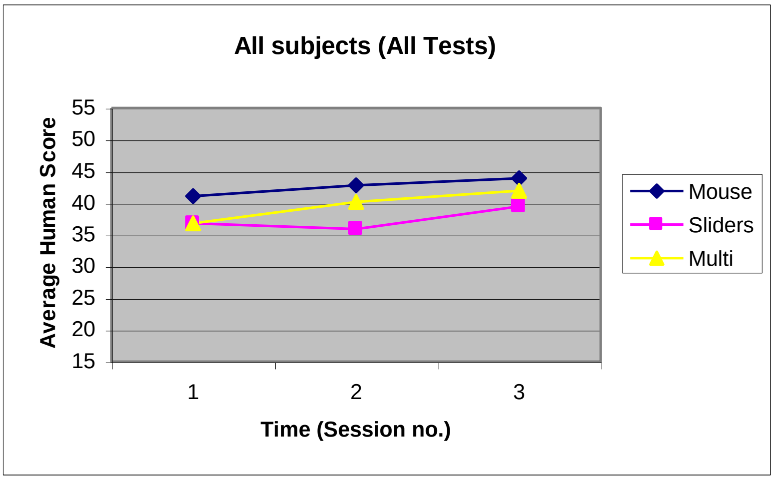


**Figure 7.6: Summary of the cross-sectional tests**

This graph appears to show that each interface overall has an upward trend over the three sessions. In other words the subjects are learning how to do the tests and how to use the interfaces. The rates of learning appear to be different for each interface. The mouse interface appears to give the best overall scores, but it is being 'caught up' by the multiparametric interface. We could speculate about whether it would match or even overtake the mouse given a few more sessions. The sliders interface gives the lowest overall marks, and in fact appears to be worse on the second session than the first! By the third session it seems to make a recovery but is still overall the least successful interface.

### 7.6.3 Results for tests of different complexity

In the above summary, we have averaged together the scores of *all* the tests, from the simplest through to the most difficult. One of the aspects to be investigated was the success of the different interfaces on the various levels of test complexity.

In section 7.2 the tests were described as fitting into three groups:

- Group A (simple step changes in one parameter).

- Group B (continuous changes, some with more than one parameter but not simultaneously).
- Group C (simultaneous parameter changes).

Figures 7.7 to 7.9 show the average results across all test subjects for tests in each of these groups.

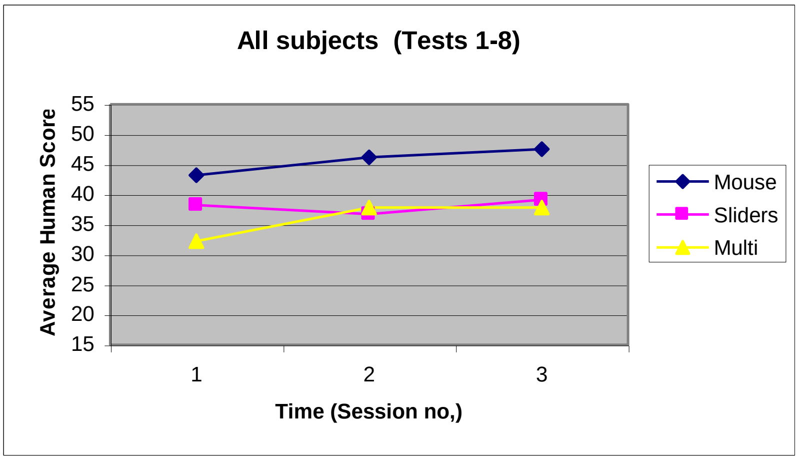


**Figure 7.7: Summary of the Group A tests**

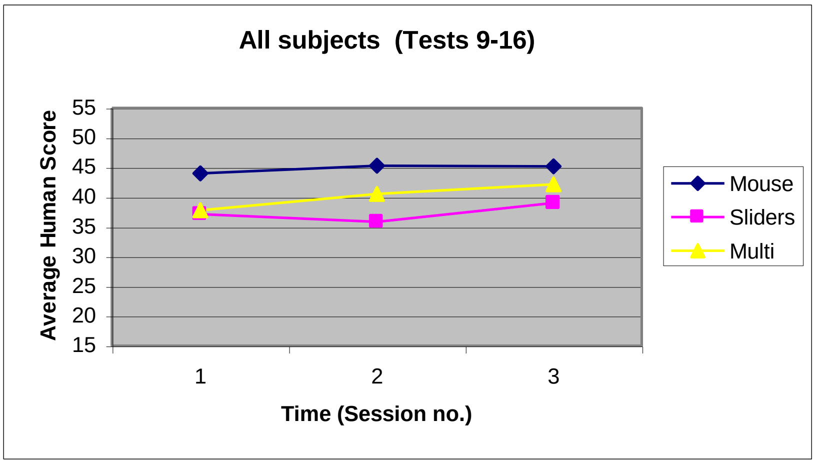


**Figure 7.8: Summary of the Group B tests**

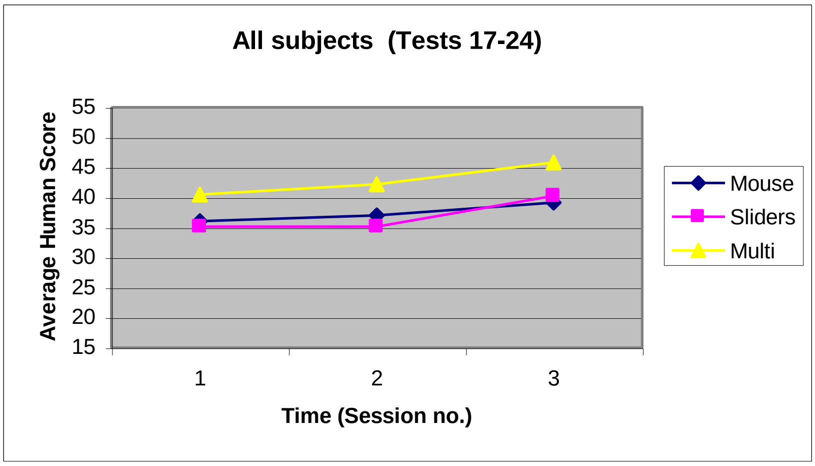


**Figure 7.9: Summary of the Group C tests**

A comparison of these graphs reveals a very different story to the overview shown in Figure 7.6. The mouse interface is the clear winner for Group A and Group B tests. However, as soon as there is more than one parameter changing simultaneously (Group C) the multiparametric interface become the best interface. A comparison of graphs for Groups A and B shows that the multiparametric interface is better than the sliders for continuous changes (Group B), but they are similar in performance for the simple step changes (Group A).

In summary these graphs show that the mouse interface is best for simple step changes, but that only the multiparametric interface allows a good average performance when more than one parameter changes simultaneously. In fact the best scores encountered for the multiparametric interface are on the most difficult tests. Perhaps a 'complex' interface is needed for a complex test! Perhaps a complex test is needed for a complex interface?! This is discussed in more detail in section 10.2.5.

### 7.6.4 Results of individual subjects

Much can be learned about individual performances by plotting the results for individual subjects. The following three graphs (shown in Figures 7.10 - 7.12) have been selected to demonstrate how three subjects reacted very differently to the same set of tests.

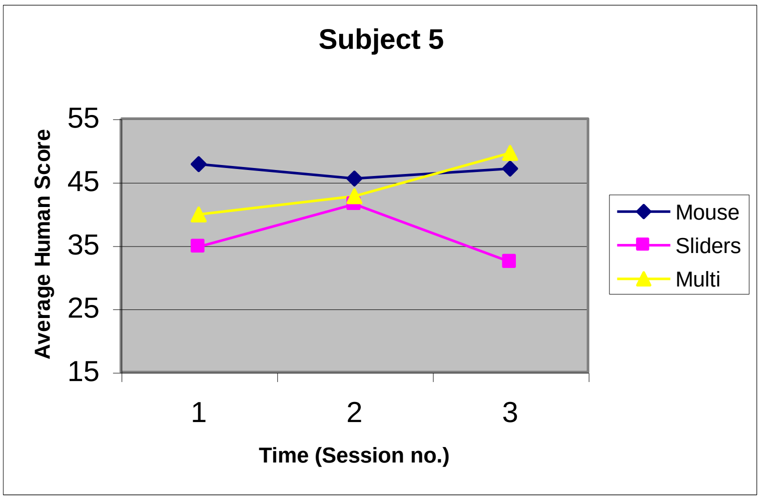


**Figure 7.10: Summary of Subject 5's results for all tests**

Subject 5's results (Figure 7.10) show that his overall performance with the multiparametric interface exceeds that with the mouse by the third session. The learning curve is steep with the multiparametric interface, and rather flat (if not downward) with the mouse. Note that the sliders have a steep downward turn by the third session.

Subject 9's results (Figure 7.11) are reasonably similar for all three interfaces. There is a slight preference for the multiparametric interface. There is an interesting drop in performance during session 2. This could simply be the result of an 'off-day', but might be because subject 9 responded negatively to encountering the test sounds in *decreasing complexity* (i.e. the hardest tests were first, so there wasn't a chance to 'practice' on the easier ones).

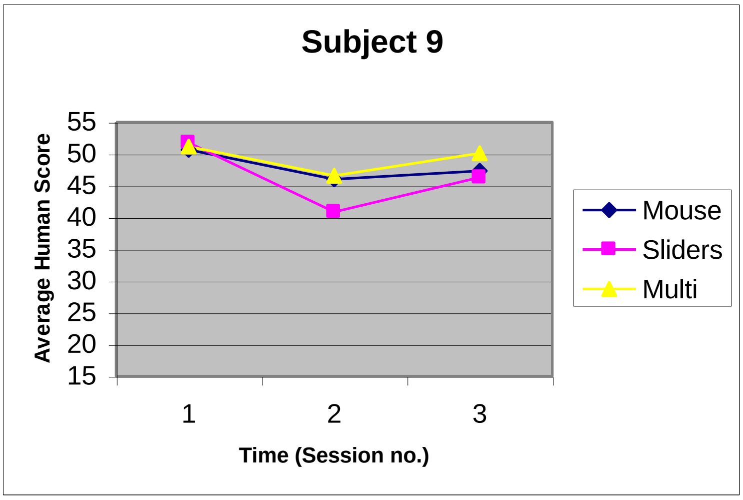


**Figure 7.11: Summary of subject 9's results for all tests**

Subject 4's results (Figure 7.12) show that he was one of the few subjects whose best interface by session 3 was the sliders (the others being subjects 11 and 6). His performances on the mouse and multiparametric both take a corresponding downward turn from the second to the third session.

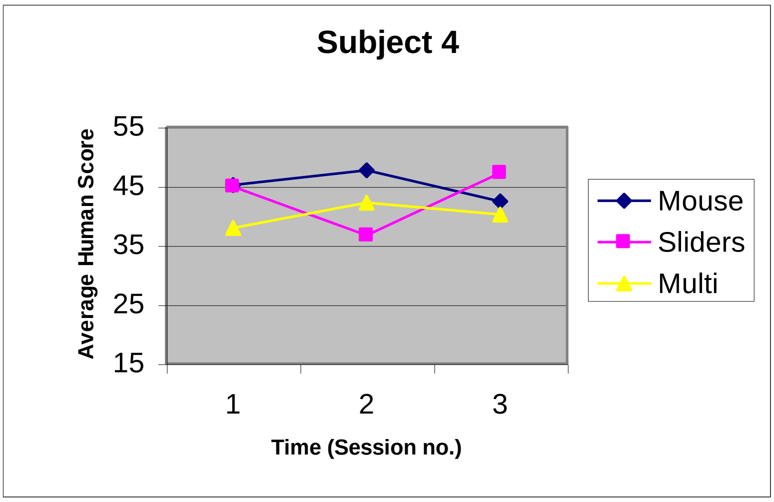


**Figure 7.12: Summary of subject 4's results for all tests**

It is perhaps not surprising to note that for most subjects their favourite interface (from the interviews) was the one they performed best with. However, this was not always the case. Subjects 4 and 11 clearly enjoyed the multiparametric interface and found the sliders annoying and frustrating, yet managed to get their highest average scores in the last session with the sliders. On the other hand subjects 7 and 16 could not decide which of those two interfaces they preferred, yet they were consistently scoring higher on the multiparametric interface.

## 7.7 Analysis of Taped Interviews

The taped interviews with the test subjects have been analysed for consistent comments given by different people. The headings for the following eight sections represent the essence of these comments, and they are each supported by some example quotes from the interviews.

### 7.7.1 The mouse is easier to start with, but has obvious limitations

*"Being restricted helps you to concentrate"* (Subject 1, session 1)

*"It's good that it 'prompts' you with where the sliders start"* (Subject 3, session 2)

*"It's easiest to use, but I've no idea how I get simultaneous moves"* (Subject 7, session 1).

*"I feel comfortable with the instrument's limitations"* (Subject 8, session 3).

*"It's easy for the single parameter changes"* (Subject 9, session 3).

*"I like the way it helps you to start off, but it definitely has much less control than the other interfaces"* (Subject 10, session 3).

*"This was fine at first"* (Subject 11, session 1), *"I've got as far as I can get with this"* (Subject 11, session 3).

*"You can learn this one fast, but there's not much depth to it"* (Subject 13, session 3).

### 7.7.2 Some users preferred the sliders interface on certain sessions

*"I would choose this one (multiparametric)"* (Subject 7, session 2). *"I'd probably choose this (sliders) as my favourite"* (Subject 7, session 3).

*"This seemed easier today. I do feel I've improved"* (Subject 12, session 3).

*"I'd definitely prefer this interface! This is my favourite."* (Subject 14, sessions 1&3)

*"I like this one best"* (Subject 16, session 2), *"I'm comfortable with this , but I prefer the (multiparametric) one today"* (Subject 16, session 3).

### 7.7.3 These same users tended not to like the multiparametric volume

*"Volume is difficult. I'd prefer it on a slider"* (Subject 7, session 2)

*"I couldn't get used to wobbling the mouse. This is really alien!"* (Subject 14, session 1)

*"The wobbling really put me off"* (Subject 16, session 1)

### 7.7.4 Sliders interface causes a mental block/confusion

*"I had to think 'what does the pitch do?', 'what does the timbre do?"* (Subject 1, session 2)

*"I don't like these! I'm having to concentrate on one or two things; I can't do all four"* (Subject 2, session 2)

*"I HATE this one! While it's counting in I'm having to think where each slider needs to be. Then I forget the sound."* (Subject 3, session 2).

*"I do find this difficult! Breaking it down into four distinct parameters"* (Subject 4, session 2)

*"I cannot get on with this at all! I'm trying to think of too many things at once; there's too many things for my mind here. The interface forces you to decompose the sound and I don't think our minds are made for decomposing stuff like that"* (Subject 5, session 3).

*"I find I have to change each of the parameters sequentially, even though they can be physically controlled together"* (Subject 8, session 2).

*"It sounds silly, but I had a panic halfway through about which slider controls what"* (Subject 8, session 3)

*"I just can't be bothered with this one. I keep looking at the sliders and thinking 'what???' "* (Subject 11, session 2).

*"I don't like this. There's too much coming at you"* (Subject 15, sessions 1& 3).

### 7.7.5 Multiparametric allows spatial thinking

*"This feels multi-dimensional, gestural. I sometimes found myself thinking of a shape"* (Subject 1, session 1)

*"The 'hard' sound sounded like shapes in the screen. It felt more natural and I could think in gestures. You can almost feel the performance space"* (Subject 4, session 1)

*"I'm not thinking of timbre as a 'parameter' like I do with the sliders - but rather that it's 'sharper' over this side and 'duller' over this side"* (Subject 4, session 2).

*"I could concentrate on the performance without worrying about the actual mechanics of it"* (Subject 8, session 1).

*"I could warm to this. I like the gesture. I found myself on a couple of those sounds thinking diagonally! I didn't do that on the other interfaces. Vibrato is circular! You start to think gesturally."* (Subject 9, session 1).

*"I felt like I was 'thinking' too much about this. It's like my fingers knew where to go but the way I was thinking got in the way"* (Subject 16, session 1)

### 7.7.6 Multiparametric interface elicits subconscious control

*"You can use your unconscious to play it after a while. You can forget about the interface so that you can concentrate on the sound"* (Subject 13, session 3).

*"This feels free - which is good - but in a way like I'm not completely in control"* (Subject 2, session 1)

*"There's less mental load. All the controls are in one, so it's easier to position"* (Subject 3, session 2)

*"Felt easy to use, but felt more out of control"* (Subject 7, session 1).

*"It became more like driving a car - in that you've got physical actions that you can sort of get on with, leaving you free to think of other things"* (Subject 7, session 2).

*"This is a lot better* [than the other interfaces] *even though I felt out of control"* (Subject 8, session 1)

*"I felt like I knew where to move it. I didn't have to think. My brain was free enough to be distracted by the workmen* [external noises which the subject did not notice when doing the other interfaces]*"* (Subject 11, session 1).

### 7.7.7 Multiparametric interface is fun

*"This is really good fun! Even when you're not doing so well!"* (Subject 5, session 2).

*"One movement controlling several things is more fun. It's not like a task - it's like playing an instrument"* (Subject 11, session 1).

*"You're not so worried about getting it right. You can just say "ah well" if it's a bit off and then adjust it. It's less technical"* (Subject 4, session 2)

### 7.7.8 Multiparametric interface has more long-term potential

*"This is intuitively easier. I appear to have more degrees of freedom. It's more closely mirroring the way the brain is working"* (Subject 5, sessions 2 & 3).

*"This interface has possibilities. I'd choose this over the long term"* (Subject 8, session 3).

*"Yeah - I like this the best; definitely! It's a lot freer - more flowing"* (Subject 1, session 3)

*"This is more of an instrument than a device. I'd like to use this not for tests but just to play around on"* (Subject 2, session 3).

*"I'd like to play this freehand - not in a test setting"* (Subject 3, session 3).

*"It's the easiest of the three. It has the musical 'edge'. I'm controlling one thing (mouse) to get more than one dimension, whereas with the sliders I'm moving 3 objects if I need 3 dimensions"* (Subject 9, session 3).

*"You have more freedom. Your mind isn't restricted in a visual way. It's a blank screen."* (Subject 10, session 3).

*"This is the best interface. You've got more freedom, and there's so much potential if I could just get the hang of it"* (Subject 15, session 3).

To sum up, the majority of users enjoyed playing the multiparametric interface and thought (quite enthusiastically) that it had the best long-term potential. This was because it caused the least mental load and allowed spatial thinking. In contrast the sliders caused mental overload and forced too much 'decomposition' (analysis) of the sound. The mouse interface was popular for the simple tasks but everyone realised its limitations. There was a small group of people who much preferred the sliders, and they tended to be the group that commented on not liking the 'wobbling' volume control of the multiparametric interface.

## 7.8 Requirement for Longitudinal Tests

The above results provide the basis for an interesting discussion about the nature of the three interfaces and how people react to them when partaking in tests of varying complexity. However the hypothesis predicts that, over time, those interfaces based on **non**-analytical modes of operation will yield a greater accuracy and flexibility of control. Based on observations of performing acoustic musicians and car drivers we would not expect the results to show themselves **instantly**. Instead, we would expect a general improvement over time, but a lot of practice would be needed. It is predicted that at some point (in a user's practice schedule) the non-analytical styles will yield more accurate results than the directly analytical paradigms. It is not necessarily within the first few sessions that this will happen.

Therefore it was decided to run a second study, using fewer test subjects, but over a longer period of time. These 'longitudinal tests' would provide a more detailed time axis on which to compare the performance on the three interfaces. They are described in detail in the next chapter.

## 7.9 Summary

The cross-sectional tests have involved 16 users participating in three sessions. In each session they have performed the same 24 sound examples on each of the three interfaces. The sound examples varied in complexity from uni-parameter step changes to multi-parameter continuous changes. The data has been gathered, stored in a spreadsheet and used to produce a series of graphs. The graphs show that different people prefer different interfaces, but that there are some general trends to be seen when averaging across all the subjects. For the most complex tests the multiparametric interface gives the best scores, but for the simpler tasks it is the mouse. The trends cannot be extrapolated on the basis of three sessions and so a series of *longitudinal* tests was deemed necessary.

Interviews with the test subjects indicated that for most of the subjects the multiparametric interface was the most fun and had the most promise, whereas for most people the sliders interface was rather frustrating. There were a few notable exceptions who preferred the sliders interface at the expense of the multiparametric. Most people acknowledged that the more conventional 'mouse' interface was clearly limited.