Copyright in relation to this thesis*

Under the Copyright Act 1968 (several provisions of which are referred to below), this thesis must be used only under the normal conditions of scholarly fair dealing for the purposes of research, criticism or review. In particular no results or conclusions should be extracted from it, nor should it be copied or closely paraphrased in whole or in part without the written consent of the author. Proper written acknowledgement should be made for any assistance obtained from this thesis.

Under Section 35(2) of the Copyright Act 1968 ‘the author of a literary, dramatic, musical or artistic work is the owner of any copyright subsisting in the work’. By virtue of Section 32(1) copyright ‘subsists in an original literary, dramatic, musical or artistic work that is unpublished and of which the author was an Australian citizen, an Australian protected person or a person resident in Australia.

The Act, by Section 36(1) provides: ‘Subject to this Act, the copyright in a literary, dramatic, musical or artistic work is infringed by a person who, not being the owner of the copyright and without the licence of the owner of the copyright, does in Australia, or authorises the doing in Australia of, any act comprised in the copyright’.

Section 31(1)(a)(i) provides that copyright includes the exclusive right to ‘reproduce the work in a material form’. Thus, copyright is infringed by a person who, not being the owner of the copyright, reproduces or authorises the reproduction of a work, or of more than a reasonable part of the work, in a material form, unless the reproduction is a ‘fair dealing’ with the work ‘for the purpose of research or study’ as further defined in Sections 40 and 41 of the Act.

Section 51(2) provides that “Where a manuscript, or a copy of a thesis or other similar literary work that has not been published is kept in a library of a university or other similar institution or in an archives, the copyright in the thesis or other work is not infringed by the making of a copy of the thesis or other work by or on behalf of the officer in charge of the library or archives if the copy is supplied to a person who satisfies an authorized officer of the library or archives that he requires the copy for the purpose of research or study’.

*’Thesis’ includes ‘treatise’, dissertation’ and other similar productions.
THE HEAD OF THE SCHOOL OF THE UNIVERSITY OF SYDNEY CONSERVATORIUM OF MUSIC HAS AFFIRMED THAT THIS THESIS HAS BEEN EMENDED AS SPECIFIED BY THE EXAMINERS. THE CANDIDATE HAS MET ALL REQUIREMENTS IN ACCORDANCE WITH THE GRADUATE STUDIES COMMITTEE REQUIREMENTS.

Errata, reported by the author. 16/05/2007

page 80 appears twice consecutively

page 30 appears before page 29

Ludwig Sugiri
Sydney Conservatorium of Music Library
Eye movement, memory and tempo in the sight reading of keyboard music

Tony Souter

A thesis submitted in fulfilment of the requirements of the degree of Doctor of Philosophy

Sydney Conservatorium of Music
University of Sydney
June 2001
Declaration of originality

I declare that this dissertation is the result of my own work and that to the best of my knowledge it contains no material, either paraphrased or by direct quotation, previously written by another person, except where due acknowledgement is made in the text.

Candidate’s signature: 

Date: 26 June 2001
Abstract

This two-stage experimental study investigated the sight reading process from a scientific perspective and developed a new method for the training of sight reading skills. The study explored two aspects of the music reading process, the memory system and the movement of the eyes over the score, as potential targets for active intervention during training. The literature on these two aspects was reviewed and a computer system called SightReader was developed for 'stretching' the minimum size of the eye-hand span, the distance on the score that the eyes read ahead of the hands, while strictly controlling performance tempo. Another effect of this 'span-stretching' mechanism was the prevention of leftward refixation, the looking back at material on the score that has already been inspected.

In the first experiment (Stage I) an eye-movement tracking device was used to gather background information for the second experiment (Stage II), a set of teaching trials using SightReader. In Stage I the eye-hand span and leftward refixation were observed in nine highly skilled keyboardists under conditions in which tempo was controlled but the eye-hand span was not. The data showed enough flexibility in the way participants adapted their eye-hand span to a doubling of tempo to allay fears that there would be significant interference between SightReader's tempo and span settings in Stage II. Leftward refixation was observed at small or negligible rates in most participants, indicating that such behaviour was unlikely to represent a contaminating factor in Stage II.

In the second experiment (Stage II) all twenty participants were moderately skilled keyboardists. Experimental participants were exposed to SightReader's unique span stretching function over a series of training sessions, with minimal changes in tempo. Control participants were exposed to a more traditional training regimen with successive increases in tempo but no control over their span size. All participants underwent this more traditional method in a pretest and posttest so that any improvement in performance that had occurred during their contrasting training experiences could be compared. This was done by measuring mean levels of action slips. It was predicted that experimental participants would improve significantly more than control participants in the sight reading of all stimulus types. The results showed that the more difficult the stimulus type, the greater the experimental group's margin of improvement over the control group, with a very marked trend for two-part readings and a significant disparity for the most difficult readings, the four-part. This suggested that computer-aided stretching of the eye-hand span is a powerful tool for the training
of sight reading. It was also predicted that experimental participants' ability to play the stimuli from memory would improve through their exposure to span stretching significantly more than for control participants, but this was not confirmed by the data.
Acknowledgments

I have been privileged to receive guidance and support of the highest quality from two scholars at the University of Sydney who have supervised the project: Dr Peter Dunbar-Hall, Senior Lecturer and Chair of Music Education, Sydney Conservatorium of Music, and Dr Ros Markham, Senior Lecturer and Coordinator of Doctoral Studies, Department of Psychology. In addition, Dr Cyril Latimer, then Senior Lecturer, Department of Psychology, offered valuable advice in relation to the experiment in Stage I, and Dr Margaret Charles, Senior Lecturer, Department of Psychology, gave assistance in relation to the statistical methods used in both stages.

I am also indebted to three officers of the University for technical support. Mr Russell Emerson, technical officer at the Centre for Performance Studies, provided technical support for the eye-tracking experiment in Stage I. Mr Andrew Martin, Visiting Programmer, wrote the prototype of SightReader under my direction. Mr Peter Thomas, Manager of the Sydney University Electronic Sound Studio, provided technical support for the experiment in Stage II, including the preparation of the stimuli for the memory tests.

Finally, I am most grateful to Dr Guenter Plum of the University of Hong Kong for editorial advice in relation to the writing of the dissertation.

Tony Souter, June 2001
Table of contents

Declaration of originality ............................................................................... i
Abstract ........................................................................................................ ii
Acknowledgments ........................................................................................ iv
Table of contents .......................................................................................... v
List of figures ................................................................................................ viii
List of tables ................................................................................................... xi

1 Introduction ................................................................................................. 1
  1.1 Aims, objectives, and structure of the dissertation ............................... 1
  1.2 What is sight reading? ........................................................................... 4
    1.2.1 Sight reading and familiarity ....................................................... 4
    1.2.2 Educational and research contexts of sight reading ...................... 11
  1.3 Technical overview of the music reading apparatus.............................. 18
  1.4 Brief description of SightReader ......................................................... 28

2 Memory and music reading ................................................................……… 31
  2.1 Review of selected literature on memory ............................................. 32
  2.2 Music reading and the memory system ................................................. 46
    2.2.1 Information sequencing, activation and the multicomponent model 46
    2.2.2 Chunking and automaticity .......................................................... 50
  2.3 Conclusions ........................................................................................... 55

3 Eye movement in the reading of language and music ............................... 56
  3.1 Background to visual input ................................................................... 56
  3.2 Review of selected literature on eye movement in language reading ..... 58
    3.2.1 The first period: unassisted observation and introspection .......... 58
    3.2.2 The second period: early tracking technology and the rise of research into eye movement in reading .............................. 60
    3.2.3 The third period: cognitive psychology, infrared tracking and computer technology .................. 62
    3.2.4 Top-down and bottom-up processing ............................................ 64
    3.2.5 Peripheral input and processing across fixations ........................... 66
    3.2.6 Refixation ...................................................................................... 69
    3.2.7 The eye-voice/eye-hand span ...................................................... 72
<table>
<thead>
<tr>
<th>Section</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.2.8</td>
<td>Conclusions</td>
<td>73</td>
</tr>
<tr>
<td>3.3</td>
<td>Review of the literature on eye movement in music reading</td>
<td>74</td>
</tr>
<tr>
<td>3.3.1</td>
<td>Equipment and related methodology</td>
<td>76</td>
</tr>
<tr>
<td>3.3.2</td>
<td>The tempo/skill/action-slip fallacy</td>
<td>80</td>
</tr>
<tr>
<td>3.3.3</td>
<td>Fixations, pauses and the number/duration relationship</td>
<td>85</td>
</tr>
<tr>
<td>3.3.4</td>
<td>Tempo</td>
<td>88</td>
</tr>
<tr>
<td>3.3.5</td>
<td>Musical complexity</td>
<td>89</td>
</tr>
<tr>
<td>3.3.6</td>
<td>Reader-skill</td>
<td>91</td>
</tr>
<tr>
<td>3.3.7</td>
<td>Stimulus-familiarity</td>
<td>92</td>
</tr>
<tr>
<td>3.3.8</td>
<td>The top/down bottom/up question</td>
<td>94</td>
</tr>
<tr>
<td>3.3.9</td>
<td>Peripheral visual input</td>
<td>97</td>
</tr>
<tr>
<td>3.3.10</td>
<td>Refixation</td>
<td>98</td>
</tr>
<tr>
<td>3.3.11</td>
<td>The eye-hand span</td>
<td>100</td>
</tr>
<tr>
<td>3.3.12</td>
<td>Conclusions</td>
<td>103</td>
</tr>
<tr>
<td>4</td>
<td>Stage I: observing tempo, span size and refixation in music reading</td>
<td>104</td>
</tr>
<tr>
<td>4.1</td>
<td>Method</td>
<td>107</td>
</tr>
<tr>
<td>4.1.1</td>
<td>Equipment</td>
<td>110</td>
</tr>
<tr>
<td>4.1.2</td>
<td>Stimuli</td>
<td>110</td>
</tr>
<tr>
<td>4.1.3</td>
<td>Tempo and output dysfunction</td>
<td>113</td>
</tr>
<tr>
<td>4.1.4</td>
<td>Participants</td>
<td>115</td>
</tr>
<tr>
<td>4.1.5</td>
<td>Procedure for conducting the trials</td>
<td>116</td>
</tr>
<tr>
<td>4.1.6</td>
<td>Data registration</td>
<td>119</td>
</tr>
<tr>
<td>4.2</td>
<td>Tempo and span size</td>
<td>121</td>
</tr>
<tr>
<td>4.2.1</td>
<td>Theoretical background</td>
<td>121</td>
</tr>
<tr>
<td>4.2.2</td>
<td>Results and discussion</td>
<td>126</td>
</tr>
<tr>
<td>4.3</td>
<td>Tempo, refixation and the O’Regan effect</td>
<td>131</td>
</tr>
<tr>
<td>4.3.1</td>
<td>Theoretical background</td>
<td>131</td>
</tr>
<tr>
<td>4.3.2</td>
<td>Results and discussion</td>
<td>137</td>
</tr>
<tr>
<td>4.4</td>
<td>Conclusions</td>
<td>144</td>
</tr>
<tr>
<td>5</td>
<td>Stage II: controlling tempo, span size and refixation with SightReader</td>
<td>146</td>
</tr>
<tr>
<td>5.1</td>
<td>Theoretical background to SightReader</td>
<td>148</td>
</tr>
<tr>
<td>5.1.1</td>
<td>Music education: looking ahead of the hands</td>
<td>149</td>
</tr>
<tr>
<td>5.1.2</td>
<td>Computer technology: the gaze-contingency paradigm</td>
<td>151</td>
</tr>
<tr>
<td>5.1.3</td>
<td>Cognitive psychology: span size, refreshment, arousal and chunking</td>
<td>152</td>
</tr>
<tr>
<td>5.1.4</td>
<td>Cognitive psychology: memorisation</td>
<td>155</td>
</tr>
<tr>
<td>5.2</td>
<td>Method</td>
<td>157</td>
</tr>
<tr>
<td>5.2.1</td>
<td>Program design of the prototype of SightReader</td>
<td>157</td>
</tr>
<tr>
<td>5.2.2</td>
<td>Equipment setup</td>
<td>160</td>
</tr>
<tr>
<td>5.2.3</td>
<td>Participants</td>
<td>161</td>
</tr>
</tbody>
</table>
List of figures

Chapter 1

Figure 1 Bars 13-20 from *Danse* (Debussy 1890) ........................................... 5
Figure 2 Bars 21-28 from *Danse* (Debussy 1890) ........................................... 5
Figure 3 Shifts in the source of information as the same text is read repeatedly .......... 8
Figure 4 Changing contributions of the three hypothesised types of familiarity as the same text is read repeatedly ........................................... 9
Figure 5 The oculomotor group and related signals ........................................ 20
Figure 6 The musculoskeletal group and related signals .................................... 20
Figure 7 The flow of information through the music reading apparatus ............... 23
Figure 8 Temporal characteristics of the signals in relation to a music score ........... 26
Figure 9 The opening window of the prototype of *SightReader* ......................... 29
Figure 10 The display of a stimulus on *SightReader* before starting to play ......... 30
Figure 11 The same display after having played the first chord .................................. 30
Figure 12 The same display after having played the seventh chord .......................... 30

Chapter 2

Figure 13 Schematic representation of the memory system based on the conception of Atkinson & Shiffrin (1968) ........................................... 35
Figure 14 Sloboda’s hypothetical durational organiser for chunking (adapted from Sloboda 1985:73) .................................................. 53
Figure 15 Excerpt from a figured-bass exercise (Morris 1933:5) .......................... 54
Figure 16 A solution for the same exercise .................................................. 54

Chapter 3

Figure 17 Schematic representation of a typical scanpath over a keyboard score .......... 58
Figure 18 Example of the early use of gaze-contingency displays, adapted from McConkie & Rayner (1975) ........................................... 62
Figure 19 A schematic example of the O’Regan effect in language reading ............... 69
Figure 20 No refixation ........................................... 69
Figure 21 Rightward refixation on the same word ........................................... 69
Figure 22 Leftward refixation to a previous word .................................................. 69
Figure 23 Upper Tempo areas for a less skilled reader  
Lower Tempo areas for a more skilled reader ..................................................... 81
Figure 24 Upper Schematic representation exemplifying the relative durations of saccades, fixations and pauses  
Lower Schematic representation exemplifying a different relativity of the durations of saccades, fixations and pauses ........................................ 87
Figure 25 Slow tempo: more fixations of slower duration and more closely spaced than in Figure 26 ................................................................. 89
Figure 26 Fast tempo: fewer fixations of shorter duration and spaced further apart than in Figure 25 ................................................................. 89
Figure 27 Weaver’s illustrations of eye movement in (a) the homophonic example and (b) the polyphonic example ........................................ 95
Figure 28 Sloboda’s schematic representations of (a) vertical saccades over homophonic music, and (b) horizontal saccades over polyphonic music .................................................. 96
Figure 29 Sloboda’s quotations of fragments of (a) vertical scanpath over homophonic music, and (b) horizontal scanpath over polyphonic music, from Weaver (1943) .................................................. 97
Figure 30 Categories of refixation in the reading of a keyboard score ........................................ 99

Chapter 4
Figure 31 Schematic depiction of the equipment setup .................................................. 108
Figure 32 Trial Stimulus A, adapted from Bartlett et al (1976:624) ........................................ 111
Figure 33 Trial Stimulus B, adapted from Bartlett et al (1976:669) ........................................ 111
Figure 34 The load/latency line \( y = \frac{x}{2} \) and its main features ........................................ 124
Figure 35 The position of each participant’s load/latency relationship  
on the line \( y = \frac{x}{2} \) ........................................ 128
Figure 36 Mean latency at slow and fast tempos ........................................ 128
Figure 37 Schematic illustration of the number/duration curve 
\[ nd = \frac{1}{2} \] ........................................ 133
Figure 38 The number/duration relationships on the curve 
\[ nd = \frac{1}{2} \] ........................................ 139
Figure 39 Mean pause duration in ms for each participant, at slow tempo (blue) and fast tempo (yellow) ........................................ 139
Figure 40 Rates of leftward (striped) and vertical (plain) refixation  
at both tempos ........................................ 141
Figure 41 Distribution of pauses over chords at slow tempo at slow tempo (right-side) and fast tempo (left-side) ........................................ 143
Chapter 5

Figure 42 Depiction of the equipment setup ........................................... 161
Figure 43 Compressed version of the Initial Questionnaire ..................... 163
Figure 44 Response template for Questions 4 and 5, including the numerical equivalents to the boxes ........................................... 164
Figure 45 Response template for Questions 6 and 7 .................................. 165
Figure 46 Example of a memory test for a one-part stimulus .................... 168
Figure 47 Example of a memory test for a four-part stimulus ................... 168
Figure 48 Posttest/pretest proportional values for mean rates of action slips for experimental (red) and control (blue) groups .......... 180
Figure 49 Compressed version of the Final Questionnaire ....................... 184
Figure 50 Response template for Questions 1-8 ....................................... 185
Figure 51 Results for experimental (blue) and control (mauve) groups for Questions 1-8 ................................................ 186
Figure 52 Response template for Questions 9 and 10 .............................. 188
Figure 53 Results for experimental (blue) and control (mauve) groups for Questions 9-12 ........................................... 186
Figure 54 Response template for Questions 11 and 12 ............................. 189
Figure 55 Thematically coded responses to Question 13 ......................... 194
Figure 56 Thematically coded responses to Question 14 ......................... 195
List of tables

Chapter 1
Table 1 Summary of the sources of musical familiarity ................................... 8
Table 2 Components and signals of the music reading apparatus ...................... 22

Chapter 2
Table 3 Summary of the components of the memory system based on the conception of Atkinson & Shiffrin (1968) .................................. 36

Chapter 3
Table 4 Mean fixation duration by encounter ............................................. 93
Table 5 Summary of the effect of four key variables on fixations .......... 94

Chapter 4
Table 6 Observation and control of the three key aspects of music reading in Stages I and II .................................................. 105
Table 7 Marks deducted for action slips ................................................... 115
Table 8 Summary of tempo and equipment conditions during each reading .......... 118
Table 9 Mean load and latency at each tempo, and the resulting load/latency relationships .................................................. 127
Table 10 Mean number of pauses per chord and mean pause duration at each tempo, and the resulting number/duration relationships ...... 137
Table 11 Rates of vertical and leftward refixation ........................................ 140
Table 12 The distribution of pauses over chords ........................................ 142
Table 13 Tempo-induced changes in refixation and the O'Regan effect .......... 144

Chapter 5
Table 14 Basic structure of the experiment in Stage II .................................. 147
Table 15 Results for Questions 2-7 of the Initial Questionnaire ...................... 163
Table 16 Guidelines for altering the tempo settings for Readings 2-6 under the normal condition .................................................. 171
Table 17 Skills targeted by various combinations of the settings on SightReader .................................................. 173
Table 18 Guidelines for altering the span setting for Readings 2-5 under the span-stretching condition .................................................. 173
Table 19 Mean tempos for two- and four-part stimuli, with proportional values .................................................. 177
Table 20  Mean levels of action slips for two- and four-part stimuli, with proportional values .................................................. 178
Table 21  Mean scores in the memory tests, with proportional values ........ 182
Table 22  Results for Questions 1-8 ................................................................. 185
Table 23  Results for Questions 9 and 10 ...................................................... 188
Table 24  Results for Questions 11 and 12 .................................................... 189
Table 25  Thematic profile of verbal responses to Questions 13 and 14 ......... 192
1 Introduction

1.1 Aims, objectives and structure of the dissertation

The turn of the twenty-first century is an unstable time for music education. Like many human activities, music education has been profoundly affected by the rapid acceleration in the development of computer technology over the last decade. Computer-aided instruction (CAI) of music skills is now available for the training of a wide variety of skills in such areas as aural perception, theory, analysis, music history, composition, keyboard and vocal performance. CAI is powerfully flexible in its ability to bring together several modes of information, including music notation, sound recordings, graphics, video and written text. This flexibility extends to its interactive potential and the fact that it can perform a variety of functions that hitherto were the sole domain of the human instructor, such as administering exercises of appropriate difficulty, monitoring students' progress and delivering customised feedback. Students can typically use it without direct supervision, allowing them to assert greater control over their learning in a self-paced environment that is more private and less stressful than classroom or one-to-one contact with human instructors.

Although it is difficult to predict exactly how computer technology will develop, it is relatively safe to assume that the coming decade will see computer technology become more sensitive to the particular abilities and requirements of individual users, and increasingly capable of speech-based interface. CAI thus represents a sudden and substantial shift in the training environment that promises to change the traditional role played by the human instructor in the acquisition of musical skills. Yet the technology now available for music training will almost certainly be regarded by future generations as crude and primitive, as mere opening steps towards the development of an educational infrastructure that enables people to acquire skills far more rapidly, deeply and easily than is currently possible. For those generations, the boundaries between pedagogy, psychology, neuroscience and computer science may have become considerably less distinct, with instructors cast as applied scientists rather as physicians apply medical science to health care.

The first appearance of many technical terms in this dissertation is in small capitals; this is a signal that a term is defined in the glossary at the end, as also noted later in this section.
The challenge for music educators in this historically unstable situation is to question, analyse and evaluate both the traditional and the new, and to be intellectually, professionally and personally open to change. If such a challenge is met, music training will be in a strong position to flourish during the coming century. That is the viewpoint from which this empirical study of musical behaviour was conceived. As such it was natural that the study should have taken on a scientific and interdisciplinary orientation, should have aimed to analyse and evaluate certain aspects of musical behaviour and pedagogy, and should have explored possibilities for further developing CAI.

One goal of the study was to analyse the process of performing music from score, to describe the various components of that task from the perspective of anatomy and information flow, and to investigate the parts of the components that hitherto had been neglected: the role of the memory system and the way the eyes move across the score. Here 'memory system' was defined broadly to encompass musicians' ability to draw on their accumulated expertise and familiarity with relevant musical styles and genres, as well as to process and store information picked up from a score in a way that results in a performance. This was seen as a foundation for understanding the music reading process from a scientific point of view, from which this study and others might proceed to meet the challenges faced by music training in a rapidly changing technological context. The other goal was to develop and trial SightReader, a unique custom-built computer system conceived by the author for improving sight reading and memorisation skills at the keyboard through strategic control of a musician's eye movement and memory system. The expectation was that SightReader would represent an effective tool for improving keyboardists' ability to promptly and efficiently familiarise themselves with a musical text, and that the system might be suitable for integration into a wider curriculum for performance training. These goals were largely achieved in the study.

The point of departure in the next section is an analysis and definition of sight reading in terms of the concept of familiarity, and an explanation of its educational and research contexts, including a discussion of why the training of sight reading should be considered important, and why it has been and remains problematic. This is followed by a technical overview of the MUSIC READING APPARATUS, a term that refers collectively to the parts of the body that contribute to the task of music reading, and the flow of information between them. Apart from serving as a foundation for understanding the music reading process, the overview introduces
numerous terms and concepts that occur frequently in the study. Chapter 1 ends with a brief description of how *SightReader* works.

Chapters 2 and 3 are essentially literature reviews, each covering a key function of the apparatus in greater detail. Chapter 2 deals with the *memory system*, emphasising its role in the reading process, while Chapter 3 concerns eye movement in the reading of both language and music. These literature reviews lay the foundation for the two empirical stages of the study. Chapter 4 is an account of Stage I, an experiment in which a tracking device was used to examine the effect of tempo on two aspects of music reading behaviour that *SightReader* alters: the distance between the eyes and the hands in terms of information on the score, and the widely observed tendency of music readers to *refixate*, or look at information on the score more than once. Chapter 5 describes Stage II, a teaching experiment in which the effects of using *SightReader* were investigated as students read short musical extracts in a series of trials. Chapter 6 draws together and summarises the findings of the study as a whole, and proposes directions for further research.

The study grew out of an interest in performance of western tonal art music from score, and although this remains the primary orientation of the thesis, its findings may be applicable to musical performance in many notation-based cultures. The dissertation is designed for a variety of readers, including practising musicians, educators and psychologists; for this reason, the register of the text is interdisciplinary. Many terms and concepts are explained in greater detail in the body of the text than would be the case if this study had been conceived entirely within a single discipline. In addition, a glossary of technical and semi-technical terms from the fields of psychology, education, statistics and music is included at the back of the dissertation; these terms have been highlighted in *small capitals* on their first appearance in the dissertation, and occasionally on subsequent appearances where this would be helpful to the reader.

Before proceeding, it is timely to state a sequence of four pedagogical principles that have guided the thesis. These principles are by no means original, and are implicit in the teaching and research of many musicians. The first principle is to identify a skill and define the objectives and problems of teaching and learning that skill. The second is to analyse the components of the skill and the conditions under which it can be observed. The third is to develop methods of practising these component-skills, typically through repetitive drilling procedures. The fourth is to integrate these
Introduction

procedures into a wider curriculum so that students are able to contextualise the skills they have acquired and apply them in real-life situations.

1.2 What is sight reading?

1.2.1 Sight reading and familiarity

A definition of sight reading is of central concern before discussing its educational and research contexts. Whether applied to keyboardists or non-keyboardists, soloists or ensemble performers, sight reading commonly refers to a particular reading condition in which musicians perform from a score they have not previously seen. As such, sight reading may be seen as a subset of music reading. Sight reading is defined in the *Grove Concise Dictionary of Music* (Sadie 1988:696) as ‘the performing of a piece of music at first sight’. In his substantial and authoritative monograph on the performance of music, Gabrielssohn (1999:501) defines sight reading as ‘performing from a score without any preceding practice on the instrument of that score’, thus excluding from the definition preparation through silent reading of a score. These simple definitions rest on the absence of familiarity with a musical text: the player either has or has not already seen it. Such an either-or definition predominates in the literature because it is easy to understand and provides a convenient criterion for assessing and researching sight reading.

In this section, however, a case will be put that it is difficult, if not impossible, to arrive at a universal definition of sight reading. It will instead be argued that the conventional notion of sight reading is simplistic and that to be meaningful, sight reading must be defined specifically for each reading situation. Moreover, when it comes to the scientific observation of music reading, it will be pointed out that the conventional notion is inadequate for the selection of participants and stimuli, and that a more detailed and analytical approach is necessary to ensure that the data generated by such experiments are reliable. Accordingly, no fewer than three working definitions of sight reading will be presented for the purposes of this study, based on the analysis below of the concept of familiarity with a musical text. The questions of whether sight reading is a desirable skill, and why there is a need to develop new ways of improving it, are also addressed in the next section.

Narmour (1999) has presented a case that familiarity with a musical text (‘stylistic expectations’) can be analysed in terms of a hierarchy from lower levels, involving scales and modes among other features, through frequently occurring melodic contours to such complex procedures as chord progressions, modulations and large-
scale structures. For the purpose of this study, however, it is regarded as more useful to approach the issue of familiarity on a simpler level, to enable the clarification of particular issues in relation to sight reading.

As pointed out above, the conventional notion of sight reading appears to presume that the performer is uniformly unfamiliar with a musical text. There are a number of problems with this. To begin with, it ignores the ubiquitous phenomenon of musical repetition. This comes in many forms, the most obvious being the repetition of large sections of text that underpin the organisation of such common structures as binary, ternary, rondo, and sonata forms. No-one would disagree that during the first reading of an unfamiliar piece readers are not, strictly speaking, sight reading as they play repeated sections, since they are no longer unfamiliar with them. The problem might be side-stepped by excluding sight reading material in which whole sections are repeated. But musical repetition usually occurs on smaller levels as well, down to short phrases and even ideas within the beat; from the smallest to the largest scale, repetition is typically what gives music cohesion, and is therefore a deeply integrated feature of most styles. Figure 1, an excerpt from Claude Debussy’s (1890) piece for piano, Danse, illustrates just how complex the business of repetition in music reading is.

![Figure 1 Bars 13-20 from Danse (Debussy 1890)](image1)

![Figure 2 Bars 21-28 from Danse (Debussy 1890)](image2)

In Figure 1, the number of times the player will have previously read the information in each bar is shown as a cardinal numeral underneath that bar. The first phrase (bars
13 to 16) is immediately repeated in whole (bars 17 to 20). Furthermore, within the first phrase bar 13 is itself repeated exactly in bar 14, and again in the second phrase, in bar 18. In addition to these whole-bar repetitions, the repeated-note motif played by the left hand is exactly the same in every bar, and presumably requires little attention by the reader once the passage has begun. To read this passage for the first time, then, is to experience a patchwork of greater and lesser familiarity.

Figure 2 shows the material that immediately follows Figure 1, that is, bars 21-28. This is more difficult to analyse in terms of familiarity, since although internally it is an exact repetition of bars 13-20, it is transposed upwards by an augmented second. The second example is therefore the same as yet different from the first. It can be assumed that the familiarity gained from playing the first example is less useful here than if the repetition were at the same pitch, but more useful than it would be if the first example had not been experienced in the first place. The traditional assumption that in sight reading the player is uniformly unfamiliar with the music, then, is at odds with the complex and multilayered phenomenon of internal repetition.

Despite the difficulty of quantifying familiarity with material that is repeated at a different pitch, repetition in these examples is a comparatively clear-cut matter involving whole bars. In much music, particularly that which is organically or developmentally organised such as the classical sonata movement, the patterns announced early in the movement (the thematic ideas) are likely to be repeated not only in exact form at various pitches during the course of the movement, but in other forms in which the relationships between the notes within the thematic cells undergo gradual alteration. The opening thematic patterns may remain familiar throughout the first playing of such music, but like the exactly transposed material in Figure 2, the extent to which such familiarity can assist players in the rest of the movement is hard to quantify.

Thus far this discussion has been concerned only with familiarity sourced from within a text currently being read (‘intraopus’ familiarity in Narmour’s model). It can be assumed, however, that readers also draw on their memory of other music that is similar to the material at hand (‘extraopus’ familiarity). Returning to Figures 1 and 2, consider the case in which players already have considerable experience of playing the piano music of Debussy, but not this particular piece. As they read through these excerpts, they have no prior knowledge of its actual patterns except for those that arise from the repeated patterns within the text, as previously discussed. But their experience in playing other music of a similar style can be assumed to be of
assistance by rendering them more sensitive to the patterns that are likely to occur here. The patterns at issue may be both compositional—including Debussy's typical harmonies, textural figurations, and indeed his penchant for small-scale repetition—and performance-related matters such as typical hand positions, fingering patterns, and characteristic ways of pedalling.

Familiarity sourced from other texts can be seen in terms of layers of increasing stylistic and technical distance from the text being read; the more distant the layer, the less specific a familiarity it is likely to provide to the patterns that are encountered in the text. In the case of Danse, the most specific layer of familiarity might arise from readers' experience of Debussy's piano works of similar conception. Less specific might be Debussy's overall compositional style, and less specific still the output of composers who share some of his stylistic features, such as Ravel. Further outwards might lie the output of French composers of the late nineteenth and early twentieth centuries, and then that of all late romantic European composers, who share common stylistic features that distinguish them from their predecessors in the classical and baroque periods. In contrast, the sources relevant to Danse would not be as useful in performing from a score by a SERIALIST composer such as Schoenberg, to which a somewhat different set of more and less specific sources of familiarity would apply.

Musical familiarity as discussed here is sourced from within the memory system, and as such can be classified as INTERNAL INPUT. This is contrasted with the other basic source of information from which a musician constructs a performance, EXTERNAL VISUAL INPUT, the stream of visual images that are picked up from a score. Performing music from score involves a constant interplay between external and internal input, and since both forms of input can be assumed to function unless a text is being played totally 'from memory', it is submitted that there is no in-principle distinction between the first reading and subsequent readings. This represents a challenge to the traditional notion of sight reading, to which this distinction is central. All that can be said is that internal input is at its weakest in first reading a musical text, exposing the music reading apparatus to the greatest amount of unfamiliarity. Each subsequent reading of the same text can logically be assumed to strengthen a performer's familiarity with it and thus the efficacy of internal input, while at the same time reducing her reliance on external visual input. This is why subsequent performances of the same text can be expected to be easier for the musician, and to produce a performance that is typically of better quality. This last point is of great importance to the experiment in Stage II, as explained in Section 5.1.3.
It can also be assumed that the repeated reading of a musical text, which is typical during its learning, is accompanied not only by a shift from external to internal input but a shift in the type of internal input. Here a new distinction is necessary, between what will be referred to here as ACTUALITY FAMILIARITY and PROBABILITY FAMILIARITY. Actuality familiarity is based on a reader's memory of the actual notes of a text, which naturally strengthens during repeated readings of the same text. Actuality familiarity is also at play while reading internal repetitions of a text, as was the case in the excerpts from *Danse*. Naturally, actuality familiarity is normally sourced only from within a text, since it would be rare for exactly the same musical patterns to exist in other texts.

By contrast, probability familiarity occurs when readers' experience of previous information in the text makes them more sensitive to the likely patterns they will encounter. This type of familiarity is more subtle. It can originate within the text currently being read as its musical logic unfolds, as clearly illustrated in the example cited earlier of gradually altering thematic material in a sonata movement from the classical period. Probability familiarity can also be sourced in other similar texts which, as pointed out above, may be viewed as lying in successive layers of stylistic and technical distance from the text at hand. The more that actuality familiarity is strengthened by repeated readings, the less players can be assumed to rely on their probabilistic knowledge of how the text might unfold. The sources of musical familiarity are summarised in Table 1 and the two qualitative shifts are shown in Figure 3.

<table>
<thead>
<tr>
<th>encounter</th>
<th>source of actuality familiarity</th>
<th>source of probability familiarity</th>
</tr>
</thead>
<tbody>
<tr>
<td>first reading</td>
<td>repetition of whole sections and smaller patterns within the current text</td>
<td>unfolding probabilities of the current text; previous readings of other texts</td>
</tr>
<tr>
<td>subsequent readings</td>
<td>previous readings of the current text</td>
<td>previous readings of the same and other texts</td>
</tr>
</tbody>
</table>

Table 1 Summary of the sources of musical familiarity

Figure 3 Shifts in the source of information as the same text is read repeatedly
Figure 4 is a hypothetical illustration of how the shifts illustrated in Figure 3 occur encounter by encounter. Each vertical bar represents an encounter with the same musical text, starting with the first encounter (1 on the horizontal axis). External visual input is represented by the yellow portion of each bar, and internal input by red (actuality familiarity) and purple (probability familiarity). The first encounter is closest to the traditional notion of sight reading, during which there is no familiarity at all with the upcoming notes in the piece based on having played the text before. The exception to this may be any actuality familiarity arising from repetition within the text; this is not shown on the graph for the sake of simplicity.

During this first encounter, players are assumed to bring to bear their probabilistic familiarity (red) based on an unfolding realisation of the parameters of the style, as explained above. Reliance on visual input (yellow) can be assumed to be at its greatest during the first reading; in other words, the first reading is likely to involve the most intense scrutiny of the score. In the second reading (2 on the horizontal axis) there is assumed to be a significant shift in the balance between information sources as the process benefits, however weakly, from the actuality familiarity that is acquired during the first reading. The balance between the two information sources thus moves from information on the page towards information that is stored in the memory system. Accordingly, it can be assumed that external visual input and probabilistic familiarity are not quite as important in the second reading as they were in the first.

![Figure 4](image-url)  
**Figure 4** Changing contributions of the three proposed types of familiarity as the same text is read repeatedly
The second and subsequent playings thus enjoy the advantage of ever greater contributions to the music reading process from internal input. The change in the balance between information sources is assumed to be greatest between the first and second encounters, and steadily less between subsequent encounters. This is because familiarisation with the notes of a text occurs most rapidly during the first few playings, then tapers off for subsequent encounters in what is known technically as a negatively accelerating curve. Past a certain number of encounters, actuality familiarity (of the actual text) may become so strong that the contribution of probability familiarity is minimal. To take the process further, it is quite possible for musicians to memorise a work so reliably that they can play without score altogether 'from memory', in which case the strength of internal input completely obviates the need for external visual input. In other words, at some point to the right in Figure 4 there would be no yellow portion at all in the vertical bars. This explanation is aligned with Goolsby's (1987) observations of the way in which eye movement changes during successive readings of the same text, discussed in Section 3.3.7.

In the light of the foregoing analysis, the traditional condition-independent definition of sight reading might be expressed as 'the first reading of a musical text with which the player has no actuality familiarity'. But such a definition of sight reading is misleading and of limited use, since each individual reading situation carries with it a different complex of familiarities and thus requires its own definition of sight reading. A definition that takes into account the condition-dependent nature of sight reading would be 'the first performance from score of a musical text with which familiarity is initially constrained in some way'. An analogous definition of sight reading ability would be 'the ability to familiarise oneself rapidly and efficiently with musical material which is initially of restricted familiarity to the player'. These definitions, however, are still misleading for situations that demand a precise meaning of the term, in particular empirical situations in which the uncontrolled influence of familiarity carries a great risk that data will be contaminated or otherwise unreliable.

Stage I of this study, for example, involves the measurement of eye movement in a single reading under strict conditions. It will be argued in Chapter 4 that participants' familiarity has a contaminating effect on the data, and thus needs to be tightly controlled. Accordingly, for Stage I sight reading is defined as 'the first reading of a musical text, (1) with which the player has no actuality familiarity, (2) where both text and player have been chosen such that probability familiarity is minimised within certain controlled limits, and (3) where the level of newness encountered by the player in reading through the text is reasonably consistent'. This requires the
generation of rules for the selection of stimuli and experimental participants that ensure reasonable consistency of familiarity between participants and minimise that familiarity in order to minimise the potential for contaminating the eye movement data. At the same time these rules must be practicable given the circumstances of the experiment, and must not create a reading situation that is substantially different from the real-life reading situations the experiment is designed to investigate.

Stage II of the study is a teaching experiment that examines sight reading as a gradual process of familiarisation with a music text. One of its foci is on the shifts in familiarity that occur in repeated readings of the same text, as illustrated in Figures 3 and 4. In Stage II there is the same need as in Stage I to control initial familiarity, so the definition is extended to ‘the first six readings of a musical text, (1) with which the player initially has no actuality familiarity, (2) where both text and player have been chosen such that probability familiarity is initially within certain limits, and (3) where the level of newness encountered by the player in reading through the text is reasonably consistent’. These definitions are revisited in Section 4.1.2 and 5.2.6 in relation to generating rules for choosing stimuli and participants for the experiments.

1.2.2 Educational and research contexts of sight reading

In the previous section, sight reading was defined as a subset of music reading in which familiarity is constrained in some way, specifically in terms of the layers of residual familiarity in the reader’s memory system. Accordingly, sight reading ability was defined in terms of the rapid and efficient familiarisation with a musical text. The focus of the discussion is now broadened to address the crucial questions of why sight reading should be worthy of attention by students, instructors and researchers, and why it should be considered problematic.

There appears to be a wide range of sight reading ability among musicians. Nowhere is this clearer than in the literature on eye movement in music reading, reviewed in Chapter 3, in which comparing groups of skilled and unskilled readers has been a central investigatory theme of over a dozen studies. Further evidence that sight reading ability varies significantly among performers and that the measurement of this variation is regarded as important lies in the existence of several standardised tests of sight reading ability. The Watkins-Farnum Performance Scale (Watkins & Farnum 1954; see Stivers 1972 for an evaluation) is the most well-known; it categorises reading ability on the basis of such components as pitch, rhythm, articulation, tempo and notational knowledge, measuring them by observing the musician’s reading of increasingly difficult stimuli.
Introduction

The assessment of sight reading is also incorporated into many institutional performance curricula including the syllabuses of major music examination bodies such as Australian Music Examinations Board (2000:14), Trinity College, London (1999:33) and the Associated Board of the Royal Schools of Music (1998:18). Sight reading is referred to as a core skill in the American *National standards for arts education* (Music Educators’ National Conference 1994), which states unequivocally that all grade school students from Grade 5 onwards should be able to sight read ‘simple melodies in both the treble and bass clefs’ (p44) and that proficient students in secondary choral and instrumental ensembles should be able to ‘sight read, accurately and expressively’ music of moderate difficulty (p61). Prominent American educators have reinforced these statements, for example Demorest (1998:183) claims that ‘choral music educators have long acknowledged the importance of teaching sight singing’. Further evidence that sight reading is treated as a serious issue in performance training can be found in the publication of numerous instructional manuals for sight reading during the 20th century. These texts were reviewed in Covington’s comprehensive (1981) survey, which is discussed below.

It is ironic that in such an energetic discourse on sight reading—empirical investigation, measurement apparatus, syllabuses, public statements and instructional manuals—there is little discussion of why sight reading is regarded as a valuable skill with which to imbue students, and what role it might have in performance tuition. Addressing such basic questions, after all, should properly have been the starting point of the discourse. For example, it needs to be pointed out widely and repeatedly among educators that fluency in sight reading, as opposed to the ability to learn a text through intensive rehearsal, can be assumed to represent a significant advantage in professional contexts such as ensemble playing, accompanying and teaching, where preview of a score is not always convenient or practicable. Fluent sight readers are also likely to be able to acquire a comparatively large performance repertoire, since they are by definition capable of rapidly and efficiently familiarising themselves with previously unknown texts. Moreover, since sight reading involves the minimising of internal input from prior rehearsal of a musical text, it is the ultimate display of the capacity of the music reading apparatus to take up information from a score, store and process it in the memory system, and from this produce a functional performance. Sight reading ability presupposes musculoskeletal agility, the capacity to make instantaneous decisions, and as Burmeister (1991:v) expresses it, ‘courage and self-reliance’ as a musician. And by virtue of her notational adeptness, the advanced sight reader is likely to be superior
Introduction

Making sense of large ensemble scores, an important skill for musicologists and conductors.

It also needed to be explained in the literature that sight reading can take on at least three quite different functions in performance training. One is as an explicit goal of the training process, where it is acknowledged that the student’s sight reading ability needs to improve. Under these circumstances, the instructor might present the student with a course of sight reading material for performance during and between lessons. Apart from selecting the material, the instructor’s role would typically be to analyse by observing the student’s strengths and weaknesses in respect of sight reading, negotiating goals for improvement with the student, and monitoring the student’s progress and provide feedback, particularly in such a way that is likely to increase the student’s self-confidence. The material would be designed to become more challenging over time to match the student’s expected increase in sight reading achievement, and might target various perceived weaknesses in the student’s sight reading ability, such as the impromptu planning of effective fingering, the recognition of frequently occurring patterns in the music and the management of accidentals.

For this purpose the instructor may or may not use an instructional manual. Numerous sight reading manuals were published during the 20th century and provide circumstantial clues as to the nature of the explicit teaching of sight reading, where it has occurred. Covington (1981) identified two basic approaches of such manuals. One is the sequencing of a range of sight reading material on the basis of difficulty and, in some cases, particular technical features. This approach, dubbed the ‘work at it’ approach by Burmeister (1991:v) is by far the most common method employed in sight reading manuals, suggesting that instruction is largely limited to controlling the level of difficulty and technical characteristics of musical examples and monitoring the student’s output.

Another approach, less widely found, is the provision of supporting activities and advice in parallel with graded sight reading material. The underlying purpose of this approach is to reinforce the simple ‘work at it’ by providing stimulation in other modes. Parallel activities include transposition, improvisation, aural training and rhythm clapping; supporting advice ranges from structural and harmonic analyses of sight reading material to explanations of notational symbols. Spillman (1990) and Burmeister (1991) are recent examples of instructional manuals that contain insightful supporting advice on sight reading. These authors break sight reading down into its component parts, including such matters as fingering practices, preparation
for leaps and stretches, pattern perception, breathing, relaxation, psychological attitude, and methods of scanning the score with the eyes. A survey of online information about sight reading instruction manuals suggests that sight reading pedagogy has remained essentially unchanged since Covington’s study.

A second function of sight reading is as a largely unacknowledged by-product of performance instruction, as opposed to an explicit goal. This can occur where sight reading improves merely through exposure to the intensive rehearsal of texts of increasing difficulty as a student progresses through a course of tuition with no explicit focus on sight reading. Improvement in sight reading skills can also be a by-product of the training of specific performance skills that are commonly included in music training, such as transposing, realising figured bass, playing from large scores, and spontaneously harmonising melodies at the keyboard. These skills are difficult to separate from that of sight reading since they partly depend on the student’s ability to perform well from an unfamiliar score.

A third function of sight reading in the performance curriculum may arise from the fact that it is the only sure method of verifying that a student has attained a prescribed knowledge of notational symbols and their meaning, based on the premise that a student could learn a set repertoire of texts without an efficient knowledge of notational symbols. While possible, this would appear to be a comparatively rare learning method that is in all cases difficult to conceal from an instructor. As a purely summative as opposed to formative function, then, using sight reading to measure such knowledge is probably of limited value.

An alternative way of regarding the function of sight reading training, particularly of the explicit, focused category discussed above, is in respect of the cognitive skills that might be associated with successful sight reading training. As is further discussed in Section 2.2.2, these skills may include the ability to chunk, the degree of automaticity in the music reading process, and the various capacities of the memory system, while Section 3.3 deals with the ways in which such skills may be manifest in a reader’s eye movement. There is, however, a basic causal problem here: it is difficult to determine whether good sight reading arises from such improvements or is responsible for them.

Not only is there in the literature a lack of analysis of the function of sight reading in the teaching and learning process and of the question of why it is important to improve musicians’ sight reading in the first place, there is also evidence that the
acknowledgment by many instructors that sight reading skills are important is not reflected in the explicit training of sight reading skills. Demorest (1998:183) has claimed that 'it seems that believing sight-singing is important and giving rehearsal time to it are two different things.' Johnson's (1987:185) study of high school choral training concluded that while 'directors agreed on the importance of sight-reading, they devoted little rehearsal time to its development'. Sloboda (1985:68) has identified a fundamental reason for avoiding the teaching and learning of sight reading skills. Most students, he argued, begin their performance studies by having to learn two complex skills simultaneously, those of reading and of performing. To acquire each skill on its own would present a challenging enough learning situation, but to acquire both at once is unusually demanding. According to Sloboda, under such circumstances poor readers can still acquire a restricted repertoire through intensive rehearsal, and typically do so rather than expose themselves more frequently to sight reading situations.

Sloboda's thesis might be elaborated by stating that to play music from score is to be two things, 'reader' and 'player', and that the tension between the two is part of the dynamics of learning to perform music in a notation-based music culture. Such tension no doubt underpins the practice of teaching performance skills before introducing western music notation in the Suzuki method of violin teaching (see Duke 1999 for a description of Suzuki teaching). The tension between 'reader' and 'player' is bound to surface when it comes to the tendency of music students to develop a small, well prepared or even memorised display repertoire, a practice clearly encouraged by the syllabuses of the music examination bodies referred to above. Such models for teaching and learning strengthen the notion that concentrating on a small repertoire of familiar works rather than exploring the wider repertoire and thereby cultivating sight reading skills is central to progress as a performing musician, and may be a strong contributing factor to the apparent neglect of the explicit training of sight reading.

The considerable amount of research conducted into sight reading over the last forty years also attests to the importance with which sight reading skills are regarded and suggests that the training of sight reading is regarded as problematic. However, a survey of the leading music education journals such as the Journal of Research in Music Education, with some 54% of citations in the area (Hamann & Lucas 1998), the British Journal of Music Education and Psychology of Music, as well as an online

2 In the next section, a technical description of the music reading apparatus, these two aspects of a person's music reading apparatus, the 'player' and the 'reader', are equated with two 'command groups'.
survey on *PSYCHNET* of the abstracts of some sixty doctoral dissertations on sight reading, revealed little information that is likely to have a direct impact on the teaching and learning of sight reading.\(^3\) This observation is consistent with Stebleton's (1987:14) suggestion, arising from his review of the literature on the predictors of sight reading skill, that focusing on merely one or a few components of the sight reading process is likely to distort the overall view of the process.

Many studies evaluated existing training methods or variations on them. For example, Anderson (1981) showed that the effects of auditioning tape-recorded models prior to sight reading had no significant effect on sight reading quality. Grutzmacher (1987) observed a greater training benefit when first-year tertiary instrumental students were familiarised with tonal patterns through harmonisation and vocalisation exercises than by merely training them through sight reading exercises. Salzberg & Wang (1989) investigated the value of counting out loud and foot tapping as prompts in rhythm sight reading, finding a significant value for novices, but not for advanced students. Cassidy (1993) demonstrated mild benefit for non-music majors' pitch accuracy when various combinations of hand signs and *SOLFÈGE* were used in the training process. Lucas (1994) compared sight reading performance when participants read melodies in a variety of harmonic and voicing contexts, but observed no useful differences. Kostka (2000) found that error-detection practice had little effect on the keyboard sight-reading performance of undergraduate music majors. Regrettably, these and other similar findings are either of limited practical use in the teaching and learning situation, or accord with what teachers instinctively understood about their art in the first place.

Some research provided mild support for the use of supporting activities in sight reading tuition, such as improvisation exercises (Montano 1983), body movement (Searle 1985) and aural skills (Fincher 1983). Other studies, such as Ludeker (1984), Harris (1971) and Mann (1991) focused on instructor-behaviour but failed to arrive at useful conclusions. Yet other investigations have developed and evaluated machines that electronically enforce tempo (Fincher 1983; Streckfuss 1984), with disappointing results. Wiley (1962), Wright (1982) and Gaynor (1995) used a TACHISTOSCOPE to examine the encoding of music notation, with inherent methodological problems that are discussed further in Section 3.2.1; and almost 20 studies into eye movement in music reading over the last seventy years have yielded

\(^3\) A full review of the general literature on sight reading was not considered to be useful and relevant enough to the objectives of this study, in which the focus was on certain component aspects of the music reading process. The literature on these aspects is fully reviewed in Chapters 2 and 3.
little of practical teaching value; these eye movement studies are fully reviewed in Chapter 3.

Other studies searched for useful correlations between sight-reading skill and other variables, either with little success or in ways that failed to provide a practical spinoff. For example, Killian (1991) found little relationship between sight singing accuracy and error detection ability in junior high school singers. Spicer (1992) found that the strongest indicator of sight reading skill was the amount of private tuition and ensemble playing that participants had experienced, hardly a surprising finding. In a similar vein, Demorest & May (1995) concluded that the number of years of school choir experience followed by the number of years of private performance lessons were the strongest predictors of individual success. Tucker (1969) also determined that there was a significant relationship between music-reading ability and years in a school choir, and in addition that school instrumental experience was more effective than vocal experience as music-reading training for choristers.

More helpful was an investigation by Elliot (1982) of the relationships among instrumental sight-reading ability and seven variables, concluding that rhythm reading ability is the single best predictor of sight-reading scores, at least among wind instrumentalists. Elliot claims that 'over two-thirds of the mistakes made by sight readers can be attributed to rhythmic problems' (p38). This was reinforced in a large study by McPherson (1994) showing 'that rhythmic errors far outweigh all other types of errors' in sight reading (p217), consistent also with Boyle's (1970) demonstration that rhythmic skills are strongly associated with sight reading skills. These findings suggest that focusing on rhythm reading and performing may be a valuable strategy in training sight reading. McPherson (1994) also addressed the issue of how closely sight reading skills are correlated with overall performance skills; he concluded that 'in the beginning stages of training, sight reading skill is not significantly correlated with the ability to perform a repertoire of rehearsed music....as instrumentalists mature, however, correlations between these two aspects of performance seem to strengthen markedly' (p217). Kornicke (1995:56) resonates with this notion by claiming that the 'finding that greater sight-reading achievement was related to a relatively late age of beginning sight-reading offers evidence that while young children do learn to read music, they may be processing the information differently from adults, according to their developmental stage'.

It is concluded that the explicit training of sight reading has changed little in practice and is still largely focused on observing and treating the end-points of the music
reading process, that is, musicians' physical engagement with the musical instrument and the resulting musical sound. It will become clearer in the light of the technical overview of the music reading apparatus below that it is worth exploring the idea of directly intervening in the functioning of eye movement and the memory system in sight reading to assist students in improving their sight reading skills. This is the rationale behind the creation of *SightReader*.

1.2 Technical overview of the music reading apparatus

Such a difficult and complex task as music reading needs to be described from the perspective of its anatomical components and the flow of information between them. This description will provide a technical background to the components that traditional teaching and learning focus on, and point the way to those parts of the apparatus that have been comparatively neglected.

Music reading, of which sight reading is a subset, is an extraordinarily intricate information-processing activity. The coincidence of three mechanical demands that it places on a musician appears to make it unique among human activities: the processing and physical implementation of complex encoded information; the strict timing of the output to an Isochronous beat; and the fact that input, processing and output function continuously. This continuity itself is challenging, since new performance instructions must be picked up from the score at the same time as old instructions are being processed and stored. Performers are thus required to process and store several instructions simultaneously in their memory system while at the same time producing a physical response to execute a musical performance. In addition to these demands, the output must adhere to a regular pulse at a predetermined tempo. Playing to the beat thus requires enough flexibility to accommodate variations in processing load, that is, more and less difficult areas of the score.

Those activities most similar to music reading, such as copy-typing and reading language aloud, are nonetheless fundamentally different from it. In copy-typing, the timing of the output is governed by global objectives such as a desire to type fast while maintaining accuracy, rather than such local constraints as an isochronous beat. Reading aloud in English also permits a greater latitude in the timing of the output (see Halliday 1994:292-294 for a discussion of pulse in spoken English). In both tasks, local variations in workload can therefore be accommodated by adjusting the
speed of the output, an option unavailable to musicians, tied as they are to a comparatively strict pulse. Furthermore, while many ball-sports and computer games require swift and precisely timed motor-reactions to changes in visual stimulus, such reactions are typically required not continuously but intermittently.

The music reading apparatus involves at least six neuroanatomical components. Each component plays a role in the elaborate task of continuously perceiving visual performance instructions encoded in music notation and transforming them into a physical response. In this section an attempt will be made to describe the basic features of the apparatus in terms of these components and the flow of information between them.

The fundamental structural components of the apparatus, 'the reader' and 'the player', are referred to here as the **oculomotor group** and the **musculoskeletal group** respectively. Both groups involve the memory system and the neurological infrastructure necessary for encoding and transmitting signals. In addition, each group involves a separate set of muscles that responds to command signals from the memory system, generates feedback to the memory system and directly impacts on other components of the apparatus. The signals that link these components fall into three categories: input, commands, and feedback. Of these, feedback can be further divided into two categories: perceptual feedback of the outside world, such as the visual images that enter the apparatus through the eyes, and proprioceptual feedback from within the body, such as the level of muscle tension. The exact function of proprioceptual feedback in the music reading process is not well known, although many questions relating to this function have been posed by Sidwell (1981).

The signals are illustrated in this section by a system of colour-coding: blue is used for input, red for commands, and green for feedback.

The oculomotor group includes the eyes and some of the oculomotor muscles, which are attached to the sides and back of the eyes and are responsible for moving them over the score. Music reading may be assumed to begin with a command from the memory system to the oculomotor muscles to begin viewing the score in a way that will support the impending start of the performance (Signal 1 in Figure 5). These muscle movements, referred to here as the **oculomotor response**, are assumed to generate two further signals that travel back to the memory system. One of these signals is **oculomotor feedback** (Signal 2), a proprioceptual signal that indicates the relative tension of the muscles and thus the position and movement of the eyes. The existence of this signal can be confirmed simply by rotating the eyes as far as
possible to one side without turning the head. According to the present model, oculomotor commands also impact directly on external visual input from the score (Signal 3), determining the sequencing and timing of the information that is taken up from information—which parts of the score the eyes perceive, in what order, and for how long. The nature of oculomotor commands is covered extensively in Chapter 3. Oculomotor feedback may well deliver information as to the position of the eyeballs, but in the light of the precise information delivered by external visual input concerning where the eyes are looking on the score, oculomotor feedback would appear to be largely redundant.

Figure 5  The oculomotor group and related signals

Figure 6  The musculoskeletal group and related signals
External visual input is encoded into a neurological signal and transmitted to the memory system, where it is assumed to be stored and processed into **musculoskeletal commands** (Signal 4 in Figure 6). These are then transmitted to the components directly involved in producing the musical sound. For keyboardists these include the muscles and joints of the fingers and hands, and to a lesser extent the arms and shoulders. These bodily components are referred to collectively as the **musculoskeletal system**, and their physical contribution to the output as the **musculoskeletal response**.

In this model, the musculoskeletal response generates three types of feedback to the memory system. One is proprioceptual **musculoskeletal feedback** (Signal 5 in Figure 6), which transmits information to the memory system concerning the ongoing position of the musculoskeletal system. Another is **tactile feedback** (Signal 6), a perceptual signal involving the sensations of touch and soft pressure arising from the player's physical contact with the musical instrument. The tactile system as applied here to keyboardists comprises the skin and subcutaneous areas of the fingers. It can readily be distinguished from musculoskeletal feedback by imagining the process of 'playing in the air', in which there is no contact with a musical instrument. In this case, there would be musculoskeletal feedback, but no tactile feedback.

Gabrielsson (1999:516) has pointed out that 'although motor processes are central to musical performance, they are still little understood'. He has summarised at least four theories of how the musculoskeletal response is controlled in terms of the various emphases they give to the central control of the response, reliance on feedback, and the structure of abstract representations of motor sequences (p518). Of considerable interest to further research in motor-control of musical performance is the 'Bernstein' approach, discussed by Wade (1990) a novel approach involving linkages between muscle groups called 'coordinative structures'. Shaffer (1980) has carried out the most work on motor-performance, in particular in relation to advanced pianists, while Sidwell (1981) has written a speculative essay inquiring into the role of the memory of musculoskeletal commands in musical performance. Much work remains to be done in respect of the role of the musculoskeletal system in

---

4 The muscles and joints of the legs and feet might also be included since these are necessary to operate the sustaining pedal and the soft pedal on the piano. In Stages I and II, however, the experiments are conducted on an electric keyboard without pedals, so for the sake of simplicity the legs and feet are excluded from this exegesis.

5 For singers and wind and brass players, a broader category would be needed to cover feedback from, for example, lung pressure and the vibration of head cavities. For these performers, 'kinaesthetic system' is probably a more appropriate term than 'tactile system'.

21
musical performance. This area, however, is considered to be beyond the scope of this thesis.

The third feedback signal generated by musculoskeletal commands is the musical sound, which is perceived by the ears and transmitted to the memory system as auditory feedback (Signal 7). This signal also represents the output of the apparatus, and represents the overall purpose of the activity. Weinberger (1999) has written an extensive overview of the physiology of the auditory system as it relates to musical performance. It is not clear whether auditory feedback alone is used to monitor the accuracy of a performance; Sloboda (1978) has argued that it is possible to realise that a wrong note has been played before hearing it, presumably through either musculoskeletal commands, musculoskeletal feedback or tactile feedback. The components and signalling functions are summarised in Table 2.

<table>
<thead>
<tr>
<th>neuroanatomical component</th>
<th>related signal</th>
<th>signal number</th>
</tr>
</thead>
<tbody>
<tr>
<td>memory system</td>
<td>• generates commands oculomotor and musculoskeletal commands</td>
<td>14</td>
</tr>
<tr>
<td></td>
<td>• receives, stores and processes external input and feedback</td>
<td>2 3 5 6 7</td>
</tr>
<tr>
<td></td>
<td>• receives, stores and processes internal input</td>
<td></td>
</tr>
<tr>
<td>oculomotor muscles</td>
<td>• receive and implement oculomotor commands</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>• generate oculomotor feedback</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>• act physically on the eyes, determining the sequence and timing of external visual input</td>
<td>3</td>
</tr>
<tr>
<td>eyes</td>
<td>• generate external visual input</td>
<td>3</td>
</tr>
<tr>
<td>musculoskeletal system</td>
<td>• receives and implements musculoskeletal commands</td>
<td>4</td>
</tr>
<tr>
<td></td>
<td>• generates musculoskeletal feedback</td>
<td>5</td>
</tr>
<tr>
<td></td>
<td>• acts physically on the tactile system, determining tactile feedback</td>
<td>6</td>
</tr>
<tr>
<td></td>
<td>• generates auditory feedback</td>
<td>7</td>
</tr>
<tr>
<td>tactile system</td>
<td>• generates tactile feedback</td>
<td>6</td>
</tr>
<tr>
<td>auditory system</td>
<td>• receives auditory feedback (output)</td>
<td>7</td>
</tr>
</tbody>
</table>

Table 2 Components and signals of the music reading apparatus

Figures 5 and 6 are conflated in Figure 7 to produce a schematic representation of the whole apparatus in relation to the performer, who is depicted side-on at a keyboard; each signal is represented by an arrow showing its direction of travel. A black arrow in Figure 7 indicates input from long-term memory to working memory.

The oculomotor and musculoskeletal groups are clearly distinguished by their temporal characteristics, a feature that is at the heart of the experiment in Stage II.
begin with, at any point in time each command group is linked to a different place on the score. This is because the oculomotor group ('the reader') must take up information from the score before it is implemented by the musculoskeletal group ('the player'), so that there is time for the information to be encoded and processed. The distance between these two locations on the score, typically of one to two seconds' duration, is known as the eye-hand span.\textsuperscript{6} The span is most obvious to the unaided observer as a musician finishes inspecting the last chord in each line of music across the score, and subsequently looks back to the start of the next line before playing the last chord.

\textbf{Figure 7} The flow of information through the music reading apparatus

\textsuperscript{6}Here the word 'hand' is used to refer to the entire musculoskeletal system. For singers, the term eye-voice span is used instead, as in, for example, Sloboda (1974) and Goolsby (1987).
On logical grounds there are natural limits to the size of the eye-hand span: it must be large enough for the information to be perceived, transmitted, and processed for implementation by the musculoskeletal system, but not so large as to place too great a burden on working memory. It is known already from the results of many studies that the cost of processing and storing information in working memory is high (see Section 2.1). For the purpose of this study, the burden on working memory during music reading is seen as taking two forms: the amount of information that must be processed and stored at any one time in working memory (load); and the amount of time for which each piece of information is stored in working memory (latency). Accordingly, the size of the eye-hand span (henceforth span size) can be expressed either in terms of load (for example, two beats at 60MM) or latency (for example, 2000ms at 60MM). The ways in which performance tempo alters the relationship between load and latency are reported in the results for Stage I. The data from Stage I show that the span typically contracts and expands in size continually as the reader progresses rightward along each line of music on the score, rather like a caterpillar in fast motion.

Having described the temporal characteristics of the two groups as a whole, the next step is to see these characteristics on a more detailed level, where each signal has its own unique temporal relationship to the music score. At any one time during a performance, the music reading apparatus can be thought of as dividing the score into three temporal domains: the musical present (what is being played at the moment), the musical past (what has already been played during the current performance), and the musical future (what is about to be played). Thus each signal exists in both a real temporal frame, since all signals in the music reading apparatus function continuously in the real present, and in one related to the musical score. Examples of these three musical domains are set out in Figure 8.

In Figure 8 the ambit of all seven signals is also shown in the context of the eye-hand span. The position of the hands is represented by a black vertical line on the left-hand side of the figure, aligned in this example with the second chord. The musical future refers to the part of the score that has not been performed in the current reading, that is, to the right of the hands on the score. Similarly, the musical past is the part of the score that has already been performed in the current reading, to the left of the hands. The position of the eyes is represented by a second black vertical line. It follows that while the eyes and the hands function continuously in real time, the eyes are always in the musical future. Although the seven signals function continuously throughout a reading, at any one time their information content is
Introduction

related to only a small portion of the score. This is due to the fact that the reader can perceive only part of the score at any one time, and to the limited capacity of working memory, discussed further in Chapter 2.

At this point a new concept is necessary to explain how the signals are temporally constrained in relation to the score. Here this concept is referred to as \textit{fixed-} and \textit{open-endedness}, and its two corollaries \textit{consolidation} and \textit{decay}. In terms of the score, the ambit of each signal is delineated at a clearly identifiable point on either its left or right side, and is fixed to either the \textit{hand-line} or the \textit{eye-line}, depending on the signal. The signals within the oculomotor group are fixed at one end to the eye-line, and the signals within the musculoskeletal group to the hand-line. In Figure 8 the fixed end of each signal is shown by a short line at a 90-degree angle to the signal line. Rather than being fixed, the other end of each signal is not clearly delineated, but rather begins to consolidate at or to decay up to an indeterminate point on the score, and for this reason is termed 'open'. Open-endedness is shown in Figure 8 by a dotted line.

Both oculomotor and musculoskeletal command signals are essentially forward plans of the movement of the eyes and the hands. At any one time it is impossible to determine exactly where the process of such planning begins in relation to the score. It is self-evident that they are indistinct in respect of music far into the future and consolidate gradually into a clearer form in respect of the immediately upcoming score. Command signals could therefore be likened to the light from a lamp penetrating forwards into thick fog. By contrast, once a command is implemented there appears to be a definite point beyond which there is logically no point in retaining it. Once the eyes have moved to a new point on the score, the command to do so becomes rapidly irrelevant to the task of continuing to read that part of the score. Likewise, once a musculoskeletal response has been executed, the commands that generated it rapidly become irrelevant to performing that part of the score; the same can be said of external visual input. Thus to retain a spent command or the external visual image that begets it would appear to place an unnecessary burden on the limited capacity of working memory and risk possible interference with upcoming signals. The concept of fixed-endedness depends on this assumption that signals rapidly drop out of the memory system at a particular point in musical time.\footnote{The concept of being fixed to a definite point on the score disregards the minute temporal discrepancies between the signals, arising from variations in transmission and processing times. For example, auditory feedback occurs several milliseconds after the hands play a chord, due to the time taken for the musical instrument to produce the sound, and for that sound to travel through the air to the ears and to be transmitted along neurological pathways to the brain. These minute variations can be regarded as inconsequential for the purposes of this analysis.}
Unlike commands, the feedback signals they generate are assumed to be fixed on their rightward side to the position of either the eyes or the hands, and to decay (rather than consolidate) indeterminately leftwards along the score, into the musical past. Auditory feedback is fixed on its leftward side to the hand-line and also decays into the musical past. Similarly, external visual input is fixed to the position of the eyes and decays leftwards from the position of the eyes on the score as it is transformed into musculoskeletal commands within the eye-hand span.

Figure 8  Temporal characteristics of the signals in relation to a music score
In performing music from score, then, the seven signals sweep slowly across the page in a constant ebb and flow, consolidating and decaying either into the musical future or past. Visual input, oculomotor commands and oculomotor feedback all have a certain ragged aspect to their progress, linked to the characteristics of eye movement across the score. The other four signals pertain to the musculoskeletal group, and thus move along the score in a more orderly fashion, since the progress of the hands is controlled by the musical beat. *SightReader* was created to intervene in this intricate travelling circus, as it were.

Since this description of the music reading apparatus illustrates its extraordinary complexity, it is timely to make a point about the ambit of the notion of sight reading. Although the thesis explicitly concerns sight reading on the keyboard, the view is taken here that there is no in-principle difference between this task and sight reading on any other instrument, or for that matter sight singing. The reason for this view is that the differences between these categories are of little significance when compared to the size and complexity of the process that is common to all musical performance regardless of the instrument. The differences that depend on the particular musical instrument being played involve first the type of musculoskeletal response required to produce the musical output. The fingers may play notes by gripping mallets, striking keys, or plucking or stopping strings; the feet, mouth, larynx and lungs may or may not be part of the musculoskeletal system.

But these matters are secondary to the fact that a sequence of complex musculoskeletal commands must be produced; it is the necessity for such a signal in the first place that is significant in defining what sight reading is, not the particular patterns and anatomical components involved. If the particular musculoskeletal response required by a particular instrument were at issue in defining sight reading, it would raise the question as to whether using the sustaining pedal while sight reading at the piano is a fundamentally different activity from not using it, or whether the finger-oriented technique that many pianists employ to perform baroque music is fundamentally different from the arm- and shoulder-oriented approach taken by many pianists in performing music of the romantic period and the twentieth century.

The second instrument-dependent difference in the performing experience concerns the oculomotor system, and arises from the number of staves that must be scanned at once while reading music. This normally ranges from one stave to three and, as will become clear in Chapter 3, alters the relationship between the vertical and horizontal dimensions of the scanpath. Again, these differences in patterning are ephemeral to
the notion of sight reading since the necessity of producing an oculomotor response signal and not the number of staves on the score is the defining structural feature of sight reading.

It is therefore contended that in this respect the patterning of the musculoskeletal command signal should be regarded as irrelevant. In the light of this, the term 'sight singing' must be regarded as unfortunate since it suggests that the structure of the sight reading apparatus is fundamentally different for singers and instrumentalists. Thus although the thesis focuses on keyboardists, its findings are applicable to a wide range of performers who read from score.

1.3 Brief description of SightReader

The prototype software of SightReader was designed by the author, who is the registered inventor of the device. The software code was written by Mr Andrew Martin, Visiting Programmer at the University of Sydney. The equipment setup for the prototype in Stage II is schematically illustrated in Section 5.2.2. It comprises an electronic musical keyboard connected through a MIDI-box to a central processing unit. Two loudspeakers are patched into the MIDI-box and located on the wall above the desk. An adjustable chair for the user is positioned in front of the musical keyboard.

SightReader works by progressively erasing the notation in each successive beat of a musical score as the user plays the music on the score. Each successive erasure is triggered instantaneously as players strike each new chord on the keyboard. When a beat is erased, it vanishes cleanly and promptly into blank stave-lines. The crux of the process is that at any one time the notation is erased ahead of the beat on the score that the hands are playing, such that the hands can never 'catch up' with the eyes. The number of beats ahead that are progressively erased is set by users on a SPAN SETTING (short for 'minimum eye-hand span setting'). The effect is to ensure that users' eyes read ahead of their hands by at least the number of beats that has been selected.

---

8 Mr Martin's contribution was funded by a research grant from the University of Sydney, which owns the intellectual property represented by SightReader and which obtained a Provisional Australian Patent for SightReader. The text of the patent is reproduced in Appendix I. The patent was largely written by the author, with advice and editing provided by Patent Attorneys BF Rice of Sydney and Mr Martin.
now the first available notation for the eyes to inspect. In this way the relationship between the musculoskeletal and oculomotor groups can be 'stretched' beyond its normal distance. A fuller description of SightReader appears in Section 5.2.

Figure 10 The display of a stimulus on SightReader before starting to play, with span setting on 'two beats ahead'

Figure 11 The same display after having played the first chord, with span setting on 'two beats ahead'

Figure 12 The same display after having played the seventh chord, with span setting on 'two beats ahead'
An exercise is selected from pull-down menus on the menu-bar in one of three musical textures, entitled respectively 'one-part', 'two-part', and 'four-part'. A chosen exercise is then displayed in its entirety, with the control panel at the bottom (Figure 9 shows a one-part exercise at approximately half size). On the control panel a span setting is selected in respect of the minimum number of beats their eyes are to be forced ahead of their hands during the upcoming performance, and a TEMPO SETTING is selected for the metronome. The 'set/reset' button on the left-side of the control panel at the bottom of the display is pressed, the metronome starts to sound, and the exercise is ready to perform, starting on any metronome beat.

Figure 9 An opening window of SightReader

Figure 10 shows a four-part exercise before the first chord is struck. Figure 11 shows how the display would look immediately after the first chord is struck, with the span setting on 'two beats ahead'. The eyes must now be inspecting the third chord, since the first two chords have been suddenly erased. An arrow has been added here to indicate the location on the score of the chord that has just been struck before it was erased. Figure 12 shows the display some seconds into the performance, immediately after the seventh chord has been struck: the ninth chord is

---

9 Up to 20 examples of each texture could be stored in the program.
2 Memory and music reading

The unanswered questions on memory abound
Despite numerous studies whose findings are sound
What's needed right now
Is for us to see how
We can put all the data on common ground.
(Adapted from Keenan (1999:xviii))

The memory system is the nerve-centre of the music reading apparatus. It generates the necessary commands, receives input and feedback, and stores and processes these signals. It is the repository of the familiarity acquired during previous performances of a musical text, in particular of the auditory patterns of those performances and the musculoskeletal commands necessary to produce them. And in what were referred to in Chapter I as ‘layers of familiarity’, the memory system encapsulates broader meanings of a musical text, such as the style and genre from which it is drawn and its cultural, psychological and social meanings. All of these aspects of memory have the power to affect the experience of reading music from score. The memory system is thus of prime importance to an understanding of the music reading process and, as will be explained in Section 5.1.3, is central to SightReader’s function of stretching the eye-hand span. This chapter lays the ground for a deeper understanding of the music reading apparatus, first through a brief account of how knowledge of certain aspects of the memory system have developed historically, and then through a focus on its specific role in music reading.

To discuss memory in such a way is not an easy task; scientific investigation into memory has tended to be a complicated affair. As Shah & Miyake (1999:2) point out, modern textbooks on cognitive psychology are typically not unanimous when defining memory. In the scientific literature it is also not always clear whether various conceptions of memory ‘are fundamentally incompatible or merely reflect differences in emphasis’ (p2). In a similar vein, Ericsson & Delaney (1999:258) admit that ‘it is rather humbling to see the controversies that still surround the appropriate definition and characterisation of those supposedly basic terms [that pertain to human memory]’. Progress in memory research has occurred in fits and starts over the last century, with a continual and by now almost predictable splintering and coalescing of received opinion on the matter. In the last few years
there appears to have been a strong tendency towards the consolidation of the various opposing views that have bedevilled the study of memory in the second half of the twentieth century. But the seeds of division are again evident as the focus shifts from the level of physical behaviour towards the neurological level. This shift is becoming increasingly possible through advances in brain imaging and computer modelling.

The next section presents an account of two major tendencies in the literature on human memory. One of these is to emphasise the fragmentation of memory into various substructures, the other to see memory as a single, unitary system. Given the vast literature on the participant, this account is necessarily selective, and is oriented towards providing a context for the theoretical background to the experiments in Stage I (Chapter 4) and Stage II (Chapter 5). Following this account, current knowledge of the relevant aspects of the memory system is used in an attempt to speculate on how the notions of automaticity, chunking and activation bear upon the task of performing music from score.

2.1 Review of selected literature on memory

For several centuries there appears to have been an awareness of the distinction between **short-term memory** and **long-term memory**. In one of the earliest known writings on memory, John Locke (1690) distinguished between a ‘temporary workspace’ for ideas within consciousness and a more permanent ‘storehouse of ideas’. Even though these concepts were expressed in the most general terms, Locke’s distinction raised issues surrounding human memory that have occupied researchers for the last century and are still unresolved: the conscious versus the unconscious and the temporary versus the permanent. It was not until the late nineteenth century that further scientific interest in the matter was documented. In an insightful passage far ahead of its time, Galton (1883), as quoted in Baddeley (1997:29), wrote:

> There seems to be a presence-chamber in my mind where full consciousness holds court, where two or three ideas are at the same time in audience, and an ante-chamber full of more or less allied ideas, which is situated just beyond the full ken of consciousness. Out of this ante-chamber the ideas most nearly allied to those in the presence-chamber appear to be summoned in a mechanically logical way, and to have their turn of audience.
Unlike Locke, who appeared to see a simple bifurcation of the memory system, Galton was implying that there is a flexible boundary between the two components with his notion that thoughts are 'more or less allied' to full consciousness, and that on this basis they can be brought into and out of conscious view. Not long after, James (1890) proposed a purely temporal distinction between short-term and long-term memory. He claimed that there is a 'primary memory' containing information that remains in the consciousness after it has been perceived, and which forms part of the psychological present. A 'secondary memory' contains information that has left consciousness and forms part of the psychological past. Locke and James can be considered among the first to attempt to analyse the structure of human memory.

Locke’s and James’s ideas, however, were essentially philosophical and introspective, and lacked the support of scientific observation. Ebbinghaus (1885) was apparently the first to attempt to scientifically measure the performance of memory, by focusing on patterns of forgetting over various intervals of time ranging from 20 minutes to 31 days. He found that the rate of forgetting is initially precipitous but tapers off more gradually, and that relearning information is easier than first learning it. This matter is taken up in Section 5.1 in relation to successive readings of the same musical text, but over much shorter delays than those used by Ebbinghaus. Another early attempt to document the capacity of memory was made by Jacobs (1887), a London school-teacher. He measured his pupils’ capacity for temporarily storing information in such tasks as repeating back sequences of numbers, technically known as digit-span tasks. Jacobs showed that it is possible to observe conscious memory in terms of its limited capacity. This was of considerable importance to the investigation of the structure and functioning of memory, which was traditionally based on observing participants’ performance in specific memory tasks.

There was a lull in such scientific observation until the middle of the twentieth century. Bartlett’s (1932) discussion of long-term memory and Tolman’s (1948) study of spatial memory and cognitive ‘maps’ were notable exceptions, although they remain highly conjectural works. In 1956 Miller published a highly influential article on the limits of short-term memory, in which he introduced a phenomenon he termed CHUNKING. The notion of chunking assumes that the capacity of conscious memory can be temporarily increased by organising incoming information into larger and fewer units. The smallest adjacent units, such as individual letters in language reading, are organised into familiar larger units, or chunks, such as words, rather than being processing separately. Readers are assumed to utilise these larger units by
accessing their memory of them. The equivalent in music reading would be to isolate such larger patterns as scales and triads rather than picking up individual notes separately (see Section 2.2.2). This explains why typographical errors persist in written text even after it has been carefully edited, since readers are accustomed to glossing over individual letters and character spaces in the chunking process.

Given the assumption that chunking cannot occur without the assistance of pre-existing information in the reader’s memory system, Miller might well have explored the possibility that the capacity of conscious memory is subtly and intimately dependent on the functioning of long-term memory. But because his data were based on span tasks, which minimise the role of long-term memory because they require prompt recall, he was able to claim simply that the capacity of short-term memory is ‘seven plus or minus two’ chunks of information. Although no-one now disputes the existence of chunking as a potent cognitive tool, his ‘7±2’ claim has since been convincingly challenged by Simon (1974), who discovered that the number of chunks that can be simultaneously held in short-term memory depends on the size of the chunks. Furthermore, Watkins (1974) and Broadbent (1977) claimed that the capacity of short-term memory in tests designed to minimise the role of long-term memory is closer to 4±1 than 7±2 chunks. Miller’s work encouraged other researchers to explore the nature of short-term memory through the measurement of its capacity.

Soon after Miller’s paper, Brown (1958) and Peterson & Peterson (1959) claimed that short-term and long-term memory may be further distinguished by how forgetting occurs. Brown and the Petersons tested the hypothesis that information decays from the short-term store rapidly and automatically unless it is refreshed by participants through conscious effort. These studies used a second task during the retention interval to suppress participants’ ability to refresh information that had already been loaded into their short-term store. Here it was assumed that interference would be more likely to occur when two units of information were similar to each other. They claimed that in their experiment, which involved counting backwards while memorising groups of letters, the two tasks were sufficiently different as to preclude interference. Brown & Peterson found that under these conditions participants were unable to retain information even over very brief intervals. By contrast, these researchers assumed that memory traces are lost from the long-term store through interference from other information stored in the long-term store, rather than from automatic decay. There is, however, other evidence that
both automatic decay and interference account for forgetting in short-term memory (Reitman 1974; Waugh & Norman 1965) and long-term memory (Bjork 1989).

Atkinson & Shiffrin (1968) proposed a model for a multicomponent memory system that has carried much weight during the second half of the twentieth century. This model combined and extended a number of pre-existing theories, among which were those of Sperling (1960), Waugh & Norman (1965) and the two studies discussed above. According to Atkinson & Shiffrin's model (see Figure 13 and Table 3) there are three major structural components of the memory system: (1) 'gateway' sensory registers, which initially receive information from the environment and are modality-specific, that is, specific to sight, hearing, smell, taste, and touch, (2) a short-term store of limited capacity, and (3) a long-term store of virtually unlimited capacity which can hold information for very long periods.

![Figure 13](image)

Atkinson & Shiffrin claimed that information passes from sensory buffers of large capacity and very short duration into the short-term store only if selectively attended by the participant, and may under certain circumstances also pass directly to long-term memory. The rest of the information is lost and has no further effect. Once in the short-term store, information can be retrieved relatively easily, but is fragile and decays within seconds unless it is refreshed. Information in the short-term store can be displaced by fresh information entering from the sensory registers. It can also be transferred from short- to long-term stores through selective attention by
the participant. Once in the long-term store, it is relatively durable but can be lost through interference from other material already residing in or subsequently loaded into the long-term store. Atkinson and Shiffrin proposed that the probability that information in short-term memory will be transferred to long-term memory depends on the time for which an item is held in the short-term store, which is connected with the rate of refreshment, the process of reinforcing the fragile memory traces in the short-term store, which is sometimes known as ‘rehearsal’. In the light of current knowledge about the memory system, this last point is regarded as an oversimplification; it is, however, of some significance to the experiment in Stage II of this study, and is re-examined in Section 5.1.3.

<table>
<thead>
<tr>
<th>component</th>
<th>capacity of storage</th>
<th>duration of storage</th>
<th>reason for information decay</th>
<th>information transfer</th>
</tr>
</thead>
<tbody>
<tr>
<td>sensory buffers</td>
<td>extremely large</td>
<td>very brief (fraction of a second)</td>
<td>non-attention</td>
<td>to the short-term store through attention or directly to the long-term store</td>
</tr>
<tr>
<td>short-term store</td>
<td>limited</td>
<td>brief (seconds)</td>
<td>displacement by other information being loaded into the short-term store; failure to refresh</td>
<td>to the long-term store through rehearsal</td>
</tr>
<tr>
<td>long-term store</td>
<td>virtually unlimited</td>
<td>extremely long (years)</td>
<td>interference from other information in the long-term store</td>
<td>to the short-term store through retrieval</td>
</tr>
</tbody>
</table>

Table 3 Summary of the components of the memory system based on the conception of Atkinson & Shiffrin (1968)

Atkinson & Shiffrin speculated that in addition to rehearsal there must be a range of ‘control processes’ for coding, decision-making and retrieval in respect of transferring information to and from the long-term store. (Retrieval here is analogous to internal input as described in Section 1.2.1). They also argued that information must be both stored and processed in the short-term store, and suggested that the capacity limitation of the short-term store involves some kind of trade-off between the functioning of the processing and the storage requirements. In other words, increased effort devoted to processing reduces the storage capacity of the short-term store.

As pointed out above, Atkinson & Shiffrin’s model contained an analysis of both the structure and functioning of memory. But it was and remains widely viewed as emphasising the structure of storage as a steady state, while leaving many questions as to processing unanswered. In a later article, Atkinson & Shiffrin (1971) responded to critics of their original model by placing greater emphasis on processing. They
proposed the term ‘working memory’ (a title which itself suggests processing) and claimed that this is central to the operation of human memory:

We tend to equate the short-term store with ‘consciousness’, that is, the thoughts and information of which we are currently aware, and because of this we consider it to be a working memory: a system in which decisions are made, problems are solved and [through which] information flow is directed. [The short-term store, then, might be considered] simply as being a temporary activation of some portion of the long-term store. (Atkinson & Shiffrin 1971:83)

Atkinson & Shiffrin further suggested that the short-term store processes and stores a number of types of information, including auditory/linguistic, visual and haptic (tactile) information, and speculated that there may be similar subsystems for other types of information. But they presented no convincing evidence for the existence of these hypothetical subsystems or for the way in which they might function. Gathering such evidence and extending the notion of fractionation within working memory became the central concern of the British psychologist Allan Baddeley and his colleagues (see below).

Atkinson & Shiffrin’s model of memory was soon challenged by other theorists, in particular Craik & Lockhart (1972), who stressed the importance of the type of processing involved in the memory system over what they saw as overly static structural divisions in Atkinson & Shiffrin’s (1968) model. In Craik & Lockhart’s model, entitled LEVELS OF PROCESSING, they proposed that the nature of information processing in the memory system, in particular whether it is in the form of ‘deep’ semantic processing or more shallow processing such as in that of physical attributes, determines memory strength; ‘shallow’ processing leads to less durable memory traces than ‘deeper’ processing. Craik & Lockhart also proposed that there are two kinds of refreshment. ‘Maintenance rehearsal’ consists of silent, typically verbal repetition of the information to oneself, whereas ‘elaborative rehearsal’ consists of deeper cognition about a stimulus than mere verbal repetition. The most serious problem for the ‘levels of processing’ model has been in defining exactly what depth of processing is. Later research (for example Baddeley 1978) has found that tasks involving apparently differing degrees of processing ‘depth’ do not always correlate well with efficiency of recall. Barnard (1999) has attempted to solve this problem by defining ‘depth of processing’ more specifically and systematically. Despite this and other drawbacks, the model has been incorporated into almost all current views of memory.
Other theorists have also questioned the notion of Atkinson & Shiffrin’s basic structural distinction between short- and long-term memory. Norman (1968) proposed that the two components might instead constitute different aspects of a single system. He claimed that short-term memory might comprise the temporary activation of traces in a single memory system and that long-term storage might be reflected in permanent structural changes in this system. This view has also been adopted in later versions of the Atkinson & Shiffrin model, for example Shiffrin (1976). More recently, several leading researchers have developed Norman’s concept of activation into a model suggesting that if there are two systems, there is an indistinct boundary between them and an essential integrity in their operation. Their opinions now vary as to whether working memory should be viewed as a currently activated part of long-term memory (for example Anderson & Bower 1973; Anderson 1983; Cowan 1988, 1993) or as structurally separate and merely associated with activation in long-term memory (for example Logie 1995, 1996). These views will be discussed in greater detail later in this section.

In a seminal study, Baddeley & Hitch (1974) proposed ‘a model of working memory in which a controlling attentional system supervises and coordinates a number of subsidiary slave systems’ (p21). Their term slave system referred to a part of working memory that is responsible for storing and processing information of a particular mode. Their initial focus was on the phonological slave system, partly because they suspected that ‘it is one of the simpler components, and partly because it is concerned with an area where considerable data already existed’ (Baddeley 1997:52). Baddeley & Hitch (1974) hypothesised that it comprises ‘a phonemic response buffer which is able to store a limited amount of speech-like material in the appropriate serial order’ (p77). They used the dual-task paradigm in an attempt to isolate their hypothetical phonemic response buffer from the general attentional resources of working memory. The central assumption was that a serial-recall task, similar to the digit-span task, would be carried out specifically by this phonemic response buffer. By contrast, activities such as reasoning, comprehension, and free recall of previously read material would absorb general attentional resources in the working memory.

Baddeley & Hitch found that when the serial-recall task required a memory load of six or more items, performance of both that task and the concurrent task (either reasoning, comprehension, or free recall) was degraded. When the serial-recall task required the retrieval of three or fewer items, however, there was little or no effect on performance in either task. Their logical inferences were, first, that the slave
component itself shows a limited capacity, of six items in this case; second, when
the capacity of the slave system is exceeded, here with a load of six or more items,
the support of the general attentional resources is required, thus degrading
performance in both tasks, implying a 'spill-over' effect from slave system to
general attentional resources. Nonetheless, they concluded that these particular
concurrent tasks appear to draw on separate cognitive resources in working memory.

Baddeley & Hitch (1974) termed the locus of general attentional resources the
CENTRAL EXECUTIVE and the slave system in question the PHONOLOGICAL LOOP
(somewhat akin to Atkinson & Shiffrin's auditory/linguistic subsystem). Baddeley
(1986, 1992) later found evidence that the phonological slave system may itself
comprise two components: a storage mechanism, and a rehearsal mechanism (an
articulatory loop) which refreshes what is stored and is closely allied to the ability to
imagine sounds and speech internally (technically known as subvocalisation). This
fractionation of the phonological slave system was later supported by detailed
analysis of the performance of brain-damaged participants, for example by Baddeley,
Lewis & Vallar (1984), and more recently by Awh, Jonides, Smith, Schumacher,
Koepe & Katz (1996), who used positron emission tomography (PET) scans to
demonstrate that phonological storage and rehearsal are carried out in anatomically
discrete parts of the brain. Baddeley & Hitch (1974) also observed behaviour
suggesting that the phonological loop holds information in acoustical rather than
semantic form.

Like Atkinson & Shiffrin (1968), Baddeley & Hitch (1974) argued that there are
other subsystems, in particular one based on the visual and spatial modes.
Preliminary evidence for a slave system that accommodates such information was
subsequently provided by Baddeley, Thomson & Buchanan (1975). In another set of
dual-task experiments, Baddeley (1986) asked participants to undertake a standard
digit-span task that was assumed to indicate capacity for storing and processing in
the phonological loop. The participants were also instructed to simultaneously
perform a tracking task involving the use of a light-pen to chase a randomly moving
icon around a display. This second task was assumed to indicate capacity for storing
and processing visual and spatial information. He observed only a small degradation
in participants' performance on each task when performed concurrently compared to
when each task was performed alone, providing support for (1) the notion that at
least two separate slave systems are involved in the working memory, one for
phonological tasks and another for visuospatial tasks, and (2) Baddeley & Hitch's
(1974) finding that there may be a small coordinating overhead in performing two
concurrent tasks. These results were supported in later studies such as Logie, Zucco, & Baddeley (1990).

There has been evidence from other dual-task experiments, among them Baddeley (1975), that visual and spatial information is stored and processed in the same slave system. Thus the term VISUOSPATIAL SKETCHPAD (or SCRATCHPAD) has been adopted by these researchers. Against this, other studies, notably Baddeley & Lieberman (1980), have suggested that visual and spatial information are processed in separate subsystems. Baddeley (1997:79) now feels that the visuospatial sketchpad may be either 'a multifaceted system, with both visual and spatial dimensions, or possibly two separate systems' and that the visual system may itself consist of 'two separable components, one concerned with pattern processing', detecting the nature of the target, and the other concerned with 'location in space', conveying information about where the target lies in the external environment.

This resonates with the work of Smyth, Pearson & Pendleton (1988), who found that subjects' memory for a sequence of body movements and for spatial position is degraded when a movement-suppression task is performed during presentation of the sequence to be remembered. The movement-suppression task, however, does not affect memory for visually presented words or a sequence of spatial movements. Smyth & Pendleton (1989) further investigated this phenomenon and concluded that memory for body movement involves different processes from those used in spatial tasks, and that there may be a need for a slave system in working memory that is specific to each. The visuospatial slave system, then, does not appear to exercise a unitary function.

The concept of slave systems continues to be one of the main strands of research into working memory. The irrelevant-speech effect, the impairment of performance by the presentation of irrelevant background speech, was one of the criteria Baddeley & Hitch (1974) had used to demonstrate the existence of the phonological loop. This effect was the subject of several studies during the 1990s. Jones, Madden & Miles (1992), Jones & Macken (1993) and Jones et al (1995) claimed that the disruptive effect of irrelevant speech is limited to tasks involving serial rehearsal, that is, tasks in which an individual repeatedly vocalises visual stimuli in sequence to remember them. Salamé & Baddeley (1990) had also found that the effect is not observable in a free-recall task, thus not requiring serial rehearsal. In the light of the fact that serial rehearsal is not relevant to most real-world tasks, Le Compte (1996) explored whether the irrelevant-speech effect has a more general suppressive effect.
on working memory. His observation of the effect in a missing-item task, also not requiring serial rehearsal, contradicts the earlier assumption.

Before completing this summary account of the two established slave systems, the opportunity will be taken to question their terminology as being misleading on a number of counts. Phonological ‘loop’ may imply that while phonological information is stored and processed in a (looped) sequence, visuospatial information is not. This is despite the fact that visual input is predominantly in the form of a sequence of fixations (even though most readers are unaware of it, and think of visual information as ‘static’ unless movement of the object is involved, as further explained in Section 3.1). ‘Loop’ may also give rise to confusion between (1) Baddeley’s articulatory control mechanism as the agent of the rehearsal process, which may well bear some resemblance to a loop, and (2) the processing and storage of information in the slave system in the absence of rehearsal which is widely accepted to decay rapidly rather than automatically ‘looping’ back. In addition, ‘phonological’ refers to vocal sounds, particularly those of a language, yet the phonological loop has been shown to admit other sounds, including those of instrumental music (Salamé & Baddeley 1989). Logie (1996:57-58), too, has also cast doubt on whether information in the phonological loop is exclusively language-based. ‘Sketchpad’ is problematic because it implies that visuospatial information is in two dimensions, whereas it is typically perceived as three. Furthermore, applying ‘sketchpad’ only to visuospatial information also implies that phonological information does not possess dimensional properties, in spite of the fact that it is binaural (in stereo, so to speak), thus giving the listener spatial cues as to the sound source. ‘Sketchpad’ also implies that other information in working memory such as musculoskeletal and tactile signals do not possess spatial properties, whereas they clearly do. In addition, the linking of spatial and visual information in the term ‘visuospatial’ is inconclusive in the light of evidence reported earlier.

Most research into fractionating working memory has concentrated on the slave systems because they can be directly related to modes of sensory input, principally the auditory and the visual. By contrast, considerably less is known about the central executive, which has been regarded as something of a default component, performing whatever tasks in working memory that are not easily attributed to these slave systems. Baddeley (1981:21) described the central executive as ‘the area of our residual ignorance’ about working memory, admitting later that it is ‘something of a ragbag for consigning such important but difficult problems as to how information

41
from the various slave systems is combined, and how strategies are selected and operated’ (Baddeley 1997:85).

Attempts have been made, however, to isolate and measure the capacity of the central executive in such tasks as generating random numbers (for example, Gilhooly, Logie, Wetherick & Wynn 1993). These tasks are assumed to involve continuous and conscious decision-making by the central executive while imposing minimal demands on the slave systems. Although there is now general consensus that the central executive performs a coordinating and attentional function, there has been and remains equivocation as to whether it is merely a coordinator of the slave systems or whether it could itself perform the functions of storage and processing, as was suggested by Baddeley (1974), Baddeley (1976, 1979), Baddeley & Hitch (1977), and Baddeley & Lieberman (1981). More recently, Baddeley has acknowledged that whether the central executive proves to be ‘a single coordinated system that serves multiple functions, a true executive, or a cluster of largely autonomous control process, an executive committee, remains to be seen’ (reported in Richardson 1996b:137).

Despite the popularity of the multicomponent model of working memory, there remains little agreement as to its structure and function. Different researchers fractionate working memory in different ways and to a different extent. Some have proposed only a few specific subsystems, such as Baddeley’s visuospatial sketchpad and phonological loop. Some have postulated that there are other subsystems, or that at a deeper level there are specific areas in working memory that represent auditory, lexical, and grammatical information. But it is now becoming clearer that the multicomponent model cannot account for many of the complexities observable in cognitive processing. Even the leading proponents of the model now appear to be stepping back from the idea of working memory as a single, isolated system that underlies cognitive capacity. As Logie (1996:40) puts it, working memory ‘is better thought of as a set of specialised mechanisms that act in concert according to the demands of the task in question.’

While the theories of Baddeley and his colleagues have undoubtedly made a significant contribution to knowledge about how working memory contains a number of specialising loci, an alternative view of memory has arisen that stresses the concept of activation over the fractionation of working memory into subsystems and the notion that working memory and long-term memory are separate stores. Earlier theorists stretching back to Norman (1968) had postulated the concept of
Memory and music reading

activation, which has brought into question the whole notion of a distinct separation between working and long-term memory. In Cowan's (1993) model of the memory system, for example, there are two levels of activation, one associated with the current focus of attention, the other with a body of information which, while not currently focused on, is relevant to it and is readily retrievable. To Cowan (1999:71), 'Baddeley's (1986) phonological loop and visuospatial sketchpad are just two varieties of memory activation, along with the processes that can be used to reactivate this memory [that is, through refreshment].' In a passage that poignantly harks back to the quotation by Galton (1893) in Section 2.1 of this dissertation, Logie (1996:37) exemplifies the Cowan's notion of activation as follows:

Thus, for example, I am writing this text while sitting on a train, and it is the current focus of my attention. I could readily call to mind information about trains, tickets, stations, and my planned destination, all of which are at activation levels just below the threshold for inclusion in my current working memory. Other information about an airplane journey later in the day is also bubbling about in the background, and it too is readily available. The rest of my knowledge base is at a much lower level of activation. Therefore, the information, say, about a particular castle in Scotland takes longer to retrieve than does information about airplanes and trains, but, after [information about the Scottish castle] has been retrieved, other things that I know about [it] become available just below the working memory activation threshold and replace information related to the current external context of the train.

Ericsson and his colleagues have questioned the extent to which Baddeley's approach of isolating components of working memory through observing performance in highly restricted laboratory tests might be applicable to everyday activities: 'It is no longer possible to isolate the critical components of working memory without first considering their functional context in the corresponding skilled activities' (Ericsson & Delaney 1999:269). Ericsson & Kintsch (1991) observed that expertise typically involves a high level of memory performance in areas that have been shown to be subject to interference in simple concurrent tasks. Furthermore, when participants are performing activities in which they are skilled and for which demands on working memory are large and complex, a more ecologically useful account of working memory capacity has been provided than when they are performing simple tasks at which they are permitted no prior practice. Practising and thereby attaining expertise at performing a task, Ericsson & Kintsch (1991) have argued, brings to bear a more efficient activation of related information in long-term memory. Concurrent memory tasks, which are central to claims that working memory is divided into specialised components, appear to cause much less impairment in highly skilled activities such as piano playing (Allport,
Antonis & Reynolds 1972), typing (Shaffer 1976), and reading (see Baddeley 1986 for a review) than in simple, unpractised activities such as digit-span tasks.

In some cases individuals have shown an extraordinary memory capacity in their area of expertise, for example in memorising huge sequences of numerals (Ericsson & Chase 1982), chess positions (Charness 1976; Saariluoma 1992), and meal orders in a restaurant (Ericsson & Polson 1992). The concept of chunking (Miller 1956) discussed earlier, is relevant here, but explanations have been more fully developed by researchers such as Ericsson & Kintsch (1991) and Ericsson & Pennington (1993), who have developed a model of working memory in which its underlying capacity is increased in an individual's areas of expertise through the activation of relevant material in long-term memory. In these studies it was postulated that information which has been coded into the long-term memory as a result of acquiring expertise is maintained at a higher base level of activation during tasks that employ that expertise, thereby increasing the capacity of working memory during the performance of those tasks.

More recently Ericsson & Delaney (1999) have proposed that there is a half-way house, as it were, between working memory and long-term memory, which they termed LONG-TERM WORKING MEMORY (LT-WM):

LT-WM is mediated by associative recall from long-term memory, and to function reliably it provides [various] mechanisms for overcoming the problems of interference resulting from repeated associations to related retrieval cues.....LT-WM reflects a complex skill acquired to meet the particular demands of future accessibility for information with tasks within a particular domain of expertise. Domain-relevant skills, knowledge, and procedures for the task are so tightly integrated into the skills for encoding information that the traditional assumption of a strict separation between memory, knowledge, and procedures is not valid for skilled performance. (p257)

Another issue that has not as yet featured in this account, that of AUTOMATICITY, is now widely accepted to have a significant bearing on the acquisition of expertise. Shiffrin & Schneider (1977) demonstrated that participants acquired automaticity in a difficult letter-recognition task after much practice. They concluded that practice streamlines both cognitive and musculoskeletal tasks and reduces the attentional resources required to perform them. Hirst, Spelke, Reaves, Charack & Neisser (1980) made similar findings in relation to a reading/dictation dual-task experiment. Norman & Shallice (1986) put forward a specific definition of automaticity, together with a detailed although untested model for how automatised processes might work. Under this model, mutually interfering activities that are carried out concurrently are
controlled by a system of relatively automatic rules acquired through practice. This is known as ‘contention scheduling’, and is accompanied by a system of triggers to alert a ‘supervisory attention system’ (analogous to the central executive) at times when these rules need to be interrupted or modified. Although there is little agreement on the specifics of automaticity, it would appear that automaticity is typically possible through access to information in long-term memory that is connected with relevant pre-learnt actions, skills and processes. Norman & Shallice’s model was later used by Truitt, Clifton, Pollatsek & Rayner (1997) as the basis for their description of automatisation in the music reading process (see Section 3.3.4).

Automaticity is used by some writers, for example Sloboda (1985:217), to refer rather generally to an activity, or part of an activity which requires little or no conscious attention to be executed. Logan (1990) suggested that there might be two alternatives to the commission of a task: fresh computation and access to long-term memory of the results of those computations based on prior experience of the task. These two alternatives may, he suggested, compete with each other every time the task is performed, such that if the residual memory traces of the relevant processing are strong enough through prior repetition of the task, they will prevail. In the words of Best (1999:72) ‘this view suggests that we set foot on the pathway to automaticity as soon as we begin to make the same type of response to specific stimuli’.

This process is central to the task of reading the same passage of music six times in Stage II of this study. Such repetition, further discussed in Section 5.1.4, may shift the reliance of the music reading apparatus from external input (and fresh computation) to internal input of memory traces of prior readings. In this way, the gradual onset of automaticity in music reading may be observed. Taking Logan’s (1988, 1991) definition of automaticity as the release of attentional capacity, it was hypothesised in Stage II that attentional resources can be redirected away from fresh computation towards maintaining either a larger minimum eye-hand span or a faster tempo as memory traces of the text are strengthened by repeated readings.

Baddeley’s (1997:90) definition is oriented towards his interest in dual-task experiments: automaticity is acquired when a given stimulus is repeatedly paired with the same response, thus presumably drawing on progressively fewer attentional resources, and interfering less and less with other concurrent tasks. The hallmarks of automaticity, he claimed, are minimal interference between the automatic process and other concurrent activities, together with an apparently unstoppable tendency for
the automatic stimulus to evoke that response. Automaticity is relevant because it depends on overcoming interference between demands that draw on similar memory resources.

2.2 Music reading and the memory system

2.2.1 Information sequencing, activation and the multicomponent model

In this section and the next, an attempt will be made to relate the foregoing accounts of the multicomponent model and the notions of activation, chunking and automaticity to the specific case of performing music from score. For this purpose it will be necessary to draw upon several other matters that have already been discussed in this dissertation. Chief among these are the notion of information sequencing and the distinct temporal qualities of the signals in the music reading process, as explained in Section 1.3 and illustrated in Figure 8. It will be suggested that a knowledge of the mechanics of the music reading process might provide an opportunity to view the memory system, in particular working memory, slightly differently from the standard set of views to be found in the literature.

The temporal sequencing of information in working memory has been assumed by psychologists for some time. Baddeley & Hitch (1974:74) claimed at the outset that their hypothetical phonemic response buffer stores information in ‘appropriate serial order’. In this study it is contended that temporal sequencing might be regarded as one of the defining elements of whether information is being stored and processed in working memory. Some psychologists may object to the notion that information passes serially through working memory, seeing an erroneous implication of physical movement through the relevant parts of the brain. In this study, however, it is argued that Baddeley & Hitch were instead using ‘serial order’ as a paradigm for the representation of information in working memory, just as they postulated that there are functional structures within the memory system without assuming that they are represented in distinct neuroanatomical loci in the brain.

There are reasons for believing that serial ordering is involved in most if not all modes of information that are stored and processed in working memory. As assumed in Section 1.3, performance instructions are visually perceived as a sequence of fixations over the score and are transformed into a sequence of commands. On logical grounds the most recently loaded chord is the least processed, and the furthest from being transformed into musculoskeletal commands. The chord that was loaded before
it, which has resided in working memory for longer, is likely to be further ‘down the track’ of being transformed into associated musculoskeletal commands; the chord before that, if as many as three chords are held within the span, is likely to be yet closer to transformation into musculoskeletal commands. There is no reason for suspecting that such a transformation occurs instantaneously or in any other order than that in which the chords are loaded into working memory. Thus it can be assumed that the transformation occurs gradually within the span according to the serial order with which each chord is loaded. In Section 1.3 this process of transformation was conceived in terms of ‘decay’ and ‘consolidation’ of the signals involved.

Music perception provides a further example of the importance of information sequencing in working memory. For example, listening to music relies on a listener’s awareness of the serial order of a sequence of auditory events, not only as they are sounded, but as they are stored and processed in working memory. Parncutt & Huron (1993) have studied the memory of music that persists for several seconds, after which it is lost unless refreshed. This type of immediate memory enables the listener to relate what is heard at any particular point of time to what has just been heard. It permits the listener to maintain a temporal ‘window’ wide enough to recognise melodic and harmonic progressions, dynamic gestures, indeed to perceive the very fabric of the music as it passes by, based on moment-to-moment sequences. In this respect such a window of memory is equivalent to decaying auditory feedback as schematically illustrated in Figure 8 in Section 1.3; it is also analogous to the eye-hand span, which similarly moves across the score as a window, although the span covers a different and contiguous window on the score at any one time.

Other modes of information would also appear to be stored and processed serially in working memory. Musculoskeletal commands must be generated in sequence to produce music that is in sequence, their window comprising consolidation rather than decay (see Figure 8 in Section 1.3). Visual information is typically perceived as a serial ordering of relatively stationary fixations; on logical grounds it is important for the memory system to maintain an awareness of the serial order of each fixation it is storing and processing, otherwise it would be difficult to make sense of visual input. Whether or not information is sequenced in such a conscious way may even represent a useful criterion for determining whether it resides in working memory, although confirmation of this point is beyond the scope of this study.
The question remains of whether the seven signals involved in the music reading process are stored and processed by the same resources in working memory, and whether such resources can be likened to the established slave systems. Deutsch (1975a, 1975b) and Martin, Wogalter & Forlans (1988) among others have demonstrated that memory for musical tones is not degraded by unattended speech, and vice versa. On the basis of their work, Bertz (1995) has put a convincing case that a 'slave system in addition to those proposed by Baddeley probably exists, one that is used to store and process musical information. .....The [musical] loop could be a totally separate component or could possibly be attached to the phonological loop....'

The primary method of isolating slave systems in the multicomponent model has been to identify interference between concurrent tasks performed by novice participants, that is, participants who were purposely not permitted to gain expertise in 'overcoming' such interference. In this context, automaticity is minimised and the potential for interference is maximised. The work of Ericsson and others over the last decade has argued that Baddeley's findings are inadequate when expert performance is of interest. Musical performance is an expert task, and a significant part of acquiring expertise may involve the 'overcoming' of interference, and the acquisition of relevant long-term working memory. Examples of such interference between simultaneous musculoskeletal demands abound in music training. It is typical, for instance, that the performance of beginner keyboard students should suffer degradation through interference when combining the hands compared to performance with separate hands. Similarly, organists typically encounter interference problems when first combining their left hand with the pedals, especially where both parts occupy a similar pitch range. But these are initial problems that students learn to overcome through practice, the very condition that has been excluded in Baddeley's experiments.

The established slave systems, therefore, may not provide a satisfactory solution for where the various signals of the music reading process are stored and processed in working memory while playing from score. It is clear from Section 1.3 that the seven information signals must all be separately sequenced and activated in working memory, regardless of any accepted slave system in which they may be represented, since they are distinguished by their perceptual/propiroceptual mode, temporal characteristics, function, and the command group to which they belong. For example, oculomotor and musculoskeletal commands may well call on similar processing
resources since both involve the operation of muscles. But as Figure 8 in Section 1.3 illustrates, they occupy different time-frames in relation to the score.

In the light of such evidence, it is proposed here that seeking to consign the storage and processing of the seven information signals involved in music reading to an analysis of Baddeley’s slave systems is only of limited use, and that it would be more helpful to assign the locus of the encoding, storage and retrieval of each signal to an ‘activation sequencer’. At least in relation to continually sequenced tasks such as music reading, it would therefore appear more useful to refer to the established slave systems not as ‘loops’ or ‘sketchpads’, but as ‘sequencers’, thus informing their names with an assumed common characteristic. The term ‘sequencer’ might raise the objection that information in long-term memory can also be coded in sequence; otherwise it would be impossible to perform music from memory. But such coding in long-term memory is not currently activated, and thus there should be no confusion over the use of the term ‘activation sequencer’ as proposed here for the slave systems, since they have been conceived only in relation to information which is the current focus of attention. In the next section this term will be assigned to the loci of storing and processing the seven signals that function in the music reading apparatus.

The definition does not imply that each activation sequencer is mutually exclusive or discretely located in the memory system, even though they may well involve those parts of working memory that Baddeley has identified as slave systems in restricted tasks. On the contrary, in the light of Ericcson’s findings in relation to expert performance, they are viewed here as likely to overlap considerably in both function and anatomical location, and to extend into long-term working memory. It is not possible at this stage to offer a more explicit definition of activation sequencers; that must wait for further research into the memory system. The current definition is proposed merely as a contribution to debate on the functioning of memory in music reading, and as a perspective on the psychological processes underlying the use of SightReader.

Activation was discussed in the previous section in the context of a growing awareness of its role in the functioning of the memory system, and as a way of explaining the interconnectedness of the traditional notions of working and long-term memory. A little work has been carried out on activation in music perception over the last few decades, summarised in Deutsch (1999) and Bharucha (1999). Deutsch (1972) investigated the phenomenon of pitch ‘priming’, the process by which
perception of one musical event, in the simplest case a single tone at a predetermined pitch, activates that pitch in the memory system and predisposes a participant to perceptually process it on subsequent sounding more quickly and accurately than if it had not been previously sounded. Tekman & Bharucha (1992) studied chord priming and concluded that evidence for the persistence of tonal activation was clear; that a chord sounded for as short a duration as 50ms could prime a subsequent chord even if the two chords were separated by as much as 2.5 seconds of silence. Barucha has used neural net modelling to investigate what he refers to as 'temporal composites' or patterns superimposed upon each other over time: 'As a piece of music unfolds, patterns can be composited over time by the accumulation of activation.' (1999:420)

2.2.2 Chunking and automaticity

In Section 2.1, chunking was discussed as being widely accepted as playing an important role in understanding the functioning of working memory. This is particularly so in respect of tasks that impose a heavy burden of processing and/or storage. For example, there is clear evidence of the importance of its role in the retention of large sequences of digits (Ericsson & Chase 1982), chess-board positions (Chase & Simon 1973; Charness 1976; Saariluoma 1992), and language (Simon 1974; Zhang & Simon 1985; Yu, Zhang, Long, Peng, Zhang & Simon 1985). There is also little doubt that the speed and accuracy of cognition required to read and perform music is partly achieved by chunking, although little direct empirical work has been done in this area, apart from Polanka (1995:182), who on the basis of their eye movement data concluded that 'better readers tended to read in larger [chunks].'

It was also pointed out in Section 2.1 that automaticity and the overriding of interference appear to be central processes in the acquisition of complex skills. These two phenomena are to some extent difficult to distinguish since chunking may well play a part in automaticity, and conversely automaticity may make chunking easier. A passage by Sloboda (1985:74) illustrates how similar these phenomena are. In relation to musculoskeletal commands in music reading, he appears to treat the 'single integrated' (that is, chunked) unit as automatised:

.... the experienced sight reader confronted with a familiar scale passage will not need to make conscious decisions about which fingers to use for which notes. His hand will automatically take up the right configuration....When we hear the phenomenal speed of a virtuoso performance, we can render it explicable....by recalling that what strikes the ear as twenty separate notes may well be, for the performer, a *single integrated* and *automated* unit. [emphasis added]
The power of chunked and automatised musculoskeletal commands is also evident to keyboardists in learning a new piece which can mostly be played using pre-learnt patterns of fingering, but which requires novel patterns to be marked on the score in a few places. These few places require particular attention and consume an inordinate amount of rehearsal time because they cannot be executed with the 'default' musculoskeletal commands. Sloboda (1976, 1978a) found that musicians are better than non-musicians at picking up and retaining, presumably in their working memory, scale and arpeggio contours from patterns of noteheads which are displayed briefly on a tachistoscope (for 20ms-2000ms), a result later confirmed by Halpern & Bower (1982). This suggests that musicians are primed to store such patterns as chunks, in a similar way to chess masters' superiority in chunking chess positions (see above). To scale and arpeggio contours might be added the harmonic formulæ that inform cadences in western tonal music; in this tradition, harmony is at its most predictable at the start and end of a phrase, where the choice of chords is most constricted. A phrase almost always ends on chord I or V (very occasionally vi), with I invariably preceded by V, and V by ii6. It is therefore likely that readers rely more on chunked internal input at the end of a phrase than in the less predictable middle of a phrase.

In relation to the auditory mode, Sloboda (1985:3) has suggested that phenomenal feats of musical memory, such as Mozart's reputed writing out from memory of Allegri's Misere after hearing only two performances of the work, are achieved by identifying and processing large patterns of individual notes as chunks. Sloboda also concluded on the basis of his (1977) study that 'music which does not contain familiar patterns and structures cannot easily be represented in a listener's memory'. More specifically he claimed that music which conforms to the rules of tonal harmony, containing many patterns which listeners are primed to chunk, is much easier to remember than nontonal music (Sloboda 1985:4). Musical composition probably also relies on the ability to manipulate chunks rather than individual notes in the memory system, in view of the speed with which some composers have written music, and the necessity of their maintaining the 'larger picture' of a work in a reasonably high state of activation as they write it. Improvisers, too, must be able to conceive the notes they play at such a speed as to make their performance extremely difficult without the use of chunking.

Exactly what constitutes a chunk has turned out to be rather more difficult to ascertain than its mere existence. This is especially problematic in music because most textures present a complex multilayered lattice of potentially chunkable
patterns, typically without clear boundaries between them. The smallest particles of a homophonic or polyphonic musical texture, the notes, are adjacent to not only their vertical neighbours in the same chord, but to their horizontal neighbours in the same part. Thus chords as well as melodic patterns could both form the basis of chunking. Furthermore, the familiar and therefore chunkable patterns of a musical style occur on a number of temporal levels. For example, there are melodic formations of single intervals, as well as larger scalar and triadic patterns. Frequently occurring chord *progressions* as well as individual chords may also be candidates for chunking.

Despite the theoretical and empirical difficulty of defining chunks, Sloboda (1985:73) has presented a convincing argument that durational structures in music (that is, those of rhythm, pulse, metre, and period) exert a strong influence on musicians' chunking procedures. One reason for this is that durational structures are usually optically distinct in music notation: individual notes of quaver-value and below are beamed together into beats, barlines provide regular boundaries marking off material that may be small enough to be chunked, and the juncture between periodic units (phrases) is typically marked with rests, slurs, and/or notes of comparatively longer rhythmic value. Another reason that duration may play a key role in chunking is that musculoskeletal commands are generated according to strict temporal values, which are therefore likely to represent a common organising feature in both visual input and musculoskeletal command sequencer.

Sloboda (1977) found that a phrase boundary just beyond a reader's average eye-hand span 'stretches' the span, and a boundary just within it 'contracts' it, suggesting that phrase boundaries mark at least some of the boundaries between effective performance units. *SightReader*'s process of stretching span size, incidentally, overrides such fluctuations with unknown consequences. But since most musical phrases contain rather more notes than could possibly be chunked, certainly more than 4 1 chunks, Sloboda later realised that 'some principled way of dividing a phrase up into smaller units' must exist (1985:72). He went on to describe a hypothetical scheme for chunking music based on 'a hierarchical binary branching' scheme of organising when each note or chord already loaded into working memory should be played, so that the performer can initiate ballistic finger movements by reference to a 'countdown' for each beat-slot. Where the music comprises notes or chords of longer duration, such as minims, the hierarchical branching scheme need only be utilised at the minim level. Where quaver movement is loaded into the working memory, a deeper level of the durational organising scheme must be brought
to bear. Figure 14 is an adaptation of Sloboda’s hypothetical durational organiser for chunking in COMMON TIME.

At least two teaching methods focus students on learning to chunk more efficiently, although this is not their explicit function. The first is technical work, that is, the rote practising of scales, arpeggios and other repetitive patterns. This is commonly viewed as a physical training, a means of acquiring musculoskeletal strength, endurance, and agility. But even though most teachers and students are not aware of it, technical work is a basic and important method of training students to chunk both musculoskeletal commands and the visual input of those patterns.

A less common method is figured-bass training, which is usually pursued to improve students’ harmony or to prepare them for professional continuo work, rather than as explicit training in chunking. Figured-bass exercises of the type quoted in Figure 15 present only the outer two parts for each chord. Students are expected to use this partial information to play full four-part chords by interpolating the two missing inner parts. In Figure 15 all the chords are in root position, so that students can identify the letter-names of the missing inner parts from the bass-note alone. The first bass note here is G, indicating a G-B-D triad. Students must mentally arrange these three tones to form the full chord, doubling an appropriate tone and taking into account the given top note. Finally they must generate the relevant musculoskeletal commands to play the full chord. In this way students are encouraged to chunk whole triads on the basis of the bass note alone, and to acquire fluency they must automatise this process. Conceptualising the notes of a chord as a single, integrated whole represents the chunking process. Figure 16 shows the solution to the exercise.

Chunking is potentially relevant to the operation of SightReader because it may be linked to the size of the eye-hand span. Sloboda (1977) demonstrated that the size of
the eye-hand span (and what he presumed is an associated propensity to chunk) is reduced in sight reading a melody when familiar patterns of pitch and rhythm are removed. He did this by adapting a technique used by Shaffer (1976), who had observed copy-typing under conditions of controlled preview using a computer monitor. Shaffer's study had suggested that copy-typists chunk words, but not larger units, since their performance speed suffered little degradation until preview was restricted to less than the immediately upcoming word. Sloboda (1985:72) concluded that 'It seems that the reader subdivides a phrase into performance units in ways dependent on the harmonic and rhythmic structure of the phrase, and that when cues to these structures are obscured, preview is not so useful and cannot, indeed, be sustained at normal levels.'

Sloboda also claimed that whereas a melody typically contains familiar structures such as scales and arpeggios which are likely to be processed as chunks, the melody probably also contains novel combinations of notes. The reader's strategy, therefore, appears to involve the chunking of frequently occurring patterns of notes, supported by some more general strategy for processing novel patterns. This explanation of chunking in music reading appears to resonate with Norman & Shallice's (1982) model of 'contention scheduling' (akin to automaticity) accompanied by a system of triggers to alert a 'supervisory attention system' (akin to the central executive) needs to intervene to provide more controlled processing (see p35).

Because SightReader forces students to hold more information than normal in their working memory, chunking is postulated here to be a readily available means of increasing its capacity while using the device. Learning to chunk more efficiently through exercises such as the one quoted in Figures 15 and 16 may be a powerful adjunct to stretching span size on SightReader, although this option was not explored in Stage II.
2.3 Conclusions

The current state of knowledge concerning the memory system could only be described as partial. It is, however, reasonably safe at this stage to make certain assumptions about the memory system. At the most basic structural level, there appears to be a functional distinction between working memory, a system responsible for the temporary storage and processing of information, and long-term memory, although the contents of working memory appear to be closely connected with relevant activated information in long-term memory, referred to as long-term working memory by Ericsson and his colleagues. The capacity of working memory is highly fluid and is almost certainly constrained by the extent of useful activation which can be brought to bear in long-term memory and by the attentional resources that are available to maintain task-relevant information and inhibit and remove task-irrelevant information.

Working memory appears to contain at least two specialised loci, referred to as slave systems, that store and process auditory/linguistic and visuospatial information respectively. A central executive may coordinate the operation of the slave systems and supervise the encoding of information in long-term memory. These structural components have been experimentally isolated in narrow tests of novice participants that may not bear a close relationship to the functioning of the memory system in complex everyday and expert activities. The exact status of the these components thus remains in doubt.

Performing music from score is an expert activity that requires the sequencing of several signals, most importantly those of visual input and musculoskeletal response. Such reliance on information sequencing provides the basis for speculating that highly activated sequencing may be a useful criterion for determining that a unit of information is being stored and processed in working memory. Acquiring musical expertise appears to require the 'overcoming' of interference between different processes that draw on similar resources in working memory, which is probably connected with the phenomenon of automaticity. Such overcoming of interference may be at least partly achieved through the process of chunking.

*SightReader*’s unique ‘span-stretching’ function requires working memory to hold more information for longer in a highly activated state, and may encourage the user to chunk more efficiently. These issues and others connected with the memory system as it functions while using *SightReader* are taken up in greater detail in several sections in Chapter 3 as well as Section 5.1.3 and 5.1.4.
Eye movement in the reading of language and music

The previous chapter, as a prelude to the two experiments in this study, dealt with issues surrounding the memory system. Chapter 3 performs the same function in relation to another part of the music reading process, the way in which the eyes move over a text during reading. The starting point is a brief explanation of the basic features of visual input. This represents important background information for understanding the rest of the chapter. Section 3.2 contains a review of the literature on eye movement in language reading. This field has historically provided the basis for the theory, equipment and methodology of most eye movement research, including that concerning such a closely related field as music reading. In this section, the theoretical and technological developments in research into eye movement in language reading are traced from the earliest surviving records to the present. Like the literature on memory, it is a vast field and the review is therefore selective, emphasising matters of relevance to this study.

Section 3.3 is a review of the literature on eye movement in music reading. Slender by comparison, this area is treated comprehensively. The current state of knowledge is evaluated, the key issues are put into perspective, and the ground is prepared for the development of some of the hypotheses, aims, and methodology for both Stages I and II. The review suggests that eye movement in music reading is poorly understood, that the methodology of its investigation has been highly problematic, and that opportunities to draw on knowledge of the memory system to explain key aspects of its data have not been adequately pursued. It also suggests that it is useful to observe eye movement with tracking equipment under reasonably natural conditions before attempting to control it with SightReader.

3.1 Introduction to visual input

External visual input is initially processed by either the central area, also known as the 'fovea', or the peripheral area, also referred to as the 'parafoveal area'. The central area delivers visual information of high resolution. Input from the peripheral
area is of markedly lower resolution than from the central area, although there is strong evidence that peripheral vision also plays an important role in reading. Where the eyes are looking is implicitly assumed in most of the literature to refer to the location of input from the central area.

Reading involves a sequence of fixations, 'snapshots' as it were, during which the eyes are relatively stationary. Fixations expose successive parts of the text to the central area as the eyes scan the page. Fixation durations in music reading are typically of 250-400ms' duration, as widely observed in the literature (see Sections 3.3 and 4.3). This is significantly longer than the 200-250ms norm in the silent reading of English text (Rayner 1995:4), but comparable to fixation durations in copy-typing English (Shaffer 1976). Fixations are linked by extremely rapid movements, flicks as it were, called saccades. Unless the eyes are following a moving target, they are virtually unable to move without saccades and fixations (Matin 1974). The alternation of fixations and saccades characterises external visual input in many human activities. Depending on the distance between the eyes and the page, saccades are on average around eight or nine letter-spaces in language reading (Rayner 1995:4).

Saccades, then, function to bring new information to the central area, and fixations to allow the perception of such information. Surprisingly, despite our experience of continuous visual perception during reading, the available evidence suggests that information is taken up only during fixations (Matin 1974) and that the perceptual system shuts down for the duration of each saccade, as frequently as several times per second. The alternation of fixations and saccades in reading traces a scanpath over the text. Figure 17 is a schematic representation of a typical scanpath over a keyboard score, showing the characteristically ragged appearance of eye movement. The representation is based on data collection in Stage I of this study, and on other studies such as Weaver (1943). Here the fixations are represented by small crosses, and the saccades by broken lines; the fixation durations are indicated in milliseconds.

10 Pronounced se-kahdz, this term comes from an old French word for a horse-rider's flicking of the reigns as a signal to the horse to change direction; in modern French, the term also refers to the flicking of a sail in the wind (Crystal 1987:208).
11 Fixations and saccades contrast with the 'smooth pursuit' of a moving target, which is executed by a quite different set of oculomotor muscles. Whereas fixating and saccading is common to all vertebrates, only primates and a few other predator vertebrates are capable of smooth pursuit (Land 1995:63).
3.2 Review of the selected literature on eye movement in language reading

From an historical perspective, eye movement research can be seen as falling into three periods. The first period stretched from early times until the second half of the nineteenth century, and was characterised by a reliance on the naked-eye observation of eye movement in the absence of technology. The second period covered from the late nineteenth century until roughly the middle of the twentieth century. During this time investigators utilised early tracking technologies to assist their observations, in a research climate that emphasised the measurement of human behaviour and skill for educational ends. Much basic knowledge about eye movement was obtained during this period. The third period, extending from roughly the middle of the twentieth century to the present day, has been characterised by three major changes: the development of non-invasive eye movement tracking equipment; the introduction of computer technology to enhance the power of such equipment to pick up, record and process the huge volume of data that eye movement generates; and the emergence of cognitive psychology as a theoretical and methodological framework within which reading processes are examined.

3.2.1 The first period: unassisted observation and introspection

Researchers in this period had at their disposal two methods of investigating eye movement. The first, the unaided observation of participants’ eye movement as mentioned above, yielded only small amounts of data and would be considered
unreliable by today's scientific standards. This lack of reliability arises from the fact that eye movement occurs frequently, rapidly, and over small angles, to the extent that it is impossible for an experimenter to perceive and record the data fully and accurately without technological assistance.\(^\text{12}\) The other method was self-observation, now considered to be of doubtful status in a scientific context. Readers need only attempt to count the number of fixations they make during this sentence to realise that self-observation cannot help but significantly alter the nature of their normal eye movement in reading. Despite these drawbacks, a surprising amount of knowledge about eye movement appears to have been derived from introspection and naked-eye observation. For example, Heller (1988:39) claims that Ibn al Haytham, a medical man in 11th century Egypt, wrote of reading as a series of quick movements and realised that readers use peripheral as well as central vision.

For the next 800 years, however, there appear to be no records of eye movement research, and it was not until the early nineteenth century that written evidence of systematic investigation emerges. At first the chief concern was to describe the eye as a physiological and mechanical moving object, the most notable attempt being Helmholtz's (1866) major work *Handbook of Physiological Optics*. The physiological approach was gradually superseded by interest in the psychological aspects of visual input, in eye movement as a functional component of visual tasks. As early as the 1840s, Hueck (1840) and Weber (1846) had speculated on the relationship between central and peripheral vision.

The following decades saw more serious attempts to observe eye movement. Cattell (1885, 1886) went one step further by looking directly at participants' eye movement and on that basis claiming that meaningful words require fewer fixations to read than random strings of letters. In what might be regarded as an anticipation of the tracking devices of the second period, a French ophthalmologist, Javal (1879) placed a hand-mirror on one side of a text so that he could see participants' eyes as they read it silently. He thus found that eye movement comprises a succession of discontinuous individual movements for which he coined the term 'saccades'. He also claimed that the eyes fixate on every tenth letter, in hindsight an implausible notion, and an example of the inherent limitations of unassisted observation.\(^\text{13}\)

\(^\text{12}\)This is not to imply that unaided observation was unique to the first period: it was indeed the method used in a preliminary investigation for this project, in which the eye movement of undergraduate keyboard students was informally observed without technological assistance as they sight read.

\(^\text{13}\)Javal (1879) appears to have been overly keen to find order in eye movement. A need to find order in seemingly chaotic eye movement is also evident in Huey's (1908) claim that both fixations and saccades are rhythmically regular. Both Huey's and Javal's claims remained influential for several decades.
Erdmann & Dodge (1898), too, used a hand-mirror to observe eye movement in reading, and estimated average fixation duration and saccade length with surprising accuracy. Their study was one of the last scientific observations of eye movement not to use mechanical or electrical recording equipment, and marks the end of the first historical period of eye movement study.

In addition to Javal’s (1879) use of a mirror, there was another precursor to the second period, the Tachistoscope. Invented by Volkman in 1850, this device displays visual images in timed succession. It spawned a voluminous amount of research on the way the eyes perceive stationary visual images, including a study by Bean (1938) of the rapid recognition of noteheads in music notation. Eye movement, however, cannot be observed with a tachistoscope, since the device does not track the position of the eyes, and displays images somewhat less frequently than the typical occurrence of fixations in reading. Since the tachistoscope requires no eye movement, it leads to an emphasis on stationary visual input when used for reading research, producing such different conditions from normal reading as to raise serious questions of ecological validity. In spite of this, many substantial but now doubtful claims have been made about the reading process on the basis of tachistoscopic studies. In the larger historical context, however, the development and use of this instrument was an important part of the growing realisation among nineteenth-century researchers that technological assistance is necessary for the systematic study of reading.

3.2.2 The second period: early tracking technology

The first devices for tracking eye movement took two main forms: those which relied on a mechanical connection between participant and recording instrument, and those in which light or some other form of electromagnetic energy was directed at the participant’s eyes and its reflection measured and recorded. Lamare (1893) was the first to use a mechanical connection, by placing a blunt needle on the participant’s upper eyelid. The needle picked up the sound produced by each saccade and transmitted it as a faint clicking to the experimenter’s ear through an amplifying membrane and a rubber tube. The rationale behind this device was that saccades are easier to perceive and register aurally than visually. Soon after, Delabarre (1898) invented a system of recording eye movement directly onto a rotating drum by means of a stylus with a direct mechanical connection to the cornea. Various other mechanical devices involving physical contact with the surface of the participant’s
eyes were developed and used from the end of the nineteenth century until the late 1920s: these included such items as rubber balloons and eye caps.

Mechanical systems suffered two serious disadvantages: questionable accuracy due to slippage of the physical connection, and the considerable discomfort caused to participants by the direct mechanical connection. Not only was it difficult to find people who would agree to be participants under such conditions, there were significant problems of ecological validity as well, since participants’ experience of reading in trials was so different from their normal experience of reading that their behaviour was likely to be affected significantly and unpredictably. Despite these drawbacks, mechanical devices were used in eye movement research well into the twentieth century.

Attempts were soon made to develop less intrusive and less ecologically problematic techniques of tracking the eyes. One solution was to use electromagnetic energy rather than a mechanical connection to measure eye movement. This initially took the form of a photographic recording technique developed by Dodge and Cline (1901), which came to be known as the Dodge technique. In this arrangement a beam of light was directed at the cornea. The reflection was focused by means of a system of lenses and then recorded on a moveable photographic plate. Erdmann & Dodge (1905, 1907) used data generated by this technique to claim that there is little or no perception during saccades, a finding that was later confirmed by Uttal & Smith (1968) using more sophisticated equipment.

The photographic plate in the Dodge technique was soon replaced with a film camera (Judd 1905, 1907). The filming of eye movement represented a significant advance but was still plagued by problems with accuracy, due to the difficulty of keeping all parts of the equipment perfectly aligned throughout a trial and accurately compensating for the distortion caused by the diffractive qualities of photographic lenses. In addition, it was usually necessary to restrain the participant’s head by using an uncomfortable bite-bar or head-clamp.

Schott (1922) pioneered a further advance called ELECTRO-OCCULOGRAPHY (EOG), a method of recording the electrical potential between the cornea and the retina. EOG requires incisions to be made on the participant’s face with the risk of minor cutaneous scarring. These incisions allow a recording to be made of the natural electrical patterns associated with oculomotor muscle movement. EOG delivered considerable improvements in accuracy and reliability which explain its continued
use for many decades. Many researchers, however, persisted with photographic techniques because of the difficulty of persuading participants to endure such an invasive technique. This invasiveness is also the reason that EOG has been largely displaced by infrared techniques in the third period.

During the 1920s, 30s and 40s, knowledge of the process of reading language progressed rather more slowly in comparison with the intensity of research in the first few decades of the second period. Venezky (1977:343) claimed that a 'near stagnation in the field for at least 30 years' was due to three factors: the domination of behaviourism in psychology, a preoccupation with assessment in educational psychology, and the difficulty of obtaining funding for basic research in the US from the time of the Great Depression until the 1950s. As a result, he concluded, very little basic or theory-building research was undertaken. The third period, stretching from the middle of the twentieth century onwards, has by contrast seen a proliferation of studies into eye movement in language reading.

3.2.3 The third period: cognitive psychology, infrared tracking and computer technology

The initial defining feature of this period was the rise of cognitive psychology during the 1950s and 1960s with its rich and productive theoretical framework for studying reading as a specific example of cognitive processing. In addition there has been an improvement in tracking technology during the third period with the development of comparatively non-invasive infrared tracking. This involves the shining of an invisible infrared light onto the limbus, the boundary between the iris and the surrounding sclera (the white of the eye). The movement of the limbus, and thus of the eyeball itself, is tracked by picking up the reflection of this light. Although accuracy and stabilisation are still typical problems with this system, it is by far the most satisfactory method to have been developed and has become the almost universal eye tracking technique in the late twentieth century. The experiment in Stage I of this study uses such a tracking method.

The most far-reaching change in the third period, however, has been the introduction of computer technology during the 1970s to the tracking and recording of eye movement data, which otherwise require cumbersome and time-consuming manual processing. By the late twentieth century, eye movement tracking systems were typically interfaced with a computer to provide continuous online recording and processing of eye movement data. Accordingly there has been a significant increase
in accuracy, expansion of empirical scope, and saving in labour. The most egregious example of the disadvantage suffered by not using computer technology in a study of eye movement in music reading was Schmidt (1981), discussed later in this chapter. Regrettably it was necessary to conduct the trials in Stage I of this study without computer assistance, although the procedure for manual registration was considerably easier to manage than Schmidt's.

A significant application of computer technology to eye movement study during this period has been the gaze-contingency paradigm, otherwise known as the moving-window technique. First developed by Reder (1973) and McConkie & Rayner (1975), the technique has enabled researchers to control stimulus displays as a function of eye location. The principle of the technique is quite simple. A computer is interfaced with both an eye-movement tracking system and a display of the text. As participants read the display their eye movement is tracked, and the computer almost instantaneously modifies the display according to where the participant is fixating at any particular moment. Figure 18 shows adapted examples of McConkie & Rayner's (1975) use of gaze-contingency displays. The window of readable text is 17 character-spaces (eight to the left and eight to the right of the fixated letter, that is, 8+1+8). All other text is masked by rows of 'Xs'. The upper window shows part of the text as it would have looked if fully displayed. The middle window shows the same text as it appeared when the reader was fixating on the letter 'r' in 'inorganic'. In the lower window, the same reader has moved on to fixate on the second 'i' in the same word, and the masking in the display has been promptly adjusted.

By far the single most abundant substance in the biosphere is the familiar but unusual inorganic compound called water. In

| XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX |
| XXXXXXXXsual inorganic compouXXXXXXX |

| XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX |
| XXXXXXXXinorganic compouXXXXXXX |

Figure 18 Example of the early use of gaze-contingency displays, adapted from McConkie & Rayner (1975)
The technique has given researchers into eye movement in language reading the ability to observe the processing of visual input in much greater detail, particularly its temporal characteristics, the perceptual span, and the nature of central versus peripheral processing in reading. Regrettably it appears that no-one apart from Truitt et al (1997), discussed in Section 3.3.4, has used the moving-window technique to observe eye movement in music reading. The idea of linking eye position to a changeable display is conceptually crucial to the devising of SightReader, in which the movement of the hands rather than the eyes is linked to the display through an electric keyboard, as explained in Sections 1.3 and 5.2.1.

During the third period researchers have pursued several fundamental questions concerning eye movement in language reading, four of which are of particular relevance to both Stages I and II of this study: (1) Is the control of eye movement in language reading based on the BOTTOM-UP processing, that is, driven by low-level graphical features in the stimulus, or on TOP-DOWN strategies predetermined by the reading apparatus? (2) What is the role of peripheral input within the span, and is the information from each fixation processed only within the duration of that fixation or is such processing integrated across fixations? (3) Why do readers REFIXATE, thereby inspecting information more than once during a single reading? and (4) What is the nature and role of the eye-voice/eye-hand span? These four issues are now discussed in turn.

3.2.4 Top-down and bottom-up processing

The psychological terms 'top-down' and 'bottom-up' have a number of different meanings depending on the context. In a strict experimental sense, for example, they mean 'concept-driven' and 'data-driven' respectively. Data-driven processes gather process information in small pieces, which are later assembled in working memory. Concept-driven processes can be likened to expectations or plans. When used in relation to the reading process, the memory system is assumed to be at the 'top' of the process, and the text at the 'bottom'. Top-down models for eye guidance in reading, for example those put forward in Smith (1971) and Rayner (1975), postulated that eye movement is controlled 'downwards' by the highest levels of the processing system, which construct textual meaning and are dependent only marginally on the visual details of the text. In Rayner's (1975) 'constant-pattern' model, one of three he proposed, the eyes move under an automatised oculomotor-command plan. This plan assumes that saccades are of roughly equal length and fixations of roughly equal duration irrespective of the size and meaning of the words.
being picked up. Such an attempt to impose theoretical order on eye movement harks back to Javal’s (1888, 1889) and Huey’s (1908) assumptions that eye movement behaviour is simple and regular in patterning (see Footnote 13). Rayner conceded in his (1978) literature review that there is too much variability in observed saccade lengths and fixation durations for the constant-pattern model to be taken seriously.

Bottom-up models were proposed by Gough (1972) among others. These models imply that eye movement patterns are largely ‘upwards’ driven by visual information from the text and are determined with minimal influence from the memory system. Rayner (1975) proposed a bottom-up ‘stimulus-control’ model, according to which eye guidance is mediated almost entirely by what the eyes see from moment to moment and owes little if anything to high-level processing. It is, however, intuitively difficult to accept the idea that a top-down strategy plays little or no part in the reading process, if only because the meaning of the textual symbols must be globally understood to make sense to the reader.

Rayner (1975) proposed a third model which is something of a compromise between his first two models, and which incidentally bears some conceptual resemblance to Norman & Shallice’s (1982) model for automaticity as explained in Section 2.1. In Rayner’s model, eye movement is under preset (top-down) control but with a gain control that can be adjusted from moment to moment to account for fluctuations in textual difficulty. This model has turned out to substantially oversimplify how the eyes move in language reading, but it embodied a more flexible view of how eye movement might be controlled. Indeed, key aspects of the original top-down and bottom-up models have also been conflated in more recent models, for example Rumelhart (1977), Just & Carpenter (1980), Rayner & Pollatsek (1989), and Kinsler & Carpenter (1995). Eye guidance in reading is now widely accepted to be under both top-down and bottom-up influence on the basis of evidence, presented below, that visual input and pre-existing information held in the memory system interact in the reading process.

Over the last three decades much detailed evidence has been gathered of the ways in which eye movement is influenced from moment to moment by bottom-up, localised features of linguistic text. Some of this influence can be accounted for by purely graphical features. For example, the eyes are less likely to fixate on blank areas in the text, including the regions between sentences, than on a word (Rayner 1975; Abrams & Zuber 1972). When fixations on blank areas do occur they tend to be shorter
Eye movement in the reading of language and music

(Rayner 1975), presumably because there is less information than normal to process. The eyes are most likely to fixate not on the centre of a word, but slightly to the left of centre (O’Regan 1981). This is almost certainly because the eyes pick up more information to the right of the central area, in the direction of travel, than back to the left (McConkie & Rayner 1976). Saccades naturally enough tend to be longer moving onto and off a larger word (O’Regan 1979).

At the same time an impressive array of evidence has accumulated that control of eye movement also involves involving higher level processing. Eye guidance appears to be influenced by the cost of accessing the meaning of the text in long-term memory. For example, the fixations of readers who have a poor technical vocabulary tend to be longer on low-frequency words (Just & Carpenter 1980; Inhoff & Rayner 1986; Rayner & Duffy 1986) and on technical words (Kolers 1976), presumably because representations of those words are not highly activated in long-term memory. The eyes pause for longer on misspelt words (Zola 1980), which presumably disrupt normal access to information stored in the long-term memory, and on words that do not make grammatical sense and thus disrupt the normal processing of the clause (Frazier & Rayner 1982). The same is true of grammatical elements that are functional but simply more difficult to process (Wanat 1971; Rayner 1977). Fixations are also longer and saccades shorter in areas of the text that the reader judges to be more important, reflecting a slowing of the rate of reading to compensate for what would otherwise be an increase in processing load.

Significantly for this study, the concept behind SightReader of pushing the eyes further ahead of the hands than they would be otherwise be involves both bottom-up and top-down control of the eye-hand span (and therefore some aspects of eye movement). A player can consciously try to look ahead further than normal, a top-down influence, or can be forced to look ahead further by the disappearing display on SightReader, a bottom-up influence. In addition to SightReader’s bottom-up influence on the reader, it appears logical that the reader must also employ a global top-down strategy to accommodate the disappearing display. At its simplest, such a strategy might be based on the attitude ‘This is going to be very different from my normal reading experience’ or ‘I’ll need to look ahead to cope with the task’.

3.2.5 Peripheral input and integration across fixations

Rayner, Carlson & Frazier (1982) used the moving-window technique to demonstrate that the integration of processing across fixations is a powerful factor in
reading. They showed that when peripheral information from the first word to the right of a fixation is withheld, especially the first two or three letters, the reading rate suffers significantly. Direct evidence against the notion that information is processed fixation by fixation has also come from Pollatsek & Rayner (1990), who claimed that the processing of information can be and almost certainly is typically integrated across fixations. In other words, the processing of information picked up in one fixation is not ‘compartmentalised’ within the duration of that fixation, but is carried over into subsequent fixation durations. They used the logical argument that since the average saccade length in reading is seven to nine character spaces and the perceptual span extends far more to the right of the fixation, the perceptual spans in successive fixations must overlap.

Still more evidence supporting the integration of processing across fixations comes from knowledge of the timing constraints of crucial events in respect of fixations, summarised by Pollatsek & Rayner (1990:148). It appears that the time taken to gain lexical access after the beginning of a fixation, together with the time needed to implement the next saccade, severely limits the possibility of immediate and exclusive control of eye movement within each fixation. Many of the observed effects of low-level textual patterns on eye movement could be described as being ‘smeared’ over a number of consecutive fixations. For example, a difficult word to process appears to be associated with a lengthening of the duration of not only the fixation on that word itself, but of fixations on subsequent words (Pollatsek & Rayner 1990:148).

The perceptual span is widely understood to be the area of text from which useful information is extracted from a fixation. Since visual input comprises distinct central and peripheral areas, the perceptual span potentially comprises both types of input. This has not always been made explicit by researchers in this field. Some have appeared to assume that the eye-hand and eye-voice spans include peripheral input (for example Sloboda 1974, 1977) in which case they are virtually equivalent to the concept of perceptual span. Others have assumed that eye-hand and eye-voice spans include only central input. In this study, the eye-hand span is defined solely in terms of central input, measured from the hands to the point of fixation rather than to the point of furthest peripheral input.

McConkie & Rayner’s (1975) moving-window technique established the extent of the peripheral span and showed that it is of considerable importance in the reading process. They did so by restricting participants’ ability to pick up information up to
only eight character spaces to the right of the fixated letter (illustrated in Figure 17), and finding that the silent reading rate is significantly slowed under this condition. Indeed it was necessary to extend the window to about 15 characters to the right of the fixated letter to avoid a slowing of the reading rate, suggesting that readers extract and use peripheral information far to the right of their central fixation area. This study showed that since the central area is normally about three or four character spaces in diameter, peripheral input is essential for fluent reading.

The role of peripheral input in language reading has also been explored in other moving-window studies. Fixation durations tend to be longer and saccade lengths shorter when letters in the peripheral area are erroneous or are replaced by a grating (McConkie & Rayner 1976; Rayner 1975; Rayner & Bertera 1979; O'Regan 1980; Rayner & Pollatsek 1981; McConkie & Underwood 1981). This suggests that fluent reading requires the reader to 'preprocess' peripherally, that is, to carry out some of the processing of the text before actually fixating on it. Accordingly, the last fixation in a line before a return sweep to the start of the next line is likely to be shorter than normal because there is no information to preprocess peripherally (Rayner 1977). Furthermore, fixation duration appears to be influenced not only by the length and frequency of words which are perceived by central vision, but also by those which are perceived peripherally (Kliegl, Olson, & Davidson 1983).

It has also been established that peripheral information is important in guiding the eyes to the next location. For example, if a longer word lies to the right of the word on which the eyes are fixating, the reader appears to perceive this fact peripherally and accordingly programs a longer saccade to reach it (O'Regan 1979). When there are two or more fixations on a word, sometimes known as 'refixation', they are each likely to be shorter than a single inspection on that word would have been. This is because the fixations presumably overlap and thus some of the information can be picked up and processed over more than one fixation (Kliegl, Olson & Davidson 1983). Refixation is discussed in greater detail in Sections 3.2.6, 3.3.4 and 4.3.1.

Another significant aspect of peripheral input is the apparent ease with which frequently occurring words are perceived and processed peripherally compared with less frequently occurring words. O'Regan (1979) found that the eyes are less likely to fixate on the frequently occurring, grammatical words such as 'the' and 'to' than on less frequently occurring, less predictable lexical words. Words that are highly predictable from their context are more likely to be skipped over altogether, and if fixated on, are inspected for a shorter duration than would otherwise be the case.
Referred to as the O'REGAN EFFECT in this study, a case will be made in Chapter 4 that the musical equivalent of this phenomenon is of considerable importance in the music reading process. A schematic example of the O'Regan effect in language reading is illustrated in Figure 19.

![Figure 19](image-url) A schematic example of the O'Regan effect in language reading

3.2.6 Refixation

The issue of why readers refixate, that is, why they inspect information on the page more than once during a single reading, is as yet only partly resolved. Kennedy (1992:385) posed a central question about refixation: 'Why, in relatively short sentences that can be readily understood if spoken [without word-repetition], should the reader look at the same word twice?' As previously pointed out, this matter is relevant to Stage II because SightReader's process of manipulating the display largely prevents readers from looking back at previous chords within the eye-hand span. Referring to left-to-right writing systems, Pollatsek & Rayner (1990:153) distinguished two types of refixation on words, both ubiquitous in language reading: SAME-WORD RIGHTWARD REFIXATION and LEFTWARD REFIXATION (sometimes referred to as 'regression'). This is exemplified in Figures 20-22, in which successive fixations are represented by crosses and are numbered in sequence. Figure 20 shows a situation in which there is no refixation at all; Figure 21 shows a same-word rightward refixation, achieved by saccading from one point in a word a small distance rightward to another point in the same word. This is likely to involve the overlapping of both central and peripheral areas of two successive fixations. Figure 22 shows a leftward refixation back to a previous word that has already been fixated on.

![Figure 20](image-url) No refixation

![Figure 21](image-url) Same-word rightward refixation

![Figure 22](image-url) Leftward refixation to a previous word
It has been widely observed that the probability of same-word rightward refixation (Figure 20) increases markedly with each additional syllable in a word (for example Kliegl, Olson & Davidson 1983). Furthermore, Inhoff & Rayner (1986) among others, have shown that this probability is inversely proportional to the frequency with which that word occurs in the language. Infrequently occurring multisyllabic words, then, are more likely to be inspected twice or more than those that occur frequently. In such cases, the first fixation on that word typically yields enough useful information from the rest of the word to gain lexical access in the memory system and thus to avoid a second fixation. In other words, the reader’s long-term memory of the probability of what the upcoming information is combined with blurred peripheral input of that information.

While same-word rightward refixation appears to be ubiquitous in language reading (for example Kliegl, Olson & Davidson 1983), there is evidence that leftward refixation in reading language occurs less frequently, particularly in skilled reading. Rayner (1978) found that leftward refixation correlates strongly with participants’ age (and therefore reading experience), with a smooth reduction of the proportion of saccades that were leftward from first grade children (52% of all saccades) to college students (15% of all saccades). Buswell (1922) had found a similar trend, with first grade children making an average of 4.0 leftward refixations per line of text, and college students 0.5 leftward refixations.

One reason for the lower frequency of leftward refixations in language reading may be that they require a much greater investment of reading resources. A leftward refixation must be followed by another rightward saccade just to return to the previous location, whereas a rightward refixation on the same word carries no such extra temporal and spatial burdens. Rayner (1978) proposed that leftward refixation results from three situations: either (1) the reader lands too far to the right of the optimum fixation point in a new word and refixates on a previous part of the same word (a proposition of as yet unresolved validity); (2) the reader has difficulty understanding the text; or (3) the reader misinterprets the text. The last two propositions are supported by Carpenter & Just (1977) and subsequent studies, which show that a refixation on a prior context sentence is more likely after encountering an anaphoric pronoun. Further support for the second and third of Rayner's propositions was provided by Shebilske & Fisher's (1981) finding that leftward refixations are more likely to occur where the text becomes more difficult.

---

14 An anaphoric pronoun refers back to some preceding item in the text, usually in the same paragraph. The context sentence, then, contains the nominal group to which the pronoun refers.
and/or more important to the reader, and Rayner's (1983) finding, supported by many subsequent studies, that lexicogrammatical ambiguities in the text encourage leftward regressions for the purpose of disambiguation.

Thus leftward refixation in language reading appears to be partly functional and partly dysfunctional, the dysfunction occurring either in the text or the reading process. Kennedy (1992:385) put forward two reasons for a functional role. The first is that the reader needs to account for alternative interpretations at a lexical level which cannot be derived from the existing trace in working memory; this would include the encountering of anaphoric pronouns that are sufficiently distant from their reference, as discussed above. The second is based on a model developed by Johnson-Laird (1983) and Hegarty (1992): leftward refixation helps readers to refresh the mental model of the text they have just read. In support of her reasoning, Kennedy draws on evidence that readers remember the physical location of key lexical items in the text (Christie & Just 1976; Zechmeister & McKillop 1972). Re-experiencing the location of a word may be even more important to the reading process than the duplicated visual input of the word itself that arises from refixation (Baccino, Pynte & Kennedy 1990; Kennedy 83). Accordingly, using the moving-window technique to prevent refixation disrupts the reader's ability to assign words to distinct spatial locations and degrades performance, provided the text is sufficiently complex (Kennedy 1992; Kennedy & Murray 1984; Pynte, Kennedy, Murray, & Courrieu 1988).

Here it is necessary to make a distinction between external refreshment, which arises from the physical (oculomotor) reinspection of information on the page, and internal refreshment, which refers to the maintenance of information in working memory without oculomotor refixation on that information, that is, solely through conscious effort within the memory system. Kennedy claims that data from the studies cited above suggest that locational information (through physical refixation and therefore external refreshment) assists comprehension and lowers the cost of maintaining information in working memory. The relevance to this study of discussing internal and external refreshment is that SightReader's process of blanking out successive chords in the display inhibits physical leftward refixation within the eye-hand span, thus forcing readers to rely solely on internal refreshment of a larger amount of material than they would otherwise carry in their working memory. The effect may be to develop readers' capacity for internal refreshment, possibly improving it. The data from Stage I of this study will also be used to show that

---

1 Refreshment is typically referred to as 'rehearsal' or 'recirculation' in the literature.
leftward and vertical refixation plays an important functional role in music reading, that of managing the competing demands of tempo and processing load (see Section 4.3).

3.2.7 The eye-voice/eye-hand span

In the context of eye movement in language reading, the eye-voice span is the distance between perceiving an item of text and reading it aloud, measured either in seconds or words. The perception of text at the 'front' end of the span can be defined in terms of either the foveal or peripheral input of the current fixation, although most research into the span fails to explicitly declare which definition it uses. Information within the span is assumed to be held in working memory, although it does not comprise the whole of working memory. The span was briefly treated in Section 1.3 in relation to SightReader and in Chapter 2 in relation to working memory. The concept of the span, and some of the findings in the literature, were central to the development of SightReader.

In the second period of eye movement research a small amount of work was carried out on the (foveal) eye-voice span. Quantz (1898) was the first to publish in this area:

In reading aloud .... words must be perceived some distance in advance of those which the voice is uttering. The rapid reader has the greatest interval between eye and voice. (p436)

Other researchers also suggested that there is a relationship between reading ability and the size of the eye-voice span. Among them were Buswell (1920), Judd & Buswell (1922), Tinker (1958) and Morton (1964). It was also found that the eye-voice span is shorter for material that requires greater concentration, such as lists of unrelated words as opposed to cohesive sentences (Buswell 1920; Lawson 1961; Morton 1964). These findings raise a key question in relation to the training of music sight reading, which will be discussed further in Chapter 5: Do skilled music readers maintain a larger eye-hand span than the unskilled?

The (peripheral) eye-voice span in the reading aloud of language was investigated during the third period by Levin & Kaplin (1970) and Levin & Addis (1980). They used a novel ‘light-out’ technique that bypasses visual input and oculomotor commands to rely solely on auditory output to generate data. Their participants were required to read aloud a paragraph of text. At a point of time unknown to
participants, the light was turned off and participants were asked to report all the words they had seen beyond the word being spoken when the light was turned off. The eye-voice span was defined as the number of words to the right of participants’ last fixation that they were able to report correctly, and thus included both foveal and peripheral input. Levin & Kaplan (1970) used this clever technique to demonstrate that the span has a significant tendency to extend to a phrase boundary, even when such a boundary falls outside the average span of a participant.

In a study on the eye-hand span in copy-typing, Shaffer (1976) varied the amount of text copy-typists could preview to the right as they typed, from as little as a single character to a whole line of text. Shaffer found that typists’ performance is degraded if preview is restricted to fewer than eight characters to the right of the character they are typing; typing speed is reduced by about 80% when only one character to the right is visible. It also appears from this study that typists need to preview one word ahead in order to plan their continuous sequence of keystrokes, but gain minimal extra benefit from previewing more than this. Shaffer (1976), Levin & Kaplin (1970) and Levin & Addis (1980) were the starting point for Sloboda’s (1974, 1977) work on the bottom-up influence of musical phrase-structure on eye movement, discussed further in Section 3.3.4.

3.2.8 Conclusion

Where, then, does research into the language reading process stand at the end of the twentieth century? In little more than a century scientists have progressed from relying solely on introspection and unassisted observation, through the introduction and gradual improvement of tracking technology, to a situation where modern scientific methodology has been brought to bear on eye movement research, allowing greater accuracy and ecological validity. Thus far, some aspects of how the reading apparatus behaves in many specific experimental situations have been investigated. These aspects are referred to by Rayner, Flores d’Arcais & Balota (1990:634) as ‘subcomponent processes’. But unfortunately such knowledge of specific eye movement behaviour has not yet led to a unified model of the reading process. In the words of Rayner et al (1990:634):

The global models that currently exist of the reading process .... all have weaknesses associated with them. Clearly, in order to develop an accurate model of reading, each of the subcomponent processes must be fully understood. .... Our best guess is that there will not be a major breakthrough in understanding reading that results from the development of a global, all-encompassing model of the reading process. Rather, continued progress will
come as a result of further research on subcomponent processes and refinement of models of such processes.

Despite its failure to arrive at a definitive model of reading, the literature on language reading contains significant findings that are relevant to both empirical stages of this study. Three matters are crucial to the conceptual development of SightReader: the concept of bottom-up and top-down control of eye movement; the technique of constraining eye position by manipulating the display, known as the gaze-contingency paradigm (Reder 1973; McConkie & Rayner 1975); and the findings that the eye-voice span in language reading is larger in skilled than unskilled reading (Quanz 1898; Buswell 1920; Judd & Buswell 1922; Tinker 1958; Morton 1964). In addition, O'Regan (1979) and studies by Kennedy and others into refixation in language reading raise the question as to whether music reading also involves skipping over and refixating information on the score, and whether this is functional or dysfunctional behaviour. This will be investigated in Stage I in order to understand more about the possible effects of SightReader's constraints on refixation.

3.3 Review of the literature on eye movement in music reading

The study of eye movement in music reading is strongly interdisciplinary, and requires a combination of knowledge, skills and instincts from quite distinct areas. Insight into music pedagogy, a grasp of scientific methodology, and experience in managing musicians under controlled conditions are all important. Eye movement in music reading has been studied by researchers from a range of backgrounds. A sizeable minority of studies has appeared as dissertations in music or music education, for example York (1952), Young (1971), Schmidt (1981), Goolsby (1987) Smith (1988), and Chang (1993). These studies reflect a curiosity among musicians about a central process in their art, and a hope, stated or unstated but in all cases apparently unrealised, that investigating eye movement might help in the development of more effective methods of training musicians' sight reading skills. To these studies can be added at least 10 by researchers whose primary field of expertise was psychology: Jacobsen (1941), Weaver (1943), Weaver & Van Nuys (1943), Halverson (1974), Sloboda (1985), Kinsler & Carpenter (1995), Truitt et al (1997), Polanka (1995), Rayner & Pollatsek (1997), Waters & Underwood (1998). One study was conducted by an ophthalmologist, Lang (1961). These studies reflect the value of music reading to scientists as an example of a complex human skill that provides rich opportunities for measurement.
Music reading may at first appear to be similar to language reading, since in both activities the eyes move over the page in fixations and saccades, picking up and processing coded meanings. It is here, however, that the obvious similarities end. Not only is music's coding system non-linguistic, thus presumably requiring very different processing, it involves what is apparently a unique combination of features among all human activities: a strict and continuous time constraint on its output, which is generated by a continuous stream of instructions. Even reading language aloud, which like reading music involves a musculoskeletal response, is comparatively free of temporal constraint. The pulse in reading aloud is a fluid, improvised affair compared to its rigid presence in most western music. As will become clear during this chapter, the strict timing requirement of musical performance is one of the main reasons that observing eye movement in music reading is fraught with more difficulties and pitfalls than observing eye movement in language reading.

Another crucial difference between reading music and reading language is the role of skill. Most people become reasonably efficient at language reading by adulthood, even though almost all language reading, as Sloboda (1985:7) points out, is sight reading. By contrast, some musicians regard themselves as poor sight readers of music even after years of study that includes sight reading training. The improvement of music sight reading and the differences between skilled and unskilled readers, then, have always been of prime importance to research into eye movement in music reading, whereas research into eye movement in language reading has on the whole been concerned with attempting to develop a unified psychological model of the reading process (see Rayner et al 1990:634 as quoted in the previous section).

It is therefore not surprising that most research into eye movement in music reading should have aimed to compare the eye movement patterns of the skilled and the unskilled. It will be argued in Section 3.3.2, however, that using skill as a variable in observing eye movement in music reading has presented serious and largely unacknowledged methodological hurdles in most studies, arising from what will be called here the \textit{TEempo/Skill/Action-Slip Fallacy}. The last section in this chapter is devoted to a discussion of the significant issues raised in the literature on eye movement in music reading that are relevant to this study.
3.3.1 Equipment and related methodology

It was not until several decades after researchers began to apply mechanical and photographic techniques to eye movement research that there were documented attempts to study eye movement in music reading. The five earliest studies (Jacobsen 1941; Weaver 1943; Weaver & Nuys 1943; York 1951; and Lang 1961) used photographic techniques, with all their attendant problems as described later in this section. The devices they employed trained either a continuous beam of visible light onto the eye to produce an unbroken line on photographic paper, or a flashing light to produce a series of white spots on photographic paper at sampling intervals of around 25ms (that is, 40 samples per second). Because the film rolled through the device vertically, the ubiquitous vertical movement of the eyes, even in single-line music reading, went unrecorded in Jacobsen (1941), York (1951) and Lang (1961). In Weaver (1943), Weaver & Van Nuys (1943) and York (1951) the vertical dimension was recorded separately with a second camera. Both film records were subsequently combined to give a two-dimensional picture of eye movement. In practice this arrangement was clumsy to operate and substantially reduced the accuracy of the data.

From the start there were yet other basic equipment problems. All of these systems were sensitive to small movements of the head or body that appear to have seriously contaminated the data in some of the studies. Jacobsen (1941) used several set-ups during the course of his trials, involving variously a bite-plate, a wooden paddle fitted tightly against the head, and no stabilisation at all. These alterations represented a significant methodological inconsistency, resulting in the ruling out of a substantial proportion of the data. Weaver (1943) and Van Nuys & Weaver (1943) were more successful in their use of a bite-board as a head stabiliser throughout their trials.

Difficulties with head stabilisation continued to plague research, in some cases prompting extraordinarily elaborate set-ups. Young (1971), for example, mounted a weighty camera on a motorcycle helmet with a periscope attached to monitor the corneal reflection in each participant's left eye. To reduce the 2.7kg weight of this apparatus, the helmet was supported by a system of counterbalancing weights and pulleys attached to the ceiling. A bite-bar was also used for head stabilisation, although Smith (1988:54) argues that despite this, the head movements of Young's participants seriously contaminated all of her data. Halverson (1974) used a more practical arrangement with a helmet in fixed position. Sources and sensors of infrared light were installed in goggles worn by the participant, with a sampling interval of
25ms. This study was the first to successfully use a computer for recording and processing the data from eye movement observation. Schmidt (1981) did not even attempt head stabilisation because his participants played wind instruments and thus could use neither bite-bar nor helmet. For this reason his spatial data were rendered inaccurate and only temporal data were used, indicating when rather than where each fixation occurred.

In Goolsby (1987) and Smith (1988), participants' heads appear to have been reasonably well stabilised by headrest and bite-plate. Polanka (1995) used an ASL 200 tracker to measure eye movement during the hummed readings of melodies. In anticipation of problems with head movement, participants read each melody silently before humming the same melody during a second reading. Despite the use of head restraint, the eye movement records produced from his participants' hummed readings of melodies were indeed marred by head movements which 'altered the calibration of the eye movement monitor and rendered much of that data unusable' so that 'only the silent reading data were analysed' (p180). As pointed out under the discussion of Lang (1961) on p78, data from silent reading is meaningless in most empirical contexts, since there is no proof of what has been perceived and processed. Chan (1996) used an ASL 3100H tracker mounted on an unfixed helmet, thus avoiding the need for stabilisation in the first place, a similar arrangement as for Stage I of this study. Head stabilisation was a problem as recently as Waters & Underwood (1998:48), whose participants 'tended to move their head a little between presentation of the two stimuli, which places some doubt on the accuracy and validity of the spatial data ....'.

In addition to extraneous head movement, research into eye movement in music reading has faced other physical, bodily problems. The musculoskeletal response required to play a musical instrument involves substantial body movement, usually by the hands, arms and torso. Such body movement is likely to be less pronounced in skilled than unskilled participants, since it can be assumed that skill is partly a matter of paring down body movement to the absolute minimum necessary to produce the desired response on a musical instrument. Extraneous body movement of the unskilled is therefore more likely to upset the delicate balance of tracking equipment, confound the registration of data, and add to the inherent uncertainty in comparing the eye movement to that of the skilled. For this reason several researchers chose to observe eye movement during singing (Halverson 1974) and humming (Goolsby 1987; Polanka 1995) only to move the source of contamination...
from extraneous torso and limb movement to interference from functional movement from the neck up.

Another problem that affects almost all unskilled keyboardists and a considerable proportion of otherwise skilled keyboardists is the common tendency to frequently glance down at the hands and back to the score during performance. The disadvantage of this behaviour is that it causes signal dropout in the data every time it occurs, sometimes up to several times per bar on top of the unavoidable dropout from participants' blinking. Weaver (1943), York (1951) and Young (1971) reported that many participants frequently disrupted the signal by looking down at the keyboard. More recently Rayner & Pollatsek (1997:49) wrote that:

> even skilled musicians naturally look at their hands at times [and that since] accurate eye movement recording generally does not allow these head movements .... musicians often need appreciable training with the apparatus before their eye movements can be measured.

In the light of such problems, researchers into eye movement in language reading may be regarded as fortunate that signal dropout is limited to participants' blinking, and does not involve participants' looking at their hands. The problem of dropout was anticipated in Stage I of this thesis and minimised by the careful selection of participants (see Section 4.1.4).

There were also basic problems with distortion caused by the lens used to focus and direct the light. In Weaver (1943) and Weaver & Van Nuys (1943), the registration of data in relation to the music score was distorted towards the edges of the page because of refraction in the prismatic lens used in the equipment set-up. This was a similar problem to the lens distortion encountered on the NAC tracker in Stage I of this study, where the problem was minimised by the use of an elaborate protocol for data registration.

Since Lang (1961), all reported studies into eye movement in music reading, aside from Smith (1988) appear to have taken advantage of the infrared tracking technology that characterises the third period of eye movement research. Young (1971) used a Westgate Eye Movement Camera EMC-2 to record her participants' eye movement, superimposed on an image of the score but at a disappointingly large sampling interval of 125ms when compared with the typical fixation duration of 250-400ms. A further problem was that the 100ms sampling interval of the oscillograph she attached to the keyboard to record the timing of the performances reduced the effective sampling interval for the eye-hand span to the lowest common
factor of both intervals, an undesirably large 500ms. Given these problems—'a step backwards' according to Goolsby (1989:114)—it is not surprising that much of Young's data had to be disregarded. Schmidt (1981) used a computer for recording the data at admirably small intervals of 16.6ms. But for want of a program to process the raw data into a useable format, an extremely laborious manual registration was necessary. This led to his decision to process data from only six participants, too small a sample to result in statistically significant data.

Goolsby's (1987) use of the newly developed Gulf Western infrared eye tracker with its sampling interval of 1ms represented a significant improvement on previous equipment. Disappointingly, like most others before it, the system failed to superimpose records of eye movement on the score. Smith (1988) chose to use the previous generation of eye-tracking technology, EOG, with reasonable accuracy and a sampling interval of 1ms, but still without a display of eye movement in relation to the score.

Research into eye movement in music reading has by and large been carried out using less than optimal equipment. This has had a pervasive negative impact on almost all research up until a few recent studies conducted by well-placed researchers. Kinsler & Carpenter (1995), for example, were able to identify eye position to within 0.25 degrees, that is, the size of the individual musical notes, at 1ms intervals. Truitt et al (1997) used a similarly accurate infrared system capable of displaying a moving window and integrated into a computer-monitored musical keyboard. Waters & Underwood (1998) used a machine with accuracy of plus or minus one character space and a sampling interval of 4ms.

In summary, the four main equipment problems in studies into eye movement in music reading have been that tracking devices (1) measured eye movement inaccurately or provided insufficient data, (2) were uncomfortable for participants and therefore risked a reduction in ecological validity, (3) did not allow for the display records of eye movement in relation to the musical score, or at least made it extremely difficult to do so, and (4) were adversely affected by most participants' tendency to look down at their hands and to move their bodies significantly during performance. Not until recently has eye movement in music reading been investigated with more satisfactory equipment. Suboptimal equipment has been the norm because most investigations have not been located in institutions devoted to eye movement research with access to the best available equipment of the day, or have been conducted by researchers who were unlikely to have attracted the
substantial funding necessary to purchase such equipment. Many studies, including Stage I of this study, have thus not been in a position to take advantage of computer technology.

3.3.2 The tempo/skill/action-slip fallacy

Of all the methodological issues in observing eye movement in music reading, the most pervasive is the inextricable relationship between tempo, skill and the level of action-slips. At the centre of this vexed issue is what is referred to here as the tempo/skill/action-slip fallacy. The ramification of the fallacy is that it is usually not possible to reliably compare the eye movement patterns of skilled and unskilled readers under the same conditions. At least for investigators, it is an unfortunate and inescapable fact that skilled musicians can perform at sight the same musical stimulus at a faster tempo than unskilled musicians. If the unskilled are forced to play at a tempo fast enough to suit the skilled, their performance will usually be marred by a significant level of action slips. Action-slips are defined here as unintended musculoskeletal responses that degrade the output. Action-slips are undesirable in the experimental context because it is impossible to measure and exclude their effect on the reading process. There will also be doubt as to whether all of the information on the score has been picked up and processed, since accurate output is the only proof of this.

Conversely, if the skilled read at a tempo slow enough to suit the unskilled, it is likely that there will be a significant amount of excess (unused) capacity in their music reading apparatus. This may result in what is referred to here as the wandering effect, an undisciplined or seemingly unfocused eye movement patterns that a number of studies have reported in the eye movement of their participants when they read at a slow tempo. Weaver (1943:15) implied the existence of the wandering effect and its confounding influence in admitting that some participants could ‘choose a [slow enough] tempo which not only permits but also requires more than the minimum number and/or duration of [fixations].’ Some writers, for example Truitt et al (1997:51), suspected that in such situations their participants were ‘hanging around rather than extracting information’. The wandering effect is undesirable because it represents an unknowable and possibly random distortion in normal eye movement patterns which necessarily reduces the scope of the interpretation.
substantial funding necessary to purchase such equipment. Many studies, including Stage I of this study, have thus not been in a position to take advantage of computer technology.

3.3.2 The tempo/skill/action-slip fallacy

Of all the methodological issues in observing eye movement in music reading, the most pervasive is the inextricable relationship between tempo, skill and the level of action-slips. At the centre of this vexed issue is what is referred to here as the tempo/skill/action-slip fallacy. The ramification of the fallacy is that it is usually not possible to reliably compare the eye movement patterns of skilled and unskilled readers under the same conditions. At least for investigators, it is an unfortunate and inescapable fact that skilled musicians can perform at sight the same musical stimulus at a faster tempo than unskilled musicians. If the unskilled are forced to play at a tempo fast enough to suit the skilled, their performance will usually be marred by a significant level of action slips. Action-slips are defined here as unintended musculoskeletal responses that degrade the output. Action-slips are undesirable in the experimental context because it is impossible to measure and exclude their effect on the reading process. There will also be doubt as to whether all of the information on the score has been picked up and processed, since accurate output is the only proof of this.

Conversely, if the skilled read at a tempo slow enough to suit the unskilled, it is likely that there will be a significant amount of excess (unused) capacity in their music reading apparatus. This may result in what is referred to here as the wandering effect, an undisciplined or seemingly unfocused eye movement patterns that a number of studies have reported in the eye movement of their participants when they read at a slow tempo. Weaver (1943:15) implied the existence of the wandering effect and its confounding influence in admitting that some participants could 'choose a [slow enough] tempo which not only permits but also requires more than the minimum number and/or duration of [fixations].' Some writers, for example Truitt et al (1997:51), suspected that in such situations their participants were 'hanging around rather than extracting information'. The wandering effect is undesirable because it represents an unknowable and possibly random distortion in normal eye movement patterns which necessarily reduces the scope of the interpretation.
Figure 23 illustrates the relationship between tempo, skill, and these two types of empirically undesirable musculoskeletal and oculomotor behaviour (action-slips and 'wandering'). Referring to the upper diagram, the horizontal line represents performance tempo for a particular player reading a particular musical excerpt. The two vertical lines represent the lower and upper boundaries of the tempo-band within which the player's reading apparatus functions neither under nor over its available capacity. This middle tempo-band is referred to here as the **peak-use area**. Tempos in the peak-use area therefore present minimal risk of contamination by either musculoskeletal action-slips or oculomotor wandering, and would therefore appear to be empirically desirable.

![Diagram showing tempo areas for a less skilled reader and a more skilled reader](image_url)

**Figure 23 (upper)** Tempo areas for a less skilled reader  
**Figure 23 (lower)** Tempo areas for a more skilled reader

The right-hand boundary of the peak-use area is the **action-slip threshold**, which represents the tempo at and above which there is an unacceptable level of action-slips. How an unacceptable level, and therefore the tempo represented by the action-slip threshold, is set depends on the objectives of the experiment; this is not important to this explanation, and will be revisited later in relation to Stages I and II. The further to the right of the peak-use area the player's tempo, the greater the deficiency of available capacity in the player's music reading apparatus and the greater the likelihood of action-slips. This range of tempos is called the **action-slip area**. The left-hand boundary is called the **undercapacity threshold**, the point at which eye movement begins to show signs of 'undisciplined' behaviour. The fact
that, in practical terms, the undercapacity threshold may be difficult to locate does not diminish its role in this theoretical explanation. The further to the left of this threshold the player's tempo, the greater the portion of available capacity in the apparatus that remains unutilised and the greater the probability that the eye movement data will be contaminated by 'undisciplined' eye movement. This range of tempos is called the **UNDERCAPACITY AREA**.

The lower diagram in Figure 23 shows the boundaries between the three areas (action-slip, peak use, and undercapacity) as lying further to the right, in faster tempo-areas, for a more skilled player. Here the boundaries might be at 80MM and 90MM, whereas for the less skilled player represented in the upper diagram, these same boundaries might be at 70MM and 80MM. The aim of any training regime, then, is to shift the action-slip threshold further to the right. It is also clear from this discussion that enforcing a single tempo on both the skilled and the unskilled risks a significant level of action-slips among the unskilled and/or undisciplined eye movement among the skilled. This is the crux of the tempo/skill/action-slip fallacy.

Unfortunately most studies have sought to compare the skilled and the unskilled in the hope of generating pedagogically useful data; and aside from Smith (1988), for whom tempo itself was an independent variable, and Polanka (1995), none has set out to control tempo strictly. Investigators have apparently attempted to overcome the consequences of the fallacy by making one or more of several compromises, in some cases in such a way as to cast serious doubt on the reliability of their data. The typical compromise has been to exercise little or no control over the tempos at which participants performed in trials. Another compromise has been the toleration of a significant disparity in the level of action slips between skilled and unskilled groups. This appears to have occurred without accounting for the possible effects that action slips or tempo itself might have on eye movement (see the discussion of Smith's (1988) findings on tempo and fixation duration in Section 3.3.4).

In Jacobsen (1941) there was no attempt at tempo control—not even a suggested tempo at the start—and participants were reported as using a wide range of tempos, some extremely slow given the low level of difficulty of the stimuli. Despite this freedom to choose tempo, the level of action-slips ranged up to a highly problematic 88% for the unskilled and approached 0% for the skilled. Like Jacobsen, Weaver (1943) accepted participants of a considerable range of skill levels and allowed them to gravitate to whatever tempo was comfortable. In his words 'an arbitrarily fixed tempo would probably be too fast for some and too slow for others of the
participants from the standpoint of obtaining records of their best possible performances....[an arrangement that might well have led to data] of doubtful value'. (p15) Tempo, then, was a confounding influence in comparing skill, since there was no control for the correlation between skill and tempo. Some participants were reported to have played at an extremely slow tempo; there was also a high average rate of action slips, estimated at 14%, both situations of potentially confounding influence on the data. The methodology for Van Nuys & Weaver (1943) was similar to that of Weaver (1943) and will not be treated separately.

York (1952) chose participants of wide-ranging skill-level and attempted a compromise in respect of the relationship between tempo, skill and action-slips. He suggested a standard tempo to participants, but subsequently allowed them to perform (by humming melodies) at any tempo they felt comfortable with and to vary that tempo at will during performance. Despite this freedom to ignore the suggested tempo, the unskilled were still reported as making significantly more action slips than the skilled. Lang (1961) divided his participants into three groups in what appears to have been a jumbled mixing of tempo- and skill-criteria: the skilled, the unskilled (fast and inaccurate), and the unskilled (slow and inaccurate). Participants were directed to read the first four of twelve examples as fast as possible and the last eight in 'normal fashion'. The unskilled were reported as varying their average tempo considerably. In any case, neither tempo nor action-slips were controlled or monitored during the trials since it appears that Lang instructed his participants to read the stimuli silently rather that to perform what they sight read. Accordingly there was no verification that all the information on the score was successfully perceived and processed into musculoskeletal commands. Perhaps this explains Lang's statement that some of the faster readers 'seemed to scan the music too rapidly, fixating rarely' (p343).

The next three studies (Young 1971; Halverson 1974; Schmidt 1981) provided participants with the beat only before they played and did nothing more than orally encourage them to adhere to it. Young (1971) wished to compare the characteristics of her nine skilled and eight unskilled participants' fixations and eye-hand spans in reading eight 'chordal' (presumably homophonic) stimuli. She tried to avoid the consequences of the tempo/skill/action-slip fallacy by in effect limiting the skill-disparity between her skilled and unskilled groups. To do this she rigorously tested all potential participants and selected those 'with the greatest differences in sight-reading ability possible' while still meeting what she referred to as the 'technical fluidity criteria', a minimum benchmark as it were (Young 1971:52). It is difficult to
gauge how successful Young was in reducing the effects of the fallacy, because in addition to tempo problems there were other serious methodological problems in her study, and little information is provided on action-slips. Schmidt's (1981) study compared skilled and unskilled readers but found no significant differences between them, almost certainly because the suggested tempo was so slow in order to accommodate the unskilled as to have almost certainly contaminated the results for the skilled.

One of Goolsby's (1987) main objectives was to compare the eye movement patterns of the skilled and the unskilled. Temporal control in the trials was disappointingly loose. Participants were given 16 counts of a metronome at MM=120 as a suggested tempo, again rather slow considering the low level of difficulty of two of the four stimuli. Participants were then instructed to maintain that tempo and to continue through the exercise without stopping. The metronome was silent during the performances, leading Smith (1988:62) to suggest that the difference in fixation durations which Goolsby found between the skilled and the unskilled (377ms vs 474ms) could have been partly or wholly due to the difference between the average performance tempos of the two groups.

Smith (1988) appears to have been more aware of the complex methodological issues surrounding tempo, skill and action-slips. His study compared the eye movement behaviour of skilled versus unskilled trumpet players at different tempos. Tempos were controlled by a metronome throughout the trials, and the rate of action-slips was monitored and reported as significant for both unskilled and skilled participants. Smith's study illustrates what might be expected: a high rate of action slips is likely to be the trade-off for requiring the unskilled to read at the same tempo as the skilled. As one of the few significant observations of eye movement in music reading under controlled tempo, Smith (1988) probably resulted in a more accurate record of the effects of the independent variables in question, tempo and skill, even if there remain doubts over exactly how tightly controlled tempo was: 'If the participant stopped playing during the middle of the etude, the investigator said "keep going" to encourage the participant to finish the etude' (p88). Smith (1988) found that an increase in tempo resulted in fewer and shorter fixation durations. This finding is supported by Kinsler & Carpenter (1995) and Stage I of this study, under what appear to be even stricter temporal conditions.

By the end of the century researchers appeared to be avoiding the worst pitfalls of previous studies in relation to the monitoring and control of tempo, skill and action-
slips, even if their results were typically somewhat marred by other methodological concerns. Polanka (1995) applied a strict tempo regime for a study into whether eye movement patterns would indicate chunking characteristics in the processing of melodies as they are hummed. But as pointed out in Section 3.2.5, only data from silent preparatory readings were used in this study, a highly questionable method of observing eye movement in music reading. Kinsler & Carpenter (1995) chose not to compare the skilled and unskilled in their study of saccades in music reading, but rather explored the influence of tempo on the saccadic behaviour of skilled readers with a view to developing a model of eye movement in music reading. Unfortunately only rhythmic patterns were used in the stimuli for this study, raising immediate questions of ecological validity. Truitt et al (1997) similarly dispensed with the unskilled and used strict tempo control to make some valuable findings in relation to the eye-hand span and perceptual span.

The tempo/skill/action-slip fallacy, then, has represented a continual methodological problem and has rendered questionable much of the data produced by tracking eye movement in music reading. Definitive evaluation of how significantly eye movement data can be contaminated by uncontrolled tempo is one of the aims of Stage I of this study, and is revisited in Section 3.3.4.

3.3.3 Fixations, pauses and the number/duration relationship

This account of the findings in the literature concerning fixations in music reading is prefaced with a theoretical explanation that may bring a new perspective to those findings. Although the explanation is elaborate, it represents a foundation for assessing the literature and describing the methodology for Stage I of this study. To begin with, by choosing not to enforce a consistent tempo in their trials, most researchers have sacrificed the opportunity to observe eye movement using a simple mathematical relationship. For this relationship to pertain, the widely used fixation duration must be substituted by a new dependent variable comprising the duration of both the saccade and the fixation it moves to. In this study the fixation/saccade pair, as it were, is referred to as the pause; pause duration is a key dependent variable in Stage I.

Merging the separate variables of fixation duration and saccade duration into a single variable might at first appear to be unwise, since it leads to a reduction in the data. But it is submitted that such a reduction carries no great disadvantage in most research contexts. It is true that if the exact duration allotted to picking up each piece
of information from the score is at issue, it may be useful to know exactly how much
time is spent fixating as opposed to saccading, since virtually all information is
picked up during fixations and virtually none during saccades (Uttal & Smith 1968).
But in all other cases the distinction between saccade and fixation durations would
appear to be redundant. Furthermore, saccade duration is typically small in relation
to fixation duration, in the vicinity of 7-10%, and while there is no reliable data on
how this proportion varies under different conditions, it is unlikely to be by more
than a few percentage points. If the relationship between saccade and fixation
durations does not vary significantly, combining both variables sacrifices little useful
data. Knowing where fixations occur, how numerous they are, and the combined
duration of each fixation and its preceding saccade (that is, each pause) would appear
to be sufficient for most purposes.

The advantage of using the pause is that there is a simple relationship between the
number of pauses, their mean duration, and the tempo of the performance. To
explain this relationship it is easiest to consider an example of two performances of
the same total duration. Since the reading process consists entirely of successive
pauses, total performance duration \( D \) equals the number of pauses \( n \) multiplied
by their mean duration \( d \); in other words \( D = nd \). This is a simple case of
multiplying the parts to make the whole. Furthermore, the two performances must
also be at the same tempo, since tempo \( t \) is inversely proportional to the total
duration (halve the tempo and the performance takes twice as long). This
relationship can be expressed as \( t \propto \frac{1}{D} \); thus it follows that \( t \propto \frac{1}{nd} \), or ‘tempo is
inversely proportional to the number of pauses times their duration’.

This relationship is exemplified schematically in Figure 24. For the purpose of
simplicity, the total reading duration in both upper and lower diagrams is a brief
780ms, and local variations in saccade and fixation durations are averaged out since
they are not relevant to this explanation. Moving from left to right, fixation durations
are represented by the unbroken lines and saccade durations by the short dotted lines
linking the fixations. Mean saccade duration in the upper diagram is 20ms, and mean

16 Before the commencement of this project, an informal study was conducted to trial an ober2 eye
tracker with a sampling interval of 1ms. Three keyboardists were observed, each sight reading three
melodies at 60MM. The melodies were made up of only crotchets and distinguished by the distance
between each note. In the first melody, the notes were so close as to be almost touching; in the
second melody they were spaced apart by 2 degrees of visual angle; in the third they were spaced 4
degrees apart, much larger than the spacing normally encountered in music notation. The same
procedure was then used at a tempo of 120MM. With little variation between the three subjects,
saccade durations varied between approximately 7% of total reading duration for the closely spaced
melody to approximately 12.5% for the distantly spaced melody. Changing the tempo appeared to
make little difference to the relationship between mean saccade and fixation duration.
3 Eye movement in the reading of language and music

fixation duration 240ms, giving a mean pause duration of 260ms. Since there are three pauses in the reading, the total reading duration is 780ms (3 x 260ms). In the lower diagram the reading is at the same tempo and therefore of the same total reading duration as in the upper diagram, that is, 780ms. Here there are fewer pauses (two rather than three), and they are of longer mean duration (390ms compared with 260ms) to compensate for their smaller number. Thus 390ms x 2 = 780ms, the same total reading duration.

Figure 24 (upper) Schematic representation exemplifying the relative durations of saccades, fixations and pauses

(lower) Schematic representation exemplifying a different relativity between saccades, fixations and pauses

The number/duration relationship is regarded in this study as one of the basic underlying oculomotor mechanisms in music reading. But it has not been possible to find any reference in the literature to the relationship, perhaps because it pertains only to reading under the strict temporal conditions lacking in most studies. Investigations into eye movement in music reading have thus far explored the effects of several key variables on fixation duration rather than pause duration. Fixation duration, as explained above, is only an approximation of pause duration as it excludes saccade duration, and thus cannot be used in the context of the simple
mathematical relationship that governs eye movement. Be this as it may, data on fixation duration and number is all that is available in the literature, and does at least give an approximate indication of how the relationship might change under various conditions.

The next three sections review studies into the effects of tempo, musical complexity, stimulus familiarity and reader-skill on fixation number and mean fixation duration. With two provisos, the effects of the tempo/skill/action-slip fallacy and the fact that fixations rather than pauses were observed, these studies give some indication as to how the underlying relationship between mean pause duration and the number of pauses might be affected by the same variables. Following this, three further issues that do not directly involve fixation number and mean duration are reviewed: the bottom-up/top-down question, peripheral input, and the eye-hand span. In many places the nature of working memory will be advanced as a possible explanation for the behavioural phenomena reviewed.

3.3.4 Tempo

Smith (1988) set out to investigate the effects of tempo on fixations in the sight reading of trumpeters. He found that when tempo is increased, eye movement patterns are affected in a similar way to that in which the lower diagram differs from the upper diagram in Figure 24: fixations are fewer in number and shorter in mean duration. In addition, Smith (1988) found that fixations tend to be spaced further apart at faster tempos. To illustrate this change in relation to a music score, Figures 25 and 26 are simplified and exaggerated representations of the effect that Smith claimed to have observed. At the fast tempo (Figure 25) the fixations are fewer, more widely spaced, and of shorter duration than at the slow tempo (Figure 26)—six fixations as opposed to twelve. Note that when forced to play faster, the scanpath could have remained unchanged, with the same high number of fixations as at the slow tempo, but to achieve this would have required a halving of fixation durations. Smith, then, was observing a particular behaviour selected from a number of alternatives available to sight readers.

Kinsler & Carpenter (1995) also investigated the effect of increased tempo, but in reading rhythmic notation rather than real melodies. They similarly found that increased tempo causes a decrease in mean fixation durations and an increase in mean saccade amplitude (equivalent to the distance on the page between fixations). The
findings of both of these studies resonate with those of Stage I, which observed similar behaviour in the sight reading of two-stave keyboard music (see Chapter 4).

Figure 25  Hypothetical representation of more fixations of slower duration and more closely spaced at slow tempo, based on Smith (1988)

Figure 26  Hypothetical representation of fewer fixations of shorter duration and spaced further apart at fast tempo, based on Smith (1988)

3.3.5 Musical complexity

From the start, researchers were interested in whether fixation durations might be influenced by the complexity of a musical stimulus. Bean (1938) pointed out that musical complexity is different from optical complexity, but his insight failed to prompt subsequent researchers to closely define what they meant by complexity and to apply that definition empirically.

At least three types of complexity need to be accounted for in music reading: the optical complexity of the notation itself, the complexity of processing visual input into musculoskeletal commands, and the complexity of executing those musculoskeletal commands. For example, optical complexity might come in the form
of the density of the notational symbols, or of the appearance of accidentals, triplet signs, slurs and other expressive markings. The complexity of processing visual input into musculoskeletal commands might involve a lack of chunkability or predictability in the music. The complexity of executing musculoskeletal commands might be seen in terms of the demands of fingering and hand position, among other considerations. These types of complexity have potentially quite independent effects in the task of performing a musical passage. It is in isolating and accounting for the interplay between them that the difficulty lies in making empirical sense of musical complexity. For this reason, little useful information has emerged from investigating musical complexity through observing eye movement.

Jacobsen (1941:213) concluded that ‘the complexity of the reading material influenced the number and the duration of [fixations]’. Where the texture, rhythm, key, and accidentals were ‘more difficult’, there was on average a slowing of tempo and an increase in both the duration and the number of fixations in all groups. This ignores the strong possibility that the characteristics of fixations were altered in reading more complex music because of slower tempo in addition to the complexity itself. Weaver (1943) claimed that (given the stimuli and skill-levels in his trials) a normal range of fixation durations was 270ms-530ms, and longer when the notation was more compact and/or complex, as Jacobsen had found. But unlike Jacobsen, Weaver did not report whether such longer fixations were associated with slower tempo.

Halverson (1974) observed that an increase in notational complexity was associated with slightly more fixations which were on average slightly shorter. This is the very opposite of the findings of Jacobsen and Weaver, which may have resulted from stricter control of performance tempo in his trials. Consistent with Halverson (1974), Schmidt’s (1981) participants used longer fixation durations in the easier melodies. Goolsby’s (1987) data mildly supported Halverson’s (1974) observations, but only in respect of his skilled participants. Indeed, Goolsby (1987:107) noted that ‘both Jacobsen and Weaver.....in letting participants select their own tempo found the opposite effect of notational complexity as the effect found in the present study.’

On balance it appears likely that under controlled temporal conditions, denser and more complex music is associated with a higher number of fixations of shorter mean duration. If the music reading apparatus does behave in this way when faced with an increase in processing demand, it might be explained as an attempt by the
oculomotor system to provide more frequent 'refreshment' of the material being held in working memory. More frequent refreshment may compensate for the need to hold more information in working memory, an issue that is further discussed in the remainder of this chapter.

3.3.6 Reader-skill

Here there is no disagreement among the major studies, from Jacobsen (1941) to Smith (1988): skilled readers appear to use more and shorter fixations across all conditions than do the unskilled. Goolsby (1987) found that mean 'progressive' (forward-moving) fixation duration was significantly longer (474ms vs 377ms) and mean saccade length significantly greater for the less skilled than for the more skilled. Although Goolsby did not report the total reading durations of his trials, they can be derived from the mean tempos of his 12 skilled and 12 unskilled participants for each of the four stimuli (Goolsby 1987:88). Provided there were no extraneous methodological factors (such as failing to exclude from total durations the 'return-sweep' saccades and non-reading time at the beginning and end of readings), his data show that the unskilled played at 93.6% of the tempo of the skilled, and that their mean fixation durations were 25.6% longer.

This raises the question as to why skilled readers should distribute more numerous and shorter fixations over the score than the unskilled. For all the work done in this area, only one plausible explanation appears in the literature. Kinsler & Carpenter (1995) proposed a model for the processing of music notation, based on observing eye movement in the reading of rhythm patterns, whereby an iconic representation of the fixated image is scanned by a 'processor' and interpreted to a given criterion of accuracy. The scan ends when this criterion cannot be reached, its end-point determining the position of the next fixation. The time taken before this decision varies depending on the complexity of a note. The processor of expert readers, they claimed, is probably far superior to that of the novice, and would therefore scan the individual notes within the iconic store more rapidly, thus promoting more numerous fixations of shorter duration than would be the case for an unskilled reader.

The model of Kinsler & Carpenter (1995), however, does not explain what advantage there is to be gained by short, numerous fixations. The answer to this question may be related to the possibility that skilled readers, who have been shown to maintain a larger eye-hand span and therefore a larger amount of information in their working memory, need to refresh that information more frequently from an
external source; this could be achieved by fixating more quickly and numerously (thereby, incidentally,refixating more frequently). Unskilled readers, on the other hand, may not need to refresh the information they hold in their working memory to the same extent because they hold less information in their working memory; this may be the reason that their fixations tend to be longer and fewer. Although this argument is conjectural, when combined with evidence from the literature as a whole it resonates with the three oculomotor imperatives that are proposed at the end of this chapter.

Finally, it should be reported that Waters & Underwood (1998) used a tachistoscope to conduct a pattern-matching experiment that required no musical performance, and only the pressing of a yes/no button as an indication of whether two successively presented melodic fragments were the same or different. In carrying out this task, experts were on average more accurate and rapid than novices, and used more and shorter fixations. Although this study used a very different method, its results would appear to support the observed association in musical performance between skill and short, numerous fixations.

3.3.7 Stimulus-familiarity

As postulated in Section 2.2, familiarity with a musical stimulus involves a functional relationship between currently activated information, and relevant information in long-term memory (or long-term working memory according to Ericsson's model). Familiarity can exist at several levels, can be sourced both within the stimulus and outside it, and can be in either actuality and probability modes, making an absolute definition of sight reading impossible. It was further postulated that in repeated readings of the same stimulus there is a shift in reliance from external to internal input, and within internal input from probability to actuality modes of information. These shifts were illustrated in Figures 3 and 4 in Section 1.2.1.

How might these shifts in the processing of information affect eye movement behaviour? As internal input strengthens with successive encounters with the same stimulus, reliance on external input will tend to be reduced. On logical grounds, it would be expected that such a reduced reliance would result in fewer and longer fixations. The data from all three studies into eye movement in the reading of increasingly familiar music support this reasoning. York's (1952) participants read each stimulus twice, with each reading preceded by a 28-second silent preview. On
average, both skilled and unskilled readers did indeed use fewer and longer fixations during the second reading.

Goolsby’s (1987) participants were observed during three immediately successive readings of the same musical stimulus. Familiarity in these trials appeared to increase fixation duration, but not nearly as much as might have been expected. The second reading produced no significant difference in mean fixation duration (from 422ms to 418ms). On the third encounter, mean fixation duration was higher for both groups (437ms) but by a barely significant amount, thus mildly supporting York’s earlier finding. Goolsby’s mean fixation durations in milliseconds are shown in Table 4 for each encounter with the same stimulus.

<table>
<thead>
<tr>
<th>encounter</th>
<th>all participants</th>
<th>skilled</th>
<th>unskilled</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>422</td>
<td>369</td>
<td>474</td>
</tr>
<tr>
<td>2</td>
<td>418</td>
<td>371</td>
<td>465</td>
</tr>
<tr>
<td>3</td>
<td>437</td>
<td>391</td>
<td>482</td>
</tr>
<tr>
<td>mean</td>
<td>426</td>
<td>377</td>
<td>474</td>
</tr>
<tr>
<td>% increase 1 to 3</td>
<td>3.55%</td>
<td>5.96%</td>
<td>1.68%</td>
</tr>
</tbody>
</table>

Table 4  Mean fixation durations in ms, by encounter

These disappointedly small changes might be explained by the unchallenging reading conditions in the trials. The tempo of MM120 suggested at the beginning of each of Goolsby’s trials appears to be rather slow for tackling the given melodies, which contained many SEMIBREVES and MINIMS, and there may have simply been insufficient pressure to produce significant results. A more likely explanation, however, is that participants played the stimuli at faster tempos as they grew more familiar with them through the three readings. (The metronome was initially sounded but was silent during the actual performances, allowing readers to vary their pace).

Thus it is quite possible that two influences were at odds with each other: familiarity may have promoted low numbers and long fixation durations, while faster tempo may have promoted low numbers and short durations. This might explain why mean fixation duration fell in the opposite direction to the prediction for the second encounter, and by the third encounter had risen by only 3.55% across both groups. Goolsby’s data, then, could be interpreted as being consistent with low numbers and high durations, but masked by tempo change. Smith’s results, published a year later
and reinforced by Kinsler & Carpenter (1995), suggest that faster tempos are likely to reduce both the number and duration of fixations in the reading of a single-line melody. If this hypothesis is correct, it may be explained by the fact that the more familiar a stimulus is, the less the workload on the memory system.

The effect of the four variables that have been reviewed here on fixation number and mean fixation duration—tempo, stimulus-complexity, reader-skill and stimulus-familiarity—are summarised in Table 5. It is possible to conclude from the table that there is an association between challenging reading conditions, which may require a larger workload in working memory, and more numerous fixations of shorter duration. The exception appears to be faster tempo, where the need to progress across the score more quickly may override the hypothesised tendency to fixate more numerously. The effect of greater reader-skill probably resembles that of a more challenging reading situation since, as will be discussed later in this section, such readers appear to hold more information in their working memory.

<table>
<thead>
<tr>
<th>change in variable</th>
<th>effect on fixations</th>
<th>sources</th>
</tr>
</thead>
<tbody>
<tr>
<td>faster tempo</td>
<td>shorter, fewer</td>
<td>Smith (1988)</td>
</tr>
<tr>
<td>less familiar stimulus</td>
<td>shorter, more numerous</td>
<td>Goolsby (1987)</td>
</tr>
</tbody>
</table>

Table 5 Summary of the effect of four key variables on fixations

3.3.8 The top-down/bottom-up question

As discussed in Section 3.2.4, there was considerable debate from the 1950s to the 1970s as to whether eye guidance in language reading is mainly under bottom-up or top-down control, or whether both directions of influence function simultaneously. Some years before this debate, Weaver (1943) set out to examine the possible (bottom-up) effects of musical texture on eye movement. He hypothesised that vertical compositional patterns would promote vertical saccades, and horizontal compositional patterns horizontal saccades. Weaver's 12 participants read a two-part POLYPHONIC stimulus, in which the compositional patterns were strongly horizontal, and a four-part HOMOPHONIC stimulus comprising plain hymn-like chords, in which the compositional patterns were strongly vertical. He included three examples of scanpaths over each stimulus, a total of six illustrations, which are reproduced here in Figure 27.
Figure 27  Weaver's illustrations of eye movement in three subjects' reading of (a) the homophonic example, and (b) the polyphonic example
It should be pointed out that the overriding problem with the notion that vertical or horizontal musical patterns promote similar patterns in scanpath is that vertical and horizontal saccades are both continually essential in the scanning of a multistave score. Therefore it is not surprising that Weaver was unsuccessful in finding that so-called vertical and horizontal textures had a bottom-up influence on eye movement characteristics. Indeed, his result was the very opposite of what might have been expected: he concluded that "the "horizontal" form of progression was used more frequently in reading [the homophonic stimulus] and the "vertical" form in reading [the polyphonic stimulus]" (p28). This conclusion is consistent with the ratio of horizontal to vertical saccades in Figure 27, the only specific evidence presented in his article: a brief analysis shows that 29% of the saccades over the homophonic stimuli and only 13% of the saccades over the polyphonic stimuli are horizontal (excluding inspections of clefs and key signatures at the start of each line).

Four decades later, when researchers were successfully finding evidence of bottom-up influence on eye movement in language reading, Sloboda (1985) was interested in the possibility that there might be an equivalent influence on eye movement in music reading. He began by claiming Weaver's original hypothesis as established:

The general strategy appears to be to identify significant structural units in successive fixations, so that in the reading of homophonic textures chords are normally inspected entirely, top and bottom or the reverse, before the eyes proceed to the next chord; whereas in polyphonic textures horizontal eye movement predominates. (p69)

Sloboda (1985:70) illustrated this with schematic representations, shown here in Figure 28.

![Schematic representations of eye movement patterns](image-url)

**Figure 28** Sloboda's schematic representations of (a) vertical saccades over homophonic music, and (b) horizontal saccades over polyphonic music
Undeterred by Weaver’s explicit conclusion and the evidence in his illustrations (Figure 27), Sloboda (1985:70) claimed that:

Weaver found that [the vertical] pattern was indeed used when the music was homophonic and chordal in nature. When the music was contrapuntal, however, he found fixation sequences which were grouped in horizontal sweeps along a single line, with a return to another line afterwards.

Sloboda then quoted two one-bar fragments (Figure 29) taken from Weaver’s illustrations of each type of stimulus. These quotations do not appear to be representative of the overall scanpaths in Weaver’s illustrations. In particular, Figure 29b shows one of the very few examples of horizontal movement in Weaver’s illustration of scanpath over his polyphonic stimulus.

![Figure 29](image)

**Figure 29** Sloboda’s quotation of fragments of (a) vertical scanpath over homophonic music, and (b) horizontal scanpath over polyphonic music, from Weaver (1943)

Although Sloboda’s specific claim relating to a confluence of textural and saccadic dimensions may be questionable, and despite Weaver’s failure to find such dimensional links, eye movement in music reading shows clear evidence in most eye-movement studies (in particular Truitt et al 1997; Goolsby 1987) of being influenced by bottom-up graphical visual features and top-down global factors related to the meaning of the symbols.

### 3.3.9 Peripheral visual input

The role of peripheral visual input in language reading has been and remains the participant of much investigation, and was briefly reviewed in Section 3.2.5. Peripheral input in music reading has been studied by Truitt et all (1997). They used the moving-window technique to measure the extent of peripheral perception to the right of a fixation. It was found that performance was degraded only slightly when four crotchets to the right were presented as the ongoing preview, but significantly
when only two crotchets were presented. Under these conditions, peripheral input extended over a little more than a four-beat measure on average. For the less skilled, useful peripheral perception extended from half a beat up to between two and four beats. For the more skilled, useful peripheral perception extended up to five beats.

Peripheral visual input in music reading is clearly in need of more investigation, particularly now that the moving-window technique has become more accessible to researchers. A case could easily be made that western music notation has developed in such a way as to allow the apparatus to utilise peripheral input in the reading process. Noteheads, stems, beams, barlines and other notational symbols are all sufficiently bold and distinctive to be useful when picked up peripherally, even at some distance from the fovea. The upcoming pitch contour and prevailing rhythmic values of a musical line can typically be ascertained ahead of foveal perception. For example, a run of continuous semiquavers beamed together by two thick, roughly horizontal beams will convey potentially valuable information about rhythm and texture, whether to the right on the currently fixated stave or above or below in a neighbouring stave. This is reason enough to suspect that the peripheral preprocessing of notational information is important, if not essential, for fluent music reading, just as it has been found to be the case for language reading (see Section 3.2.5).

If the music reading apparatus utilises the O'Regan effect to skip over information on the score, peripheral input can be assumed to be at play, combining with the player's long-term memory of the probabilities of what the unfixated information is, as has been demonstrated in language reading (see Section 3.2.5). The effect is consistent with the findings of Smith (1988) and Kinsler & Carpenter (1995), who reported that the eyes did not fixate on every note in the reading of melodies. The effect of tempo on the musical equivalent of the O'Regan effect in the reading of two-stave scores has remained unexplored. This gives rise to one of the hypotheses of the experiment in Stage I of this dissertation, that the eyes skip over more material at faster tempos than slower tempos, as they are under greater pressure to maintain progress across the page.

3.3.10 Refixation

As discussed in Section 3.2.6, why readers refixate in language reading is still largely an unsolved puzzle. Two main types of refixation have been identified: same-word rightward refixation, which assumes a degree of overlap in the foveal areas of
successive fixations; and leftward refixation to a previous word. Studies on eye movement in music reading, including Weaver (1943) and Young (1971) have tended to discuss leftward movement, typically referring to them as ‘regressions’, but not the musical equivalent of same-word rightward refixation, despite the fact that closely spaced rightward fixations in the reading of single-line melodies have been widely reported.

The situation in relation to the reading of (two-stave) keyboard music is quite different. Here, there is a constant need for vertical movement between the two staves. Thus three types of refixation are likely to occur in the reading of a multi-stave score: (1) rightward along the same stave to the next chord, if the chords are closely enough spaced for foveal overlap, (2) vertically within the same chord, after that chord has already been inspected on both staves, and (3) leftward back to a previous chord, either on the same stave or diagonally to the other stave. Type (1) is difficult to demonstrate from scanpath alone, since actual perception, central and peripheral, are at issue. Types (2) and (3), however, are dependent on scanpath and are indicated with numerals in Figure 30. Weaver’s (1943) examples, quoted in Figure 27, also provide examples of Types (2) and (3).

![Figure 30](image)

**Figure 30** Schematic illustrations of two categories of refixation in the reading of a keyboard score

Leftward refixation occurs in music reading at all skill-levels, as reported by Goolsby (1987) and Smith (1988) among others. It involves a saccade back one or two notes or chords, followed by at least one saccade then forward again to regain lost ground. It therefore requires a greater investment of time, and is likely to be considerably less common than the other types of refixation, a point made on p67 in relation to language reading. Because it is more costly than the other two types, it is logical to assume that leftward refixation occurs less frequently than vertical refixation, which is theoretically a much less resource-intensive way of refreshing the contents of
working memory. Both of these types of refixation are also likely to occur less frequently at fast tempo, under pressure to make urgent progress across the page. These logical inferences form the basis for two of the hypotheses in Stage I (see Sections 4.3 and 4.3).

What, then, can be found concerning leftward refixation in the literature on eye movement in music reading? There are claims from Jacobsen (1941) onwards that the unskilled refixate leftwards at a significantly higher rate than the skilled, although the supporting evidence was possibly contaminated by the slower tempo of the unskilled. Weaver (1943) reported that leftward regressions ran from 7% to a substantial 23% of all saccades. Goolsby (1987) and Smith (1988) also reported significant levels of leftward refixation across all skill-levels. It should be noted, however, that none of these studies appears to have observed eye movement at fast tempos, where the occurrence of leftward refixation might have been minimised through the need to progress across the page more rapidly. These findings prompt the investigation of leftward refixation in Stage I of this dissertation, specifically its relationship to tempo. Refixation may also be linked to the size of the eye-hand span, a matter which is discussed in the next section.

3.3.11 The eye-hand span

A brief account of the work done on the eye-voice span in the reading of language was presented in Section 3.2.7. The main findings were that (1) a larger span is associated with faster/more skilled readers, (2) a shorter span is associated with greater stimulus-difficulty, and (3) the span appears to vary according to linguistic phrasing. At least eight studies into eye movement and music reading have investigated similar issues. Jacobsen (1941), for example, measured the average span to the right in the sight singing of melodies as 0-2 notes for the unskilled and 1-4 notes for the skilled, whose faster average tempo in this study raises doubt as to whether skill alone was responsible for this difference. In Weaver (1943:28), the eye-hand span varied greatly but never exceeded ‘a separation of eight successive notes or chords’ a figure which seems impossibly large for the reading of keyboard scores. Young (1971) found that both skilled and unskilled participants previewed approximately one chord ahead of their hands, an uncertain finding in view of the methodological problems that were raised in Sections 3.3.1-3.3.2.

Goolsby (1994a) found that skilled sight singers’ eyes were on average about four beats ahead of their voice, and less for the unskilled. He claimed that when sight singing, ‘skilled music readers look farther ahead in the notation and then back to the
point of performance' (p77). To put this another way, skilled music readers maintain a larger eye-hand span and are more likely to refixate within it. This association between span size and leftward refixation could arise from a greater need for the refreshment of information in working memory, and will be investigated in Stage II of this dissertation. This finding is of major importance in the pedagogical rationale of *SightReader*, to be treated further in Chapter 5.

Sloboda (1974; 1977) cleverly applied Levin & Kaplin’s (1970) ‘light-out’ method (discussed in Section 3.2.7) in an experiment designed to measure the size of the span in music reading. Sloboda (1977) asked his participants to sight read a melody and turned the lights out at an unpredictable point during each reading. The participants were instructed to continue playing correctly ‘without guessing’ for as long as they could after visual input was effectively removed, giving an indication as to how far ahead of their hands they were perceiving at that moment. Here, the span was defined as including peripheral input. Participants were allowed to choose their own performing speed for each piece, introducing a layer of uncertainty into the interpretation of the results. Sloboda reported that there was a tendency for the span to coincide with the musical phrasing, so that ‘a boundary just beyond the average span “stretches” the span, and a boundary just before the average “contracts” it’ (as reported in Sloboda 1985:72). Good readers, he also found, maintain a larger span size (up to seven notes) than poor readers (up to four notes).

Truitt et al (1997) found that in sight reading melodies on the electronic keyboard, span size averaged a little over one beat and ranged from two beats behind the currently fixated point to an incredibly large 12 beats ahead. The normal range of span size, however, was rather smaller: between one beat behind and three beats ahead of the hands for 88% of the total reading duration, and between nought and two beats ahead for 68% of the duration. (Such large ranges, in particular those that extend leftwards from the point of fixation, were probably due to the wandering effect. This occurs where the music reading apparatus is operating significantly under its available capacity, allowing the eyes to wander about the score.) For the less skilled, the average eye-hand span was approximately half a crotchet-beat. For the more skilled, eye-hand span averaged about two beats and useful peripheral perception extended up to five beats. This, in the view of Rayner & Pollatsek (1997:52) suggests that:

> a major constraint on tasks that require translation of complex inputs into a continuous motor ‘transcription’ is [the limited capacity of] short-term memory. If the encoding process gets too far ahead of the output, there is likely to be a loss of material that is stored in the queue.
In the opinion of Rayner & Pollatsek (1997:52), the data from Truitt et al (1997) is 'contrary to [the] conventional wisdom', that is, of encouraging top-down effort by musicians to increase their span size. The observation by Truitt et al (1997) of a 'cut-off' point, as it were, beyond which the eye-hand span can rarely be extended, resonates with the possibility that span size can be increased, whether by the conventional top-down method or by using a device such as SightReader.

Rayner & Pollatsek (1997) report on and further interpret the previous experiment by Truitt et al (1997) into the nature of both the eye-hand span and the perceptual span. The authors appear to have used these two terms to refer to the central and peripheral components of the eye-hand span respectively. Rayner & Pollatsek (1997) explained the size of the eye-hand span as a continuous tug-o-war, as it were, between two forces: (1) the need for material to be held in working memory long enough to be processed into musculoskeletal commands, and (2) the need to limit the demand on span size and therefore the workload in the memory system. They claimed that 'most music pedagogy supports the first aspect [in advising] the student that the eyes should be well ahead of the hands for effective sight reading' (p52). They also held that despite such advice, for most readers the second aspect prevails, that is, the need to limit the workload of the memory system. This, they claimed, results in a very small span under normal conditions. The authors presumably meant that asking students to look ahead further, the top-down counterpart to SightReader's bottom-up effect, is usually not a successful strategy. This lack of success is the prime reason behind the development of SightReader, which exerts bottom-up control over span size.

The eye-hand span in both its foveal and peripheral guises was discussed by Kinsler & Carpenter (1995) in purely speculative terms. They reasoned that music notation conveys information in a more concentrated form than language, with a closer connection between the symbols on the page and the musculoskeletal commands they represent. This is a puzzling assumption, since to read language aloud also requires a precise connection between symbols and commands, and there is no evidence as to the relative closeness in respect of linguistic and musical performance. Nevertheless, Kinsler & Carpenter (1995) went on to claim that 'one might expect to find a rather more precise control of gaze while reading music as compared with the typically imprecisely targeted saccades that are found in [language] reading' such that 'there is a closer link between the immediate stimulus and the associated eye movement' (p1448). They go on to suggest that because 'musicians are explicitly trained to read well ahead of the point in the music corresponding to the notes being executed at any one moment', eye movement would be expected to be influenced by
the 'buffer' (eye-hand span) rather than by the immediate meaning of the notes on the page. (p1448) This statement apparently implies that the need to maintain an appropriate span size prevails over other bottom-up influences on eye movement. Such an either-or approach does not seem to be likely in the light of the wealth of evidence that eye movement patterns are the result of many influences in the reading process.

Surprisingly, the effect of tempo on span size has not yet been explored. This is despite the fact that tempo is a key variable in musical performance, and plays a major part in determining the workload of the memory system. The effect of tempo on span size in the reading of two-stave music is the participant of several key hypotheses in both Stages I and II.

3.3.12 Conclusions

Both logical inference and evidence in the literature point to the fact that there are three oculomotor imperatives in the task of eye movement in music reading. The first imperative seems obvious: the eyes must maintain a pace across the page that is appropriate to the tempo, and they do this by manipulating the number and duration of fixations, and thereby the scanpath across the score. The second imperative is to provide an appropriate rate of refreshment of the information being stored and processed in working memory by manipulating the number and duration of fixations. Such workload appears to be related to tempo, stimulus-complexity and stimulus-familiarity, and there is strong evidence that the capacity for high workload in relation to these variables is also connected with the skill of the reader. The third imperative is to maintain a span size that is appropriate to the reading conditions. The span must not be so small that there is insufficient time to perceive visual input and process it into musculoskeletal commands; it must not be so large that the capacity of the memory system to store and process information is exceeded. These three imperatives, concerned with the need for sufficient pace, refreshment, and span size respectively, were explored in detail in the remainder of this study.

The music reading apparatus, then, appears to use oculomotor commands to address all three imperatives simultaneously, which are in effect mapped onto each other in the reading process. Eye movement thus embodies a fluid set of characteristics that are not only intimately engaged in engineering the right visual input to the apparatus, but in servicing the processing of that information in the memory system. The next chapter explores these three imperatives by observing the eye movement of skilled keyboardists in a prelude to the teaching experiment on SightReader that is reported in Chapter 5.
Stage I: observing span size and refixation in music reading

*SightReader* could not be fully understood or evaluated without some knowledge of three aspects of music reading behaviour over which it exerts control: performance tempo, controlled by *SightReader’s* tempo setting; and minimum span size and leftward refixation, controlled by *SightReader’s* unique process of chord erasure. Despite the numerous studies into eye movement in music reading that were reviewed in the previous chapter, little is known about span size and refixation and the way these two variables might be influenced by tempo. Accordingly, the main purpose of Stage I was to observe how these phenomena interact in a comparatively normal reading situation. This was achieved by using an infrared tracking device to measure eye movement in the sight reading of keyboard music. The current chapter is an account of this first experimental stage in the study; as anticipated, its findings assisted in the design and conduct of Stage II, a teaching experiment using *SightReader*.

Stages I and II were structured around two distinct reading conditions. The first is referred to here as the NORMAL CONDITION, under which tempo was controlled but span size and leftward refixation were not. Traditional musical rehearsal can be thought of as occurring under the normal condition, since it involves the control of tempo (either internally by the performer, or externally by a metronome, conductor’s beat or the demands of ensemble performance) but does not directly concern itself with span size or leftward refixation. The normal condition is contrasted with the SPAN-STRETCHING CONDITION, in which *SightReader’s* unique mechanism of ongoing chord erasure was used in addition to tempo control. The span-stretching condition thus exerted control over all three variables, tempo, leftward refixation and minimum span size.

The observation and control of these variables in each stage is summarised in Table 6. The normal condition was used to observe the eye movement of all participants in Stage I. Stage II presents a more complex arrangement, comprising six trial sessions for each participant. The first trial session comprised a PRETEST and sixth a POSTTEST; in both of these sessions the normal condition was used to compare the performance.
of all participants. In the middle four sessions, however, participants in the EXPERIMENTAL GROUP were exposed not only to tempo control, but to SightReader's process of chord erasure that controls span size and leftward fixation (see the second-bottom row in Table 6). By contrast, during these sessions participants in the CONTROL GROUP were exposed to the normal condition. In this way it was possible to compare two contrasting methods of training sight reading: a traditional method of training sight reading under the normal condition (control group), and the novel span-stretching condition (experimental group).

<table>
<thead>
<tr>
<th>stage</th>
<th>portion of trial</th>
<th>participants</th>
<th>tempo</th>
<th>span size &amp; leftward refixation</th>
<th>reading condition</th>
</tr>
</thead>
<tbody>
<tr>
<td>I</td>
<td>whole trial</td>
<td>all</td>
<td>controlled</td>
<td>observed</td>
<td>normal</td>
</tr>
<tr>
<td></td>
<td>pretest (1)</td>
<td>all</td>
<td>controlled</td>
<td>—</td>
<td>normal</td>
</tr>
<tr>
<td>II</td>
<td>trial sessions</td>
<td>control group only</td>
<td>controlled</td>
<td>—</td>
<td>normal</td>
</tr>
<tr>
<td></td>
<td>(2), (3), (4), (5)</td>
<td>experimental group only</td>
<td>controlled</td>
<td>controlled</td>
<td>span stretching</td>
</tr>
<tr>
<td></td>
<td>posttest (6)</td>
<td>all</td>
<td>controlled</td>
<td>—</td>
<td>normal</td>
</tr>
</tbody>
</table>

Table 6 Observation and control of three aspects of music reading in Stages I and II

In Stage I, four HYPOTHESES arose from this need to know more about the eye movement behaviour that SightReader would act upon by observing how span size and leftward refixation interact with tempo under the normal condition. Skilled sight readers and stimuli of consistent texture and difficulty were used in this experiment, to avoid the methodological problems surrounding the relationship between tempo, skill, action slips and stimulus-difficulty that were discussed in Section 3.3.2. Further information about participants and stimuli are provided in Sections 4.1.2 and 4.1.4. The INDEPENDENT VARIABLE (that is, the variable manipulated by the experimenter, and 'independent' of the data output) was tempo, with a control for performance accuracy. The five DEPENDENT VARIABLES (that is, the variables 'dependent' on the data output) were (1) mean LATENCY, (2) mean LOAD, (3) mean pause duration, (4) mean number of pauses per chord, and (5) the rate of leftward refixation. Latency and load are different ways of measuring the eye-hand span, in terms of duration and information content respectively; they are explained in greater detail in Section 4.2.1.
Hypothesis 1, discussed in Section 4.2, concerned the effect of tempo on span size and was of immediate interest to the development of a strategy for manipulating SightReader's tempo and span settings in Stage II. The goal of such a strategy was to maximise any training benefit and the consistency with which participants in the experimental group were treated during the four training sessions (2, 3, 4 and 5) in Stage II during which they were exposed to the span-stretching condition. Hypothesis 1 correctly predicted that two components of the span, load and latency as defined in Section 4.2, contributed roughly equally to the task of tempo adaptation. Observing the effect of tempo on span size also represented a chance to explore the third imperative for eye movement in music reading as proposed in Section 3.3.5: to maintain a span size that is appropriate to the reading conditions.

The remaining hypotheses are covered in Section 4.3. Hypothesis 2 concerned the relationship between another pair of related variables, pause number and mean pause duration, with reference to the first and second oculomotor imperatives, i.e. that the eyes maintain a pace across the page appropriate to the given tempo, and that they provide sufficient refreshment of information in the eye-hand span. It was predicted that pause number and duration, first treated in Section 3.3.3, would contribute roughly equally to the task of tempo adaptation. As explained later, this finding represented important background information for the investigation of the remaining two hypotheses.

Hypothesis 3 predicted that the level of leftward refixation would be lower than that for vertical refixation, based on the relative costs of each type of refixation in respect of the first oculomotor imperative (pace). This prediction was relevant to Stage II in that it involved investigating how disruptive to normal eye movement patterns SightReader's restrictions on leftward refixation would be in Stage II. Hypothesis 4 explored the possibility that refixation, the O'Regan effect, pause duration, pause number, and tempo are part of a single, interdependent mechanism for accommodating the first and second oculomotor imperatives.

At this point a note of caution must be sounded concerning the limited nature of the data in Stage I. Although the data strongly supported three of the four hypotheses, the results were based on a sample of only nine participants. It is argued, however, that since the participants were carefully selected on the basis of a narrow range of skill-levels, their performance can be generalised to the population of all highly skilled keyboard sight readers with reasonable confidence.
In the next section the equipment and methodology used in the Stage I trials are described and three protocols are presented that were employed to maximise the likelihood that the trials would run smoothly, consistently and ethically. Following this, the theoretical background to each hypothesis is explained and the results presented and discussed.

4.1 Method

4.1.1 Equipment

The equipment for tracking and recording eye movement in Stage I was located in a small, quiet office in the Centre for Performance Studies at the University of Sydney. It consisted of 11 components, listed below and identified by number in Figure 32, a schematic depiction of the equipment setup: (1) a pair of goggles from an NAC video eye-tracking system; (2) a video camera, permanently affixed to the top of the goggles; (3) the free-standing microphone attachment to the video camera; (4) a portable musical keyboard with built-in speaker; (5) a second, freestanding video camera; (6) a portable electronic digital timer; (7) a pocket-sized electronic metronome with an auditory tick and a small readout displaying a simultaneous visual tick; (8) a control box linked to (1), (2), (3), (4), (8), (9) and (10); (9) a television monitor; (10) a video cassette recorder (VCR); and (11) associated cabling and office furniture including an adjustable ‘crane’ document holder for displaying the stimuli, a desk, an adjustable chair and a trolley for accommodating items (9) and (10). In addition, a rubber bathing cap was worn by participants to stabilise the goggles. The bathing cap was used only from the third participant onwards, after it was discovered that slippage during the first two trials was causing the task of accurately registering the data to be more difficult and elaborate than the fishbowl effect had already made it.

These components are now described in greater detail. The most specialised item in the setup was the pair of goggles housing the eye movement tracking devices (Item 1). The frame of the apparatus was constructed of metal, with plastic inlays for strength and solidity. Metal wings on either side of the frame fitted over the participant’s ears. Velcro straps attached to each side of the frame stretched around the participant’s head and were joined firmly at the back to stabilise the fitting. (Without stability, the impending calibration of the cursor position would have been corrupted, a problem that was discussed Section 3.3.1 in relation to earlier studies.)

17The system had been purchased second-hand without its video camera component; a replacement camera was subsequently fitted to the goggles.
Each eye looked through one of two circular apertures approximately 60mm in diameter. Tiny devices embedded around each aperture shone (invisible) infrared light at the LIMBUS, the boundary between the iris and the sclera (the white of the eye). Nearby sensors then picked up and recorded the reflection of this light at intervals of 1ms. Because each sensor could assign only a one-dimensional value to the position of the eyes at any one time, separate sources and sensors of infrared light were needed to pick up the vertical and horizontal dimensions of eye position. Thus there were eight tracking devices in all: four served each aperture, of which two were assigned to each dimension, one device sending infrared light, the other receiving it.

![Figure 31](image.png)

Figure 31  Schematic depiction of the equipment setup

When connected to the television monitor (Item 9), the goggles produced a small illuminated cursor shaped like the letter V against a blank background; the point at the bottom of the V marked the position of the centre of the fovea at any one time. A separate video camera (2) was permanently mounted onto the frame of the goggles, bringing the weight of the apparatus to approximately 850gm. This enabled the participant’s eye movement to be displayed against a background representation of her/his visual field, that is, the music score. The camera picked up the stimulus from a point about 70mm above the participant’s eyes using the standard VCR sampling interval of 40ms (25 measurements per second). The camera also provided a synchronised audio track through the VCR microphone-attachment (3) that was set...
up on the main table and directed towards the sound source in the portable keyboard (4). The second VCR camera (5) sat on an adjacent table, focused on the display of the portable electronic digital timer (6) on the right and the readout of the metronome (7) on the left of its field. The timer updated every 10ms and provided a convenient and reliable way of verifying the sequence and timing of the frames, saving much frame-by-frame counting during the registration process.

The signals from the infrared sensors in the goggles and from both video cameras were transmitted to the custom-made control box (8) sitting on the lower tier of a moveable trolley. Here they were combined electronically and transmitted to the television monitor (9) on the upper tier of the trolley for immediate display, and to the VCR (10) on the lower tier to be recorded for later playback. The resulting composite display comprised a moving cursor against a background representation of the stimulus, with a small inset in the bottom-right corner showing a digital time and metronome displays. The experimenter calibrated the system by adjusting the position of the cursor on the stimulus to match the exact position of the participant's gaze. This was done by manipulating two controls on the control box (8), one for each dimension, while watching the cursor position on the monitor. The electronic keyboard (4) sat on the desk in front of the participant, a battery-operated, 5-octave Roland with standard key-width, without touch sensitivity, and set on its basic pianoforte tone-colour. Hard copy of the stimulus was mounted on the document holder affixed to the side of the main desk. Participants sat on a height-adjustable office chair arranged so that a distance of 22-27cm could be comfortably achieved between their eyes and the stimulus. The monitor faced away from participants and was used only by the experimenter to verify and adjust the signal during the trials.

The setup offered two clear advantages over the equipment used in most previous studies of eye movement in music reading: the participant's scanpath could be viewed against the stimulus in real time and frame by frame; and the freedom of head and body movement permitted by the mounted video camera avoided the ecological problems caused by the constricting, disconcerting bite-bars and headrests that have typically been a necessary part of eye-tracking equipment (see Section 3.3.1). These advantages must be balanced against two shortcomings. One was that the background display on the monitor was rather smaller and of poorer resolution than was desirable for efficient and accurate registration. This drawback was partially addressed by reducing the allowable range of distance between participants' eyes and the stimulus from the planned 30-35cm to 22cm-27cm, in order to produce a larger
image on the television monitor. A second problem was that the display on the monitor was horizontally reversed, requiring the use of a mirror to re-reverse the image during registration.

The most significant deficiency in the equipment, however, was the fishbowl-like distortion of the cursor position in relation to the background score. The further a fixation lay from the position at which the cursor was calibrated with the participant's gaze, the greater the disparity between the positions of the actual and the represented fixation. Given the paucity of funding for equipment, it was not possible to have this fault corrected. Instead, the 'fishbowl' distortion was addressed primarily by developing and implementing an elaborate procedure for data registration, explained in Section 4.1.6.

### 4.1.2 Stimuli

Four practice stimuli and two trial stimuli were used in Stage I, all hymn settings in **four-part vocal style** adapted from Bartlett (1977). Four-part vocal style is one of two standard textural/notational modes used for hymn settings, the other being **pianoforte style**. Four-part vocal style is a standard tool for teaching music theory (Forte 1982:175; Aldwell & Schachter 1989:63). In this style, two notes of each four-note chord appear in the bass stave (the tenor and bass notes) and two appear in the treble stave (the soprano and alto notes); in each chord, the stems of notes on the same stave are arranged in opposite directions to avoid their jostling with each other. The second-lowest note, the tenor, is typically lower in pitch than would be the case in pianoforte style, and some grammatical rules are more strictly applied. Although originally conceived for vocal realisation, music notated in four-part vocal style is often performed on keyboard, as in this experiment. The two trial stimuli are shown in Figures 32 and 33 at approximately two-thirds of the size used in the experiment. The 32 beats of each stimulus are arranged in two double lines (an upper and a lower) across the page, as they were displayed in the trials.

Hymn settings in four-part vocal style were chosen for a number of reasons. First, distributing the information for each chord over two staves offered considerable advantages over a single-line melody. It is evident from studies as far back as Weaver (1943) that a keyboard score typically requires readers to fixate alternately on the treble and bass staves, or the reverse (see the illustration from Weaver's study in Figure 27 in Section 3.3.4). It can be assumed that unless the notation is unusually large or the eyes unusually far from the stimulus, it is impossible to fixate on both
staves simultaneously. This is because the distance between the two staves in a keyboard score exceeds the foveal diameter, and while readers might be capable of picking up information from large-scale patterning on the other stave through their peripheral vision, such as whether there are notes or rests at a particular point, peripheral input from the opposite stave is not useful enough of itself to avoid the need to fixate alternately on each stave. For these reasons the information on each stave, each half-chord as it were, can be considered as an optically isolated unit.

Figure 32 Trial Stimulus A, adapted from Bartlett (1977:624)

Figure 33 Trial Stimulus B, adapted from Bartlett (1977:669)
By contrast, melodies are typically notated on a single stave as a closely spaced, more or less continuous stream of visual information in roughly the same horizontal direction. Consequently, little or no information in a melody is optically isolated, making it difficult to determine how many times each piece of information has been exposed to the fovea. This is analogous to the difficulty of measuring same-word rightward refixation in language reading, noted in Section 3.3.6. The advantage of using keyboard music in Stage I was that it afforded a greater ability to clearly gauge the extent of both the O'Regan effect and refixation.

The second reason for choosing hymn settings was that they were likely to provide a means of successfully controlling the level of participants' familiarity with the stimuli. In this respect such a choice is consistent with the definition of sight reading for Stage I as first explained in Section 1.2.1: Sight reading is defined as the first reading of a musical text (1) with which the player has no actuality familiarity, (2) where both text and player have been chosen such that probability familiarity is minimised, and (3) where the level of newness encountered by the player in reading through the text is reasonably consistent. In respect of Point (1), there is such a large repertoire of this genre that there was a minimal risk that a participant would already have possessed actuality familiarity with the stimuli, that is, would already have read, performed or listened to them. In respect of Point (2), it was highly likely that all participants would have gained a basic experience of the genre, including playing examples of it on the piano. Accordingly, all participants were questioned before the trials to confirm that they had a basic familiarity with the genre, while the fact that they had no prior experience of the actual stimuli was confirmed after the trials. In respect of Point (3), hymn settings tend to be 'plain' in texture, comprising a succession of unornamented chords, that is, chords with minimal rhythmic activity and melodic character. It is also easy to segment hymn settings to avoid internal repetition of phrases. Thus music of this genre permitted a high level of internal consistency, and made it easy to avoid the 'bumps' and 'surprises' that might have complicated the eye movement data. A further reason for choosing hymn settings was their rhythmic uniformity, which was predicted to minimise action slips on the basis of the suggestion in Section 1.2.2 that rhythmic errors outweigh all other errors in music reading.

Despite the suitability of this body of hymn settings to the objectives of the experiment, the chosen stimuli still required modification. The tenor and bass parts were rewritten in the few places where they were more than an octave apart. This avoided the tricky re-allocation of the tenor part to the right hand and back again to
the default left hand that would otherwise have been required. The sudden demand for re-allocation of a part was likely to represent a challenge, even for skilled players, and may have encouraged them to glance down at their hands with resulting discontinuity in the flow of eye movement data. Potential stimuli were ruled ineligible if they included crossed parts, that is, bass moving above tenor or tenor moving above alto, because although this may not have challenged the musculoskeletal apparatus, it would have risked a sharp increase in the processing demand.

The two-beat cadential chords in the stimuli were regarded as 'dead' rhythmic areas whose relatively lower processing demands may have upset the calculation of one of the dependent variables, the mean number of pauses per chord. For this reason, these chords were activated by providing movement in at least one of the four parts in each beat. Such changes are quite standard ornamental procedures and were therefore regarded as causing little or no weakening of ecological validity. Quaver movement was reduced in three places in the lower three parts of the stimuli for the sake of uniformity of processing demand, but was left intact in the soprano line to preserve the melodic flow of the excerpt, again for ecological reasons.

Some commonplace differences between the two trial stimuli were regarded as permissible, indeed desirable, in the light of the need to present a certain level of newness in each successive reading during a trial. To this end, the two trial stimuli differed in key, periodicity (phrase length) and superficial aspects of localised ornamentation (quaver movement). The two trial stimuli were nonetheless presented in COUNTERBALANCED ORDER to account for any differences between the two passages that might have masked the effects of tempo. In other words, half of the participants read Trial Stimulus A first, and the other half read Trial Stimulus B first. The first trial stimulus, whether A or B, was read at 60MM, and the second at 120MM. This slow-fast order of tempos was used under the assumption that it would result in a lower overall rate of action slips compared to a fast-slow order.

4.1.3 Tempo and output dysfunction

The two contrasting tempos for the Stage I trials were established with reference to the discussion of functional and dysfunctional tempo areas for a particular reader of a particular stimulus (see Figure 23 in Section 3.3.2): the undercapacity area, the peak-use area, and the action slip area. These areas are separated by two thresholds, the action slip threshold and the undercapacity threshold. It was regarded as
advantageous for both tempos to lie within the peak-use area of most if not all participants, to avoid undesirable oculomotor and musculoskeletal effects. Accordingly, the fast tempo needed to be a small but significant distance below a participant’s action slip threshold, so that the apparatus could be observed working at a level that was close to capacity but without the risk of a significant level of action slips. Conversely, the slow tempo needed to be fast enough so as not to risk contaminating the data with significant levels of oculomotor ‘wandering’. In addition, both tempos needed to be sufficiently contrasting as to produce significant differences in the dependent variables. On the basis of the preliminary investigation, 60MM and 120MM were chosen as the tempos most likely to fulfil these conditions.

As noted on in Section 3.3.2, the action slip threshold for any particular reading situation is set at an arbitrary point based on what is an acceptable level of action slips given the objectives of the situation. In Stage I there was a need to guarantee a high level of accuracy in participants’ musical output to ensure that the information in the stimuli had been properly perceived and processed, and to avoid feedback from action slips of which the effect was hard to predict. To this end, the threshold of an acceptable level of action slips was set at 15%, the point beyond which was deemed to represent an impractical level of dysfunction in a training environment. In other words, if a player omitted or played incorrectly more than 15% of the 32 chords in either trial stimulus (that is, five or more chords), that player was disqualified as a participant.

To calculate the level of action slips, each chord was allocated one mark out of a possible 32 marks, one mark for each chord. One mark was deducted for a substantial action slip in a chord, and a half-mark for a minor action slip. Two minor action slips in the same chord counted as one mark, with one mark the maximum that could be deducted for any one chord. The sum of all deducted marks in a reading was then calculated to arrive at the level of action slips. Categories of action slip and the marks deducted when they occurred are set out in Table 7. It is necessary to explain two terms that appear in this table. ‘Voicing’ refers to the arrangement of the tones of a triad above a bass, specifically their presence, doubling or omission, and their octave placement. The most frequently occurring minor errors of voicing involve spacing and chord construction. An example of a minor harmonic error is the incorrect addition or omission of a seventh above a root, provided such an addition or omission makes musical sense. A note was deemed to be ‘out of time’ if it occurred more than 250ms either side of the metronome beat. This threshold was measured by
the experimenter to what might be described as a professional musician's accuracy. During the registration process, the required judgement was periodically tested against SightReader's SPLAY-THRESHOLD DEVICE (see Section 5.2.1).

<table>
<thead>
<tr>
<th>type of action slip</th>
<th>marks deducted</th>
</tr>
</thead>
<tbody>
<tr>
<td>minor errors of VOICING and/or chord construction</td>
<td>$\frac{1}{2}$</td>
</tr>
<tr>
<td>major errors of voicing and/or chord construction</td>
<td>1</td>
</tr>
<tr>
<td>one note smudged or omitted</td>
<td>$\frac{1}{2}$</td>
</tr>
<tr>
<td>two or more notes smudged or omitted</td>
<td>1</td>
</tr>
<tr>
<td>one or more notes clearly foreign to the triad</td>
<td>1</td>
</tr>
<tr>
<td>one note out of time, the others in time</td>
<td>$\frac{1}{2}$</td>
</tr>
<tr>
<td>two or more notes out of time</td>
<td>1</td>
</tr>
</tbody>
</table>

Table 7 Marks deducted for action slips

Overall the levels of action slips and dysfunctional chords were extremely low, confirming the advantage of using only skilled readers as participants, as discussed in the next section. Participant 5 was almost excluded for playing four chords out of time in each of the two trial stimuli; Participant 9 was also close to the action slip threshold for tempo in the second trial stimulus. During the first three practice readings two of the nine participants needed encouragement to blink less often, and three to play precisely in time with the metronome, one of these participants common to both groups.

4.1.4 Participants

Thirteen people were identified as potential participants; all were professionally known to the experimenter as good sight readers, and were initially approached over the telephone. Three were excluded on the basis of information supplied during the initial telephone interaction; one was later excluded on the basis of the global level of action slips during the trials. The nine remaining participants ranged in age from 22-49 years. Five were university lecturers in music, or high school or private studio music teachers, who could play the piano at a professional standard; two were professional repetiteurs (rehearsal coaches at the piano); and three were superior undergraduate majors in piano performance.

An initial conference was conducted either over the telephone at that time or at a subsequent time convenient to the potential participant. Potential participants were asked whether they were interested in participating as volunteer participants in a single half-hour experiment in which their eye movement would be tracked as they
sight read keyboard music. Those who expressed interest were then asked whether they would be willing to answer several questions about their keyboard playing, so they could be assessed for suitability. Those who expressed a willingness to proceed were asked whether they believed they could sight read with reasonable accuracy a straightforward four-part hymn setting in 3/4 or 4/4 time with up to three sharps or flats, keeping strictly to a metronome at 120MM. It was made clear to potential participants that in the experiment they would be required to do this without looking at their hands, that is, after establishing their initial hand position. It was also confirmed that each potential participant had good natural or contact-lens-corrected eyesight for music reading, since spectacles would not fit under the goggles.

Potential participants who were deemed suitable on the basis of their responses to the previous questions were then informed of a number of matters connected with their possible participation in the trials: that if they participated in the experiment their eye movement would be recorded and analysed, but all records of the trials, including the fact that they had participated, would remain strictly confidential; that it would be necessary to wear a cumbersome and mildly uncomfortable pair of goggles for up to 15 minutes as they read several unfamiliar but straightforward hymn settings on an electronic keyboard at 60MM and 120MM; that it might be necessary to exclude the data generated by their reading if their level of action slips were greater than a certain arbitrary level; that during the warm-up session the experimenter might need to encourage participants to keep time and to address the electronic keyboard effectively, and that participants would be given a minimum of information about the study until after the trials. Those who agreed to these conditions were engaged as participants, and a date and time for the trial negotiated.

4.1.5 Procedure for conducting the trials

Trial session

Introduction
The participant was greeted and asked to sit in the appropriate seat. The position of the seat was adjusted so that the keyboard was within easy range of the participant's hands. The position of the document holder was adjusted so that the participant was able to comfortably maintain a 22-27cm range between the eyes and the stimulus.
Briefing
During and after the adjustment, the participant was briefed about the procedures for fitting and adjusting the equipment, cooperating in the calibration of the machine, and sight reading the test stimuli. Specifically the participant was asked to read in a natural and relaxed way as they s/he would away from the experimental situation, reminded of the importance of playing to the metronome as accurately and consistently as possible, and in particular, of the need to play through to the end without 'rallentando', that is, slowing. The participant was also reminded that it would not be possible to look down at his/her hands once the goggles had been fitted.

First three practice readings
While the participant was still unencumbered by the equipment, s/he played through three practice readings to familiarise him/herself with the touch and audio output of the musical keyboard, the style and texture of the stimuli, and the task of playing the stimuli accurately to the metronome. The participant was asked to play through (i) Test Stimulus 1 'at a leisurely but consistent tempo', initially of the participant's own choice, but if it appeared to the experimenter to be inappropriately slow or fast, it was spontaneously modified by further verbal intervention by the experimenter. (ii) Test Stimulus 2 to the metronome at 60MM, and (iii) Test Stimulus 3 to the metronome at 120MM. (A summary of the tempo and headgear requirement for each reading appears in Table 8.)

Fitting and calibrating
(i) The bathing cap and goggles were fitted and adjusted.
(ii) A point on the stimulus display close to the centre of the display on a bass note in the upper system was chosen for the purpose of calibrating the cursor position. The participant was asked to gaze at this point. The cursor position was then calibrated by adjusting the controls for vertical and horizontal dimensions, to match the position of the cursor to the point at which the participant was gazing.

Fourth practice reading
The participant was asked to play through Test Stimulus 4 while wearing the headgear, with the metronome sounding at 60MM.
Stage I: observing span size and refixation in music reading

First trial reading
The participant played through Trial Stimulus A or B in counterbalanced order at 60MM.

Second trial reading
The participant played through the other trial stimulus at 120MM.

Equipment
The participant was assisted in taking off the headgear and thanked for participating.

Familiarity check
The participant was asked whether s/he had been familiar with either of the two trial stimuli. No familiarity was established for any of the participants. (If prior familiarity had been established, the entire data for the participant would have been deleted.)

<table>
<thead>
<tr>
<th>stimulus</th>
<th>tempo (MM)</th>
<th>headgear</th>
</tr>
</thead>
<tbody>
<tr>
<td>Test Stimulus 1</td>
<td>participant's choice</td>
<td>off</td>
</tr>
<tr>
<td>Test Stimulus 1</td>
<td>60</td>
<td>off</td>
</tr>
<tr>
<td>Test Stimulus 1</td>
<td>120</td>
<td>off</td>
</tr>
<tr>
<td>Test Stimulus 1</td>
<td>60</td>
<td>on</td>
</tr>
<tr>
<td>Trial Stimulus A or B</td>
<td>60</td>
<td>on</td>
</tr>
<tr>
<td>other trial stimulus</td>
<td>120</td>
<td>on</td>
</tr>
</tbody>
</table>

Table 8 Summary of tempo and equipment conditions during each reading

Individual readings

1 It was confirmed that the participant was comfortable and ready to start the reading.

2 The display on the monitor was checked to ensure that it was in order and that the video player was recording.

3 The metronome was turned on and the setting checked and adjusted if necessary.

4 A hard-copy of the stimulus was placed on the document holder and the participant was given approximately five seconds to pick up the basic information at the start of the excerpt (key, metre, etc).

5 The participant was counted in, synchronous with the appearance of cardinal seconds on the timer readout. For this purpose, the experimenter uttered 'one, two, three, four' or 'one, two, three', depending on whether the metre of the
stimulus and whether or not it began with an anacrusis. A hand-signal was made on camera in the participant’s peripheral visual area, synchronously with the first beat after the counting in.

6 During the reading of the four practice stimuli, the levels of blinking and action slips were observed. If the levels of blinking or errors of pitch were undesirably high during the reading of any of the four practice stimuli, the participant was encouraged in the moments after that reading to perform more functionally. If the level of temporal errors was undesirably high, the experimenter intervened during the reading by counting aloud the beats in time. There was no such intervention during the readings of the trial stimuli.

7 After the reading the metronome was switched off and the experimenter offered positive reinforcement to the participant with a view to maximising his/her performance throughout the trial.

4.1.6 Data registration

Terminology

‘Video playback’ is specified as one of two modes: ‘in real time’, that is, with the audio component, and ‘frame by frame’, which occurred necessarily without the audio component.

‘The fishbowl effect’ refers to the discrepancy between the actual location of a fixation and its representation by the cursor in relation to the score.

‘Chord-slots’ refer to the assumed position of each chord in the sequence of fixations represented by the cursor, and thus distorted by the fishbowl effect. The locations of the chord-slots outlined by the cursor were therefore out of phase with the actual chords as they appeared in the background of the display.

Procedure

Calibrating the video frames with the start of the performance

The recording of the start of the performance was repeatedly played back in real time until the start-time could be accurately matched with both (a) the nadir of the experimenter’s manual cue on the first beat of the participant’s performance and (b) the timer readout on the display. The frame at this point was registered as Frame 1, occurring at 0ms.

The point at which eye movement began as a cohesive sequential scanning pattern from the start of the score was established so that the eye-hand span at the start of
the performance could be calculated. This point always occurred before the performance start-time, invariably soon after the start of the experimenter's countdown. Accordingly, the moments before the start-time were repeatedly played back until the point was identified; negative time-values and frame-numbers were then assigned to this frame, and to the frames that recorded the start of subsequent pauses before the music started.

**Recording the data**

The scanpath was viewed over the entire upper double-stave several times in immediate succession both in real time and frame by frame, to assess and check the number of chord-slots in the cursor's representation of the scanpath.

The video records for the whole system were played through frame by frame, and the following information was recorded by hand for each new pause: (1) the timer readout, frame number, and pause duration in milliseconds, (2) the chord, numbered from the start of the system, and (3) the stave, treble or bass, on which the pause occurred.

The data were keyed into *Excel 5.0* for the Macintosh. The procedure described above was repeated for the lower double-stave, and the data were keyed into *Excel 5.0* for the Macintosh, joining the data with that from the upper double-stave to form a continuous stream from the start to the finish of the reading.

**Unusual saccadic timings, leftward refixations, and interruptions to the data**

*Frames that captured saccades in mid-flight*

Between 5% and 10% of saccades occurred at such a point in time that the image of the cursor was captured in flight between fixations, resulting in a blurred cursor-image. If the cursor-image was located between one quarter and three quarters of the way to the next fixation, the start of the next pause was registered as having occurred halfway between frames, that is, 20ms later than would otherwise have been the case. If the middle of the blurred cursor-image was located before one quarter of the way to the next fixation, the start of the next pause was registered as having occurred at the following frame. If the middle of the blurred cursor-image was located in the last quarter of the way to the next fixation, the start of the next pause was registered as having occurred on that frame itself. In this way, the effective resolution of the data was increased.
**Leftward refixations**

When registering the proportion of saccades that were leftward, all saccades involved in a leftward refixation were regarded as leftward, including rightward recovery saccades, since all are inevitable ramifications of the decision to refixate on information in a previous chord.

In calculating the number of fixations a chord received, the total number of times the chord had been paused on were counted. The sequence in which the fixations on a chord occurred, that is, whether or not they were leftward refixations, was therefore regarded as irrelevant.

*Interruptions to data caused by blinks and off-score glances*

Where the signal returned to the same location as before the interruption, the interruption and the fixations either side of the disruption were all regarded as part of the same fixation; accordingly their durations were combined.

Where the signal returned to another part of the stimulus, the interruption was regarded as part of the immediately preceding fixation; accordingly their durations were combined.

**4.2 Tempo and span size**

**4.2.1 Theoretical background**

Span size is commonly assumed to refer to the distance between the current fixation point of the eyes and the current position of the hands at any one point in a musical performance (assumed in such studies as Weaver 1943; Young 1971; Truitt et al 1997). Defining span size merely in relation to the music score, however, is inadequate for observations in which tempo is strictly accounted for. This is because span size can be measured in two quite different ways: in terms of the amount of musical material between these two points on the score, and in terms of the amount of time for which each piece of material remains in the span.

In this study the two ways of measuring span, the information-related and the time-related, are called respectively **load** and **latency**. At any point in a reading at a particular tempo, then, span size can be expressed in terms of load (for example 2 beats at 120MM) or latency (1000ms at 120MM). Load is defined as *the number of beats from and including the next beat to be played up to and including the most rightward beat that has thus far been fixated on*. Latency is defined as *the number of*
The existence of these two distinct ways of measuring span size is not mentioned anywhere in the literature. Most studies (for example Weaver 1943; Young 1971; Goolsby 1987) refer to span size implicitly in terms of load, that is, the number of notes or chords in the span, and make no reference to latency. This is surprising, since load and latency may represent more than mere theoretical constructs: they may well constitute somewhat different capacities in a musician’s working memory, the ability to hold a large amount of information in the span, and the ability to hold information in the span for a long period of time. Moreover it is not beyond the bounds of possibility that some musicians have a more efficient ability to increase load than latency, or vice versa, and that it may be possible to isolate and train each capacity with SightReader. Indeed, in the light of data from Stage I it was decided to include an exercise that isolated latency in the training sessions for participants in the experimental group in Stage II. Measuring load and latency in Stage I was also important for developing a strategy for manipulating SightReader’s tempo and span size settings in Stage II. But before discussing these matters further, it is necessary to explain the relationship between load, latency and tempo. This relationship, along with that between pause number and mean pause duration (discussed in Section 4.3), is the basis of what might be referred to as the mathematics of music reading.

The relationship between load, latency and tempo can be expressed in terms of several mathematical equations. It must first be pointed out that any given tempo implies a corresponding beat duration. Since tempo is expressed in terms of the number of beats per minute, and there are 60,000ms in a minute (that is, 1000ms x 60 seconds), beat duration is calculated by dividing 60,000 by the tempo in beats per minute (MM):

Equation 1: \[\text{beat duration} = \frac{60000}{\text{tempo}}\]

For example, if the tempo is 60MM, then beat duration is 1000ms (\(\frac{60000}{60}\)); if tempo is doubled to 120MM, beat duration is halved to 500ms (\(\frac{60000}{120}\)). Beat duration is the most basic way of relating load and latency, thus:

18Strictly speaking, under this definition latency moves through a constant background cycle of gradual reduction throughout each beat simply due to the effluxion of time, followed by a sudden increase in latency by the exact value of one beat as each new chord is struck. These rapid cyclical fluctuations in latency between arrivals at each successive chord are a highly predictable microfeature that is of no concern in this study; here only latency-on-arrival at each chord is at issue.
Equation 2: \[ \text{latency} = \text{load} \times \text{beat duration} \]

To exemplify Equation 2, if tempo were 120MM, beat duration were 500ms, and the eyes were three beats ahead of the hands, then latency would be 3 beats \( \times \) 500ms, or 1500ms. Combining Equations 1 and 2 makes it possible to relate latency directly to load using tempo rather than beat duration:

Equation 3: \[ \text{latency} = \frac{\text{load} \times 60000}{\text{tempo}} \]

and conversely:

Equation 4: \[ \text{load} = \frac{\text{latency} \times \text{tempo}}{60000} \]

It is evident from Equations 1 to 4 that the relationship between latency and load changes when tempo changes. For example, if latency is 1000ms and tempo is 60MM, load is \( \frac{1000 \times 60}{60000} \), or 1 beat. If, however, tempo is doubled to 120MM and latency remains unchanged at 1000ms, load becomes \( \frac{1000 \times 120}{60000} \), or 2 beats. To accommodate the doubling of tempo in this example, load has doubled (that is, by a factor of 2.0) and latency remains unchanged at 1000ms (a factor of 1.0). But the same result might just as well have been achieved conversely, with load unchanged at 1 beat (a factor of 1.0) rather than doubling it to 2 beats, and instead halving latency from 1000ms to 500ms (a factor of 0.5): in other words, at the original tempo, load was \( \frac{1000 \times 60}{60000} \), or 1 beat; now tempo is doubled to 120MM and load remains unchanged at 1 beat, and to maintain the equation latency is halved; accordingly \( \frac{500 \times 120}{60000} = 1 \) beat. This is a mathematical way of describing the simple fact that when tempo is doubled, chords will pass through the eye-hand span twice as quickly if the same number of chords are held within the span. As will be further discussed below, the changeable relationship between load and latency was a matter of prime importance to the experimental training sessions in Stage II, where participants' tempo and minimum span size were constantly altered. In addition, it was necessary to program Equations 3 and 4 into SightReader to enable it to calculate the precise timing of the blanking out of each chord in its display. This is fully explained in Section 5.2.1.

Whenever tempo changes, then, load and latency form a theoretical relationship in respect of the extent to which they each contribute to the necessary adaptation to
the new tempo. This will be referred to as the LOAD/LATENCY RELATIONSHIP. The two examples above represent extreme forms of adaptation to a new tempo, involving the complete adaptation by one variable and no adaptation at all by the other. But when tempo is altered, as it was in Stage I of this study, both load and latency, rather than the completely one-sided arrangement described above, may change to accommodate the new tempo, in which case their relationship is less extreme.

The load/latency relationships for all changes in tempo fall on a LOAD/LATENCY LINE \( y = \frac{x}{t} \), where \( y \) is the change in latency, \( x \) the change in load, and \( t \) the change in tempo between the two readings. For Stage I, \( t \) is a constant valued at 2.0 since tempo was doubled, so the line was described by \( y = \frac{x}{2} \). In other words, when tempo doubles, the change in latency will always be half the change in load, such that if latency increases by 60% (1.6), load shrinks by 20% (0.8). This point, (1.6, 0.8), like those of all possible relationships in Stage I, then, falls as a point on \( y = \frac{x}{2} \), shown in Figure 34 along with its main structural features.¹⁹

![Figure 34](image)

**Figure 34** The load/latency line \( y = \frac{x}{2} \) and its main features

¹⁹If the tempo had been tripled, the load/latency line would have been described by \( y = \frac{x}{3} \), or \( y = 2x \) if tempo had been halved. These lines, like \( y = \frac{x}{2} \), pass through the point \((0, 0)\) but are of different angles.
The point at which load and latency adapt equally, the EQUAL-CONTRIBUTION POINT is \((\sqrt{t}, \sqrt{\frac{t}{2}})\).\(^{20}\) This is represented here by the black dot in the middle of Figure 34, here at \((\sqrt{2}, \sqrt{\frac{1}{2}})\), or \((1.414, 0.708)\) since \(t = 2\). To the bottom-left of this point, latency changes more than load in the process of adaptation (red); to the top-right, load changes more than latency (green). The disparity in the relationship increases from the equal-contribution point up to the two SOLE-CONTRIBUTION POINTS: the point in the upper-left part of the figure at which only load adapts \((2.0, 1.0)\) and the point in the lower-right at which only latency adapts \((1.0, 0.5)\). These points represent the load/latency relationship for the two examples cited above.

In the light of this exegesis of the mathematics of span size, it is now appropriate to return to the third imperative in music reading that was proposed in Section 3.3.12:

The third imperative is to maintain a span size that is appropriate to the reading conditions. The span must not be so small that there is insufficient time to perceive visual input and process it into musculoskeletal commands; it must not be so large that the capacity of the memory system to store and process information is exceeded.

What, then, is the appropriate mix of load and latency in the task of arriving at a span size of appropriate size at a new tempo? Hypothesis 1 posited a general principle that when tempo is altered in the sight reading of excerpts of similar keyboard music, the mean load/latency relationship would lie close to the equal-contribution point. The reasoning behind Hypothesis 1 was that the further a relationship falls from the equal-contribution point, the less room there is for the apparatus to shift the relationship further in the same direction should this be subsequently required. In other words, a relationship that falls close to the equal-contribution point provides the greatest likelihood that the apparatus will easily be able to adapt to a further change in the reading condition.

The need for such flexibility and adaptability in music reading should not be underestimated, since music reading typically presents constantly fluctuating demands on the memory system. There is a large body of evidence that the workload imposed on the memory system is reflected in both span size and other eye-movement patterns: that reader-skill (Quantz 1898) and stimulus difficulty and familiarity (Buswell 1920; Lawson 1961; Morton 1964) can affect span size in

---

\(^{20}\)This is derived from the basic algebraic fact that \(t = \sqrt{\frac{t^2}{1}}\).
4 Stage I: observing tempo, span size and refixation in music reading

language reading; and that the same variables, with the addition of tempo, affect span size in music (Goolsby 1987; Smith 1988).

**Hypothesis 1**

When tempo is increased, the mean load/latency relationship will be in the vicinity of the equal-contribution point.

In addition, Hypothesis 1 is relevant to Stage II in that it may shed light on how to manipulate the tempo and span settings on *SightReader* in an educationally and empirically effective way. For example, if *SightReader*’s tempo setting had been altered for successive readings while the span setting had been left untouched, the load/latency relationship would have been skewed entirely in the direction of either the red or green *sole-contribution points* in Figure 34.21 This would have imposed an entirely one-sided adaptation to the new tempo, a potentially contaminating factor. A simple solution would have been to apply the same load/latency relationship to all tempo changes; the equal-contribution point would thus have been considered for use as the default relationship for all participants in Stage II if Hypothesis 1 were supported in Stage I. If, however, Stage I showed participants’ load/latency relationships to be consistently and strongly skewed towards either the green or red areas in Figure 36, a default skewed in that direction may have been considered to be more appropriate in Stage II.

**4.2.2 Results and discussion**

Hypothesis 1 was supported by the data, which are displayed in the last two columns of Table 9 and in Figure 35. The mean load/latency relationship was (1.337, 0.668), SD (0.292, 0.146).22 In other words, when tempo was doubled, on average participants’ span size adjusted so that just over a third more chords were held in it (33.7% more, to be precise) but each chord passed through the span almost exactly a third more quickly (66.8% of the mean time previously held in the span). The mean load/latency relationship was therefore in the vicinity of the equal-contribution point (1.414, 0.707), in the red area of the load/latency line slightly favouring a reliance on reducing latency over increasing load to adapt to the faster throughput of chords.

---

21 Whether red or green would depend on whether span size is calibrated in beats or milliseconds. Either way, one factor would bear all of the adaptation to the new tempo and the other would bear none.

22 Standard deviation (SD) is a basic indication of the dispersion of values around the mean. Approximately 68% of values typically lie one SD above and below the mean. The remaining 32% of values lie yet further above and below the mean. The greater the SD, the more dispersed the values. All data in this study were analysed using the computer application *SPSS 6.1*. 

126
Despite the overall support for Hypothesis 1, what was surprising about this result was the extent of the dispersion of individual participants’ load/latency relationships. The relationship of Participant 1 (0.510, 1.020) was almost exactly on the sole-contribution point for latency, and for Participant 5 (0.459, 0.918) was beyond that point. Participant 5 thus reduced load more than the 50% that was necessary to completely adapt to the doubling of tempo, such that in addition latency needed to be reduced by nearly 9%, an apparently extreme and costly strategy. Two other relationships, those of Participants 3 and 8, also fell in the ‘red’ zone by a significant distance from the equal-contribution point. Against this, the relationships of two Participants, 6 and 9, fell on the ‘green’ side of the equal-contribution point; one, that of Participant 2 (0.956, 1.912) was close to the other sole-contribution point. Altogether only three of the nine relationships could be said to have fallen in the vicinity of the equal-contribution point, while three were at moderate distance and three were at an extreme distance from it.

<table>
<thead>
<tr>
<th>partici</th>
<th>slow tempo</th>
<th>fast tempo</th>
<th>load/latency relationship</th>
</tr>
</thead>
<tbody>
<tr>
<td>mean</td>
<td>load (beats)</td>
<td>latency (ms)</td>
<td>mean load (beats)</td>
</tr>
<tr>
<td>1</td>
<td>1.365</td>
<td>1365</td>
<td>1.392</td>
</tr>
<tr>
<td>2</td>
<td>1.629</td>
<td>1629</td>
<td>3.114</td>
</tr>
<tr>
<td>3</td>
<td>1.219</td>
<td>1219</td>
<td>1.548</td>
</tr>
<tr>
<td>4</td>
<td>1.497</td>
<td>1497</td>
<td>2.034</td>
</tr>
<tr>
<td>5</td>
<td>1.164</td>
<td>1164</td>
<td>1.068</td>
</tr>
<tr>
<td>6</td>
<td>1.205</td>
<td>1205</td>
<td>1.748</td>
</tr>
<tr>
<td>7</td>
<td>1.434</td>
<td>1434</td>
<td>1.994</td>
</tr>
<tr>
<td>8</td>
<td>1.123</td>
<td>1123</td>
<td>1.336</td>
</tr>
<tr>
<td>9</td>
<td>1.306</td>
<td>1306</td>
<td>1.986</td>
</tr>
<tr>
<td>mean</td>
<td>1.327</td>
<td>1327</td>
<td>1.802</td>
</tr>
<tr>
<td>stdev</td>
<td>0.169</td>
<td>169</td>
<td>0.596</td>
</tr>
</tbody>
</table>

Table 9 Mean load and latency at each tempo, and the resulting load/latency relationships
Stage 1: observing tempo, span size and refixation in music reading

Figure 35  The position of each subject’s load-latency relationship on the line $y = \frac{x}{2}$

Figure 36  Mean latency at slow and fast tempos
The ‘within-participants’ data for mean load and latency that produced the relationships discussed above are also presented in Table 9; the data for latency are illustrated graphically in Figure 36. Mean latency for all participants fell from 1327ms (SD=169ms) at slow tempo to 901ms (SD=298ms) at fast tempo, $t(8)=7.20$, $p=0.002$. Mean load for all participants, directly arising from the values for latency, increased from 1.327 beats, that is, $\frac{1327 \text{ms}}{\text{beat duration}}$, or 1327ms (SD=0.169) at slow tempo to 1.802 beats, that is, $\frac{901 \text{ms}}{500 \text{ms}}$ (SD=0.596) at fast tempo, $t(8)=3.1$, $p=0.015$. Only one of the nine participants (Participant 2) failed to reflect the general fall in mean latency and rise in mean load. Participants’ mean latencies ranged from 1123ms to 1629ms at slow tempo, and from 534ms to 1557ms at fast tempo; loads ranged from 1.123 beats to 1.629 beats at the slow tempo and from 1.170 beats to 2.034 beats at the fast tempo, with one outlier, Participant 2, on an uncommonly large 3.19 beats. Surprisingly, the dispersion of values for load and latency were more dispersed at the fast tempo than at the slow tempo.

Such a wide dispersion of values for span size at both tempos, and of the load/latency relationships, is surprising considering that it occurred under conditions in which all participants were highly skilled and managed to produce almost exactly the same musical output. Several explanations are possible. The first is that the dispersion may have arisen from individuals’ preference for using load or latency to adapt to a change in tempo, whether such preference is acquired through experience, including training, or whether it is a ‘deep’ capacity of the memory system, or a combination of both. In other words, individuals may not necessarily possess equal capacity or preference in respect of load and latency, and that inequality may be strong enough to override any strategic advantage in striking a load/latency relationship close to the equal-contribution point. Participant 2 may be ‘naturally’ better at holding a large load at a fast tempo than that of Participants 1 and 5, whose strengths may lie instead in processing a smaller load more quickly. A second possible reason for the dispersion is that it arose from random, stress-induced or arbitrary factors brought about by the conditions of the experiment, and that such ‘noise’ in the data was pronounced because of the small sample size. A third possible reason is that individual differences may have resulted from the

---

23 The expression ‘$p = 0.002$’ indicates that there is a 0.002 (one in 500) chance that the difference between the means of subjects’ latencies at different tempos was due to factors other than the difference in tempo. The threshold for significance for all data in this study was $p<0.05$. In other words, there needed to be less than a one in twenty chance that the difference between the two means arose from extraneous factors. In psychology it is also normal to express the ‘degrees of freedom’, here 8, and a $t$-value that must be evaluated against a table of sampling distributions of $t$ to ascertain significance. The $p$-value, however, is sufficient to indicate whether the data are significant. All calculations in this study were performed using the statistical package SPSS 6.1.
superimposition of natural strengths and weaknesses on top of the random, stress-indued or arbitrary (a combination of (1) and (2), as it were).

At this point it would appear unlikely that the individual differences that were observed resulted purely from random factors, although numerous tests on the same participants would have been required to rule out this possibility. Accordingly, the following conclusions were drawn from this part of Stage I: First, the fact that on average the mean load latency relationship was close to the equal-contribution point and that on average there was no significant bias towards relying on load or latency to accommodate tempo change suggested that it would not be necessary to build such a bias into the manipulation of SightReader's tempo and span settings in the experimental training sessions in Stage II; any fears that there would be significant interference between span and tempo settings in Stage II were thus regarded as unnecessary. Second, it appeared likely fluent sight reading can occur over a significant range of span sizes, and that participants in the experimental group in Stage II would need to begin their training at a variety of span settings. This was in fact the case, as explained in Section 5.2. Third, the divergent values for load and latency among the participants suggested that it might be useful to try to isolate each variable with SightReader in the experimental training sessions in Stage II. This was achieved by setting tempo low and minimum span size high at a certain point in the training sessions in Stage II, and thus forcing the load/latency relationship closer to the sole-contribution point (1.0, 0.5) in Figure 36, as more fully explained in Section 5.2.6.

The data from this part of Stage I are consistent with the findings of Weaver (1943) and Young (1971) that eye-hand span is between one and two chords in the sight reading of keyboard music, although the association may be coincidental considering the loose tempo regime that pertained in those studies. When the mean values for load in Stage I are factored by the four notes in each beat of this texture, the mean load of 1.327 and 1.802 beats observed here comes to 5.3 and 7.2 notes respectively. This resonates with Goolsby's (1987) finding that the eye-hand span is a mean of 5-7 notes in the skilled sight reading of a melody, and Sloboda's (1974) finding of a mean perceptual span of 7-8 notes in melody reading. But again, the association may not be significant, since the processing of single-line melodies is probably different from that of four-part vertical chords and, in respect of Sloboda's finding, comprised peripheral as well as foveal perception and processing.
4.3 Tempo, refixation and the O'Regan effect

4.3.1 Theoretical background

Refixation has already been discussed in relation to language reading (Section 3.2.6) and music reading (Section 3.3.10). Refixation in both types of reading is a poorly understood phenomenon, although Kennedy (1992) and her colleagues have made a valuable contribution to the area, advancing the understanding of refixation as the refreshment through external input of information that is stored and processed in working memory. The nature and function of refixation in music reading is similarly unresolved, even though it has been a dependent variable in several previous studies of eye movement studies (for example Young 1971; Goolsby 1987; Smith 1988).

It is evident from these studies, as well as from the data from Stage I of this study, that there are two distinct types of external refreshment in the reading of keyboard music: vertical refixation, resulting from either two or more vertical saccades within a chord, and horizontal refixation, in which there are one or more leftward saccades to a previous chord and back again to recover or pass over the former position. These types of refixation were schematically illustrated in Figures 20-22 in Section 3.2.6. In addition, there would appear to be a musical equivalent of rightward same-word refixation that was identified in language reading by Kennedy (1992:153) as reviewed in the same section. The last type of refixation was not of concern in this study, since it would appear to be impossible to isolate and measure.

Apart from investigating span size, Stage I was used to observe the level of external leftward refixation under the normal condition to assess how disruptive SightReader would be to the music reading process in Stage II. If a significant proportion of saccades in Stage I were shown to be leftward, this would have been regarded as a potential problem for some users of SightReader; on the other hand, if the presence of leftward refixation were insignificant in Stage I, this behaviour would have been considered to be less likely to contribute to problems in using SightReader or in interpreting the results for Stage II.

The data from Stage I also afforded an opportunity to investigate the theoretical likelihood that refixation in music reading is part of a tempo-adaptive mechanism in which the O'Regan effect plays a complementary role. It is argued in this study that refixation and the O'Regan effect together are the chief mechanism of determining the relationship between the number of pauses and their mean duration in a reading at any particular tempo. In order to pursue this argument, the investigation into refixation in music reading starts with an examination of the theory behind the
number/duration relationship. Just as the previous section began by presenting equations that relate load, latency and tempo, here the first task is to directly relate tempo with two other variables: the number of pauses and their mean duration. Equation 5 has already been substantively discussed in Section 3.3.3 and illustrated in Figure 24.

**Equation 5:** number of pauses \( \times \) mean pause duration = total reading duration

Since total reading duration is inversely proportional to tempo, it follows that tempo is inversely proportional to the number of pauses, times their duration. Another way of expressing this is to use letters to represent the change in each value from one reading to the next, as was done in the previous section. The expression is now one of equality rather than proportion:

**Equation 6:** \[ nd = \frac{1}{t} \]

where \( n \) is the change in the number of pauses, \( d \) is the change in their mean duration, and \( t \) the change in tempo. In Stage I, tempo was doubled from 60MM to 120MM in the reading of the second trial stimulus, so the relevant equation is:

**Equation 7:** \[ nd = \frac{1}{2} \]

This equation describes what will be referred to as a NUMBER/DURATION CURVE, just as the line \( y = \frac{x}{2} \) in the previous section described the load/latency line for a doubling of tempo. Whereas the load/latency relationship is linear, intersecting the point \((0, 0)\), pause number and mean duration form a hyperbolic relationship, in that neither \( n \) nor \( d \) theoretically ever reaches the value of 0. Figure 37 illustrates this curve along with its key structural features, the equal-contribution point (black), and the sole-contribution points at which number alone adapts (blue), and duration alone adapts (orange). At the equal-contribution point \( n \) equals \( d \) (that is, both variables change equally to accommodate the change in tempo); thus \( n^2 \) and \( d^2 \) both equal \( \frac{1}{2} \), and \( n \) and \( d \) both equal \( \sqrt{\frac{L}{2}} \), or 0.708. The equal-contribution point is therefore at \((0.708, 0.708)\). The sole-contribution points are at \((1.0, 0.5)\) and \((0.5, 1.0)\) respectively.
Stage 1: observing span size and refixation in music reading

Figure 37  Schematic illustration of the number/duration curve \( nd = \frac{1}{2} \)

Like Hypothesis 1, Hypothesis 2 predicts that on average the two variables, in this case pause number and mean pause duration, will contribute roughly equally to a forced change in tempo. Hypothesis 2 is based on the same reasoning as Hypothesis 1: a number/duration relationship that lies close to the equal-contribution point allows the apparatus the greatest flexibility to adapt to further changes in reading conditions.

**Hypothesis 2**

*Tempo, pause number and mean pause duration*

When tempo is increased, the mean number/duration relationship will be in the vicinity of the equal-contribution point.

The theoretical background to Hypothesis 2, however, is more involved than this implies, due to the connection noted earlier between refixation, the O'Regan effect and the three variables at issue in Hypothesis 2. Every time a refixation occurs at a fixed tempo, more pauses must fit into the same total reading duration; to compensate for this increase in number, the pauses must be of shorter mean duration. Conversely, every time the O'Regan effect is at play and a piece of information on the score is bypassed, fewer pauses are used in the scanpath and their mean duration will be longer. Refixation and the bypassing of information on the score, at least a
keyboard score in which the parcels of information are optically isolated, determine the number/duration relationship. Logically, the number/duration relationship struck between an existing and a new tempo will depend on the change in the levels of refixation and the O’Regan effect.

In the previous section it was argued that a load/latency relationship located on or near one of the sole-contribution points in Figure 34 would be disadvantageous in that it renders the apparatus less flexible in the face of further changes in reading conditions. This underlay the hypothesis that the equal-contribution point is a safer and more likely strategy. In stark contrast to this, the number/duration relationship located at (0.5, 1.0) in Figure 37, the sole-contribution point at which pause duration alone carries the burden of adaptation to the new tempo and the number of pauses is unchanged, has a distinct theoretical advantage over any other point on the number/duration curve. This is because (0.5, 1.0) is the only point on the number/duration line that permits the reader to maintain a standardised scanpath irrespective of tempo. On logical grounds a single, unchanging scanpath design would appear to be a simpler, less costly approach to designing the eyes’ journey across the score. To strike a number/duration relationship anywhere else but at the sole-contribution point for mean duration ((1.0, 0.5) in Figure 37) sacrifices the saving that would be afforded by a fixed, predictable scanpath irrespective of tempo.

Later in this section it will be proposed that the O’Regan effect acts on pause number in order to reduce what would otherwise be the required variation in mean pause duration when changing tempo; that refixation performs this role as tempo slows, and the O’Regan effect performs it as tempo increases. The effect would be a shifting of the number/duration relationship closer than it would otherwise have been to the equal-contribution point in Figure 37. In other words, both behaviours, by changing the number of pauses, act as buffers to allow greater stability of pause duration when tempo changes. The other effect, it will be proposed, is to regulate the progress of the eyes across the score in relation to the demands of tempo.

All the evidence, however, suggests that scanpaths in music reading are not fixed, orderly and predictable, but rather are inherently changeable and exhibit a certain ragged, ad-hoc characteristic (for example Weaver 1943, from which selected scanpaths were quoted in Figure 27 in Section 3.3.8 of this dissertation; Goolsby 1987; Smith 1988; and Truitt et al 1997). Added to this, Smith (1988) observed a fall in both fixation duration and number in melody reading at faster tempos, although the reliability of Smith’s data may have suffered from tempo/skill/action-slip problems.
In the light of these points, Hypothesis 2 is in effect predicting that to sacrifice a standardized approach to scanpath design (by using only pause duration to adapt to tempo change) is preferable to an extreme number/duration relationship.

There are two further implications of Hypothesis 2. First, finding a functional link between tempo and scanpath in Stage I would challenge the notion that scanpath reflects the horizontal or vertical emphasis of the musical texture, as proposed by Sloboda (1985) and Weaver (1943). Second, a significant association between tempo and pause number and mean duration would indicate that the tempo/skill/action slip problems are likely to have caused considerable contamination of the data in numerous previous studies into eye movement in music reading (see the review in Section 3.3).

Hypotheses 3 and 4 arose from the need to know the extent to which the natural patterning of leftward refixation would be disrupted by SightReader. As previously noted in Section 3.3.10, refixation is costly in that it inhibits progress across the page. Leftward refixation is particularly expensive, since it requires not only a leftward saccade but a subsequent rightward saccade to return to or pass back over the original point of reversal. Vertical refixation, by contrast, requires no recovery saccades; it allows the eyes to remain at all times within easy reach of the next new chord to the right, and requires less investment of time and sacrifice of progress while at the same time still refreshing information within the eye-hand span.24 For these reasons leftward refixation is likely to be significantly less common than vertical refixation at both tempos. While Hypothesis 3 drew on the imperative to make sufficient progress across the page, it was the only hypothesis that did not directly concern the influence of a doubling of tempo.

Hypothesis 3

*Rates of leftward vs vertical refixation*

At both tempos the number of saccades involved in leftward reinspection (including 'recovery saccades') will be lower than the number of saccades involved in vertical reinspection.

Hypothesis 4 concerned both the connection between the number/duration relationship, refixation and the O'Regan effect, and the oculomotor imperatives to progress across the score at a pace that is appropriate to the tempo and to service

---

24SightReader’s process of erasing successive chords on the display does not directly impede vertical refixation, but it makes leftward refixation extremely difficult. In order to look back and view a chord to the left before it disappears into empty stave-lines, the eyes need to be ahead of the hands by the minimum span size that has been set. If indeed the eyes are far enough ahead to be able to refixate to the left, there is logically still room for the upward adjustment of the span setting.
the requirements of working memory (see Section 3.3.12). Hypothesis 4 was tested by using a stimulus-oriented paradigm that involved categorising the 32 chords of each stimulus into those that were (1) fixated on only once, in which case the O’Regan effect can be assumed to be at work, since one of the two staves has been ignored; (2) fixated on twice, in which case it can be assumed that each stave has been visited once, and only once, and (3) fixated on three times or more, in which case at least one of the two staves must have been visited more than once. By comparing these proportions, it would be possible to see this tempo-adaptive mechanism at work in terms of way each piece of information in the stimuli was treated.

**Hypothesis 4**

Tempo, refixation and the O’Regan effect

When the tempo is faster, the mean rates of both vertical and leftward refixation will be lower and the mean level of the O’Regan effect will be higher.

The primary rationale behind Hypothesis 4 was this: the costs of both types of refixation suggest that their rates would fall as tempo increased, along with a correspondingly greater need to make progress across the page. Similarly, the O’Regan effect was predicted to become more frequent as tempo increased, since it would have the opposite effect to refixation in terms of progress across the stimulus. An additional reason for Hypothesis 4 would have been provided if Hypothesis 1 was confirmed, that is, that an increase in tempo reduces latency, which in turn may have reduced the need to refresh the information that was being stored and processed in working memory. Confirmation of Hypothesis 4 would also resonate with Goolsby’s (1987, 1994) and Smith’s (1988) findings of reduced levels of leftward refixation in the sight reading of melodies at increased tempos, as reviewed in Section 3.3.10.

It was hypothesised that in reading music at the fast tempo, two factors may be associated with the skipping over of information on the score for which nevertheless a corresponding functional musculoskeletal response: (1) a more urgent need to make progress across the page, and (2) a likely reduction in the need to refixate, brought about by the reduction in latency predicted in Hypothesis 1. If Hypothesis 1 were confirmed, it could have been assumed that mean latency increases as tempo slows. At the same time there is less pressure to make progress across the page, more room for costly refixation, and the increase in pause duration is limited by additional refixations. At a faster tempo the situation is reversed: latency is smaller thus requiring less refreshment, and the need to progress across the page is more urgent. At a faster tempo, skipping over information to make swifter progress across the
4.3.2 Results and discussion

Hypothesis 2 was strongly confirmed by the data: when tempo doubled, both the number of pauses per chord and the mean pause duration fell such that the mean number/duration relationship was \((0.705, 0.709)\), close to the equal-contribution point of \((0.708, 0.708)\). The standard deviation of the mean relationship was \((0.138, 0.118)\). It appears that stability of scanpath (tenable only when the number/duration relationship is located at \((0.5, 1.0)\), one of the sole-contribution points as illustrated on Figure 35 in the previous section) was sacrificed for the sake of maintaining a comparatively stable mean pause duration. A repeated-measures \(t\)-test was carried out on the data. Table 10 shows that mean pause duration was 368ms at slow tempo (SD=45ms) falling to 263ms at fast tempo (SD=42ms), \(t(8)=5.75, p<0.001\), that is, with an extremely small probability that the ‘slow’ and ‘fast’ samples observed were not representative of their respective populations. Correspondingly, the mean number of pauses per chord fell from a mean of 2.75 (SD=0.31) at slow tempo to 1.94 (SD=0.27) at fast tempo, \(t(8)=6.97, p<0.001\). Figure 39 shows mean pause durations for each subject at each tempo.

<table>
<thead>
<tr>
<th>subject</th>
<th>slow tempo</th>
<th>fast tempo</th>
<th>number/duration relationships</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>mean no. of pauses/duration (ms)</td>
<td>mean no. of pauses/duration (ms)</td>
<td>fast/slow factor: number</td>
</tr>
<tr>
<td>1</td>
<td>2.611</td>
<td>1.634</td>
<td>0.626</td>
</tr>
<tr>
<td>2</td>
<td>2.725</td>
<td>1.812</td>
<td>0.665</td>
</tr>
<tr>
<td>3</td>
<td>2.967</td>
<td>2.183</td>
<td>0.736</td>
</tr>
<tr>
<td>4</td>
<td>2.551</td>
<td>1.429</td>
<td>0.560</td>
</tr>
<tr>
<td>5</td>
<td>2.681</td>
<td>1.894</td>
<td>0.706</td>
</tr>
<tr>
<td>6</td>
<td>2.865</td>
<td>2.155</td>
<td>0.752</td>
</tr>
<tr>
<td>7</td>
<td>2.141</td>
<td>2.183</td>
<td>1.020</td>
</tr>
<tr>
<td>8</td>
<td>3.155</td>
<td>2.058</td>
<td>0.652</td>
</tr>
<tr>
<td>9</td>
<td>3.058</td>
<td>2.101</td>
<td>0.687</td>
</tr>
<tr>
<td>mean</td>
<td>2.750</td>
<td>1.939</td>
<td>0.705</td>
</tr>
<tr>
<td>stdev</td>
<td>0.304</td>
<td>0.280</td>
<td>0.138</td>
</tr>
</tbody>
</table>

Table 10 Mean number of pauses per chord and mean pause duration at each tempo, and the resulting number/duration relationships.
Figure 38 shows the number/duration curve $nd = \frac{1}{2}$, where $n$ is the change in pause number, $d$ is the change in pause duration, and $2$ is a constant representing the doubling of tempo. The average relationship is represented by the blue point at $(0.705, 0.709)$. Like the load/latency relationships displayed in Figure 35, the number/duration relationships showed a diversity among the subjects even though the mean relationship, and five to six of the nine relationships, were close to the equal-contribution point. Again, one subject (Subject 7) exhibited a highly eccentric strategy by producing a number/duration relationship $(1.020, 0.490)$ that lay beyond the sole-contribution point for duration; in other words, at the fast tempo this subject's mean pause durations were less than half their value at the slow tempo, necessitating a small reduction in number, strongly defying the hypothesis. On the other side, the relationship for Subject 4 was significantly dispersed towards the opposite sole-contribution point, at $(0.560, 0.893)$.

The confirmation of Hypothesis 2 challenges the notion that scanpath (largely or solely) reflects the horizontal or vertical emphasis of the musical texture, as proposed by Sloboda (1985) and Weaver (1943), since it has now been shown that the horizontal and vertical dimensions of scanpath are highly dependent on tempo. Such confirmation also strengthens the case that the tempo/skill/action slip fallacy is likely to have caused considerable contamination of the data in numerous previous studies into eye movement in music reading.

Hypothesis 3 predicted that at both tempos the rate of leftward refixation would be lower than that of vertical refixation. This was borne out by the data, as set out in Table 11 and graphically illustrated in Figure 40. At slow tempo, a mean of 23.13% of saccades were involved in vertical refixation ($SD=5.76\%$) compared to a mean of 5.05% in leftward refixation ($SD=4.81\%$) $t(8)=5.37, p<0.001$. At fast tempo, the rates were 8.15% (SD=4.41%) versus 2.41% (SD=2.37%), $t(8)=3.32, p=0.011$. These significant differences occurred even though recovery saccades were included in the count for leftward refixations. Hypothesis 3 was thus confirmed. In addition, it is noted that at slow tempo there was a significant negative correlation between leftward and vertical refixation ($r=-0.826, p=0.006$). This suggests that subjects were using leftward and vertical refixation for the same purpose, but favouring one or the other type, such that relatively high levels of one type were associated with relatively low levels of the other. This phenomenon may not have been evident in the data for fast tempo because of a ‘floor’ effect, with a considerable proportion of results for leftward refixation at zero ($r=-0.84, p=0.830$).
Stage 1: observing span size and refixation in music reading.

Figure 38  The number/duration relationships on the curve $nd = \frac{1}{2}$

Figure 39  Mean pause duration in ms for each subject, at slow tempo (blue) and fast tempo (yellow)
Stage 1: observing span size and refixation in music reading

<table>
<thead>
<tr>
<th>subject</th>
<th>slow</th>
<th>fast</th>
<th>slow</th>
<th>fast</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>18.78</td>
<td>8.32</td>
<td>10.80</td>
<td>0.00</td>
</tr>
<tr>
<td>2</td>
<td>20.63</td>
<td>6.12</td>
<td>8.84</td>
<td>2.02</td>
</tr>
<tr>
<td>3</td>
<td>25.42</td>
<td>6.75</td>
<td>2.04</td>
<td>0.00</td>
</tr>
<tr>
<td>4</td>
<td>14.07</td>
<td>4.71</td>
<td>13.80</td>
<td>2.33</td>
</tr>
<tr>
<td>5</td>
<td>26.05</td>
<td>3.76</td>
<td>2.40</td>
<td>5.75</td>
</tr>
<tr>
<td>6</td>
<td>26.59</td>
<td>9.25</td>
<td>2.25</td>
<td>3.70</td>
</tr>
<tr>
<td>7</td>
<td>16.94</td>
<td>18.75</td>
<td>3.21</td>
<td>1.76</td>
</tr>
<tr>
<td>8</td>
<td>31.18</td>
<td>6.29</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>9</td>
<td>28.50</td>
<td>9.38</td>
<td>2.12</td>
<td>6.16</td>
</tr>
</tbody>
</table>

Table 11 Rates of vertical and leftward refixation

Hypothesis 4 predicted among other things that the rate of both types of refixation would be reduced at the fast tempo. While the means do bear this out, the differences were significant only for vertical refixation, where only Subject 7, the outlier, went against the trend. For vertical refixation, 23.12% (SD=5.75%) at slow tempo was reduced to 8.15% (SD=4.41%) at fast tempo, t(8)=5.56, p<0.001, and for leftward refixations, 5.05% (SD=4.81%) at slow tempo was reduced to 2.41% (SD=2.37%) at fast tempo, t(8)=1.37, p=0.209. While there were sharp falls in leftward fixation for three subjects when tempo was doubled, and mild falls for another two subjects, three of the remaining four subjects increased their rates, thus increasing the variance for this comparison such that the difference in the means was below the threshold for significance. Accordingly, this part of Hypothesis 4 was only partially confirmed. There are at least two possible reasons for the unexpected increase in leftward refixation. First, the increases were small and from a low base-line, for example, from 2.40% to 5.75% (Subject 5) and from 2.25% to 3.70% (Subject 6). Second, such increases were in all three cases vastly overshadowed by sharp falls in vertical refixation. In summary, it is still not clear why leftward refixation was not uniformly influenced by a doubling of tempo, however the prevention of leftward refixation was by this stage known to be unlikely to present a serious problem to the user of SightReader, since it appears to occupy only a small fraction of saccades.
Hypothesis 4 also tested the notion that refixation and the O’Regan effect both play a part in shifting the number/duration relationship towards the number-only point in Figure 37 (1.0, 0.5), and are thus part of an interdependent mechanism for controlling progress across the score and the rate of refreshment of the contents of working memory. As shown in Table 12 and Figure 41, the total number of chords in each reading was classified according to the number of fixations it received: one, two, or three or more pauses; this was done at both slow and fast tempos. Chords receiving only one pause (green in Figure 41) involved the O’Regan effect, since half of their information (two notes, in either treble or bass clef) was omitted from the sequence of fixations yet was functionally performed. The information in chords receiving two pauses (yellow), one fixation on each stave, was inspected once, and once only. These chords thus involved neither behavioural phenomenon; all their information was fixated on, once and once only. Chords receiving three pauses (red) received a second inspection on one stave, except for a small number of cases in which both staves were reinspected (see Footnote 25).

25In practice, only five of a total of 288 possible chord inspections involved four pauses, two involved five pauses, and one involved seven pauses, all at slow tempo. At fast tempo, no chord was inspected more than three times.
### Table 12

<table>
<thead>
<tr>
<th>Subject</th>
<th>Slow Tempo</th>
<th></th>
<th></th>
<th>Fast Tempo</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>% of Chords Receiving 1 Pause</td>
<td>% of Chords Receiving 2 Pauses</td>
<td>% of Chords Receiving ≥3 Pauses</td>
<td>% of Chords Receiving 1 Pause</td>
<td>% of Chords Receiving 2 Pauses</td>
<td>% of Chords Receiving ≥3 Pauses</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>12.50</td>
<td>43.75</td>
<td>43.75</td>
<td>53.13</td>
<td>34.38</td>
<td>12.50</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>12.50</td>
<td>37.50</td>
<td>50.00</td>
<td>28.13</td>
<td>59.38</td>
<td>12.50</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>0.00</td>
<td>31.25</td>
<td>68.75</td>
<td>3.13</td>
<td>81.25</td>
<td>15.63</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>15.63</td>
<td>53.13</td>
<td>31.25</td>
<td>62.50</td>
<td>31.25</td>
<td>6.25</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>18.75</td>
<td>18.75</td>
<td>62.50</td>
<td>18.75</td>
<td>75.00</td>
<td>6.25</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>6.25</td>
<td>21.88</td>
<td>71.88</td>
<td>12.50</td>
<td>68.75</td>
<td>18.75</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>25.00</td>
<td>40.63</td>
<td>34.38</td>
<td>31.25</td>
<td>31.25</td>
<td>37.50</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>0.00</td>
<td>9.38</td>
<td>90.63</td>
<td>12.50</td>
<td>75.00</td>
<td>12.50</td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>3.13</td>
<td>18.75</td>
<td>78.13</td>
<td>6.25</td>
<td>71.88</td>
<td>21.88</td>
<td></td>
</tr>
<tr>
<td>Mean</td>
<td>10.41</td>
<td>30.56</td>
<td>59.03</td>
<td>25.35</td>
<td>58.68</td>
<td>15.97</td>
<td></td>
</tr>
<tr>
<td>Stdev</td>
<td>8.70</td>
<td>14.30</td>
<td>20.40</td>
<td>20.70</td>
<td>20.66</td>
<td>9.56</td>
<td></td>
</tr>
</tbody>
</table>

The influence of the doubling of tempo was very marked on the distribution of pauses over the chords in the trial stimuli. At slow tempo, a mean of 10.41% of all chords received only one pause (SD=8.70%). When the pressure to make progress across the page was doubled, this proportion rose to 25.35% (SD=20.70%). A repeated-measures t-test resulted in t(8)=2.62, p=0.031. Similarly, at slow tempo 30.56% of chords were fixated on in both staves once only, (SD=14.30%), rising to 58.68% at fast tempo (SD=20.66%), t(8)=2.51, p=0.036. At slow tempo 59.03% of chords fell into the third category, in which half the information was refixated (SD=20.40%) falling to 15.97% at fast tempo (SD=9.56%), t(8)=5.50, p<0.001.26

26 These data do not account for the five instances in which chords were refixed more than once (see the previous footnote).
Figure 41 Distribution of pauses over chords at slow tempo (left side) and fast tempo (right side)
From these proportions it was possible to arrive at a measure of the levels of refixation and the O’Regan effect, and how these two phenomena changed upon a doubling of tempo. Levels for the O’Regan effect were calculated by halving the values in the one-pause category. In other words, at slow tempo 5.21% of the information on the score was unfixated (that is, half of 10.41%) rising to 12.68% at fast tempo. Similarly, at slow tempo at least 29.52% of the information was refixated, falling to 7.99% at fast tempo. This information is set out in Table 13, showing that both the O’Regan effect and refixation appear to be part of an interconnected mechanism for adapting to changed tempo in music reading. This part of Hypothesis 4, then, was strongly confirmed.

<table>
<thead>
<tr>
<th>phenomenon</th>
<th>slow mean</th>
<th>SD</th>
<th>fast mean</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>refixation</td>
<td>29.52</td>
<td>10.20</td>
<td>7.99</td>
<td>4.78</td>
</tr>
<tr>
<td>skipping over</td>
<td>5.21</td>
<td>4.35</td>
<td>12.68</td>
<td>10.35</td>
</tr>
</tbody>
</table>

Table 13 Tempo-induced changes in refixation and the O’Regan effect

4.4 Conclusions

In summary, Stage I has provided valuable background information for the upcoming teaching experiment on SightReader. It has suggested that a narrow range of highly skilled readers use a variety of strategies to adapt to different tempos. Approximately half of the subjects adjusted their span size in an even-handed manner, striking a load/latency relationship close to the equal-contribution point; the average for all subjects was also close to this point. On the basis of this result, in particular the fact that there was no clear favouring of load or latency in tempo adaptation under the normal condition, it was decided that the strategy for manipulating the settings in Stage II should not build in a bias towards either variable, and that there was not a significant risk of interference between the two settings in respect of load and latency. In view of the range of individual load/latency relationships, however, it would appear that load and latency embody distinct capacities in working memory; for this reason, Stage II included an activity that isolated latency (see Section 5.2.6).

Similar results were found for the number/duration relationship: there was considerable diversity among individual relationships, although the mean, and approximately half of the individual relationships, were located close to the equal-
contribution point. *SightReader*’s prevention of leftward refixation appeared to be unlikely to present a serious problem to the user of *SightReader*, since under the normal condition it occupied only a small fraction of saccades, and in any case was highly changeable as a proportion of all saccades when tempo was doubled. Subjects appeared to use both vertical and leftward refixation for the same purpose, to control their pace across the page, although they varied considerably in their use of both components. The use of vertical refixation and the O’Regan effect correlated strongly with tempo, suggesting that both phenomena are central to the process of tempo adaptation in music reading.

With hindsight, three rather than two contrasting tempos may have permitted greater confidence in the data on the effects of tempo on the dependent variables, as would a repetition of the observations with different stimuli. These restrictions are cause for recommending that the issues discussed here might be investigated with a larger sample in the future.
5 Stage II: controlling span size andrefixation with SightReader

A case was put in Chapter 1 that the teaching of sight reading skills is and has for some time been widely perceived among music educators as problematic, as a matter deserving of attention and effort by students, instructors and researchers. It was also proposed that in many contexts it is more useful to define sight reading not just as the first encounter of an unfamiliar text, but as the process of familiarisation that occurs gradually over the first several encounters. While Stage I explored the traditional notion of sight reading as first encounter in order to observe certain characteristics of the reading process while minimising stimulus familiarity, in Stage II, a teaching experiment and thus closer to the practical task of improving sight reading, the focus shifted to observing the first six readings of a text and the possibility of maximising the efficiency of acquiring familiarity during those readings.

SightReader was developed as one possible solution to the search for more powerful methods of improving musicians’ ability to rapidly and efficiently familiarise themselves with a musical text. The primary objective of Stage II was to evaluate SightReader as a tool for improving such skills, with the expectation that if its utility were confirmed, the device might represent a valuable part of the performance curriculum. The experiment involved comparing the performance of 10 experimental and 10 control participants who worked with SightReader. The experimental group was exposed to the unique span stretching function of the device, with little change in tempo during those sessions. By contrast, the control group experienced the normal condition during all trials, in other words, a similar training regimen as for the experimental participants, but based on progressive increments in the tempo setting.

The structure of both Stages I and II was set out in Table 6 at the start of Chapter 4 in terms of the observation and control of three key reading variables: tempo, span size and leftward refixation. Table 14 sets out the structure of the trials in terms of both the reading condition and structural function of each of the six sessions that each participant underwent. It was predicted that experimental participants would on average experience significantly greater improvement than control participants in
their mean level of action slips (Hypothesis 1) and memorisation skills (Hypothesis 2) when their performances in the pretest and posttest were compared. (Here the hypotheses were numbered separately from those in Stage I.) The secondary aim of Stage II was to gather information on the use of SightReader that would suggest ways of optimising the training outcome as opposed to evaluating the effectiveness of the device. Both of these aims were largely achieved in the experiment.

<table>
<thead>
<tr>
<th>stage of experiment</th>
<th>condition for experimental group</th>
<th>condition for control group</th>
</tr>
</thead>
<tbody>
<tr>
<td>Trial Session 1: Initial Questionnaire then pretest</td>
<td>normal</td>
<td>normal</td>
</tr>
<tr>
<td>Trial Session 2</td>
<td>span stretching</td>
<td>normal</td>
</tr>
<tr>
<td>Trial Session 3</td>
<td>span stretching</td>
<td>normal</td>
</tr>
<tr>
<td>Trial Session 4</td>
<td>span stretching</td>
<td>normal</td>
</tr>
<tr>
<td>Trial Session 5</td>
<td>span stretching</td>
<td>normal</td>
</tr>
<tr>
<td>Trial Session 6: posttest then Final Questionnaire</td>
<td>normal</td>
<td>normal</td>
</tr>
</tbody>
</table>

Table 14 Basic structure of the Stage II experiment

The trials produced three sets of data. The first set was procedural, arising from participants' responses to an Initial Questionnaire completed before the start of the pretest. Since participants had been randomly assigned to experimental and control groups upon recruitment, the chief purpose of this questionnaire was to confirm that the distribution of participants' self-perception of their overall performance, sight reading and memorisation skills, and their motivation to improve in the last two skill categories, were approximately equal between the groups. Knowing that this was the case would increase the reliability of any significant differences between further data produced by the two groups. The results of the Initial Questionnaire, which confirmed that the groups were approximately equally matched, are presented and interpreted in Section 5.2.1.

The second source of data were measurements of participants' performances in the pretest and posttest, during which both groups were observed under the normal condition, that is, by progressively increasing their tempo but with no span stretching function. Of prime importance was a statistical comparison of the changes in their mean level of action slips between pretest and posttest, representing participants' improvement in sight reading skills. An evaluation of SightReader as a training environment against the normal condition was based on this comparison. The results showed that the more difficult the stimulus type, the greater the experimental groups' margin of improvement over the control group, with a
significant disparity for the most difficult type of stimulus used, the four-part chorale harmonisation. At the same time, no significant differences between experimental and control groups were observed in relation to their mean tempo. Surprisingly, no significant differences between experimental and control groups were observed in relation to participants' scores in the memory tests. The results are presented and discussed in Sections 5.3.4 and 5.3.5.

The third source of data came from a Final Questionnaire, administered to each participant at the end of the posttest. Here participants provided a self-evaluation of the improvement of their skills during the trials, and made written comments about the best and worst aspects of their experience. As predicted, experimental participants rated the improvement in their sight reading skills significantly higher than control participants rated their own improvement. Some indication was also given of the relative merits of training sight reading and memorisation skills using SightReader's span setting as opposed to a more traditional training method using controls on tempo alone. The results are presented and interpreted in Section 5.3.6.

The next section discusses the theoretical background to SightReader. The brief description of the operation of the device in Section 1.4 is expanded into a fuller exegesis of the way SightReader works, the user interface, and some of the intricacies of the programming task. Following this, the equipment setup and methodology for Stage II are described. Finally, the hypotheses, aims and results are presented and discussed.

5.1 Theoretical background to SightReader

The concept of SightReader is rooted in three areas of knowledge, those of music education, computer technology, and cognitive psychology. The ideas that generated the device are not new, having largely arisen before and during the 1970s. In the context of their own areas they are simple, even basic, and it was only in combination that they suggested a computerised span-stretching mechanism and the hypothesis that it should be useful in the training of sight reading and memorisation skills. The three areas are now examined in relation to their contribution to SightReader.
5.1.1 Music education: looking ahead of the hands

The earliest conceptual ingredient was the belief among music teachers that encouraging students to keep their eyes further ahead of their hands on the score would improve their sight reading capacity. There is no way of knowing when the eye-hand span was first mentioned or explicitly treated in music lessons, but it is clear that the idea of manipulating it for training purposes received endorsement in print as long as half a century ago. In the widely circulated American music-education periodical, The Philadelphia Étude, the idea of enlarging the eye-hand span was promoted by three respected pedagogues, although only in brief articles and without scientific support. The earliest mention is by Nash (1953:19), referring to sight reading at the piano:

To develop discipline of the mind and eyes, read at least one measure ahead. Read the first measure silently, test the skill by covering that measure while the student plays it. Read the second measure while playing the first measure, and so on through the piece...until reading ahead becomes a habit.

In the light of Stage I of this study, where even at the slow tempo span sizes under the normal condition seldom exceeded two beats, to read one bar ('measure') ahead appears to be an ambitious goal unless the music is uncommonly simple. It is also not clear whether Nash was advocating a continuous process or a 'stop-start' process of locally memorising individual bars. Three editions later, Berkeley (1953:25) was more explicit in his advice to students and teachers of the violin, although considering the minimum span sizes achieved by experimental participants in reading one-part stimuli in Stage II, probably errs in the other direction in recommending a minimum span size of only one beat in reading a single musical line:

Whenever you play anything from the music, keep your eyes at least one beat ahead of what is being sounded. At first they will constantly fall back, but you must just as constantly push them forward again. In a week or two you should be able to master the knack of it, especially if you have plenty of opportunity to sight read. But it is not only when sight reading that you can practice reading ahead: whenever you play anything from the notes—no matter how well you know it—your eyes should be ahead of your fingers.

Berkeley was advocating a process of manipulating the span through conscious 'top-down' effort. Four years later, this idea was fundamentally extended to a 'bottom-up' display-based process by Jones (1957:56):

The development of eye-span [is an] important factor in piano sight reading. Turning a page several measures from the end for an expert accompanist illustrates how a good sight reader uses and needs this skill.... Another help is to use a small card to cover a half-measure at a time forcing
the pupil's eyes to read the next measure while playing the first. This can be increased to several measures depending on musical complexity, tempo, and level of progress. Doing this at the weekly piano lesson is better than nothing, but the real problem is to enlist a family member to help at a short daily practice in reading.

In advocating that teachers regularly move a blank card along the score as their students sight read, Jones comes closest to anticipating *SightReader*. Jones's statement is interesting on two other counts. She points out that a player's span is most obvious at the turn of a page. (Indeed, if carefully observed, the span is evident yet more frequently than this, on the move from one line to the next, not just from one page to the next.) She also implies that minimum span size is dependent on musical complexity, tempo and reader-skill. Whether this represents an unfounded assumption or is based on actual teaching experience, however, is not clear.

The advice of these pedagogues resonates with anecdotal evidence accumulated through the author's personal interaction over the period of a decade with music students at the University of Sydney. The evidence suggests that a substantial minority of instrumental and vocal teachers at one time or another advise their students when sight reading to look ahead further than they otherwise would. Such advice has a *prima facie* logic that requires no particular psychological expertise to understand: looking further ahead allows more time for processing the musical information in working memory and provides a larger buffer against unexpected processing demands. But it was also apparent that such teachers seldom employ Jones's idea of the moving card, and typically refer to 'looking ahead' only on a few occasions during a course of tuition: a sustained program of exerting either top-down or bottom-up influence on the eye-hand span to improve students' sight reading appears to be rare if it occurs at all.

There are several possible reasons that the systematic manipulation of the eye-hand span is not practised more widely. First, instrumental pedagogy has by and large focused on musculoskeletal output, on what instructors can readily see and hear, rather than on components of the music reading process that are more difficult for both instructor and student to perceive, such as eye movement and the workings of the memory system. Instructors are thus understandably less likely to be attracted to the idea of systematically manipulating the span. The second reason is that while the intuitive logic of some instructors may point to its validity as a training process, there is no scientific evidence that such manipulation is helpful. Sloboda (1985:68) went as far as to caution against assuming that merely practising the characteristics of good reading will produce good readers:
.... the finding that fluent readers typically look further ahead than poor readers does not automatically yield the prescription that poor readers should practise looking further ahead. It may be that increased ability for preview is the result of some other skill, such as the ability to detect pattern or structure in the score, and that simply trying to look ahead will not improve this skill.

The moving-card technique also has several practical disadvantages that may discourage its regular use by instructors. These disadvantages are listed here on the basis of the author's own experience in applying the technique. It is an awkward, labour-intensive process requiring the instructor to lean over the student's shoulders or in some other way to position himself/herself close to the score without obstructing the student's overall vision of the music. At the same time the instructor must enforce a consistent performance tempo; this is more important than might at first be thought, given that there is a tendency for students to unconsciously play ahead of the tempo to reduce their span size, that is, so that their hands might 'catch up' with their eyes. Moving a card over a score manually in time with the beat is also difficult to achieve precisely and consistently by comparison with the finely graded calibration of the eye-hand span that is possible on SightReader, as discussed below.

5.1.2 Computer technology: the gaze-contingency paradigm

The second contributing source to SightReader comes from computer technology, specifically the gaze-contingency paradigm, otherwise known as the moving-window technique. Developed by Reder (1973) and Rayner (1975), and reviewed in Section 3.2.3, the technique has since been used in many studies on eye movement in language reading, and by Truitt et al (1997) in relation to eye movement in music reading. In the moving-window technique, a display is connected through a computer system to an eye movement tracking device. The system can be programmed for experimental purposes to respond to the position or movement of the eyes by changing the display strategically and almost instantaneously. SightReader utilises the similarities between this paradigm and Jones's manual moving-card technique in music training to produce a computerised version of the manual technique. In SightReader the display is linked not to the movement of the eyes but to the musculoskeletal response on the musical keyboard: spontaneous changes in the display are thus triggered by the striking of a note or chord on the keyboard rather than by oculomotor commands. In this way, the musculoskeletal and oculomotor groups that were proposed in Chapter 1 are forced apart during performance (see Figure 8 in Section 1.3).
Controlling span size and leftward refixation with SightReader

*SightReader* transcends several of the shortcomings of the pre-existing manual techniques that were discussed above. First, the timing of *SightReader*'s erasure of material on the score is to an accuracy of several milliseconds, which is considerably more precise than human temporal judgement is likely to be. The second advantage concerns the calibration of the span setting. Under the normal condition in Stage I, span sizes for individual participants typically fluctuated between approximately one and two beats at 60MM, and between 1.5 and 2.5 beats at 120MM. Such a small range of span sizes suggests that, at least in sight reading four-part chorale harmonisation on *SightReader*, there is a considerable difference in processing demand between a minimum span setting of one and two beats, or two and three beats. Accordingly, to increase the span setting on *SightReader* by whole beats would appear to be too sudden a change in training conditions, risking frequent user 'crashes'. For this reason a time-delay mechanism was built into *SightReader* that provides for gradation in the minimum span-size between whole beats. On the prototype, for example, the setting for each reading of a stimulus could be progressively increased from 1 to 1.25 beats, 1.25 to 1.5, 1.5 to 1.75, and 1.75 to 2 and so on. The time-delay mechanism is explained in greater detail in Section 5.2.1.

Another advantage of *SightReader* is that since it can be self-operated and self-paced, it is more student-oriented than the manual technique applied during performance lessons. Like many music-training applications that have been developed over the last decade, *SightReader* allows the student to take control of the training process, unconstrained by the need for an instructor. *SightReader* could also eventually be programmed to provide motivational features such as automatic feedback to the user on performance quality, advice on what tempo and span settings to use, and a record of the user's progress. These student-oriented functions are likely to promote regular practice at span stretching, whereas the manual technique is likely to occur not more frequently than during the weekly lesson which, as Jones (1957:56) implies, is inadequate for a sustained improvement in sight-reading skills.

### 5.1.3 Cognitive psychology: span size, refreshment, arousal and chunking

The third theoretical source for *SightReader* lies in psychological studies of the reading process, many of which have already been extensively reviewed in Chapters 2 and 3. One concept in this literature, span size, is of central importance to the functioning of *SightReader* while three other concepts, refreshment, chunking and Arousal, may be relevant to such functioning. In relation to span size, studies
including Goolsby (1987), Sloboda (1977), Truitt et al (1997), Jacobsen (1941) and Weaver (1943) have given scientific backing to the intuitive knowledge of pedagogues such as Jones (1953), Nash (1957) and Berkeley (1953) that span size tends to be larger for skilled than for unskilled readers. Similar findings exist in relation to language reading (Quantz 1898; Buswell 1920; Judd & Buswell 1922; Tinker 1958; Morton 1964). In addition, there is other evidence that span size is an important variable in the reading of both music and language, that it responds to various reading conditions such as musical phrase boundaries (Sloboda 1977; Levin & Kaplin 1970) and to the difficulty of processing linguistic text (Buswell 1920; Lawson 1961; Morton 1964). Work by Shaffer (1976) in relation to copy-typing, and Kinsler & Carpenter (1995) and Rayner & Pollatsek (1997) in relation to music reading, also give weight to the importance of the span in the reading process.

There are also signs that the other physical aspect of reading over which *SightReader* exerts control, refixation, may have a beneficial effect on the music reading process. In Chapters 3 and 4 refixation was treated under the assumption that it provides refreshment of information in working memory, and it was on this basis that the third oculomotor imperative was developed in Section 3.3.12. (The inherently limited capacity of working memory and the need to refresh its contents if they are to remain activated for more than a few seconds had already been discussed in Chapter 2.) At this point, however, it is necessary to refer back to the earlier discussion of refreshment. The work of Kennedy (1992), Johnson-Laird (1983), Hegarty (1992), Christie & Just (1976) and Zechmeister & McKillop 1972), reviewed in Section 3.2.6, suggests that it may be useful to think of refreshment as existing in two forms. The first, EXTERNAL REFRESHMENT, comprises refreshment through physical refixation; until now this has mostly been referred to simply as ‘refreshment’ in this study. The second type of refreshment, as explained in Section 3.2.6, is INTERNAL REFRESHMENT, in which information is re-activated through internal imaginative effort rather than through external visual input by physical refixation.27

By preventing external leftward refixation through its unique process of chord erasure, *SightReader* may encourage readers to rely solely on the internal refreshment of previous chords within their span, whereas under the normal condition they might rely either on external refreshment alone or a combination of external and internal refreshment to maintain the contents of their working memory. In addition, the demand for refreshment might logically be considered to be greater than normal when the span is stretched, since an increased amount of information must be stored and

27 Kennedy and his colleagues, however, did not use the terms external and internal refreshment.
processed in working memory for a longer period. It is, however, impossible to prove experimentally that *SightReader* promotes internal refreshment, and if it does, whether such promotion enhances reader-skill in Stage II; this question awaits further research into the role of the memory system in music reading. All that can be said at this stage is that the promotion of internal refreshment *may* occur with the use of *SightReader*, and *may* be beneficial. The notion of refreshment will be returned to in the next section in relation to memorisation.

The second behavioural aspect that may be connected with the functioning of *SightReader* is that of chunking. As was explained in Section 2.1, chunking is the processing in the memory system of small units of information into groups using to the reader’s pre-existing knowledge for doing so, and may be carried out with little or no awareness. Chunking is widely regarded as an important part of the music reading process. Gabrielsson (1999:511), among others, believes that ‘the good sight reader works with rapid and effective chunking using short-term memory’. Although there is no evidence in the literature that forcibly increasing span size promotes chunking, this may well be the case on logical grounds since chunking, as has been shown in numerous studies since Miller (1956), is one of the few ways of increasing the capacity of working memory. *SightReader* may therefore encourage chunking, in reverse as it were.

The effects of span stretching may also extend to arousal. This is suggested by the comments of experimental participants in Questions 13 and 14 of the Final Questionnaire concerning stress levels when using *SightReader* (see Figures 54 and 55 in Section 5.3.6). Kahneman (1973), in a major work on attention and effort, claimed that an increase in task demands can increase arousal, which is in turn associated with an increase in the ability to focus attentional capacity on a task. It is therefore possible that an increase in arousal among experimental participants in Stage II may increase the capacity of experimental participants for ever higher span settings during their reading of each stimulus (see Section 5.2.6). It may also contribute to the predicted greater improvement in sight reading skills over that of control participants.

While span stretching was a central focus of Stage II, none of the three additional behavioural aspects—internal refreshment, chunking and arousal—can be directly investigated in the present study. They are, however, submitted as possible contributing factors to the training advantage observed in experimental participants over control participants in Stage II.
5 Controlling span size and leftward refixation with *SightReader*

### 5.1.4 Cognitive psychology: memorisation

It was suggested in the previous section that *SightReader* may promote internal refreshment by preventing refixation and thus external refreshment, and by encouraging readers to attempt to hold more information for longer than they normally would in their working memory. If span stretching does indeed increase in the level of refreshment, the question arises of whether this might improve not only readers' sight reading ability, but their ability to subsequently play the music from memory. Atkinson & Shiffrin's groundbreaking (1968) paper on memory, reviewed here in Section 2.1, claimed that long-term coding of information is at least partly dependent on the amount of (internal) refreshment it receives when initially stored and processed in short-term memory. This part of Atkinson & Shiffrin's speculative theory was reinforced by Rundus & Atkinson (1970) in relation to the role of refreshment in the free recall of word lists. But the link between refreshment and long-term coding of information was overshadowed by claims that the efficient transfer of information from short-term to long-term memory depends more on the type of processing rather than the amount of refreshment it receives, chiefly in Craik & Lockhart's (1972) 'levels of processing' theory, reviewed in Section 2.1.

While it is not within the ambit of the present study to investigate this general issue, the use of *SightReader* does put an interesting perspective on the relationship between span stretching and playing music from memory. To approach the question of this relationship, an extreme case might be considered in which a reader's span is stretched so that 10 beats at a time must be held within it while reading through a melody of fairly consistent difficulty, with one note per beat. (A span setting of 10 beats is significantly higher than was ever achieved in the 720 readings in Stage II in which experimental participants' spans were stretched.) If the entire melody in this extreme example is only 15 notes long, after the fifth note is played the entire melody will have been erased from the display (since 5+10=15); thus the remaining two-thirds of the melody in effect must be played from memory. It is clear that span stretching when taken to this extreme is difficult to distinguish from playing from memory. The question then arises as to whether less extreme span settings, for example of five beats, would also aid memorisation, since they would impose the same type of memory demand on a more localised level: a window sift ing through the text in which five notes rather than 15 must be temporarily held in working memory.

This argument that span stretching is not much different in principle from playing music from memory formed the basis of Hypothesis 1, a prediction that reading a
Controlling span size and leftward refixation with SightReader

passage of music using SightReader’s span stretching function would improve a participant’s ability to subsequently play the passage from memory (see Section 5.3.4). As stated earlier, Kahneman (1973) argued that task demands can increase arousal and thus available memory capacity, although he was not concerned with the consequences of arousal for long-term retention. However, a substantial literature on the effects of arousal on memory (for example Eysenck 1982) has proposed that unless arousal is at extreme levels, it is likely to increase the retention of material in long-term memory.

Finally, opinions are divided on whether the ability to play music from memory is related to sight reading ability. Gabrielson (1999:511), apparently assuming that there is only one method of memorising music, points out that ‘in memorizing music one works slowly with awareness and control of each note until the procedures to a large extent can be automatic and stored in long-term memory. The goals as well as the means are thus different [from those of sight reading].’ Wolf (1976:167) claimed that good sight readers may be poor at memorising, and vice-versa. On the other hand, Nuki (1984) found a positive correlation between sight reading and memorisation ability in pianists, Thompson (1987) observed that the sight reading ability of flautists was correlated positively with their scores in a test of music recall, and McPherson (1995) found a positive correlation between sight reading and memorisation ability in clarinet and trumpet players.

The question of whether sight reading and memorisation ability are related informed an aim in Stage II, to investigate whether there is a strong correlation between these two skills, at least for keyboardists in this particular context. Since the investigation of a long-term memory effect was of interest, all participants were directed to read each stimulus not once but six times in immediate succession. Repeated exposure to the same material, it was reasoned, would strengthen traces of that material in experimental participants’ long-term working memory or long-term memory. This would allow the span setting to be progressively increased without undue degradation in performance quality. Repeated readings of the same material were also employed because it accorded more closely with the favoured definition of sight reading in this study, of not merely the first reading of an unfamiliar text, but the ability to rapidly familiarise oneself with a musical text over several readings (see Section 1.2.1). This gave rise to Hypothesis 2, that it would be possible to progressively increase the span setting without overly damaging the musical output while repeatedly reading the same text.
The weight of evidence and logical inference from music educators, computer technology and cognitive psychology thus pointed to a need to develop and evaluate SightReader as a precise, effective computerised tool for improving musicians' sight reading, in the expectation that when used as part of an overall program to improve musicians' skills at sight reading and rapid familiarisation with music, it would represent a significant pedagogical advance. The remainder of Chapter 5 concerns the design, execution and results of Stage II.

5.2 Method

5.2.1 Program design of the prototype of SightReader

SightReader's mechanism and user-interface were briefly explained in Section 1.3. The prototype of the system was written for Stage II using the multimedia programming software MAX 3.0 ( Opcode Systems, Inc). MAX was used because, given the equipment and programming limitation at the time, it offered the best facility for creating precise time-based function. The GRAPHICAL USER INTERFACE used in the programming consisted entirely of MAX 'objects' and 'messages' (analogous to lines of writing in some computer languages). In addition, some eight custom-built objects were written to accommodate the particular needs of the program. Each stimulus for Stage II was authored (entered) into the program by a four-stage method. A stimulus was first converted into FINALE 3.0, a music notation program. A pixel version was then created, and the application NIH ANIMATOR was used to produce one frame for every beat of the stimulus. The first frame contained the entire stimulus; the second frame contained the entire stimulus reduced by the notation in the first beat of the stimulus, which had been converted into blank stave-lines using NIH Animator's cut and paste function. For the third frame, the notation in the second beat, as well as the first, was converted into blank stave-lines, and so on. Thus a 24-beat stimulus required 25 frames, the last comprising only blank staves plus key and time signatures.

The frames of this graphical animation file were then loaded into SightReader using its custom-built authoring function. As the reader played through the stimulus, each successive chord struck on the musical keyboard would trigger an instantaneous change to the next frame. SightReader executed the change from one frame to the next smoothly and cleanly such that apart from the erasure, nothing on the display appeared to move, jump or change. The impression was one of a continuous display. The pulse from the musical keyboard could be represented by the striking of a chord
Controlling span size and leftward refixation with *SightReader*

or of a note in a melody. There was always a danger that a sloppily played chord, where not all notes were struck simultaneously, would trigger the system to change more than one frame. For this reason a **splay-threshold setting** was built into *SightReader* so that a time-interval between a pulse and its immediate predecessor could be set, below which the pulse would fail to trigger a change of frame. In other words, if two pulses were sent from the keyboard closer together than the splay threshold, as could easily occur in the untidy playing of a chord, the second pulse would be prevented from triggering an additional and unwanted change of frame. The threshold was set on a default of 200ms, which worked satisfactorily throughout the trials in Stage II. This level was chosen since it was below the 250ms duration of the quaver at 120MM, the fastest tempo used in Stage II.

Quaver movement was also related to another significant technical problem. Without quaver movement, the stimuli would have presented serious ecological problems, since music without at least some rhythmic activity is rarely found. Because frame changes were made on the basis of receiving one trigger per beat from the keyboard, when beats containing quaver movement were played, the second quaver would have triggered an unwanted second change of frame within that single beat, pushing the eyes ahead by two beats. Thereafter the eyes would be unintentionally forced ahead of the hands by the extra beat. Similarly, minims would have had the effect of reducing the span setting by one beat, since they failed to provide a trigger on the beat that fell half-way through their duration. If the program had not taken account of such rhythmic variation, the effective minimum span setting would have varied constantly according to the rhythmic values in each beat.

To solve this problem, a custom-built object called the **rhythm compensator** was created so that the second of two consecutive quavers would be ignored as a trigger for a change of frame. Similarly, a function was inserted into the program such that an additional change of frame would be automatically triggered one beat after a two-beat chord (minim) was struck. Other rhythmic values larger than a crotchet, such as the dotted crotchet and dotted minim, were treated analogously. To enable this process to occur, it was necessary for *SightReader* to have access to the rhythmic values in each stimulus. This was not possible with the graphical animation described above. Since MIDI-files provided such access, during the authoring process a parallel stimulus was saved as a MIDI-file and was also loaded into *SightReader*.

---

28 Setting the splay-threshold at a level above the 250ms duration of quavers may have solved the problem if tempos slower than 120MM had not been used in Stage II, but 120MM was in fact the fastest tempo employed.
In addition to a facility for compensating for rhythmic variation, a mechanism for initially establishing the minimum span setting had to be programmed into *SightReader*. This mechanism was named the INITIAL CLEARANCE DEVICE. Since it was essential that readers be able to view the opening chords before starting to play, *SightReader* was programmed to display the entire stimulus before the reader began to play, and to skip over *more than one* frame rather than a single frame in response to the striking of the first chord. For example, if a minimum span size of two beats had been set, two frames rather than one were turned over when the first chord was struck. In other words, the display would skip straight from Frame 1 to Frame 3 leaving the third chord as the left-most notation on the display. The third chord was two chords ahead of the struck chord, giving a minimum span-size of two beats. At this point the eyes must already have arrived at the third chord or at least have left the second chord and be moving towards the third chord. The striking of the second and subsequent chords would have prompted the normal single turnover of frames so that after striking the third chord, Frame 4 would be displayed. If minimum span size had been set on three beats, the initial clearance device would have skipped straight from Frame 1 to Frame 4, removing the first three chords of the stimulus in response to the first pulse from the keyboard. At this setting the musician should therefore have been ready to look three chords ahead at the fourth chord while striking the first.

A need for the span setting to be calibrated in smaller units than whole beats was identified in the previous section. This was achieved by programming another device into *SightReader*: the TIME-DELAY CALIBRATOR. Gradations between whole beats were achieved by temporal subtraction, as it were, from the next highest whole-beat setting. For example, if the span setting were 2.25 beats, the initial clearance device would erase two chords immediately on the striking of the first chord; it would then automatically erase the third chord .75 beats later, that is, without further triggering from the keyboard. This would assume that the eyes were already inspecting the fourth chord one quarter of a beat before playing the second chord. The duration of these ‘calibration delays’ depended on the performance tempo and the related beat duration, and therefore the time-delay calibrator was linked to the metronome. If the example given above occurred at 60MM, the beat duration would be 1000ms, and the delay of the automatic frame change would have been 750ms (.75 x 1000). As pointed out in Section 5.1.2, this is an essential advance over the manual technique since to move straight from one to two, or from two to three whole beats is typically too sudden a gradation in reading four-part musical textures.
As explained in Section 5.1.2, minimum span size could be set on fraction of a beat, for example 1.5 beats rather than 2 beats as exemplified above. The first chord played would then trigger the disappearance of the first two chords, but half a beat later the third chord would disappear without further cue from the keyboard. Similarly, the fourth chord would disappear not as the third chord is played, but again, half a beat after it has been played. If the minimum span size were set at 1.75 beats, each chord on the display would disappear a quarter of a beat after being triggered by the keyboard. This may appear to be anomalous until it is pointed out that the concept of beat-fraction involves subtracting from the timing of what would be the next chord disappearance (that is, 2.00 - 0.75 = 1.25), rather than adding to the current moment.

For settings on beat-fraction to work, the tempo and span settings must be linked. Without such a link, SightReader would not be able to calculate when to erase chords on the display to enforce a fractional minimum span size, since it would not ‘know’ the time-value of a whole beat (in milliseconds). Double the tempo, and the time-value of the beat is halved; SightReader must therefore adjust the time-delay accordingly. In other words, a tempo of 60MM gives a beat-value of 1000ms. With a span setting of 1.75 beats, the display must change 250ms after the trigger from the keyboard. By contrast, if the tempo is 120MM, the beat-value is only 500ms, and the display must change 125ms (a quarter of 500ms) after receiving the trigger from the keyboard.

5.2.2 Equipment setup

The equipment setup was located in a small annex to the Sydney University Electronic Sound Studio. The setup is depicted in Figure 42. The prototype of SightReader comprised an electronic musical keyboard (1) connected through a MIDI-box (2) to a Power Macintosh 7100/66AV (3). In front of the monitor and CPU on the same desk was an alpha-numeric keyboard (4) controlling the system. On an adjacent desk a portable cassette recorder (5) was positioned, on which the trials were recorded. Two loudspeakers (6), located on the wall above the desk, were patched into the system. An adjustable chair for the participant was positioned in front of the musical keyboard; a second chair, for the experimenter, was positioned in front of the adjacent desk, but does not appear in Figure 42.
5.2.3 Participants

20 participants were engaged for the experiment and randomly assigned to control and experimental groups. All participants were undergraduate music majors at the University of Sydney between the ages of 19 and 23 years who had a stated interest in improving their sight reading and memorisation skills and who agreed to participate in a series of six half-hour coaching trials over three weeks. Candidates were informed that they would be paid a fee of A$80 at the conclusion of the trials. Candidates were warned before engagement that they may be exposed to mildly stressful reading conditions during the trials, and that if they participated it would be necessary to attend six one-to-one trial sessions over three weeks, that is, two sessions per week, as they played several unfamiliar but straightforward melodies, two-part inventions and four-part hymns on an electronic keyboard. They were informed that their tempos and certain other aspects of their performances would be manipulated by the experimenter, and that their performances during the trials recorded and analysed, but that all records of the trials would remain strictly confidential, including the fact of their participation. They were also informed that they would be given a minimum of information about the study until after the trials. The trials were conducted with the approval of the Human Ethics Committee at the University of Sydney (reference number 98/3/30).
Since Stage II was a teaching experiment premised on the improvement of participants' skills during the trials, the skill-level required at the beginning was lower than that for Stage I. To be eligible, candidates had to express confidence that they could sight read on the keyboard with reasonable accuracy what would be the most difficult stimulus-type, a straightforward four-part hymn setting in 3/4 or 4/4 time with up to two sharps or flats, keeping strictly to a metronome at least 25MM. It was anticipated that the common behaviour of repeatedly looking down at the hands while sight reading at the keyboard would be overly disruptive when using SightReader. Candidates who were not confident that they could play from score without looking down at their hands were therefore given a brief test to exclude those who were unsatisfactory in this respect. Although this restriction may be viewed as affecting the ecological validity of the experiment, it is contended that playing without looking at the hands does not alter the music reading mechanism in principle, but represents a more efficient process that any musician can adopt with a small amount of training.

Candidates were informed that they would be asked to complete an Initial Questionnaire (see Figure 43) before the start of the first trial to confirm their suitability. This questionnaire also functioned to confirm that the experimental and control groups were not significantly different in terms of the background, skill and motivation of the participants. As a method of assessing participants' skill-levels, the questionnaire was regarded as more focused on the requirements of the experiment than a standardised measurement of sight reading skill such as the Watkins-Farnum Performance Scale (Watkins & Farnum 1954) would have been.

The questionnaire consisted of seven questions concerning any training in sight reading and memorisation that participants had received and may have been currently receiving, their self-perceived sight reading and memorisation abilities compared to players of similar performance standard, and their current motivation to improve those abilities. In Figure 43, horizontal rows of boxes for participant responses that appeared below each of Questions 3-7 are omitted to save duplication; the font-size is also smaller than that which was actually used. The numerical results of Questions 2-7 of the questionnaire are set out in Table 15. Copies of participants' individual responses may be obtained from the author by request.
Initial Questionnaire

Participant number: ____________

1 Are you undergoing any other specific training in sight reading or memorisation at the moment? Circle one: yes no

2 For how many years have you undertaken formal keyboard studies (private lessons, regular practice at the keyboard, etc)? ____________

3 Give an accurate estimation of your performance standard in terms of the AMEB grades for piano, or the equivalent: ____________

4 Compared to other keyboardists you consider to be of similar overall performance standard to you (same AMEB grade equivalent), how do you rate your sight reading ability? Tick one box.

5 Compared to other keyboardists you consider to be of similar overall performance standard to you (same AMEB grade equivalent), how do you rate your ability to play from memory a short passage of previously unfamiliar music after repeatedly reading it from score? Tick one box.

6 How important is it to you to improve your sight reading? Tick one box.

7 How important is it to you to improve your ability to memorise music? Tick one box.

Figure 43 Compressed version of the Initial Questionnaire

Question 1 required the only non-numerical response, a yes/no polarity. Here, none of the participants indicated that they were currently undergoing any other specific training in sight reading or memorisation. Participants had already been informed that it was a requirement of the study that they not be simultaneously studying sight reading or memorisation elsewhere. An affirmative response to this question would therefore have disqualified a participant. Questions 2-7 required participants to respond numerically, or in a form that could be readily transformed into numerical values. The responses to these questions revealed no significant distinctions between the two groups in respect of relevant skills and attitudes, thus confirming that the two groups were roughly equal for empirical purposes.

<table>
<thead>
<tr>
<th></th>
<th>all mean</th>
<th>stdev</th>
<th>exp mean</th>
<th>stdev</th>
<th>cont mean</th>
<th>stdev</th>
<th>p≤</th>
</tr>
</thead>
<tbody>
<tr>
<td>2 years of formal study</td>
<td>9.50</td>
<td>3.32</td>
<td>9.00</td>
<td>3.40</td>
<td>10.00</td>
<td>3.33</td>
<td>0.257</td>
</tr>
<tr>
<td>3 AMEB grade equivalent</td>
<td>7.85</td>
<td>1.57</td>
<td>8.10</td>
<td>2.02</td>
<td>7.50</td>
<td>1.90</td>
<td>0.252</td>
</tr>
<tr>
<td>4 self-rated SR ability</td>
<td>4.70</td>
<td>1.63</td>
<td>4.60</td>
<td>1.78</td>
<td>4.80</td>
<td>1.55</td>
<td>0.396</td>
</tr>
<tr>
<td>5 self-rated memorisation ability</td>
<td>4.45</td>
<td>1.73</td>
<td>4.80</td>
<td>1.81</td>
<td>4.10</td>
<td>1.66</td>
<td>0.190</td>
</tr>
<tr>
<td>6 motivation to improve SR</td>
<td>7.55</td>
<td>2.06</td>
<td>7.60</td>
<td>2.07</td>
<td>7.50</td>
<td>2.17</td>
<td>0.459</td>
</tr>
<tr>
<td>7 motivation to improve memorisation</td>
<td>7.80</td>
<td>2.21</td>
<td>7.90</td>
<td>2.38</td>
<td>7.70</td>
<td>2.16</td>
<td>0.423</td>
</tr>
</tbody>
</table>

Table 15 Results of Questions 2-7 in the Initial Questionnaire.
In Question 2, participants indicated the number of years for which they had undertaken formal keyboard studies, including private lessons and/or regular keyboard practice. Experimental participants declared a mean of 9.0 years’ study against 10.0 years for control participants, with an overall mean of 9.5 years.

Question 3 concerned participants’ self-perceived performance standard in relation to the graded public examinations administered by the Australian Music Examinations Board, a state-run system with which virtually all tertiary music students and professional musicians in Australia are acquainted. In this system, akin to the Associated Board system in the United Kingdom, there are ten grades, First to Eighth Grades leading to two diploma levels, the Associate and Licentiate. For these purposes the Associate was assigned the value of 9, and the Licentiate the value of 10. The lower diploma level (9) is regarded as ‘a test of executive ability’ (Australian Music Examinations Board 2000:60). For the upper diploma level (10) ‘a concert standard of performance’ is demanded (p62). Experimental participants declared a mean of slightly above Eighth Grade, and control participants half-way between Seventh and Eighth Grades, giving an overall mean of 7.85. Participants could therefore be considered to be of fairly uniform skill and experience, and skilled and experienced enough to cope with the demands of SightReader.

For Questions 4-7, participants were asked to tick one of ten boxes arranged horizontally across the page below the text of each question, under which three or four descriptors appeared, as shown in Figures 44 and 45. Each of these boxes was assigned a value from 1 (right side) to 10 (left side) so that the responses could be statistically processed.

```
[10 9 8 7 6 5 4 3 2 1]
```

```
excellent average poor
```

Figure 44 Response template for Questions 4 and 5, including numerical equivalents to the boxes

Question 4 asked participants to rate their sight reading ability compared to other keyboardists they considered to be of similar overall performance standard to themselves, in other words, equally ranked in terms of AMEB examination standards. The results were fairly tightly clustered around 4.7 on the low side of ‘average’, with a mean of 4.6 for experimental participants and 4.8 for control participants. None of the participants regarded themselves as ‘poor’ in this respect. The groups were slightly more distinct in their responses to Question 5, which asked
for a rating of participants’ ability to play a short passage of music from memory after playing it repeatedly from score, again in relation to keyboardists of similar standard. The mean for experimental participants was 4.8, and for control participants 4.1, giving an overall mean of 4.45, on the low side of ‘average’, slightly lower than for the ratings for sight reading ability. The difference between the groups, however, was not statistically significant.

![Figure 45 Response template for Questions 6 and 7](image)

Question 6 concerned participants’ motivation to improve their sight reading skills and Question 7 their motivation to improve their memorisation abilities. Here the mean for improving both skills was on the high side of ‘quite important’ (7.55 for sight reading and 7.80 for memorisation), with little difference between the groups. Motivation to improve sight reading was one of the criteria for accepting potential participants into the trials, as earlier stated.

In summary, the Initial Questionnaire provided useful information about the participants that suggested that the experimental and control groups were approximately matched in skill and experience. The range of relevant skills was reasonably narrow, with the mean of AMEB grade equivalents lying just over one grade short of diploma level, suggesting that all participants would be capable of learning and performing one of the easier piano sonatas by Beethoven. Participants were well motivated to improve their sight reading and memorisation skills, another requirement of Stage II, and none was undertaking other tuition in the skills at issue.

### 5.2.4 Stimuli

The selection of stimuli was based on the definition of sight reading for Stage II as discussed in Section 1.2.1: the first six readings of a musical text (1) with which the player initially has no actuality familiarity, (2) where both the text and the player have been chosen such that probability familiarity is initially within certain limits, and (3) where the level of newness encountered by the player in reading through the text is reasonably consistent. In accordance with this definition, all stimuli were TONAL, with keys restricted to major and minor modes with up to two flats or sharps. This range of keys was chosen so as to provide a certain level of variety in the stimuli.
without presenting a significant challenge in respect of participants’ familiarity with keys and the inherent difficulties of fingered in some keys. To provide a minimal level of variety, most stimuli contained a modulation or tonicisation to a closely related key, and were in 4/4 or 3/4 metre. (Compound time, in which the beat is typically divided into three or six notes, would have been used but for the fact that the rhythm compensator was programmed to accommodate only simple time, in which the beat is typically divided into two or four notes.) The length of the stimuli ranged from eight to 16 bars; those in 3/4 time or of less dense texture tended to comprise more bars; those in 4/4 time or of denser texture also tended to comprise fewer bars.

To maximise the consistency of the experience for both groups, all readings took place using the same equipment, in other words, all participants read from SightReader’s monitor and musical keyboard and under the ‘normal’ condition, where span stretching was not at issue, minimum span size was set at ‘0’ and there was thus no functional connection between keyboard and monitor. In each of six sessions for each participant there were 21 readings: participants read first a one-part, then a two-part, and finally a four-part stimulus, each six times in immediate succession followed by a seventh reading from a modified score, comprising a memory test. A total of 36 stimuli were therefore used in the trials, 18 full versions of the 18 stimuli (three stimuli in each of six trials), and 18 modified versions for the memory tests. The intention was that all six stimuli of each type were of approximately equal difficulty, although this is a very complicated aim to achieve with absolute confidence given the number of variables that must be accounted for.

The three types of stimulus were chiefly distinguished by the amount of information per beat they required the reader to encode, store, process, and express as a musculoskeletal response. This difference in information density (and thus performance difficulty) was one of the independent variables as discussed in the next section. The one-part stimuli were the least dense in terms of the number of musical notes per bar, and were written on one stave rather than two, thus simplifying the scanpath. They also required the use of only the right hand, thus avoiding the challenge of coordinating the hands. The two-part stimuli were more demanding in respect of both oculomotor and musculoskeletal function, being written on two staves and requiring the use of both hands, although each hand was required to execute only a single part. The four-part stimuli, by contrast, required each hand to execute not one but two separate musical parts, representing a significantly greater
challenge in respect of fingering and hand position, and presumably resulting in a greater processing workload in the memory system.

Copies of all stimuli used in Stage II appear in Appendix II at the end of the dissertation. The one-part stimuli bore some resemblance to the stylistic characteristics of the English folk song. Rather than using existing folk songs, the melodies were composed by the experimenter so that several typical features of their style could be excluded on the basis that they might have been overly unfamiliar to participants and represented potentially confounding factors, for example the use of non-tonal modes or local melodic ornaments such as acciaccaturas; or on the basis that they involved excessive internal actuality familiarity, chiefly the repetition of the medium- and large-scale structures. The use of fresh composition also excluded the possibility of participants' prior actuality-familiarity with the stimuli, and made it possible to restrict rhythmic values to those that could be accommodated by SightReader's rhythm compensator, that is, quavers in groups of two or more, crotchets, dotted crotchets, minims and dotted minims.

The two-part stimuli could be characterised as two-part inventions for bass and soprano voices. They were extracted by the experimenter from existing four-part chorale harmonisations by JS Bach and his contemporaries, and slightly modified to avoid medium- and large-repetition and unacceptable rhythm durations. The rhythmic interaction between the parts was mostly crotchet against crotchet with some crotchets set against two quavers. Unlike the one- and two-part stimuli, which were freshly composed or arranged for the purpose of the trials, the four-part stimuli were existing four-part chorale harmonisations. These stimuli underwent minor modification by the experimenter in similar ways as for the stimuli in Stage I (see Section 4.2.1). Tenor and bass parts were rewritten in the few places where they were more than an octave apart. Potential stimuli were ruled ineligible if they included crossed parts, that is, bass moving above tenor or tenor moving above alto.

Immediately after the sixth reading a memory test was administered to all participants. This test involved their reading a static display of a specially prepared version of the stimulus in which every even-numbered bar was replaced by blank stave-lines. Participants were asked to play through the stimulus continuously, playing the erased bars from memory as best they could. The reason for using this type of stimulus was to avoid the situation in which a participant would 'lose track' comparatively early in the stimulus, and be unable to regain their place. This was anticipated to represent a high risk if participants had been required to play from
memory with no stimulus at all. Participants were advised that it was acceptable to play wrong notes or chords, or nothing at all, if their memory of the erased bars failed them, but to attempt to play the whole stimulus perfectly if they could, as for the previous six readings. Examples of memory tests appear in Figures 46 and 47.

Figure 46 An example of a memory test for a one-part stimulus

5.2.5 Temporal structure of the trials

Participants attended six one-to-one trial sessions, each of approximately 30 minutes' duration. The six sessions were held over a three-week period, with two
sessions per week. The trial periods for each participant were staggered such that not more than eight participants were being trialed at any one time; the total duration of the experiment was ten weeks. The overriding structure of these six sessions was set out in Table 6 at the start of Chapter 4. As far as could be arranged, the session times were at the choice of participants, except that to ensure a minimum dispersion of the training experience, they were not scheduled on consecutive days of the week (for example Tuesdays and Wednesdays). Participants were ill or otherwise indisposed for 11 of the 180 sessions in Stage II. In these cases an extra session was substituted in the same week or the following week (provided that a time-slot on a non-consecutive day was available), or the three-week trial period was extended by two or three working days to accommodate the sixth session.

5.2.6 Developing guidelines for adjusting the settings

Before each of the 21 readings in each session, the experimenter needed to make an on-the-spot decision as to the appropriate level of the span and tempo settings for the upcoming reading. Guidelines for this purpose were developed as a result of informal preliminary trials with a different set of participants, and took into account the objectives that the difficulty of the reading conditions was to be gradually increased and that participants should be reasonably challenged by the reading condition without experiencing an unpleasant amount of stress. Since no set of rules can take account of variables such as participants' mood and stress levels, and since such factors can have a significant bearing on their performance capacity, the guidelines were regarded as a set of flexible principles rather than strict rules. In operation they were contingent on negotiation with participants and were strictly followed in only approximately 80% of cases. Participants were thus asked to inform the experimenter if they felt that the settings were too low or too high at any stage. The guidelines were also expressed in flexible terms such as increasing the tempo setting by 5-15MM, rather than by a precisely specified amount.

The guidelines for setting minimum span size and tempo were complex. To explain them requires an account of how the level of action slips was arrived at for each reading and a description of the separate decision-making processes for experimental and control participants, pretest, posttest and trial sessions, and for Reading 1, Readings 2-5, Reading 6, and Reading 7 of each stimulus.29 The rate of action slips was calculated as a percentage of correctly played notes or chords by the same

---

29 It should be clarified here that there were seven 'readings' of each stimulus, three stimuli in each 'trial session', and six 'trial sessions' for each subject.
method as that which was described in relation to Stage I, in Section 4.2.2. Half-marks and whole marks were deducted for chords that were not played accurately, and a percentage score arrived at by dividing the accumulated marks deducted by the number of active beats in the stimulus. Inactive beats, that is, beats in which no musculoskeletal response was required, were excluded from this calculation. To expedite this decision-making process, the thresholds for decision-making as described in the next section were calculated beforehand for each stimulus in terms of numbers of marks; in this way the experimenter was able to promptly decide the percentage equivalent of the previous level of action slips, and either to apply the guidelines or modify them through negotiation with the participant.

Separate guidelines were developed for each of the two basic training conditions pertaining in the trials. One was the ‘normal’ condition, in which the gradual increasing of the tempo setting was the main focus from reading to reading of the same stimulus, and in which span stretching played no role; this situation pertained to all participants in the pretest and posttest, and to control participants in Trial Sessions 2-5. In the other ‘span stretching’ condition, gradually increasing the span setting was the main focus; this pertained only to experimental participants in Trial Sessions 2-5.

5.2.7 Adjusting the tempo setting under the normal condition

The simplest guidelines for the tempo setting concerned one-part readings. Except for the memory tests that are discussed below, these were all read at a fixed tempo of 120MM for both normal and span-stretching conditions. This was due first to an assumption on the basis of the preliminary trials that all participants would be capable of reading one-part stimuli at this tempo, an assumption that was later borne out in the data. Second, it was predicted that some participants would be able to reach much faster tempos than 120MM given the relative ease of reading a single line. Such fast tempos may have posed a problem for SightReader’s splay threshold mechanism (see Section 5.2.1), since quaver durations at 120MM are 250ms, only slightly above the splay-threshold of 200ms. For the same reason, an upper tempo limit of 120MM was also enforced for two- and four-part stimuli, although this was only rarely necessary due to the considerably greater task demand represented by these stimulus types.

The rest of this explanation thus pertains only to readings of two- and four-part stimuli which, unlike one-part readings, were subject to changes in the tempo setting.
from reading to reading of the same stimulus. In the pretest, the opening tempo setting for Reading 1 was established by starting at 60MM for the two-part stimulus and 50MM for the four-part stimulus. If the participant was unable to cope with the given tempo, evidenced by either a breakdown or a rate of action-slips in the first three bars that appeared to be below the 80% threshold of acceptability, the experimenter stopped the reading, reduced the tempo setting to a more appropriate level, normally by 5-10MM, displayed the first frame and asked the participant to start again. This procedure was repeated if necessary until an appropriate tempo was established, and was employed throughout the trials for any reading that was unsuccessful under the normal condition, except the memory tests.

In subsequent sessions, the guidelines for the level of the tempo setting for Reading 1 of the two- and four-part stimuli were that the highest setting attained during the reading of that stimulus-type in the previous trial should be reduced by 5-10MM, although the new setting was not to be lower than the lowest setting for that stimulus-type in the previous trial. For example, if the tempo setting for the two-part reading in the previous session had begun at 60MM and had been progressively raised to 80MM, the new starting point would normally have been 70 or 75MM; if the tempo had previously begun at 60MM and had been raised to 65MM, the new starting point would normally have been 60MM again.

For Readings 2-6, the tempo setting was altered according guidelines that took into account the rate of action slips in the immediately preceding reading (see Table 16). For example, if the rate of action slips had been 8.5% in Reading 1 of a two-part stimulus, the tempo setting for Reading 2 would normally have been increased by 5-15MM. Guidelines for tempo changes were expressed in terms of ranges rather than specific quantities to allow for negotiation with the participant as to his or her stress levels in the previous reading. This was a practical measure to minimise the number of mid-reading 'crashes' in the trials that might have occurred if participants' stress levels had not been taken into account in this way.

<table>
<thead>
<tr>
<th>rate of action slips</th>
<th>0-10%</th>
<th>11-25%</th>
<th>26%+</th>
</tr>
</thead>
<tbody>
<tr>
<td>one-part stimulus</td>
<td>no change</td>
<td>no change</td>
<td>no change</td>
</tr>
<tr>
<td>two-part stimulus</td>
<td>increase by 5-15MM</td>
<td>no change</td>
<td>reduce by 5-10MM</td>
</tr>
<tr>
<td>four-part stimulus</td>
<td>increase by 5-10MM</td>
<td>no change</td>
<td>reduce by 5-10MM</td>
</tr>
</tbody>
</table>

Table 16 Guidelines for altering the tempo settings for Readings 2-6 under the normal condition.
5.2.8 Adjusting tempo and span settings under the span stretching condition

Although the span setting was the main focus under the span stretching condition, in addition a tempo setting had to be established for experimental participants in the Sessions 2-5, and after establishment was subject to minor adjustments. In Session 2, the first time the span setting was brought into play, the aim was to establish a slightly slower tempo setting than that which pertained for the participant during the pretest, when s/he was free of any span stretching, and to keep to that tempo throughout Sessions 2-5. The tempo setting was altered only where it was felt by the experimenter and/or the participant to be necessary, typically when the participant complained that is was too slow or fast; such alterations were kept to a minimum since the aim of the experiment was to give experimental and control participants contrasting experiences based on either progressive increases in tempo or minimum span size.

The exception to this desire for stability of tempo under the span stretching condition was Reading 6 of each stimulus, for which the tempo was set 10-30MM slower than the slowest tempo used during the previous five readings. To explain the rationale for this, it is necessary to make two points. First, SightReader is quite capable of targeting certain aspects of the music reading process by employing simple relationships between the settings. Four such relationships, three of which were used in Stage II, are set out in Table 17. One relationship (Row 1 in Table 17) could target the capacity of the musculoskeletal response with fast tempo, while at the same time a low (or absent) span setting meant that load and latency were not directly challenged; this relationship pertained for all readings in the normal condition except for the memory tests. Another relationship (Row 2) could target the capacity of the musculoskeletal response with a fast tempo and in addition could target load with a high span setting since, while at the same time not directly challenging latency, since at a fast tempo a comparatively large amount of information moves through the span rapidly. A third relationship (Row 3) could moderately target all three aspects, load, latency and musculoskeletal response with moderate tempo with a high span setting; this relationship pertained for Readings 1-5 under the span stretching condition. And finally a fourth relationship (Row 4) could target the capacity for latency with a slow tempo and a high span setting, since under these conditions a comparatively small amount of information moves through the span slowly; this relationship pertained for Reading 6 under the span stretching condition.
Controlling span size and leftward refixation with *SightReader*

<table>
<thead>
<tr>
<th>tempo setting</th>
<th>span setting</th>
<th>targeted skill</th>
<th>readings</th>
<th>condition</th>
</tr>
</thead>
<tbody>
<tr>
<td>high</td>
<td>nil</td>
<td>musculoskeletal response</td>
<td>1-6</td>
<td>normal</td>
</tr>
<tr>
<td>high</td>
<td>high</td>
<td>load, musculoskeletal response</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>moderate</td>
<td>high</td>
<td>load, latency, musculoskeletal response</td>
<td>1-5</td>
<td>span stretching</td>
</tr>
<tr>
<td>low</td>
<td>high</td>
<td>latency</td>
<td>6</td>
<td>span stretching</td>
</tr>
</tbody>
</table>

Table 17  Skills targeted by various combinations of the settings on *SightReader*

Second, there is no reason for believing that load and latency are inseparable skills in music reading, although confirming this hypothesis is beyond the scope of the present study. The significant lowering of the tempo setting for each Reading 6 under the span stretching condition was thus an attempt to isolate and challenge the capacity for latency in experimental participants with a view to increasing the training efficacy of span stretching for experimental participants.

The arrangement for establishing an opening span setting for experimental participants at the start of Session 2 was analogous to that for establishing the opening tempo setting for all participants in Session 1: a minimum span size of 2.0 beats was set for the first reading of the one-part, 1.5 beats for the two-part, and 1.0 for the four-part stimulus. Thereafter the guidelines as set out in Table 18 were used, with the same role given to negotiation with the participant as for the tempo setting under the normal condition. In subsequent sessions the guideline for the level of the span setting for the first reading of each stimulus was that the highest setting attained during the reading of that stimulus-type in the previous trial should be reduced by 0.5-1.0 beats. Table 18 sets out the guidelines for altering the span setting for Readings 2-5 under the span-stretching condition in relation to the rate of action slips in the immediately preceding Reading.

<table>
<thead>
<tr>
<th>rate of action slips</th>
<th>0-10%</th>
<th>11-25%</th>
<th>26%+</th>
</tr>
</thead>
<tbody>
<tr>
<td>one-part stimulus</td>
<td>add 1.0 beat</td>
<td>no change</td>
<td>reduce by 0.5-1.0 beats</td>
</tr>
<tr>
<td>two-part stimulus</td>
<td>add 0.5-1.0 beats</td>
<td>no change</td>
<td>reduce by 0.25-0.75 beats</td>
</tr>
<tr>
<td>four-part stimulus</td>
<td>add 0.25-0.5 beats</td>
<td>no change</td>
<td>reduce by 0.25-1.0 beats</td>
</tr>
</tbody>
</table>

Table 18  Guidelines for altering the span setting for Readings 2-5 under the span-stretching condition

5.2.8 Procedure for conducting the trials

At the start of the pretest participants were greeted, invited to sit in the participant’s seat and asked to complete the Initial Questionnaire. After completing the questionnaire, they were informed that they would be asked to read three
rate of action slips | 0-10% | 11-25% | 26%+
--- | --- | --- | ---
one-part stimulus | add 1.0 beat | no change | reduce by 0.5-1.0 beats
two-part stimulus | add 0.5-1.0 beats | no change | reduce by 0.25-0.75 beats
four-part stimulus | add 0.25-0.5 beats | no change | reduce by 0.25-1.0 beats

Table 18 Guidelines for altering the span setting for Readings 2-5 under the span-stretching condition

5.2.9 Procedure for conducting the trials

At the start of the pretest participants were greeted, invited to sit in the participant’s seat and asked to complete the Initial Questionnaire. After completing the questionnaire, they were informed that they would be asked to read three different excerpts from the display, each a total of seven times in succession, and that between each reading the experimenter might adjust the metronome. Participants were also informed that one of the purposes of the trials was to increase the difficulty of the reading conditions in each successive reading, such that they felt challenged without being exposed to an unpleasant level of stress, and for this reason from time to time they would be asked whether they felt that it would be overly challenging to increase the difficulty of the reading conditions. Participants were also asked to attempt to perform strictly at the given tempo with as few wrong notes and vagaries of tempo as possible, to try to avoid looking down at their hands while performing, and to concentrate on the sight reading task during the first six readings, rather than specifically attempting to memorise the stimulus. Participants were warned that after the sixth reading of each excerpt they would be asked to read it again in a memory test, and that the display for the memory test would appear with full notation alternating bar by bar with blank stave-lines. Thus as they read through the entire stimulus for the last time, they would be attempting to play every second bar from memory. The height of the seat was adjusted so that the keyboard was within easy range of the participant’s hands.

At the start of Trial Session 2, control participants were informed that subsequent sessions would be substantially the same as the pretest, although with different stimuli. By contrast, experimental participants were informed in Trial Session 2 that the upcoming trials would be different from the pretest in that their eyes would be pushed ahead of their hands through a process of disappearing chords on the display. Experimental participants were also informed that the tempo might be changed occasionally, but the emphasis would instead be on increasing the distance between their eyes and their hands. At the start of Trial Session 6, experimental
marks and horizontal (half) marks and summed on the spot at the end of the reading. The experimenter also made notes concerning any apparent problems during the playing. Such notes covered, for example, the need to reinforce the metronome beat vocally at any particular point during the session, or the experimenter's perception that the participant's stress level was higher than normal. For Readings 1-5, on completion of the reading the experimenter calculated the level of action slips in that reading, displayed the first frame again, made any necessary adjustments to the settings, and instructed the participant to read the stimulus again. This procedure usually took between seven and fifteen seconds to complete. After Reading 6, the experimenter informed the participant that s/he would now be required to complete the memory test, comprising the same stimulus display with alternate bars blanked out. The experimenter then displayed the memory-test stimulus, set the metronome at a lower tempo, normally 10MM less than the maximum that had been achieved during the previous six readings, and asked the participant to proceed.

5.3 Hypotheses, aims, results and discussion

As previously explained, Stage II generated three sets of data. One set, arising from the Initial Questionnaire, was dealt with in the previous section. It remains to present the hypotheses, aims, results and discussion for the experiment in Stage II. The measurement of participants' performance indicators is treated first. This is followed by the presentation and interpretation of the data from the Final Questionnaire, in order that participants' self-perceptions of the trials be triangulated with their performance indicators. The main aim of analysing the reading variables was to assess whether exposure to span stretching during the course of the trials offered any advantage in terms of improved sight reading and memorisation skills over a training experience based purely on incremental increases in tempo. It was predicted that experimental participants would show greater improvement in both sight reading and memorisation skills as a result of their exposure to span stretching.

5.3.1 Tempo

The level of action slips in a performance is partly dependent on the tempo: increase the tempo and the level of action slips will typically rise; slow the metronomic rate
and the level of action slips can be expected to fall.\textsuperscript{30} Accordingly, before discussing the data for action slips it was necessary to confirm that the mean change in tempo from pretest to posttest were not significantly different for the two groups, in one-, two- and four-part readings. If this were confirmed, the improvement in each group's mean level of action slips could be compared knowing that tempo was not a significant confounding influence. For one-part readings there was no change of tempo at all, since all such readings were at 120MM as previously explained. In respect of two- and four-part readings, however, it was necessary to conduct two-tailed 'between-participant' \textit{t}-tests to gauge the difference between the groups' performance statistically.\textsuperscript{31}

Table 19 shows the data for each participant's mean tempo for Readings 1-6 in the pretest and Readings 1-6 in the posttest, for both two- and four-part stimuli. Participants 1-10 are experimental, and Participants 11-20 are control. In addition, a proportional value (\textit{post-}\textsuperscript{\textit{pre-}}) indicates the change in mean tempo between the tests for each participant. This value was arrived at by dividing the value for the posttest by the value for the pretest. Proportions rather than subtractions were used since the mean tempos for each stimulus category varied markedly; subtractions would have allowed this difference to be a distorting factor; for example, mean tempos in all one-part readings in pretests and posttests were 120.00MM, for two-part readings were 75.54MM, and for four-part readings were 54.57MM. Using proportions avoided the distortions that would have arisen from the other method of calculation, subtraction.

Means and standard deviations for each group appear below the data for Participants 10 and 20 respectively. Means and standard deviations for the proportions were calculated from participants' individual proportional values in the column above rather than a simple division of the mean posttest value by the mean pretest value for each group as they appear to the left in the same row. Experimental participants on average read the two-part stimulus in the posttest 15.7\% faster than they read the two-part stimulus in the pretest, that is, a mean proportion of 1.157 (SD=0.076); for control participants the corresponding mean improvement was 21.0\%, or 1.210 (SD=0.154), \textit{t}(18)=1.00, \textit{p}=0.164. For four-part readings the mean proportional change for experimental participants was 1.239 (SD=0.214), and 1.254 for control

\begin{footnotesize}
\textsuperscript{30} Although Handel (1986) provides some evidence that below a certain tempo performance quality can be degraded, in a reversal of the trend through higher tempos.
\textsuperscript{31}One-tailed \textit{t}-tests are used where the direction of the data (which group will show higher values than the other) is predicted. Here this was not the case, and therefore a two-tailed test was used, thereby reducing the power of the statistics.
\end{footnotesize}
participants (SD=0.182), t(18)=0.17, p=0.432. There was thus no significant difference in tempo improvement between the groups in either stimulus category, and tempo was not considered as an important contaminating factor in assessing the relative improvements in performance quality through the mean levels of action slips, presented in the next section.

<table>
<thead>
<tr>
<th>participant</th>
<th>two-part</th>
<th>four-part</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>pre-</td>
<td>post-</td>
</tr>
<tr>
<td>1</td>
<td>76.67</td>
<td>83.33</td>
</tr>
<tr>
<td>2</td>
<td>87.50</td>
<td>95.00</td>
</tr>
<tr>
<td>3</td>
<td>80.00</td>
<td>91.67</td>
</tr>
<tr>
<td>4</td>
<td>81.67</td>
<td>92.50</td>
</tr>
<tr>
<td>5</td>
<td>65.83</td>
<td>83.33</td>
</tr>
<tr>
<td>6</td>
<td>60.83</td>
<td>74.17</td>
</tr>
<tr>
<td>7</td>
<td>73.33</td>
<td>92.50</td>
</tr>
<tr>
<td>8</td>
<td>56.67</td>
<td>60.00</td>
</tr>
<tr>
<td>9</td>
<td>77.50</td>
<td>85.00</td>
</tr>
<tr>
<td>10</td>
<td>74.17</td>
<td>90.00</td>
</tr>
<tr>
<td>mean</td>
<td>73.42</td>
<td>84.75</td>
</tr>
<tr>
<td>stdev</td>
<td>0.963</td>
<td>10.69</td>
</tr>
<tr>
<td>11</td>
<td>51.67</td>
<td>51.67</td>
</tr>
<tr>
<td>12</td>
<td>76.67</td>
<td>85.83</td>
</tr>
<tr>
<td>13</td>
<td>66.67</td>
<td>85.00</td>
</tr>
<tr>
<td>14</td>
<td>72.50</td>
<td>102.50</td>
</tr>
<tr>
<td>15</td>
<td>72.50</td>
<td>82.50</td>
</tr>
<tr>
<td>16</td>
<td>66.67</td>
<td>94.17</td>
</tr>
<tr>
<td>17</td>
<td>47.50</td>
<td>50.00</td>
</tr>
<tr>
<td>18</td>
<td>68.33</td>
<td>90.00</td>
</tr>
<tr>
<td>19</td>
<td>67.50</td>
<td>71.67</td>
</tr>
<tr>
<td>20</td>
<td>59.17</td>
<td>77.50</td>
</tr>
<tr>
<td>mean</td>
<td>64.92</td>
<td>79.08</td>
</tr>
<tr>
<td>stdev</td>
<td>0.937</td>
<td>17.13</td>
</tr>
<tr>
<td>p</td>
<td>0.164</td>
<td>0.432</td>
</tr>
</tbody>
</table>

Table 19 Mean tempos for two- and four-part stimuli, with proportional values, for experimental participants (1-10) and control participants (11-20)

5.3.2 Action slips

Hypothesis 1 predicted that the reduction in the rate of action slips for the experimental group would be greater than that for the control group. The hypothesis was in three parts, (a), (b) and (c), relating to each of the three stimulus types.

Hypothesis 1

The proportion of action slips made in the posttest compared to the pretest will be less for experimental than control participants in (a) one-part, (b) two-part, and (c) four-part readings.
The data for each participant’s rate of action slips for Readings 1-6 of the pretest and posttest are shown in Table 20. Again Participants 1-10 were experimental and Participants 11-20 were control. Values for the proportional change between the tests are also indicated. Means and standard deviations are highlighted at the bottom of the data for each group. Participant 1, for example, performed much better on average in the six readings of the one-part stimulus in the posttest than s/he did in the same stimulus-type in the pretest, with approximately a tenth (0.107) of the action slips. In the two-part readings the same participant performed slightly worse in the posttest than in the pretest, with the level up by nearly 26% (1.259), slightly improved in the four-part, with a reduction of 20% (0.800).

<table>
<thead>
<tr>
<th>participant</th>
<th>one-part</th>
<th>two-part</th>
<th>four-part</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>pre</td>
<td>post</td>
<td>pre</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>pre</td>
</tr>
<tr>
<td>1</td>
<td>2.869</td>
<td>0.308</td>
<td>1.017</td>
</tr>
<tr>
<td>2</td>
<td>0.820</td>
<td>0.155</td>
<td>0.189</td>
</tr>
<tr>
<td>3</td>
<td>1.230</td>
<td>1.942</td>
<td>1.579</td>
</tr>
<tr>
<td>4</td>
<td>2.186</td>
<td>0.000</td>
<td>0.000</td>
</tr>
<tr>
<td>5</td>
<td>8.130</td>
<td>0.000</td>
<td>0.000</td>
</tr>
<tr>
<td>6</td>
<td>8.470</td>
<td>0.320</td>
<td>0.038</td>
</tr>
<tr>
<td>7</td>
<td>4.065</td>
<td>0.000</td>
<td>0.000</td>
</tr>
<tr>
<td>8</td>
<td>4.878</td>
<td>2.564</td>
<td>0.526</td>
</tr>
<tr>
<td>9</td>
<td>0.813</td>
<td>0.833</td>
<td>1.025</td>
</tr>
<tr>
<td>10</td>
<td>0.273</td>
<td>0.000</td>
<td>0.000</td>
</tr>
<tr>
<td>mean</td>
<td>3.373</td>
<td>0.612</td>
<td>0.346</td>
</tr>
<tr>
<td>stdev</td>
<td>1.610</td>
<td>0.433</td>
<td>0.545</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>4.098</td>
<td>1.984</td>
<td>0.484</td>
</tr>
<tr>
<td>12</td>
<td>0.407</td>
<td>0.000</td>
<td>0.000</td>
</tr>
<tr>
<td>13</td>
<td>3.968</td>
<td>0.642</td>
<td>0.162</td>
</tr>
<tr>
<td>14</td>
<td>9.146</td>
<td>4.487</td>
<td>0.491</td>
</tr>
<tr>
<td>15</td>
<td>1.503</td>
<td>0.320</td>
<td>0.213</td>
</tr>
<tr>
<td>16</td>
<td>1.503</td>
<td>0.320</td>
<td>0.213</td>
</tr>
<tr>
<td>17</td>
<td>1.016</td>
<td>0.000</td>
<td>0.000</td>
</tr>
<tr>
<td>18</td>
<td>0.820</td>
<td>0.962</td>
<td>1.173</td>
</tr>
<tr>
<td>19</td>
<td>2.049</td>
<td>0.481</td>
<td>0.235</td>
</tr>
<tr>
<td>20</td>
<td>3.279</td>
<td>0.641</td>
<td>0.196</td>
</tr>
<tr>
<td>mean</td>
<td>2.779</td>
<td>0.984</td>
<td>0.317</td>
</tr>
<tr>
<td>stdev</td>
<td>2.589</td>
<td>1.357</td>
<td>0.343</td>
</tr>
<tr>
<td>p</td>
<td>0.443</td>
<td>0.074</td>
<td>0.044</td>
</tr>
</tbody>
</table>

Table 20 Mean levels of action slips for two- and four-part stimuli in pretest and posttest, with proportional values, for experimental participants (1-10) and control participants (11-20)

Both groups, on average, reduced their mean level of action slips from pretest to posttest in all three stimulus types, although the degree of change was variable as reflected in the standard deviations. One-tailed ‘between-participant’ t-tests were performed on the proportional values for both groups’ readings of each of the three...
stimulus types. The resulting p-values for the differences between each of the three sets of two proportional values appear at the very bottom of Table 20.

In one-part readings, the mean rate of action slips for experimental participants fell from 3.373% to 0.612%, proportionately 0.346 (SD=0.545) and for control participants from 2.779% to 0.984%, proportionately 0.317 (SD=0.343). This proportional difference was well below the .05 significance threshold: \( t(18)=0.15, p=0.443 \). An analysis of statistical power using the computer application Gpower 2.1.2 showed that more than 100 participants would have been required to achieve an 80% probability of significance.

In two-part readings the rates of action slips for each test were clearly higher for both groups than in one-part readings. The results for improvement between the tests were similar: for experimental participants the mean rate fell from 5.399% to 2.865%, proportionately 0.642 (SD=0.573) and for control participants from 6.159% to 5.186%, proportionately 1.026 (SD=0.563), \( t(18)=1.51, p=0.074 \), a very marked tendency but still below the .05 threshold for significance. While on average the experimental group tended to improve more than the control group, the large standard deviations affected the level of statistical significance of the proportional differences. A calculation of statistical power showed that 42 subjects would have been required to achieve an 80% probability of significance.

In the four-part readings the difference between the groups was statistically significant at the 0.05 level: the mean rate of action slips for experimental participants was considerably higher for this more difficult texture: 13.611% falling to 8.776%, proportionately 0.615 (SD=0.34), with corresponding mean values for control participants of 14.028% falling to 11.000%, proportionately 0.954 (SD=0.479), \( t(18)=1.81, p=0.044 \). Hypothesis 1 was therefore confirmed only in the reading of four-part stimuli (c). Figure 48 illustrates the proportional values for the mean levels of actions slips as they changed from pretest to posttest; red represents the values for the experimental groups, blue for the control group. The further below the horizontal axis a bar reaches, the greater the improvement over the course of the trials.

5.3.3 Discussion of the results for action slips

It would appear from the data that the training advantage enjoyed by the experimental group became progressively more pronounced with the difficulty of the
reading task, from the easiest task, the one-part, through the two-part to the most difficult task, the four-part. While it would have been more strongly confirmatory of Hypothesis 1 if the reading of all three stimulus types had yielded significant disparities between the groups, the actual result suggests that the one-part stimuli were simply too easy for both groups to represent the kind of challenge that might have differentiated their contrasting training experiences. When the task demands are raised, it appears, the advantages of span stretching are more evident. Stage II, then, has demonstrated that SightReader has the potential to be a valuable tool for training sight reading skills as measured by increased accuracy of the experimental group, provided the task demands are high enough.

![Figure 48 Posttest/pretest proportional values for mean rates of action slips for experimental (red) and control (blue) groups](image)

It is, however, worth approaching the results from a negative stance by asking whether, apart from the lower task demands, there might have been other contributing factors to the absence of significant differentials between the groups in the one- and two-part readings. Some of these factors may be related to the design of the experiment. There was, for example, a clear bias towards the control group in that the criterion for sight reading skill was the mean rate of action slips as measured under the normal condition rather than under the span stretching condition. One or
other condition had to be chosen for the assessment process at the start and finish of
the trials to be credible, and it appeared to be preferable to use the normal condition.
The reason for this decision was that using span stretching on SightReader would
have been a totally foreign experience to which to expose control participants during
assessment, whereas all experimental participants would have experienced
progressive tempo increases during rehearsal, on which the normal condition was
based, even if they were not primed for such assessment during Sessions 2-5 as had
control participants. In other words, a bias towards one group or the other was
inevitable, and the distortion was less on logical grounds under the normal condition.
Thus control participants were inherently favoured by the fact that the condition
used in their training was the one used to measure the dependent variable for
improvement; by contrast, experimental participants did not enjoy this advantage.

Other aspects of the design of Stage II may also have affected the potential
differences between the groups. The relatively brief period of exposure, only four
sessions over two weeks in which the training differed for the two groups, may have
been too brief for span stretching to have had a major impact on the experimental
group, particularly as participants were exposed to only two sessions per week.
(One participant in the control group suggested that there were not enough sessions
to 'make a difference' to participants' skill-levels, as reported in Figure 58). The
small sample size of 10 participants in each group may have made significance at the
0.05 level more difficult to achieve than with larger groups; although a slightly larger
sample may have yielded significance for two-part readings, the data for one-part
readings suggest that even a considerably larger sample would have been unlikely to
result in significant differences.

5.3.4 Memorisation
Hypothesis 2 predicted that experimental participants would show a significantly
greater improvement than control participants in their scores for the memory tests
that came at the end of each set of six readings of each stimulus in each trial session.
In the memory tests, all participants played from a specially prepared score that
presented in alternate bars the notation of the stimulus and blank stave-lines. No
span stretching was used, and tempo was reduced by 10MM from the fastest tempo
attained in the previous six readings. Participants’ performance was measured as the
proportion of erased notes or chords that were omitted or incorrectly played, with
half-marks awarded on the same basis as for the scoring of action slips in the six
readings leading up to the memory test (see Section 5.3.7).
Hypothesis 2

The proportion for scores in the memory tests in the posttest compared to the pretest will be greater for experimental than control participants, for (a) one-part, (b) two-part, and (c) four-part readings.

The reasoning behind Hypothesis 2 was that span stretching can be considered a form of localised memorisation, as explained in Section 5.1.4. Experimental participants experienced such localised memorisation in reading each stimulus in this way not once but six times in succession before each memory test. Thus it was hoped that this training would have advantaged experimental participants over control participants in relation to each individual memory test, and that in the course of completing 15 such memory tests in Trial Sessions 1-5 a clear improvement in experimental participants’ ability to memorise short passages of music would be reflected in a greater improvement than for control participants.

| participant | one-part | | | | two-part | | | | | four-part | | | |
|-------------|---------|--------|----|-----------------|---------|--------|----|-----------------|---------|--------|----|-----------------|---------|--------|
|             | pre     | post   | post/ | pre            | post    | post/  | pre     | post   | post/ | pre     | post  | post/  | pre     | post   | post/ |
| 1           | 31.03   | 8.33   | 0.268 | 39.13          | 25.00   | 0.639  | 17.86   | 8.33   | 0.466 | 22.41   | 20.83 | 0.929  | 67.39   | 50.00  | 0.742 |
| 2           | 22.41   | 20.83  | 0.929 | 67.39          | 50.00   | 0.742  | 75.00   | 79.17  | 1.056 | 3.45    | 3.70  | 1.072  | 45.65   | 28.13  | 0.616 |
| 3           | 14.75   | 12.00  | 0.814 | 32.61          | 18.75   | 0.575  | 55.77   | 27.08  | 0.486 | 24.00   | 10.71 | 0.446  | 28.95   | 34.38  | 1.188 |
| 4           | 34.48   | 17.86  | 0.518 | 56.52          | 71.88   | 1.272  | 100.00  | 72.92  | 0.774 | 20.00   | 28.57 | 1.429  | 17.39   | 65.63  | 3.774 |
| 5           | 60.00   | 39.29  | 0.655 | 97.83          | 68.75   | 0.703  | 100.00  | 68.75  | 0.688 | 22.00   | 16.07 | 1.730  | 63.04   | 31.25  | 0.496 |
| 6           | 24.14   | 44.64  | 1.849 | 60.87          | 62.50   | 1.027  | 46.88   | 54.17  | 1.156 | 24.53   | 20.20 | 0.971  | 50.94   | 45.63  | 1.071 |
| 7           | 60.00   | 39.29  | 0.655 | 97.83          | 68.75   | 0.703  | 100.00  | 68.75  | 0.688 | 0.24    | 0.06  | 0.082  | 65.22   | 75.00  | 1.150 |
| 8           | 22.00   | 16.07  | 1.730 | 63.04          | 31.25   | 0.496  | 64.29   | 66.67  | 1.037 | 24.14   | 44.64 | 1.849  | 60.87   | 62.50  | 1.027 |
| 9           | 22.00   | 16.07  | 1.730 | 63.04          | 31.25   | 0.496  | 64.29   | 66.67  | 1.037 | 24.14   | 44.64 | 1.849  | 60.87   | 62.50  | 1.027 |
| 10          | 24.14   | 44.64  | 1.849 | 60.87          | 62.50   | 1.027  | 46.88   | 54.17  | 1.156 | 25.63   | 20.20 | 0.971  | 50.94   | 45.63  | 1.071 |
| mean        | 25.63   | 20.20  | 0.971 | 50.94          | 45.63   | 1.071  | 67.45   | 56.04  | 0.818 | 14.75   | 13.44 | 0.544  | 23.23   | 20.32  | 0.976 |
| stdev       | 21.99   | 13.44  | 0.544 | 23.23          | 20.32   | 0.976  | 24.55   | 22.39  | 0.238 | 14.75   | 13.44 | 0.544  | 23.23   | 20.32  | 0.976 |

Table 21 Mean scores in memory tests, with proportional values, for experimental participants (1-10) and control participants (11-20)
The results of the memory tests are set out in Table 21 in the same manner as for tempo and action slips above. High values represent a large proportion of wrong or omitted notes or chords; low values represent a small rate of dysfunction. It is noted that in the pretest the experimental group tended to score higher than the control group in one- and four-part stimuli, and scored almost exactly the same for two-part stimuli. Be that as it may, one-tailed t-tests were carried out on the proportional values representing changes in scores from pretest to posttest, and yielded no significant differences between the groups: for one-part readings the mean proportional change for the experimental group was 0.971 (SD=0.544) and for the control group 1.0710 (SD=0.588), t(18)=0.39, p=0.349; for two-part readings the mean for the experimental group was 1.071 (SD=0.976) and for the control group 1.248 (SD=0.614), t(18)=0.40, p=0.349; and for four-part readings the mean for the experimental group was 0.818 (SD=0.238), and for the control group 0.9147 (SD=0.208), t(18)=0.96, p=0.124. Calculations of statistical power showed that well over 100 participants would have been required to achieve an 80% probability of significance. Hypothesis 2 was therefore not confirmed.

5.3.5 Discussion of the results for memorisation

The fact that the scores in the memory tests did not significantly favour experimental participants resonated with the results for Question 12 in the Final Questionnaire, where experimental and control groups both rated the difficulty of the memory tasks with a mean of 6.8 and similar SDs, t(18)=0.00, p=0.500. In the same questionnaire, experimental participants, with a mean of 5.30, rated the usefulness of the memory tasks less than did control participants, with a mean of 6.90, a difference that approached significance, t(18)=1.63, p=0.061 (see the next section). The results for memorisation scores were the only major surprise in the empirical results of this study, and are difficult to explain. It would appear that more empirical work is necessary before it can be shown whether there are circumstances in which SightReader has a positive training effect on users’ ability to memorise a short passage of music, or whether the results attained in the present study are robust under a variety of procedures, such as when different types of memorisation tests are used.

5.3.6 The Final Questionnaire

At the end of the posttest the Final Questionnaire was administered to each participant. A compressed version of the questionnaire is shown in Figure 49,
omitting the answer templates to avoid duplication. The data it generated were intended to provide information of a ‘participantive’ quality that could be triangulated with the data from the measurement of reading variables as presented and discussed in the previous section. The objective here was to arrive at a richer and more reliable interpretation of the experiment than would have been possible with the measurement of reading variables alone.

### Final Questionnaire

**Participant number:**

1. To what extent do you feel there has been an improvement in your overall sight reading ability during the course of the trials? Tick one box.

For Questions 2-7, indicate the extent to which you feel there has been an improvement in the following sight reading skills during the course of the trials. Tick one box for each skill:

2. Being confident and relaxed:
3. Playing the correct notes:
4. Keeping to a steady beat:
5. Performing with musical flow:
6. Keeping your place on the score:
7. Playing at a fast tempo:
8. To what extent do you feel there has been an improvement in your ability to memorise a short passage of music after repeatedly playing it from score? Tick one box.

9. [Experimental group only] How difficult or easy did you find the increases in the span setting for successive readings? Tick one box.

9. [Control group only] How difficult or easy did you find the increases in tempo for successive readings? Tick one box.

10. How difficult or easy did you find playing the memory exercise after the six readings of each excerpt? Tick one box.

11. [Experimental group only] How useful did you find the increases in the span setting for successive readings? Tick one box.

11. [Control group only] How useful did you find the increases in tempo for successive readings? Tick one box.

12. How useful did you find playing the memory exercise after the six readings of each excerpt? Tick one box.

13. What were the two best aspects of the training process? [verbal response]

14. What were the two worst aspects of the training process? [verbal response]

### Figure 49 Compressed version of the Final Questionnaire

Three hypotheses (3, 4 and 5) prompted the first eight questions and predicted that experimental participants would report a greater improvement in various specified skills during the course of the trials than would control participants. One-tailed t-tests were therefore conducted on these data. Aims 1 and 2 concerned further
questions (9-12) for which it was difficult to predict participants' responses, the data for which were processed with two-tailed $t$-tests. Aim 3 was related to verbal responses prompted by the last two questions (13 and 14) in the questionnaire.

The first 12 questions required participants to tick one of 10 horizontally aligned boxes with descriptors appended (as was the case in Questions 4-7 in the Initial Questionnaire). These boxes, later assigned values from one on the left to 10 on the right for statistically processing, are not shown in Figure 49, as stated above. For Questions 1-8, participants ticked one box according to the descriptors shown in Figure 50.

```
[10 9 8 7 6 5 4 3 2 1]

very strong improvement  moderate improvement  slight improvement  no improvement
```

**Figure 50** Response template for Questions 1-8

Question 1 was related to Hypothesis 3, which concerned participants' perceptions of the extent to which their overall sight reading ability had improved over the course of the trials.

**Hypothesis 3**

Experimental participants will report a significantly greater improvement in sight reading skills than will control participants.

<table>
<thead>
<tr>
<th></th>
<th>all mean</th>
<th>SD</th>
<th>exp mean</th>
<th>SD</th>
<th>cont mean</th>
<th>SD</th>
<th>$t$-value</th>
<th>$p$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 overall sight reading</td>
<td>6.60</td>
<td>1.47</td>
<td>7.40</td>
<td>0.97</td>
<td>5.80</td>
<td>1.48</td>
<td>2.87</td>
<td>0.005</td>
</tr>
<tr>
<td>2 confidence &amp; relaxation</td>
<td>6.20</td>
<td>1.70</td>
<td>6.40</td>
<td>1.07</td>
<td>6.00</td>
<td>1.83</td>
<td>0.51</td>
<td>0.307</td>
</tr>
<tr>
<td>3 playing the right notes</td>
<td>6.15</td>
<td>1.14</td>
<td>6.40</td>
<td>1.17</td>
<td>5.90</td>
<td>1.10</td>
<td>0.98</td>
<td>0.170</td>
</tr>
<tr>
<td>4 keeping to the beat</td>
<td>6.05</td>
<td>2.31</td>
<td>5.70</td>
<td>2.91</td>
<td>6.40</td>
<td>1.58</td>
<td>-0.67</td>
<td>0.256</td>
</tr>
<tr>
<td>5 musical flow</td>
<td>4.95</td>
<td>2.04</td>
<td>5.40</td>
<td>2.17</td>
<td>4.50</td>
<td>1.90</td>
<td>0.99</td>
<td>0.169</td>
</tr>
<tr>
<td>6 keeping place on score</td>
<td>6.45</td>
<td>2.06</td>
<td>7.00</td>
<td>2.45</td>
<td>5.90</td>
<td>1.52</td>
<td>1.21</td>
<td>0.122</td>
</tr>
<tr>
<td>7 playing fast enough</td>
<td>6.90</td>
<td>1.89</td>
<td>7.50</td>
<td>1.90</td>
<td>6.60</td>
<td>1.58</td>
<td>1.15</td>
<td>0.132</td>
</tr>
<tr>
<td>8 memorisation</td>
<td>4.45</td>
<td>2.26</td>
<td>4.90</td>
<td>2.77</td>
<td>3.90</td>
<td>1.60</td>
<td>0.99</td>
<td>0.168</td>
</tr>
</tbody>
</table>

Table 22 Results for Questions 1-8

Question 1 was the most important in terms of the objectives of Stage II, and the only question to yield a significant difference between experimental and control groups. The responses from the experimental group averaged 7.4 (greater than a 'moderate improvement') and those of the control group averaged 5.8 (below the
same descriptor). This difference was statistically significant, $t(18)=2.87, p=0.005$. Participants who were trained by the process of span stretching thus rated the improvement in their sight reading skills significantly better than participants trained under the normal condition rated their own improvement in sight reading. Hypothesis 3 was therefore confirmed.

Hypothesis 3 was therefore confirmed.

![Figure 51](image)

**Figure 51** Results for experimental (blue) and control (mauve) groups for Questions 1-8

Questions 2-7 were related to Hypothesis 4. These questions asked participants to rate their improvement in six skills that might be considered to be associated with sight reading ability. Confidence and relaxation, and their ability to play the correct notes, perform with musical flow, keep their place on the score, and play at a fast tempo while sight reading, experimental participants rated their improvement at a higher level than did control participants (a mean of 6.54 vs 5.78). Although these data were in the predicted direction, the differences were not significant at the .05 level for any individual skill. Hypothesis 4 was therefore not confirmed. The reason for this is not clear; it may have been the case that participants were more uniform across groups in their self-appraisal when it came to specific, identifiable skills; it is plausible that more confidence or technical insight might have been required to give an accurate self-appraisal in such specific skills.
Hypothesis 4
Experimental participants will report a significantly greater improvement in specified skills associated with sight reading than will control participants.

The data for one skill, that of ‘keeping to a steady beat’, tended to lie in the opposite direction to the prediction; here, control participants on average rated themselves more highly than experiment participants rated themselves (6.40 against 5.70). One explanation for this apparent discrepancy might be that during the trials control participants were exposed to greater metronomic discipline through frequent gradual increases in tempo; experimental participants experienced comparatively few changes in tempo. The reversal in the direction of these data might also be due to the marked variance in experimental participants’ responses (SD=2.91).

While Question 1 concerned participants’ perceived increase in their sight reading ability over the course of the trials, Question 8 applied to their perception of any increase in their ability to memorise a short passage of music. Question 8 was connected with Hypothesis 5:

Hypothesis 5
Experimental participants will report a significantly greater improvement in their ability to memorise a short passage of music than will control participants.

Although this was not a significant difference, the higher rating of experimental participants was in the expected direction, consistent with span stretching as a form of localised memorisation as explained in Section 5.1.4. It is interesting to note that the mean rating by both groups in response to Question 8 tended to be lower than for Question 1.

Figure 52 shows the answer template, and Table 23 and Figure 53 the results for Questions 9-12 of the Final Questionnaire. These questions concerned participants’ appraisals of their experience of the span setting (experimental group) and tempo setting (control group) during the six successive readings of each stimulus, and the subsequent memory tasks. Aim 1 concerned participants’ perception of the usefulness of several key aspects of the training experience, and was pursued in Questions 9 and 10, concerning the perceived difficulty of the setting and memory tests respectively. Here the use of separate questionnaires for each group was required so that the appropriate variable could be specified.
Aim 1

To discover whether the usefulness of (a) the training method that each group was exposed to (that is, span stretching for the experimental group and tempo increases for the control group), and (b) the memory test that both groups were exposed to, were judged differently by each group.

In relation to Question 9, the experimental participants reported that they found successive increases in the span setting slightly more difficult on average to negotiate than the control participants found the increases in tempo (5.8 vs 5.0, both above ‘a little difficult’), but the difference was not statistically significant. It should be pointed out that these data depended on the extent to which span stretching (for experimental participants) and tempo increases (for control participants) were varied by the experimenter during the trials. The result for Question 9 is thus desirable, since it indicates that each group was ‘pushed’ approximately equally in the teaching situation. The two groups were virtually identical in their responses to Question 10, which required an assessment of the difficulty of the memory tasks. This identity was surprising, since it was expected that the localised memory demands of the span stretching during the six preceding readings of each stimulus would have primed experimental participants for the memory test significantly more efficiently than the increased tempo settings experienced by control participants, thus making the memory task appear easier to them.

Aim 2 concerned participants’ perception of the difficulty (as opposed to usefulness) of several key aspects of the training experience, and was connected with Questions 11 and 12, for which the answer template is shown in Figure 54 and the numerical data are set out in Table 24.
Aim 2

To discover whether the difficulty of (a) the training method that each group was exposed to (that is, span stretching for the experimental group and tempo increases for the control group), and (b) the memory tasks that both groups were exposed to, were judged differently by each group.

<table>
<thead>
<tr>
<th>[10]</th>
<th>9</th>
<th>8</th>
<th>7</th>
<th>6</th>
<th>5</th>
<th>4</th>
<th>3</th>
<th>2</th>
<th>1</th>
</tr>
</thead>
<tbody>
<tr>
<td>[ ]</td>
<td>[ ]</td>
<td>[ ]</td>
<td>[ ]</td>
<td>[ ]</td>
<td>[ ]</td>
<td>[ ]</td>
<td>[ ]</td>
<td>[ ]</td>
<td>[ ]</td>
</tr>
</tbody>
</table>

very useful moderately useful a little useful not very useful of no use at all

Figure 53 Response template for Questions 11 and 12

<table>
<thead>
<tr>
<th>Question</th>
<th>Mean</th>
<th>SD</th>
<th>Experiment Mean</th>
<th>SD</th>
<th>Control Mean</th>
<th>SD</th>
<th>t-value</th>
<th>p =</th>
</tr>
</thead>
<tbody>
<tr>
<td>11 Usefulness of increased settings</td>
<td>7.90</td>
<td>1.41</td>
<td>7.70</td>
<td>1.70</td>
<td>8.10</td>
<td>1.10</td>
<td>-0.62</td>
<td>0.540</td>
</tr>
<tr>
<td>12 Usefulness of memory tasks</td>
<td>6.10</td>
<td>2.29</td>
<td>5.30</td>
<td>2.75</td>
<td>6.90</td>
<td>1.45</td>
<td>-1.63</td>
<td>0.121</td>
</tr>
</tbody>
</table>

Table 24 Results for Questions 11 and 12

Figure 54 Results for experimental (blue) and control groups (mauve) for Questions 9-12
For Questions 11 and 12, two-tailed $t$-tests were performed. In Question 11, participants indicated that the setting they were exposed to was 'moderately useful', with a mean of 7.70 for experimental participants, which was not significantly different from the mean of 8.10 for control participants. The slightly lower mean rating given by experimental participants to the usefulness of increasing minimum span size may have been due simply to a perception among all musicians that increased tempo equates with progress. Manipulating span size, by contrast, is a novel technique. It should be pointed out that in the verbal responses at the end of the Final Questionnaire, nine of the ten experimental participants made favourable comments about SightReader's span stretching function, whereas only three of the ten control participants made favourable comments about the increasing tempo settings they experienced. This does not appear to be borne out in the responses to Question 11.

It was in their appraisals of the usefulness of the memory tasks that the groups tended to a stronger divergence from each other, $t(18)=1.63, p=0.121$. Control participants appeared to find the tasks more useful (a mean of 6.90, just below 'moderately useful') than experimental participants (5.30, just above 'a little useful'). The reasons for this are not at all obvious. It may be that the memory tasks loomed larger in the minds of the control participants because they were set against a traditional and unremarkable process; by contrast, experimental participants were participanted to a novel process that by itself already manipulated the memory system. This may have led experimental participants to place relatively less emphasis on the memory tasks.

Aim 3 concerned the interpretation of participants' free (written) impressions of the trials, with the expectation of generating valuable information on the experience of participants in both groups. Accordingly, participants were asked in Questions 13 and 14 to specify the best two and the worst two aspects of the training process. These questions were related to Aim 3:

**Aim 3**
To discover how participants viewed the training experience by eliciting from them qualitative verbal comments on their choice of the best and worst aspects of the trials.

Participants' responses to Question 13 ('the two best aspects') are reproduced in Figure 55; their responses to Question 14 ('the two worst aspects') are reproduced in Figure 56. Here the data are coded into themes; coding refers to a stage in the analysis of data that is in the form of a set of verbal comments ranging over a number
of themes. On the left-hand side of both figures appear comments by experimental participants; on the right-hand side appear comments by control participants on the same theme. Where the same, or virtually the same comment was made by more than one participant in a group, this is indicated by a number in square brackets after the comment representing the number of times it appeared, for example ‘[x 2]’. Table 25 sets out the number of comments made by participants in each group under six main themes that are evident from the data. Experimental participants wrote a mean of 1.9 comments on the best aspects of the trials, and 1.5 comments on the worst aspects. Control participants wrote fewer comments, a mean of 1.1 and 0.9 comments respectively, possibly because their experience was less novel and closer to traditional musical rehearsal than that of the experimental participants.

Since participants were free to choose the themes of their comments, their choice in this respect gave some indication as to what they regarded as important about their experience in the trials. For experimental participants, the process of stretching the span size was the most frequently occurring theme, with nine favourable comments, for example, ‘Being made to look several beats ahead and memorise’ and ‘The blanking out helped me to keep looking forward and stopped me from looking back.’ Although no unfavourable comments were made concerning the use of span stretching in the training process, three experimental participants mentioned problems with the span stretching mechanism under the theme of equipment. One experimental participant found the operation of the rhythm compensator on minims and dotted minims ‘a distracting problem.’ Most participants, however, appeared to have rapidly adapted to this aspect of the mechanism. Another participant complained of a certain bumpiness in the mechanism of erasing successive chords or notes. The participant was referring to the occasional tendency of the erasure mechanism to be momentarily late in its operation, arising from the fact that SightReader’s operation required more speed and memory of the computer than was readily available. This problem did not occur frequently enough to be considered a contaminating influence on the data. A third participant rightly pointed out that when a note or chord is accidentally double-struck, the size of the span setting is altered. Experimental participants were alerted to this problem at the beginning of Trial Session 1, and advised to avoid such double-striking. The problem would not have occurred if the mechanism had been programmed in another language, such as C++, rather than with Opcode MAX.
Control participants were more occupied by the issue of memory, which probably loomed larger in their impression since they had no opportunity to experience *SightReader*'s span stretching function. Six control participants made favourable comments about the memory tests. One claimed that knowing s/he would have to attempt the memory exercise made her/him attend to the (pitch) contour of the music during the six successive readings. Yet another thought the readings were a good preparation for the memory task, yet another liked the form of the memory task, with alternate bars excised. Two experimental participants commented favourably on this, one of them identifying the process of isolating latency on the sixth reading by forcing a slower tempo. The theme of memory also drew two unfavourable comments from each group, all pointing to the difficulty of the memory tasks. These comments resonate with the higher mean numerical response to Question 10 compared with Question 9.

Participants appeared to be mixed in their attitudes to the use of three types of stimulus, or to have no strong feelings either way. Two participants liked the variety of stimulus texture used in the trials, but the use of four-part stimuli drew criticism from four participants as being too great a leap in difficulty compared to the one- and two-part exercises. One participant felt that there were not enough sessions to 'make a difference' to participants' skill-levels, a matter that was raised in Section 5.3.3.

In relation to the theme of stress, two experimental participants made unfavourable comments in relation to the theme of stress, and one made a favourable comment about stress under 'Other matters' concerning the challenging conditions that were experienced in the trials. Although this is consistent with the expectation that span stretching caused more arousal in participants than did increases in the tempo.
setting, evidence for differences in arousal was not strongly exhibited in participants' comments. Under the theme of tempo, two experimental and four control participants made favourable comments about the discipline of keeping strict time to a metronome; two comments by control participants cited the benefits of the gradual increases in tempo they experienced. These comments supported the suspicion that comparing the quantitative variables of both groups using only the tempo setting in the pretest and posttest favoured the control group, as discussed in Section 5.3.3.

5.4 Conclusions

The most important result of Stage II was that repeated exposure to SightReader's span-stretching appeared to bestow a greater advantage to the experimental participants than to the control participants. This result was achieved despite the fact that the experimental design appeared to favour control participants as explained previously. The responses to Question 1 in the Final Questionnaire resonated with these data: experimental participants rated the improvement in their sight reading skills significantly higher than control participants rated their own improvement.

Some data, however, were not readily explicable. In the Final Questionnaire, there were no significant differences between the improvement-ratings of the two groups in respect of 'subskills' that might be associated with strong sight reading. The failure of this experiment to show that repeated exposure to span stretching improved memorisation skills was counter-intuitive in the light of the theoretical background set out in Section 5.1.4. It is contended that span stretching as localised memorisation practice therefore requires more research, perhaps with a larger sample and different methodology.

SightReader's span stretching proved to be reasonably easy to administer to experimental participants after they became used to the process; the method of negotiating the span setting with the participants was, in retrospect, a successful teaching strategy, auguring well for the prospects of self-use. Most experimental participants expressed a positive attitude towards SightReader, and some were enthusiastic about its potential to improve their sight reading skills. There is little doubt that span stretching caused more arousal for experimental participants than the did traditional tempo increases for control participants. For this reason, it is recommended that careful consideration be given by instructors as to the role of stress in an individual student's rehearsal when introducing SightReader is contemplated.
Q13 The two best aspects
Experimental group

**Tempo**
- Working to a steady metronome beat very useful.
- Keeping to the metronome, as I am normally bad at keeping time.

**Stimuli**
- Training with stimuli of different difficulty (ie melody/2-pt/4-pt).

**Span size**
- Increase in the span setting.
- Being forced to look ahead. [x2]
- Learning to read ahead rather than reading the very notes that I’m playing.
- Blanking out notes forced me to keep my eyes on the music.
- Being made to look several beats ahead and memorise.
- Extension of span.
- Increasing the span setting over the duration of training.
- The blanking out helped me to keep looking forward and stopped me from looking back.

**Memory**
- Slowing tempo down to test my [working] memory rather than my ability to play quickly.
- Testing my memory to the limits.

**Other matters**
- Being forced to look at the music rather than my hands.
- Feeling yourself improve.
- Gave me practice, training, and experience in sight reading.
- The nature of the exercise as a whole, being somewhat challenging.

Control group

- Managing the tasks at fixed tempos.
- Keeping to the metronome helped to improve my rhythm.
- Increasing the tempo gradually. [x2]

- The use of various textures (single, 2-pt, 4-pt).

- Training your memory to work.
- Memory exercises: helped me look at the contour/shape of the notes.
- Gradual increase of tempo and six playings before memory test. A good preparation for the memory test.
- The encouragement to attempt playing from memory early in the learning process.
- The memory test.
- The memory test with alternate bars missing.

- Concentrating on finding patterns during the first six playings.
- The discipline of not looking at my hands, and hitting the correct note first time (not smudging).
- It was fun. The session were of an appropriate length.
- Having to practise sight reading at regular intervals.
- Having someone watch and monitor as I sight read.

Figure 55 Thematically coded responses to Question 13
Q14  The two worst aspects

**Experimental group**

<table>
<thead>
<tr>
<th>Tempo</th>
<th>Control group</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Stimuli</strong></td>
<td></td>
</tr>
<tr>
<td>• Being able to constrain myself to some of the slow tempos.</td>
<td>• Putting the 4-part exercise last, when concentration was lapsing (although it could hardly have been put first).</td>
</tr>
<tr>
<td>• Four-part harmony exercises were substantially more difficult than two- and one-part—losing my concentration! [x 2]</td>
<td>• Because some pieces were harder than others, it was difficult for me to judge whether I improved or not.</td>
</tr>
<tr>
<td></td>
<td>• Unchanging format: always had problems with 4-pt so couldn’t feel the improvement after each session.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Memory</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>• There wasn’t time to concentrate on memorising while concentrating on the span stretching exercises—it just didn’t seem to sink in [to the memory].</td>
<td>• The memory tests.</td>
</tr>
<tr>
<td>• Not being able to remember the missing bars in the memory tests.</td>
<td>• Some excerpts were easier to memorise than others.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Equipment</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>• Display changes on long notes/ chords were a distracting problem.</td>
<td>• The use of the electronic keyboard rather than piano.</td>
</tr>
<tr>
<td>• Blanking out of notes seemed bumpy at times.</td>
<td></td>
</tr>
<tr>
<td>• When you accidentally double-strike a note and it puts the span setting out.</td>
<td></td>
</tr>
<tr>
<td>• The electronic keyboard brought out mistakes that wouldn’t normally be obvious on an acoustic piano (eg bumping neighbouring keys); also the difficulty of executing phrasing, dynamics and the lack of a sustaining pedal made the keyboard a little disconcerting, but not a big problem.</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Stress</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>• My own lapses in concentration; reaching the limit of my capability.</td>
<td></td>
</tr>
<tr>
<td>• As the settings increased, and in the memory task, feeling my adrenalin increase, feeling nervous, tense, loud heart beat.</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Other matters</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>• Trying to figure out the fingering while playing.</td>
<td></td>
</tr>
<tr>
<td>• I don’t know whether six sessions are really enough for some sort of improvement to show. [x 2]</td>
<td></td>
</tr>
</tbody>
</table>

---

Figure 56  Thematically coded responses to Question 14
6 Conclusions and directions for further research

At the start of this dissertation it was argued that the training of musical skills is likely to undergo considerable change over the next few decades as rapidly accelerating advances in information technology are applied to teaching and learning. It was suggested that such advances offer the prospect of powerful new diagnostic and training methods that can only be dreamed of at the present time, and that the application of technology to music training is as yet crude and cumbersome. Motivated by a desire to improve on the existing resources for training music reading, this study took up the challenge to make a substantive contribution towards the development and application of technology for the teaching and learning of musical skills. Specifically this challenge concerned a search for ways of improving sight reading by applying scientific methods to its teaching and learning. The search covered several relevant areas such as the definition and importance of sight reading and role of the memory system and eye movement in the music reading process; it culminated in the development of SightReader and the experimental observation of its use as a tool for improving sight reading skills. This was based on the expectation that the device would be confirmed as an effective tool for the training of music performance skills. In the light of the success of Stage II, methods by which the use of SightReader might be integrated into a wider curriculum are explained later in this chapter.

Although this study might have been contextualised in a number of ways, for example in terms of computer programming and the psychology of tasks involving a complex musculoskeletal response, it was primarily located in the art and science of teaching and learning musical performance as discussed in Chapter I. The study began by analysing sight reading in terms of layers of familiarity, and problematising the training of sight reading skills in relation to their educational and psychological contexts. The music reading process was then analysed in terms of its anatomy and patterns of information flow. Reviews were conducted of selected literature on the memory system and on eye movement in reading. These reviews uncovered many facts and suppositions about the music reading process, both directly in relation to music reading and by analogy with language reading and other tasks, and revealed the
particular combination of ideas that led to the development of SightReader. The reviews also highlighted the extent to which the music reading process is not understood, and the problematic nature of experimental methodology in relation to the scientific observation of music reading, in particular of eye movement in music reading.

This lack of understanding was the reason for conducting Stage I of this study, an experimental investigation under the normal condition of the two key behavioural phenomena that SightReader was to alter in Stage II, the eye-hand span and leftward refixation. In addition to providing important background knowledge about these phenomena for Stage II, numerous findings were made in Stage I about oculomotor behaviour in music reading that are interesting in themselves. In Stage II, SightReader was designed and written, and used in an experiment in which these phenomena were manipulated in a teaching situation. Thus it was possible to measure the effects of such manipulation on skill acquisition, and to determine whether the device is an effective tool for improving sight reading skills. In Stage II the findings suggested that SightReader can be a powerful tool in the teaching and learning of sight reading skills.

The next section presents a summary of the findings of the study in greater detail. The summary emphasises the most original and innovative parts of the study and draws them together to present what is now known about eye movement, memory and tempo in the sight reading of keyboard music that was not known before the study. Particular prominence is given to findings that might lead to improvements in the teaching and learning of music reading skills. Following this, proposals are made for future research into the matters at issue in this dissertation.

6.1 Summary of findings

6.1.1 Foundations

The opening analysis of sight reading in terms of familiarity with a musical text highlighted the importance of memory in the music reading process. Familiarity was likened to the player’s long-term memory of the actual notes of a text gained from previous readings of that text (actuality familiarity), and the probability of what those notes are, based on the player’s memory of the broader style and genre from which the text is drawn (probability familiarity). It was explained that the increase in familiarity with a text that is typically gained through repeated readings involves a gradual shift from probability to actuality familiarity, and a simultaneous shift in
reliance from external visual input of information from the score towards internal input from long-term memory. Since these shifts occur gradually during the learning process, it was argued that there is no in-principle difference between the first reading of a text (traditionally regarded as sight reading) and subsequent readings during the stage of learning to perform that text.

For this reason, it was proposed that for experimental purposes the traditional notion of sight reading is inadequate, that it should be closely defined for each particular learning situation, and that it is useful to include the first several readings in the definition of sight reading unless there is a reason not to do so. Accordingly, a specific definition of sight reading was provided for each of the two experiments in this study, influencing the selection of both participants and stimuli. Thus it is essential for the reliable observation of music reading in experimental situations to establish the objectives of a sight reading situation and to define sight reading accordingly.

The music reading process, however, consists of far more than the functioning of the memory system. A case was put that the process involves five anatomical components (the eyes, oculomotor muscles, musculoskeletal system, tactile system, and ears) in addition to the memory system, and that these components are linked through the memory system by seven continuous signals. The signals were divided into three functional categories, input, command and feedback. It was also posited that each anatomical component and signal can be thought of as belonging to either the oculomotor and musculoskeletal group, each of which is related to a different location on the music score at any one time. The fact that each signal is clearly distinguished by its temporal characteristics was later used to infer on logical grounds that the signals are stored and processed by different resources in working memory.

This analysis of the music reading apparatus made it easier to identify which of its components presented new possibilities for active intervention in the training of sight reading. It became clear that traditional training methods have largely focused on the musculoskeletal group—by monitoring and directly intervening to improve body movement and auditory output—and that it would therefore be profitable for this study to explore the possibilities of observing and intervening in the functioning of comparatively neglected parts of the apparatus, the oculomotor group and the memory system. For this reason, the primary foci of investigation in this study were the way the eyes move over the score, and the nature of the eye-hand span. Stage I was oriented towards the observation of these phenomena, and Stage II towards their
control. It might be added that the analytical approach taken in Chapter 1 might have been well placed towards the beginning of Galbrielssohn's (1999) laudable account of the psychology of musical performance, acting as an introductory map of the music reading apparatus and as a foundation for future research into it.

6.12 Literature reviews

As pointed out at the start of Chapter 2, the field of human memory is notoriously unstable and complex. Some facts about memory that are relevant to this study are, however, reasonably certain. In music reading, the contents of the eye-hand span are stored and processed in working memory, a highly activated part of the memory system that is of limited capacity. This limited capacity appears to be an important factor in music reading skill: a large capacity has been shown to be associated with fast and skilled reading of both language and music, and has a number of advantages that were explained in Section 5.1. This was one of the main justifications for hypothesising that the strategic stretching of the span by SightReader in Stage II would be of greater benefit than traditional training.

Two claims in Atkinson & Shiffrin's (1968) seminal study of human memory are relevant to SightReader's span stretching function. First, Atkinson & Shiffrin held that information can be transferred from short- to long-term stores through selective attention. It may be the case that stretching the span *per se* promotes such attention, and may in turn have improved readers' capacity for memorisation through mere arousal, as it were. Atkinson & Shiffrin also proposed that the probability that information in the short-term store will be transferred to long-term memory depends on the time for which an item is held in the short-term store. Although this claim has subsequently been questioned by theorists (for example Craik & Lockhart 1972), it was argued in Chapter 5 that span stretching is closely related to the notion of localised, temporary memorisation, and therefore should aid participants' ability to play the text from memory after it has been practised on SightReader. This reasoning and the opening analysis of sight reading in terms of gradually shifting categories of familiarity provided theoretical support for the exploration in Stage II of the relationship between working and long-term memory during successive encounters with a musical text.

The review of the literature on eye movement in language reading in Chapter 3 brought to light several matters that were crucial to the conceptual development and functioning of SightReader: the concept of bottom-up and top-down control of eye
movement; the technique of constraining eye position by manipulating a display, known as the gaze-contingency paradigm; the finding that the eye-voice span in language reading is larger in skilled than unskilled reading; and the phenomenon of refixation, which SightReader would act upon. The last two points formed the basis of the four hypotheses in Stage I, and are further discussed below.

The methodological implications of the review were also important for the study. Chapter 3 demonstrated that observing eye movement in music reading is considerably more difficult than in language reading, and provided the impetus for the development of a methodological framework for Stage I. It was found that the main methodological problem in observing eye movement in music reading concerns the contaminating effects of the troublesome tempo/skill/action-slip fallacy, that is, the widespread presumption that it is valid or useful to compare the eye movement patterns of the skilled and the unskilled, and to interpret temporal patterns of eye movement in readings in which tempo was not strictly controlled. Specifically, it was argued that insufficient control over tempo, skill and action slips can have potentially undesirable effects on musculoskeletal and oculomotor behaviour. This led to the coining of the terms 'peak-use area', the range of tempos in which there is minimal risk of contamination by either musculoskeletal action-slips or oculomotor wandering, 'action-slip area', in which there is an unacceptable level of action-slips, and 'undercapacity area' in which there is a probability that the data will be contaminated by undisciplined eye movement, referred to here as 'the wandering effect'. It is submitted that these tempo ranges, which played an important role in choosing tempos, participants and stimuli in Stage I, should be taken into account in the design of all experiments in which eye movement in music reading is observed.

Another theoretical innovation to arise from the review of eye movement in music reading was the idea of merging fixation and saccade durations into a single measure, pause duration. This made it possible to measure eye movement in terms of a simple relationship between the number of pauses, their mean duration, and the tempo of the performance. Accordingly, it was proposed that the number/duration relationship is one of the basic underlying oculomotor mechanisms in music reading; the relationship was therefore used as a key dependent variable in relation to Hypothesis 2 in Stage I. Unfortunately, however, pause duration does not appear to have been used as a dependent variable in any of the studies reviewed, rendering their findings on the characteristics of fixations less meaningful in most cases.
The review focused on the effect of four variables on fixation number and mean fixation duration in music reading: tempo, stimulus-complexity, reader-skill and stimulus-familiarity. Despite the methodological problems in previous studies on eye movement in music reading, they contain enough evidence to suggest that challenging reading conditions, which presumably make larger task demands of working memory, require a greater level of refreshment of its contents, which is reflected in more numerous fixations of shorter duration. The exception appears to be faster tempo; although this represents an increase in task demands, it appears that the need to progress across the score more quickly may override the tendency to fixate more numerously to service, as it were, the increased load in working memory.

As a logical extension of this and other reasoning in Chapter 3 arose the suggestion that there are three oculomotor imperatives in music reading. The first is that the eyes must maintain a pace across the page that is appropriate to the tempo, and that they do this by manipulating the number and duration of fixations. Since any change in number means a commensurate change on balance in the rate of refixation and the level of the O'Regan effect (see below), the characteristics of any scanpath are at least partly a product of this imperative. The second imperative is to provide an appropriate rate of refreshment of the information being stored and processed in working memory by manipulating the number and duration of fixations and the rate of refixation. The level of refreshment required appears to be related to tempo, stimulus-complexity, stimulus-familiarity and possibly reader-skill. Since the number of fixations is a variable here, this imperative also impacts on the features of scanpath. The third imperative is to maintain an appropriate span size for the reading conditions, large enough to perceive visual input and process it into musculoskeletal commands, but not so large as to exceed the capacity of the memory system to store and process the input. It was proposed that the music reading apparatus fashions its oculomotor commands in such a way as to address all three imperatives simultaneously. Eye movement thus embodies a fluid set of characteristics that are not only intimately engaged in engineering the right visual input to the apparatus, but in servicing the processing of that information in the memory system. The results of Stage I were consistent with these three imperatives, although it was beyond the scope of this study to offer a convincing proof of their validity.

The review also covered two behavioural aspects that were relevant to Stage II: (1) refixation, together with what was argued to be a related phenomenon, the O'Regan effect, and (2) span size. Refixation was found to be a common behaviour in the
reading of both language and music, but just why it occurs at all was found to be as yet an unsolved puzzle. Since SightReader would prevent leftward refixation within the span, it was important to know whether refixation is functional or dysfunctional behaviour in music reading, making this one of the aims of Stage I. The term ‘the O’Regan effect’ was coined in the course of this review to describe the phenomenon of skipping over some words (or musical notes) in a text, established by O’Regan (1979) in relation to language reading. The effect is presumed to occur through readers’ combining of peripheral input of certain information on the page with their knowledge of the likelihood that such information will occur in the given context. Both phenomena were the participant of predictions in Stage I, as explained in the next section.

6.1.3 Stage I

Since manipulating tempo is one of the key variables available to musicians as they learn a musical text both under the normal condition and in using SightReader, understanding the effect of tempo on eye movement and the memory system was of considerable importance for the conduct of Stage II and the interpretation of its data. Chapter 3 indicated that no reliable data existed on how tempo change would impact upon the two key behavioural aspects of music reading over which SightReader would exert control, span size and leftward refixation. Accordingly, gathering such data was the primary objective of Stage I. Such an investigation required the development in part of what might be termed the mathematics of music reading, involving several equations of temporal and spatial variables that govern span size and the characteristics of fixations.

It was argued that maintaining an eye-hand span may involve separate capacities in working memory, one for storing and processing a large amount of information in the span at any one time (load), and the other for storing and processing information in the span for a long duration (latency). This was of immediate interest to the development of a strategy for manipulating SightReader’s tempo and span settings in Stage II, since it would be easy to inadvertently favour either load or latency in the strategy, which it was feared may have contaminated the data. Would a load/latency bias have an adverse affect on some or all participants, and in the event that readers tend to favour load or latency when adapting to a new tempo, should such a bias be built in to the strategy for manipulating the tempo and span settings during Stage II?
Hypothesis 1 of Stage I concerned the effect of a doubling of tempo on load and latency. It was explained that the load/latency relationship of each participant would fall on a load/latency line, and it was predicted that the mean load/latency relationship would lie close to the equal-contribution point, the point at which there is no bias towards either load or latency. This prediction was made on the basis that equal contribution by load and latency to the task of adapting to a new tempo would permit the music reading apparatus the greatest flexibility to adapt to further changes in the reading conditions. Hypothesis 1 was supported by the data: when tempo was doubled, participants’ span sizes adjusted so that on average just over a third more chords were held within the span (1.802 vs 1.327 chords), and each chord remained in the span for almost exactly two-thirds of the duration (901ms vs 1327ms). The mean load/latency relationship was therefore in the vicinity of the equal-contribution point, slightly favouring a reliance on reducing latency over increasing load to adapt to the faster throughput of chords. Despite the confirmation of Hypothesis 1, three of the nine participants recorded significantly one-sided load/latency relationships. It was decided, however, that since six of the nine participants did adapt by changing both components substantially, there would be no particular concern in Stage II in using a uniform strategy for coordinating the span and tempo settings for the experimental participants, and that interference between the settings was unlikely to be encountered.

The remaining three hypotheses in Stage I concerned the question of refixation. If leftward refixation had been found to be dysfunctional, experimental participants in Stage II would in effect have been prevented from carrying out dysfunctional behaviour. This would be expected to have a beneficial effect. On the other hand, if leftward refixation were shown in Stage I to be functional behaviour, it might have explained why participants experienced serious difficulties in using SightReader, if this turned out to be the case. It was first argued that the rate of refixation, regardless of whether it is leftward or vertical, is closely connected with another two variables in the reading process, pause number and mean pause duration. Like load and latency, it was pointed out that these variables also form a relationship when tempo is changed, and that all such relationships will fall on a number/duration curve. It was also pointed out that a number/duration relationship depends on the change in the levels of refixation and the O’Regan effect; that the O’Regan effect acts on pause number and by doing so reduces what would otherwise be the required variation in mean pause duration when tempo is increased. Conversely, refixation performs this role as tempo slows. In other words, both behaviours, by inducing a change in the number of pauses, could be seen as acting as buffers to allow greater stability of
pause duration in the process of adapting to a change in tempo. Their other effect, it was proposed, may be to regulate the progress of the eyes across the score in relation to the demands of tempo; in relation to the second oculomotor imperative, the O’Regan effect speeds progress across the page, while refixation slows it. Like Hypothesis 1, Hypothesis 2 predicted that on average each variable would lie close to the equal-contribution point, thus contributing roughly equally to the forced change in tempo. The reason for this prediction was the same as for Hypothesis 1, that equal contribution will allow the apparatus the greatest flexibility to adapt to further changes in reading conditions. This prediction, however, was different from Hypothesis 1 in that here a relationship located anywhere but on the sole-contribution point for duration would sacrifice a standardised approach to scanpath design (by requiring a change in pause number as well as duration to adapt to tempo change). In effect, then, Hypothesis 2 predicted that it is preferable to change the scanpath rather than to rely on change in duration alone to adapt to a new tempo, and that both refixation and the O’Regan effect are functional behaviour in music reading.

Hypothesis 2 was strongly confirmed by the data: when tempo doubled, mean pause duration fell from 368ms to 265ms, while the mean number of pauses per chord fell from 2.75 to 2.01; in other words, readers responded by increasing their pause rate by roughly 30%, close to the equal-contribution point. Stability of scanpath was thus sacrificed for the sake of maintaining a comparatively stable mean pause duration. Like the load/latency relationships measured in Stage I, the number/duration relationships showed a diversity among the participants even though the mean relationship, and five to six of the nine relationships, were close to the equal-contribution point. By contrast, two participants’ relationships were extremely one-sided, lying close to a sole-contribution point.

Hypothesis 3 also drew on the first oculomotor imperative, to make sufficient progress across the page; it predicted that at both tempos the rate of leftward refixation would be significantly lower than that of vertical refixation, since on logical grounds leftward refixation is a considerably more costly behaviour than vertical refixation. Importantly for Stage II, if the level of leftward refixation was shown to be minute, there would be minimal chance that it would be a contaminating factor in using SightReader. Hypothesis 3 was confirmed by the data: at the slow tempo a mean of 23.00% of saccades were involved in vertical refixation compared to a mean of 5.10% in leftward refixation. At the fast tempo the rates were 8.35% versus 2.30%, both significant differences. Leftward refixation was thus shown to occur
rarely enough to present little problem in the use of *SightReader*. Individual participants' vertical and leftward refixation were negatively correlated at slow tempo ($r = -0.8261$); that is, participants who showed high rates of vertical refixation tended to show low rates of leftward refixation, and vice versa. At fast tempo there was no discernible correlation between vertical and leftward refixation ($r = -0.0838$). It is not clear why this pattern was evident only at slow tempo. Participants appeared to treat both forms of refixation together as a device for controlling progress across the page and for refreshing the contents of working memory.

Hypothesis 3 was partially confirmed. The differences were significant only for vertical refixation, with only one participant of the nine going against the trend. But while there were sharp falls in leftward fixation for three participants when tempo was doubled, and mild falls for another two participants, three of the remaining four participants increased their rates against the prediction, such that the difference merely approached significance. It is not clear why leftward refixation was not uniformly influenced by a doubling of tempo; based on these data, however, it was decided that the prevention of leftward refixation would be unlikely to present a serious problem to the user of *SightReader* since it occupied only a small fraction of saccades.

Hypothesis 4 concerned both the connection between the number/duration relationship, refixation and the O'Regan effect, and the oculomotor imperatives to progress across the score at a pace that is appropriate to the tempo and to service the requirements of working memory. Hypothesis 4 thus drew on the first oculomotor imperative: it predicted that since both vertical and leftward refixation are costly strategies in terms of making progress across the page, their rates would be reduced when tempo was doubled. The precise levels of refixation and the O'Regan effect were observed by measuring the number of times chords were fixated on at each tempo. In other words, the costs of both types of refixation suggested that their rates would fall as tempo increased and with it the need to make progress across the page increased. Similarly, the O'Regan effect was predicted to become more frequent as tempo increased since, as explained above, it would have the opposite effect to refixation in terms of progress across the stimulus. The confirmation of this hypothesis suggested that refixation in music reading is a tempo-adaptive function, a form of pacing, involving refixation and the O'Regan effect. It also resonated with the second oculomotor imperative, to refresh the contents of working memory at a rate that is appropriate to the tempo.
In addition, the confirmation of Hypotheses 2, 3 and 4 challenged the notion that scanpath reflects the horizontal or vertical emphasis of the musical texture, as proposed by Sloboda (1985) and Weaver (1943), and strengthened the likelihood that the tempo/skill/action fallacy is likely to have caused considerable contamination of the data in numerous previous studies into eye movement in music reading. Confirmation of Hypothesis 3 also resonated with Goolsby's (1987) and Smith's (1988) findings of reduced levels of leftward refixation in the sight reading of melodies at increased tempos, as reviewed in Section 3.3.4.

In summary, the data for Stage I showed enough flexibility in the way participants adapted their eye-hand span to a doubling of tempo to allay fears that there would be significant interference between SightReader's tempo and span settings in Stage II. Leftward refixation was observed to occur at low or negligible rates in most participants, indicating that such behaviour was unlikely to represent a contaminating factor in Stage II.

6.1.4 Stage II

The idea of SightReader came from three sources. Music educators had pointed instinctively to the notion of forcing students' eyes ahead of their hands as a training technique. The gaze-contingency paradigm served as the technical model for SightReader's mechanism. And cognitive psychology provided the scientific background for analysing the effects that SightReader might have on participants; not only did it support the notion that good readers' spans are larger than those of poor readers, it suggested that chunking and arousal might play a role in the use of the device.

The primary objective of Stage II was to evaluate SightReader as a tool for improving keyboardists' sight reading and memorisation skills. This was done by exposing two groups of moderately skilled keyboardists, an experimental and a control group, to a series of training sessions on SightReader. The distribution of sight reading and memorisation skills, and the motivation to improve those skills, were confirmed in an Initial Questionnaire as being approximately equal among the two groups. The experimental group trained with the unique span stretching function on the device, while the control group experienced the normal condition during the trials, based on progressive increments in tempo. For both groups, the independent variable was stimulus type, involving one-, two- and four-part stimuli.
The efficacy of *SightReader* as a training tool for sight reading was measured by observing the relative improvements between the groups’ mean levels of action slips between a pretest and a posttest, both tests conducted under the normal condition to attain consistency in the measurements. Hypothesis 1 of Stage II predicted that the mean level of action slips (the dependent variable) would improve significantly more between the tests for experimental participants than for control participants. To ensure that tempo was not a confounding factor in the measurement of the dependent variable, however, it was first necessary to confirm that any proportional increase in mean tempo was not significantly different between the two groups. This was achieved using a two-tailed t-test. The results for the dependent variable showed that the more difficult the stimulus type, the greater the experimental groups’ margin of improvement over the control group, with a very marked trend for two-part readings and a significant disparity for the most difficult readings, the four-part.

Hypothesis 2 predicted that experimental participants’ ability to play the stimuli from memory would improve through their exposure to span stretching significantly more than for control participants. In an unexpected result, however, the groups’ scores in tests of their memorisation ability showed no significant difference, belying the assumption that span stretching represented a form of localised, temporary memorisation training for the experimental group.

A complementary source of data in Stage II came from a Final Questionnaire, administered at the end of the posttest. This yielded numerical indications of self-perceptions by the participants on their experience of the trials. The results showed that experimental participants were significantly more likely to feel that their overall sight reading ability had improved during the trials than control participants, thus resonating with the data gathered in pretest and posttest. In the Final Questionnaire the two groups were virtually identical in their assessment of the difficulty of the memory task. This was also consistent with the failure to observe significant advantage in the training of memorisation skills with *SightReader*’s span stretching mechanism, but went no further in explaining why Hypothesis 6 was not confirmed.

The Final Questionnaire also asked participants to write their impressions of the trials. For experimental participants, the process of stretching the span size was the most frequently occurring theme, with nine favourable comments. Although no unfavourable comments were made concerning the use of span stretching in the training process, three experimental participants mentioned problems with the span stretching *mechanism* under the theme of equipment. Control participants were more
occupied by the issue of memory, which probably loomed larger in their impression since they had no opportunity to experience SightReader's span-stretching function. Six control participants made favourable comments about the memory tests. Participants appeared to be mixed in their attitudes to the use of three types of stimulus, or to have no strong feelings either way. Experimental participants appeared to find the experience slightly more stressful than did control participants.

6.1.5 Training on SightReader

It is suggested that the use of SightReader should follow four pedagogical steps on the part of the instructor: diagnosis, prescription, support, and assessment. Diagnosis might be based on three activities: observing the way in which a student learns to play a previously unfamiliar score under the normal condition, observing the same using SightReader's span stretching, and by conferring with the student about their experience of carrying out these tasks. A key problem in diagnosing a musician's suitability for training on SightReader and arriving at a skill profile so that an appropriate rehearsal method can be prescribed is the number of variables that probably impact on such ability and the difficulty of isolating and measuring them. Among these variables are a musician's ability to (1) concentrate, related to the high task demand and consequent arousal experienced by users, (2) maintain adequate latency and load, as discussed in Chapters 4 and 5, (3) chunk effectively, since this is clearly beneficial when it comes to increasing the capacity of working memory, (4) efficiently code and retrieve information from long-term memory, involved in repeated readings of the same material with incremental increases in the span setting, and (5) produce a desired musculoskeletal response. In addition, other specific skills are necessary if a musician is to handle SightReader without undue stress: (6) the ability to keep to a metronome, necessary if the span is to be stretched precisely, and (7) the ability to read without frequently glancing down at the hands, a behaviour that disrupts the span stretching mechanism and results in early fatigue.

The knowledge so gained concerning a student's strengths and weaknesses should inform the next step, of prescribing a pattern of usage on SightReader. Instructors have at their disposal several factors that can be manipulated on the basis of the diagnosis to formulate and recommend the most appropriate practice regimen to the student. These factors include (1) the duration, frequency and distribution of practice sessions, (2) the type and difficulty of the stimuli to be used, (3) the number of times each stimulus should be read in immediate succession, (4) the starting point for the settings, and (4) intermediate goals for the settings. The structure of the
experiment in Stage II of this study might serve as a reference point for prescribing such patterns. In addition, the possibility of combining regular rehearsal on SightReader with exercises that may enhance students' ability to chunk should not be overlooked (see Section 2.2.2 for a discussion of chunking and the related text, Morris 1933). Such a combination may be mutually reinforcing, given that chunking is probably both an advantage in and an outcome of span stretching.

All students will need support from time to time during their training on SightReader, since the human interaction has the power to enrich and quicken the process of skill acquisition, particularly when it centres on a mechanical function such as span stretching. Support can take the form of discussion of any problems encountered, reminding students of their goals in using SightReader, regular monitoring of their motivation, rehearsal patterns, and the levels of settings and stimuli used, and observation of students' rehearsal together with informal feedback. Support can also take the form of assessment, in which students are able to review their progress in the light of more formal feedback.

6.2 Further research and development

This investigation into eye movement, memory and tempo in the sight reading of keyboard music has contributed to knowledge about musical skills and pedagogy in three basic areas: the analysis of the music reading process and of the concept of sight reading; the effect of tempo on eye movement patterns in music reading, including the methodology of eye movement observation; and the development and evaluation of SightReader, and how its use might be integrated into a wider performance training curriculum. As well as being useful or interesting in their own right, these substantive findings are capable of spawning further research and development through enhancement and extension.

Section 1.2 presented an analysis of the music reading process by treating it as an organic machine with neuroanatomical and informatic components. Although this analysis merely scratched the surface of the process from a neurophysiological viewpoint, it proved to be most useful to the study in clearly laying out the parts of the process that have traditionally been targeted in teaching and learning, and in pointing to the parts that have been comparatively neglected. It is contended that such an analysis could act as a valuable foundation to much research into performance training, whether of musical performance or of other, non-musical tasks.
that demand a complex musculoskeletal response. Similarly, the vagueness of the received definition of sight reading and the virtual absence in the literature of arguments for the value of explicitly targeting it in the training process made timely a thorough analysis of exactly what sight reading is. The notion of sight reading in terms of layers of familiarity should prove to be useful in the selection of stimuli for experiments on sight reading, and of material for sight reading manuals and assessment tasks.

The second area in which this study has made a significant contribution is that of eye movement in music reading. There was an obvious need for a thorough critique of the methodology of previous investigations in this area, so that the serious and widespread inadequacies in this hazardous and demanding area can be avoided in future research. There is now no excuse for observing eye movement in music reading without accounting for the potentially contaminating effects of tempo, skill and action slips. Furthermore, the overriding difficulties that have arisen from inadequate observational hardware and software should deter any researchers who may subsequently wish to embark on experimental work in this area without careful consideration of the capabilities of the available equipment.

There is no doubt that as a result of the theoretical background expounded in Chapter 3 and of the experiment in Stage I, considerably more is now known about the basic mathematics of music reading and the effects of tempo on eye movement patterns and the eye-hand span, load and latency, number and duration as related variables, refixation and the O'Regan effect. Further research is needed to investigate the influence upon these variables of stimulus familiarity and stimulus type, among other matters, with more sophisticated equipment and more participants than were available to this study. The three oculomotor imperatives also need to be investigated directly, since their existence is still speculative despite their apparent resonance with the data from Stage I of this study.

The third area to which this dissertation has contributed is in the development of SightReader. In relation to further research, the most obvious development would arise from the need to improve on the quality of the prototype of SightReader. The prototype was designed and written with minimal funding, and for this reason consisted of the bare minimum functionality to conduct Stage II. While it was adequate for this task, the prototype suffered from several serious drawbacks. Opcode MAX was chosen for its precise temporal capacity from a restricted number of languages that at the time could be used in-house. One of the disadvantages of
using MAX was that the process of authoring in stimuli was extremely cumbersome and time-consuming. Another was that an elaborate ‘rhythm compensator’ function had to be written to allow the system to accommodate rhythmic variety in the stimuli without distorting the span setting. *SightReader* needs to be rewritten in a standard programming language such as C++ for both Macintosh and Window platforms, with a new MIDI-based system of authoring in stimuli. This would obviate the need for the rhythm compensator and make it possible to use stimuli of an unlimited textural and rhythmic variety on *SightReader*.

The prototype also lacked the kind of accessories that are now a feature of many effective commercial software training packages. The addition of a facility for automatic written feedback on a user’s action slips, and consequent advice as to the appropriate level of the settings for the upcoming exercise, would enhance the self-paced potential of the program. A user-database file could also be included in the package, recording a student’s history of use, including performance quality, the level of settings for each exercise completed, and an appraisal of reading standard and strategies for improvement. A commercial version of *SightReader* would be accompanied by a user’s manual with a description of the software, who will benefit from using it, information on how to use the program effectively, and advice on how to integrate such use into a wider curriculum of performance training. This last matter is now discussed in greater detail in respect of directions for further research by individual instructors.

In the light of this study, it is submitted that performance training should aim to reach a point where it is possible to measure as many of the variables at play in a student’s reading and performance to arrive at a skill-profile of the individual. This point will not be reached until the relevant technology is developed and becomes widely available and inexpensive. Even when such technology is at hand, it will be necessary to conduct research into how to use it and how to interpret the data it produces. This would probably include the identification of dysfunctional and functional traits and the development of ways of intervening in the music reading process for the purpose of teaching and learning.

The possibilities of diagnostic technology can be exemplified with eye movement tracking technology. Such technology, while continually being improved, is still at an immature stage of development, and is typically too costly for either individual educators or training institutions. It is instructive to imagine how sophisticated the equipment might become: a less intrusive method of observation than is currently
possible, with small, light-weight goggles not bigger than standard spectacles; a generously-sized monitor that displays musical text at high resolution and at any desired size; software that plays back a representation of the scanpath over the text, in real time and slow-motion as desired, leaving a trace of the saccades and fixations with information as to their durations and load and latency at the time; integration with the moving-window technique to assess an individual's peripheral acuity; and summary data on mean pause duration and saccade amplitude, span size, and rates of refixation and the O'Regan effect in relation to an established norm and the student's own records of reading similar stimuli. These functions might eventually be available at a modest cost that is within the budget of schools, colleges and studio teachers. Under these circumstances, the boundaries between teaching, diagnosis, treatment and research would be far less distinct than they currently are.

Perhaps by that stage pedagogy will have become a true science. The challenge for music educators in respect of the impending technological changes is to promote technological development, and to be open to a scientific view of teaching and learning in general, and of the music reading process in particular. It is hoped that this study will encourage musicians and music educators to meet this challenge.
Bibliography

Abrams SG & Zuber BL (1972) Some temporal characteristics of information processing during reading Reader Research Quarterly 8, 46-51


Baccino T, Pynte J & Kennedy A (1990) Spatial coding of word position in text during mouse operations (paper presented at the Fourth Conference of the European Society for Cognitive Psychology, Como, Italy)

Baddeley AD (1976) The psychology of memory New York: Basic Books

Baddeley AD (1978) The trouble with levels: a re-examination of Craik & Lockhart's framework for memory research Psychological Review 85, 139-152


Banton LJ (1995) The role of visual and auditory feedback during the sight reading of music *Psychology of Music* 23, 3-16

Bartlett FC (1932) *Remembering: a study in experimental and social psychology* Cambridge UK: Cambridge University Press


Bean K (1938) An experimental approach to the reading of music *Psychological Monographs 17*, 1-80

Berkley H (1953) Reading ahead an aid to sight-reading *Philadelphia Étude* 71 (August), 25


Boyle (1970) The effects of prescribed rhythmical movements on the ability to read at sight *Journal of Research in Music Education* 18, 307-318


Buswell GT (1920) An experimental study of eye-voice span in reading *Supplementary Educational Monographs 17*

Carpenter PA & Just MA (1977) ‘Reading comprehension as the eyes see it’ in Carpenter PA & Just MA (eds) *Cognitive processing in comprehension* Hillsdale, NJ: Erlbaum


Cattell J (1885) Über die Zeit der Erkennung und Benennung von Schriftzeichen, Bilders und Farben *Philosophische Studien 2*, 635-650

(1886) Psychometrische Untersuchungen (Part 2) *Philosophische Studien 3*, 452-492


Chase WG & Simon HA (1973) Perception in chess *Cognitive Psychology 4*, 5-81

Christie J & Just MA (1975) Remembering the location and content of sentences in a prose passage *Journal of Educational Psychology 68*, 702-710


Cowan N (1988) Evolving conceptions of memory storage, selective attention, and their mutual constraints within the human information-processing system *Psychological Review 104*, 163-191


Craik FIM & Lockhart RS (1972) Levels of processing: a framework for memory research *Journal of Verbal Learning and Verbal Behavior 11*, 671-684


Delabarre EB (1898) A method of recording eye-movements *Psychological Review 8*, 572-574


Bibliography

Deutsch D (1972) Effect of repetition of standard and comparison tones by recognizing memory for pitch *Journal of Experimental Psychology* 93, 156-162
(1975a) Facilitation by repetition in recognition memory for tonal pitch *Memory and Cognition* 3, 263-266

Dodge R & Cline TS (1901) The angle velocity of eye-movements *Psychological Review* 8, 145-157


Elliot CA (1982) The relationships among instrumental sight-reading ability and seven selected predictor variables *Journal of Research in Music Education* 30(1), 5-14

Erdmann B & Dodge R (1898) *Psychologische Untersuchung über das Lesen auf experimenteller Grundlage* Niemeyer: Halle

Ericsson KA & Chase WG (1982) Exceptional memory *American Scientist* 70, 607-615


Fincher BJ (1983) The effects of playing the melody by rote during the pre-study procedure upon sight-reading skill development of beginning class piano students (doctoral dissertation, University of Oklahoma, AAC8403971)


216
Bibliography


Goolsby TW (1987) *The parameters of eye movement in vocal music reading* (doctoral dissertation, University of Illinois at Urbana-Champaign AAC8721641)

(1989) Computer applications to eye movement research in music reading *Psychomusicology* 8(2), 111-126

(1994a) Eye movement in music reading: effects of reading ability, notational complexity, and encounters *Music Perception* 12(1), 77-96


Handel S (1986) Tempo in rhythm: comments on Sidwell *Psychomusicology* 6, 19-23

Harris BP (1971) *Comparisons of attained ratings to instructional behaviours and techniques exhibited by band directors in sight-reading performance situations*, (doctoral dissertation, Florida State University, AAC9202299)


217
Bibliography

Helmholtz H (1866) *Handbuch der physiologischen Optik* Voss: Hamburg


Huey EB (1908) *The psychology and pedagogy of reading* New York: Macmillan


Jacobs J (1887) Experiments in ‘prehension’ *Mind* 12, 75-79

Jacobsen OI (1941) An analytical study of eye movements in reading vocal and instrumental music *Journal of Musicology* 3, 3-32, 69-100, 133-164, 197-226

James W (1905) *Principles of psychology* (Vol 1) London: Methuen


Javal E (1879) Essai sur la physiologie de la lecture *Annales d’Oculistique* 80, 61-73, 157-162, 72-81, 159-170, 242-253

Johnson GJB (1987) *A descriptive study of the pitch-reading methods and the amount of time utilized to teach sight-singing by high school choral teachers in the north central region of the American Choral Directors Association* (unpublished masters dissertation, University of Nebraska)


Jones MD (1957) Sight-reading all-important *Philadelphia Étude* 75 (March), 17 and 56

Judd CH (1905) Movement and consciousness *Psychological Review Monograph Supplements* 29

Judd CH (1907) Photographic records of convergence and divergence *Psychological Review Monograph Supplements* 34

Judd CH & Buswell GT (1922) Silent reading: a study of the various types *Supplementary Educational Monographs*, 23


218
Bibliography


Locke J (1690) *An essay concerning humane understanding* London: Thomas Bassett

Logan GD (1988) Toward an instance theory of automatization *Psychological Review* 95, 492-527


Lucas KV (1994) Contextual condition and sightsinging achievement of middle school choral students *Journal of Research in Music Education* 42/3, 203-216

Ludeker CT (1984) A simulated controlled reader utilization for increased sight reading accuracy at the piano (EDD dissertation, West Virginia University, AAC8417538)

Mann RG (1991) The use of Kodaly instruction to develop the sight-reading skills of undergraduate flute students (DMA dissertation, Arizona State University, AAC9124822)

Martin RC, Wogater MS & Forlano JG (1988) Reading comprehension in the presence of unattended speech effect and music *Journal of Memory and Language*, 382-398

Matin E (1974) Saccadic suppression: a review and analysis *Psychological Bulletin* 81, 899-917


McConkie GW & Rayner K (1975) The span of the effective stimulus during a fixation in reading *Perception & Psychophysics* 17, 578-586

(1976) Asymmetry of the perceptual span in reading *Bulletin of the Psychonomic Society* 8, 365-368
McConkie GW & Underwood NR (1981) Some temporal characteristics of perception during a fixation in reading (unpublished manuscript)


Miller GA (1956) The magical number seven, plus or minus two: some limits on our capacity for processing information Psychological Review 63, 81-97


Montano DR (1983) The effect of improvisation in given rhythms on rhythmic accuracy in sight reading achievement by college elementary group piano students (DMA dissertation, University of Missouri, AAC8323620)

Morris RO (1933) Figured harmony at the keyboard: part 1 Oxford: Oxford University Press


Music Educators’ National Conference (1994) National standards for arts education: what every young American should know and be able to do in the arts Reston Virginia: author


Nash GC (1953) The importance of sight-reading Philadelphia Étude 71 (May), 19 & 51

Norman DA (1968) Toward a theory of memory and attention Psychological Review 75, 522-536


Bibliography

(1984) ‘How the eye scans isolated words’ in Gale AG & Johnson F (eds) Theoretical and applied aspects of eye movement research Amsterdam: North-Holland

Ortmann O (1937) Span of vision in note reading Yearbook of the Music Educators’ National Conference, 88-93


Pollatsek A & Rayner K (1990) Eye movements, the eye-hand span, and the perceptual span in sight-reading of music Current Directions in Psychological Science, 49-53


Quantz PA (1898) An experimental approach to the eye-voice span Supplementary Psychological Monographs 17

Rayner K (1975) The perceptual span and peripheral cues in reading Cognitive Psychology 7, 65-81


(1978) Eye movements in reading and information processing Psychological Bulletin 85, 618-660


Rayner K & Bertera JH (1979) Reading without a fovea Science 206, 468-469


Rayner K & Duffy SA (1986) Lexical complexity and fixation times in reading: effects of word frequency, verb complexity, and lexical ambiguity Memory & Cognition 14, 191-201


222
Bibliography

(1997) Eye movements, the eye-hand span and the perceptual span during sight-reading of music *Current Directions in Psychological Science* 6(2), 49-53


Saariluoma P (1992) Visuospatial and articulatory interference in chess players’ information intake *Applied Cognitive Psychology* 6, 77-89


Schmidt FO (1981) *The eye movement patterns of woodwind instrument performers while sight reading* (doctoral dissertation, Ohio State University, 42, 2929-A: University Microfilms No. MBB84-01549)

Schneider W & Shiffrin RM (1977) Controlled and automatic information processing I: detection, search and attention *Psychological Review* 84, 1-66

Schott E Über die Registrierung des Nystagmus und anderer Augenbewegungen vermittels des Saitengalvanometers *Deutsches Archiv für klinische Medizin* 140, 79-90

Scripture EW (1905) *The new psychology* London: Scott

Searle JW (1985) An investigation of movement to music and rhythmic pattern reading flash-slide training in fifth- and sixth-grade music classes (doctoral dissertation, University of Michigan, AAC8520979)

Shaffer LH (1976) Intention and performance *Psychological Review* 83, 375-393

Bibliography


Shiffrin RM & Schneider W (1977) Controlled and automatic human information processing: II. Perceptual learning, automatic attending, and a general theory Journal of Experimental Psychology: Learning, Memory, and Cognition 14, 763-769


Simon (1974) How big is a chunk? Science 183, 482-488


(1977) Phrase units as determinants of visual processing in music reading British Journal of Psychology 68, 117-124

(1978) Perception of contour in music reading Perception 6, 323-331


Smith DJ (1988) An investigation of the effects of varying temporal settings on eye movements while sight reading trumpet music and while reading language aloud (doctoral dissertation, UMI No. 890066)


Sperling G (1960) The information available in brief visual presentations Psychological Monographs: General and Applied 74, 1-29

Spicer JK (1992) A study of keyboard proficiency requirements of non-keyboard majors in universities and conservatories (DMA dissertation, University of Washington, AAC9223033)

Spillman R (1990) Sight reading at the keyboard New York: Schirmer
Bibliography

Stebleton E (1987) Predictors of sight-reading achievement: a review of the literature Update: the applications of research in music education 6(1), 11-15


Streckfuss RJ (1984) The effect of a sight reading pacer machine upon the sight reading ability of college wind instrumentalists (DMA dissertation, Catholic University of America, AAC8405277)


Tinker MA (1958) Recent studies of eye movements in reading Psychological Bulletin 55, 215-231

Tolman EC (1948) Cognitive maps in rats and men Psychological Review 55, 189-208


Tucker DW (1969) Factors related to the music reading ability of senior high school students participating in choral groups doctoral dissertation, University of California, Berkeley


Van Nuys K & Weaver HA (1943) Memory span and visual processes in reading rhythms and melodies Psychological Monographs 55(1), 33-50


Wanat S (1971) Linguistic structure and visual attention in reading Newark, Delaware: International Reading Association


Waugh NC & Norman DA (1965) Primary memory Psychological Review 72, 89-104
Weaver HA (1943) A survey of visual processes in reading differently constructed musical selections Psychological Monographs 55(1), 1-30


Wetherick NE (1975) The role of semantic information in short-term memory Journal of Verbal Learning and Verbal behavior 14, 141-480

(1976) Semantic information in short-term memory: effects of presenting recall instructions after the list Bulletin of the Psychonomic Society 8, 79-81


Wiley CA (1962) An experimental study of tachistoscopic techniques in teaching rhythmic sight-reading in music (ADD dissertation, University of Colorado, AAC6302029)


Wright TO (1982) A study of the tachistoscope in teaching rhythmic sight-reading (doctoral dissertation, Louisiana State University, AAC8312110)

York R (1952) An experimental study of vocal music reading using eye movement photography and voice recording (doctoral dissertation, Syracuse University)

Young LJ (1971) A study of the eye-movements and eye-hand temporal relationships of successful and unsuccessful piano sight-readers while piano sight-reading (doctoral dissertation, Indiana University, RSD72-13141)


Zechmeister EB & McKillop J (1972) Recall of place on the page Journal of Educational Psychology 63, 446-453


AUSTRALIA

Patents Act 1990

THE UNIVERSITY OF SYDNEY

PROVISIONAL SPECIFICATION

Invention Title:

Sight reading trainer

Inventor:

Tony Souter

The invention is described in the following statement:
Technical Field

This invention concerns a Sight Reading Trainer, that is a device for use in training people to read fluently and at the same time to implement a series of instructions.

One significant application for such an invention is in training musicians to sight read unfamiliar music, and although the invention has been described almost exclusively with reference to this application, it should be appreciated that it has other applications. For instance, it might be useful for training the operators of key operated machines, such as in training copy typists or the operators of mail sorting machines.

Background Art

Musicians must read ahead of the note or chord they are playing so that there is time to turn the information picked up from the score into the corresponding body movements which will produce the musical performance. Skilled sight readers pick up information significantly further ahead of their hands than do the unskilled. In other words, the skilled hold more information in their eye-hand span as they sight read.

Summary of the Invention

As currently envisaged, the invention is a sight reading trainer for training people to read fluently and at the same time to implement a series of discrete instructions, the trainer including:

- instruction output means to output discrete instructions for visual display;
- performance input means to receive item by item, input related to the serial execution of those instructions;
- blanking means to control visual display of the discrete instructions such that, in use at any given time, an instruction that has just been executed and a predetermined extent of instructions yet to be executed are not visually displayed.
The effect of using the invention is to require the user to read ahead of the instruction currently being executed. This process has the capacity to stretch the size of the user's working memory.

The extent of instructions that are blanked out may be determined in a variety of different ways and may be selectable by the user or a supervisor. For instance, a selected number of instructions may be blanked out, or all the instruction occurring over a period of time may be blanked out.

In use the trainer will usually be connected to a visual display device on which the instructions are displayed. The trainer will also usually be connected to equipment which will provide feedback concerning the execution of the instructions.

At the start, the first instructions of the series may be displayed, then when the first instruction is executed the first instruction and a preselected range of instructions yet to be executed are immediately blanked out. Then as further instructions are executed, further instructions are blanked out, keeping constant the range of instructions yet to be executed that are blanked out.

The range of blanked out instructions need not be a whole number of instructions; it may be a fraction of an instruction, or a whole number plus a fraction of an instruction. One way of achieving this is to delay the blanking out of each instruction by an appropriate time period.

When the sight reading trainer is used to train the sight reading of music, the instructions are musical notations of notes or chords. The music is displayed on a VDU. Playing of the first chord, which may be only a single note, causes the first few notes, or chords, to be blanked out. As each subsequent chord is played, another is blanked out. The effect is that the musician is always reading a few chords ahead of the music that is being played.

The range of music that is blanked out is selectable, and the trainer can be used to stretch the musician's working memory by gradually extending the load setting. It may be too great an increment to stretch the load setting by a
whole number of beats; in order to stretch the distance by smaller amounts it is possible to delay the time at which blanking occurs by an amount less than the duration of a beat.

The musician may choose the tempo and minimum load independently, in other words the musician may control how fast to play and how many beats, or seconds, of music are blanked out ahead of each note or chord during a performance. The trainer may calculate and apply a load setting dependant on the tempo, so that the load setting is automatically adjusted as the tempo changes.

In one embodiment the trainer relates the tempo to the span by:

\[ L = T^k \]

where \( L \) is the minimum eye-hand information load, calibrated in beats; \( T \) is the tempo setting, calibrated in beats per minute; and \( k \) defines the relationship between eye-hand information load, the amount of information stored, and eye-hand latency, the amount of time it takes this information to pass.

Thus when the user changes the tempo setting and specifies \( k = 0.5 \), both load and latency will be proportional to the square-root of \( T \), and the balance between load and latency will be maintained at the new tempo setting. If the user wishes to favour one factor over the other when changing tempo, he or she may choose a value for \( k \) other than 0.5.

The trainer may both display the music and receive an electronic pulse from the musical keyboard each time the user plays a musical beat on it. In a preferred example, the display is prepared by authoring music notation into an animation file, and the electronic pulse from the keyboard is a command to change the display from one frame of the animation to the next.

Musical beats are typically inconsistent with respect to the time-values of the notes they contain. For instance, crotchets are twice as long as quavers. In a preferred example, the trainer delineates the musical beats by a process of ignoring some pulses from the keyboard or adding extra pulses. For example,
the device ignores the pulse which is generated by the second of two quavers within a beat. In the case of a minim (a two-beat note), the trainer automatically generates an extra pulse to compensate for the absence of a pulse halfway through the duration of the minim. To achieve this, the music notation may be authored into MIDI files so that the trainer can distinguish the note-values in the display.

The trainer allows the user to isolate and challenge three of the components of the sight-reading process by manipulating the tempo and load setting as follows:

- eye-hand latency (low tempo/high load);
- eye-hand information load (high tempo/high load); and
- musculo-skeletal response (high tempo/low load).

Using a minimum load setting at any tempo stops the reader from reinspecting back within the span. This isolates and challenges the reader's ability to recirculate (refresh) the information in the eye-hand span.

The relationship between the user's working memory and long-term memory can be isolated and challenged when the user repeatedly plays through the same material with ever-increasing load settings.

The trainer has the following advantages over the manual technique:

- temporal consistency and precision to within a few milliseconds;
- numerical feedback and records of usage as motivating factors;

and

- a high degree of control over the rehearsal process.

The trainer may also be used to "squeeze" the eye-hand span, by animating in reverse so that each chord appears a split second before it has to be played, thus challenging the speed of the whole reading/performing apparatus.
Appendix I: Copy of Provisional Australian Patent for *SightReader*

**Brief Description of the Drawings**

An example of the invention will now be described with reference to the accompanying drawings, in which Figure 1 is a schematic diagram of a sight reading trainer embodying the invention.

**Best Modes for Carrying out the Invention**

The sight reading trainer 1 comprises a programmed microprocessor 2, command input means 3, music pulse input means 4, video output means 5, audio output means 6 and beat output means 7. In use the trainer 1 is connected to a control input keyboard 8, a VDU 9, a musical keyboard 10, a speaker system 11 and a metronome 12.

A MIDI connection 13 operates so that every time a chord is played on the keyboard 10 the corresponding music is output by speakers 11. The music is authored into animation files 14 so that it can be displayed on the VDU 9.

To train, a musician selects a piece of music from a library of authored pieces, and then loads the corresponding MIDI and animation files into the trainer 1. The musician then operates the control keyboard 8 to select the appropriate tempo, and the extent of the music that is to be blanked out ahead of being played (the load). The span may be selected either in terms of beats (load) or seconds (latency), and in this case is two beats with one chord for each beat. Before playing commences the trainer 1 displays the first page of music to be played:

![Music notation image]

232
The metronome 12 begins, to set the tempo for the musician, and when the first chord is struck on keyboard 10 a pulse is sent from the musical keyboard 10 to the trainer 1 and the first two chords disappear:

At this point the musician's eyes must already have arrived at the third chord, or at least have left the second chord and be moving towards the third chord.

As the user strikes the second chord another pulse is sent from the keyboard 10 and one more chord is removed, to maintain the established span of two beats:

To achieve fine gradations in load setting the trainer subtracts time from the next highest whole-beat setting. A load setting of 2.25 beats causes the trainer to take a whole beat setting of 3 beats and delay the blanking of each chord by 0.75 beats. The disappearance of the third beat in front of the struck chord is always delayed by .75 beats after the pulse is sent. The duration of the calibration delays depends on the performance tempo, or the duration of
the beat. For this reason the trainer is linked to the metronome to give a delay of 750ms at 60MM (at which a beat is 1000ms).

The musical keyboard 10 generates a pulse, and the trainer receives the pulse as every chord is struck. A steady pulse is only produced by playing a rather dull succession of crotchet-chords (or worse, a melody of crotchets only). The slightest rhythmic elaboration will cause the effective minimum load to gyrate in response to the unequal pulses coming in from the keyboard. The trainer preserves the minimum load by modifying the uneven pulses which arise from rhythmic activity.

The rhythmic values of the music are read from a MIDI file, where they are stored numerically, and this information is used by the trainer to preserve the load setting by ignoring the pulse of the second of two consecutive quavers (half-beat notes), and internally generating a pulse one beat after a minim (two-beat note) has been struck.

For example, the trainer ignores the arrowed note in Beat 4 because it will cause an extra whole beat ahead to disappear when it is played:

```
BEAT NUMBERS:
1 2 3 4 5 6 7 8 9 10
```

Also, when the hands play the long two-beat chord in Beats 7 and 8, they fail to deliver a pulse in Beat 8 and the minimum span-size suddenly shrinks by one beat. In this case a pulse is automatically generated one beat
after the pulse of the minim, to make up the deficit. Other note values are accommodated in similar ways.

To maximise training value the musician must strategically manipulate the tempo and span settings. This is not a simple matter. Two variables are at play when it comes to the performance tempo and the size of the eye-hand span: the amount of information stored in the span, or the load, and the amount of time it takes this information to pass through the span, or the latency. The relationship between these variables is governed by performance tempo, such that when users change the tempo setting they also need to change the span setting to preserve the same mix of load and latency in their span.

For example, if the tempo is doubled from 60MM to 120MM and the span setting remains unchanged at two beats (two chords), the latency will automatically fall from two seconds to one second because the chords are passing through the span twice as quickly. Conversely, if the span setting is fixed at two seconds as opposed to two beats, a doubling of tempo from 60MM to 120MM will cause the load to double from two chords to four.

The trainer maintains a relationship between the load setting $L$ and the tempo setting $T$ according to the following formula:

$$L = Tk$$

where $L$ is the minimum eye-hand information load, calibrated in beats; and $T$ is the tempo setting, calibrated in beats per minute; and $k$ defines the relationship between eye-hand information load, the amount of information stored, and eye-hand latency, the amount of time it takes this information to pass. $k = 0.5$ has been found to preserve the balance between load and latency as the tempo changes.

If a user wishes to favour one factor over the other at the new tempo, he or she must choose a setting other than 0.5.

Although the invention has been described with reference to a particular example it should be appreciated that it may be exemplified in other ways. For instance, the invention is not restricted to operation with a musical
keyboard, and it is possible to envisage operation with a woodwind instrument. In this case the trainer may interpret the sound generated by the instrument to determine when each chord is played.

It will be appreciated by persons skilled in the art that numerous variations and/or modifications may be made to the invention as shown in the specific embodiments without departing from the spirit or scope of the invention as broadly described. The present embodiments are, therefore, to be considered in all respects as illustrative and not restrictive.
CLAIMS

1. A sight reading trainer for training people to read fluently and at the same time implement a series of discrete instructions, the trainer including:
   - instruction output means to output discrete instructions for visual display;
   - performance input means to receive item by item, input related to the serial execution of those instructions;
   - blanking means to control visual display of the discrete instructions such that, in use at any given time, an instruction that has just been executed and a predetermined extent of instructions yet to be executed are not visually displayed.

2. A sight reading trainer according to claim 1, wherein the extent of instructions that are blanked out is selectable.

3. A sight reading trainer according to claim 2, wherein a selected number of instructions are blanked out.

4. A sight reading trainer according to claim 2, wherein all the instructions occurring over a period of time are blanked out.

5. A sight reading trainer according to claim 1, wherein, at the start, the first instructions of the series are displayed, then when the first instruction is executed the first instruction and a preselected range of instructions yet to be executed are immediately blanked out.

6. A sight reading trainer according to claim 5, wherein as further instructions are executed, further instructions are blanked out.

7. A sight reading trainer according to claim 5 or 6, wherein the extent of blanked out instructions need not be a whole number of instructions.

8. A sight reading trainer according to claim 7, wherein the blanking out of each instruction is delayed by a selected time period.

9. A sight reading trainer according to any preceding claim, wherein the trainer is used to train sight reading of music and the instructions are musical notations of notes or chords.
10. A sight reading trainer according to claim 9, wherein playing the first chord causes at least the first chord to be blanked out, and as each subsequent chord is played, another is blanked out.

11. A sight reading trainer according to claim 10, wherein the tempo and load are selected independently.

12. A sight reading trainer according to claim 10, wherein the load setting is automatically calculated and applied dependant on the tempo.

13. A sight reading trainer according to claim 12, wherein the trainer relates the tempo to the span by:

\[ L = T^k \]

where \( L \) is the load setting and is calibrated in beats; and \( T \) is the tempo setting and is calibrated in beats per minute.

14. A sight reading trainer according to claim 13, wherein \( k = 0.5 \).

15. A sight reading trainer according to claim 10, wherein the trainer compensates for the different beat values of notes by ignoring the second of two consecutive quavers or doubling the beat value of a minim.

16. A sight reading trainer according to claim 10, wherein musical notation is authored into animation files for display.

17. A sight reading trainer according to claim 16, wherein the trainer receives a pulse as every chord is played, the trainer allocates a beat value from MIDI files in which the musical notation is authored, and then determines how much of the display to blank out.

18. A sight reading trainer substantially as herein described with reference to the accompanying drawing.

Dated this sixteenth day of September 1997

THE UNIVERSITY OF SYDNEY
Patent Attorneys for the Applicant:
F.B. RICE & CO.
Appendix II

Copies of stimuli used in Stage II

NB Copies are at approximately two-thirds of the size that was presented in trials
Appendix II: Copies of stimuli used in Stage II

One-part stimuli

Top: Pretest (1)  Middle: Trial Session 2  Bottom: Trial Session 3
Appendix II: Copies of stimuli used in Stage II

One-part stimuli

top: Trial Session 4  middle: Trial Session 5  bottom: Posttest (6)
Appendix II: Copies of stimuli used in Stage II

Two-part stimuli

top: Pretest (1)  middle: Trial Session 2  bottom: Trial Session 3
Appendix II: Copies of stimuli used in Stage II

Two-part stimuli

top: Trial Session 4  middle: Trial Session 5  bottom: Posttest (6)
Appendix II: Copies of stimuli used in Stage II

Four-part stimuli

top: Pretest (1)    middle: Trial Session 2    bottom: Trial Session 3
Appendix II: Copies of stimuli used in Stage II

*Four-part stimuli*

top: Trial Session 4
middle: Trial Session 5
bottom: Posttest (6)
Glossary of terms

The glossary consists of selected technical and semi-technical terms used in this dissertation. The terms have been highlighted in small capitals on their first appearance in the dissertation, and in some cases on subsequent appearances where it appeared that this would be helpful to the reader. The terms in the glossary are drawn from the fields of music, music education, psychology, neuroanatomy and statistics. They appear in their singular nominalised form without deictic ('the'). Where a definition includes a term that is itself defined in the glossary, that term appears in italics. Terms are marked with an asterisk if they are unique to this study or have been used in a unique way.

acciacciatura A localised melodic ornament of no defined rhythmic value that typically moves extremely swiftly a step up or down to a longer note; characteristic of many western and non-western folk styles.

action slip An imperfection in a reader's musical output, whether a 'wrong note', rhythmic error, or any other degradation in output resulting from insufficient control by the reader over the music reading process. See also action-slip area.

action-slip area* The set of tempos at which a performance is degraded to the extent that it is regarded as being unsuccessful. For the purpose of this study, an unsuccessful performance was defined as one in which the level of action-slips significantly obstructs the learning benefit of the performance. See also undercapacity area and peak-use area.

action-slip threshold* The tempo at and above which a performance is degraded to the extent that it is regarded as being unsuccessful; the boundary between peak-use and action-slip areas. See also undercapacity threshold.

activation The notion that some information stored in the memory system can reside on a closer level of accessibility to conscious thought than other information. Information that is stored and processed in working memory may be considered to be in a highly activated state.

actuality familiarity* A category of familiarity with a musical text, comprising a reader's memory of the actual notes of a text or of a portion of it, arising from previous exposure to the text and/or internal repetition of recognisable patterns such as occurs in the repeated sections of binary and ternary forms. See also probability familiarity.

arousal The level of alertness and activation; physiological excitation; general non-directional drive-state.

attention The selective concentration and focusing of mental activity.

auditory feedback* One of the seven signals proposed to function in the music reading process. The signal exists in two forms, one being the musical sound as
produced by the musical instrument, the other being the more abstract coding of that sound after it has been perceived by the ears.

**auditory system** The part of the body that participates in perceiving and transmitting auditory feedback to the memory system.

**automaticity** Automatic as opposed to controlled processing of an activity; the situation in which an activity, or a component of an activity, requires little or no conscious attention to be executed.

**bottom-up** An epithet used to describe processing that draws on the low-level features of perceived information rather than its context. Another way of describing this term is 'data-driven' rather than 'concept-driven'. Data-driven processes gather process information in small pieces, which are later assembled in working memory. See also top-down.

**central area** The small area in the middle of the visual field that delivers information of high resolution. The central area is also known as the 'fovea'.

**central executive** The component of working memory assumed to function as a supervisory attention system, coordinating the operation of the slave systems and being the locus of consciousness.

**chunking** The process of combining units of information according to a rule or correspondence to some pre-existing pattern stored in the memory system.

**command** One of the three proposed signal categories in the music reading process, comprising the two signals, musculoskeletal and oculomotor, that produce a functional physiological response in the music reading process.

**consolidation** The coalescing of the continuous stream of oculomotor and musculoskeletal commands into a functional form in the music reading process. See also decay.

**control group** A group of subjects who are exposed to different conditions in an experiment to those to which the experimental group are exposed, to serve as a base-line with which to compare the experimental group.

**counterbalanced order** An established method of rotating the order of presenting stimuli in an attempt to avoid any confounding influence on the data that might arise from differences between the stimuli.

**decay** The degeneration of a signal trace in the memory system, used here specifically to refer to such degeneration of input and feedback signals in the music reading process. See also consolidation.

**dependent variable** An aspect of the environment that is empirically investigated for the purpose of determining whether it influences behaviour. See also variable and independent variable.

**ecological validity** The degree to which experimental conditions resemble the real-life situation being studied.

**electro-oculography (EOG)** A reasonably accurate and reliable method of tracking eye movement by recording the electrical potential between the cornea and the retina, and requiring incisions to be made on the subject's face.
equal-contribution point* The point at which two variables contribute equally to an adaptation. This study involves two such points, one on the load/latency line, the other on the number/duration curve. See also sole-contribution point.

experimental group A group of subjects who are exposed to the conditions under investigation in an experiment, as opposed to the control group, who are exposed to different conditions and whose data are used as a base-line against which to compare the results of the experimental group.

external refreshment* The fortifying of information in working memory through physical refixation to prevent the decay of that information.

external visual input* The stream of visual images that are picked up from a score and from which a musician constructs a performance; one of the two sources of information for constructing a musical performance, the other being internal input.

eye-hand span (abbr. span) The distance between the eyes and the hands in performing music from score, measured in this study in terms of either load or latency. The span is typically assumed to include central input only, that is, to stretch between the note or chord currently being played and the lateral location of the current fixation. Many authors refer to the span including peripheral input as the perceptual span. See also eye-voice span.

eye-line* An imaginary vertical line on a music score indicating the hitherto most rightward fixation on the current line during a reading. The eye-line is thus the right-hand boundary of the eye-hand span. See also hand-line.

eye-voice span A term analogous to the eye-hand span, but used in relation to singing and reading language aloud.

feedback* One of the three proposed signal categories in the music reading process, comprising the four signals that transmit to the memory system perceptual and proprioceptual information concerning the music reading process from the oculomotor, musculoskeletal, tactile and auditory systems.

fixation The point between any two saccades during which the eyes are relatively stationary and during which virtually all visual input occurs.

fixed-endedness* A characteristic of each of the seven proposed signals in the music reading process such that the ambit of the signal is delineated at a clearly identifiable point on either its left or right side. The signals within the oculomotor group are fixed at one end to the eye-line, and the signals within the musculoskeletal group to the hand-line. See also open-endedness.

four-part vocal style A basic notational and textural style in western tonal music in which two parts, a soprano and an alto, are notated on a treble stave, and two parts, a tenor and a bass, are notated on a bass stave. See also pianoforte style.

fovea See central area.

gaze-contingency paradigm (Also known as the moving-window technique) The technique of electronically manipulating a stimulus display in response to the position of a subject's eyes. The technique was developed in the 1970s and was one of the contributing ideas to SightReader.

hand-line* An imaginary vertical line on a music score indicating the note or chord currently being played. The hand-line is thus the left-hand boundary of the eye-hand span. See also eye-line.
homophony One of the fundamental musical textures, comprising a succession of chords with minimal rhythmic interaction between the voices.

hypothesis A formal prediction of the data that will be produced in an experiment.

independent variable Some well defined aspect of the environment that is measured in a study. See also variable and dependent variable.

initial clearance device* A function on SightReader that erases a set number of beats from the start of the stimulus display as the first note or chord is struck, in order to establish the set minimum span size.

input* One of the three proposed signal categories in the music reading process, input comprises the perception and transmission to the memory system of visual information from the score that assists in the production of musculoskeletal commands.

interference The inhibition that some memories can impose on others; the displacement or reduction in retrievability of items in memory by more recently perceived items; the reduction in activation of some information due to an increase in activation of other information.

internal input The source of information from long-term memory, or long-term working memory, from which a musician constructs a performance; with external visual input, one of the two sources of information for constructing a musical performance.

internal refreshment The fortifying of information in working memory solely through internal imaginative effort to prevent the decay of that information.

isochronous Of equal time, applicable to the musical beat.

latency* The duration between any two events. In the case of music reading in this study, the number of milliseconds it will take before the hands play the chord the eyes are currently fixating on.

leftward refixation The category of refixation in which the eyes look back to the left to previously inspected information. See also vertical refixation.

levels of processing Craik & Lockhart's (1972) notion that the strength of memory depends on the nature of its processing.

load* The amount of information held in the eye-hand span at any one time. In this study, in which reading, processing and performance difficulty vary little from chord to chord, load was measured in terms of the number of notes or chords between the eyes and the hands. See also load/latency relationship, load/latency line.

load/latency line* The line that describes the relationship between load and latency in music reading when the tempo is altered. See also load/latency relationship.

load/latency relationship* The relationship between load and latency when tempo is altered in music reading, described by the line $y = \frac{x}{t}$, where $y$ is the proportional change in latency, $x$ the proportional change in load, and $t$ the proportional change in tempo between the two readings.

long-term memory The part of the memory system that provides more permanent storage of, and less easy access to, information than working memory.
long-term working memory (LT-WM) A set of acquired strategies that allows information that has previously been processed in ‘short-term working memory’ to be readily re-activated and retrieved from long-term memory.

memory system The sum of all memory functions, including working memory, long-term memory, short-term working memory and long-term working memory.

MIDI A device used to connect a computer system with a sound production system.

millisecond (ms) A thousandth of a second, used as a standard temporal calibration of microbehaviour such as eye movement.

minim A two-beat note in simple time (US = half-note).

MM (Maelzel metronome) A calibration of beats-per-minute on a metronome, used internationally by musicians to indicate tempo.

modulation the progression of a passage of music from one key to another. See also tonicisation.

moving-window technique See gaze-contingency paradigm.

musculoskeletal command* One of the seven signals functioning in the music reading process that directs the movement of the musculoskeletal system in such a way that produces a musical performance.

musculoskeletal feedback* One of the seven signals functioning in the music reading process that communicates to the memory system the position of the musculoskeletal system.

musculoskeletal group* One of the two fundamental structural components of the apparatus, ‘the player’ as opposed to ‘the reader’, involving the set of muscles that responds to musculoskeletal command signals from the memory system, and sensory components that generate musculoskeletal feedback.

musculoskeletal response* The physical contribution by the musculoskeletal system to the musical output.

musculoskeletal system* For keyboardists, the muscles and joints of the fingers and hands, and to a lesser extent the arms and shoulders.

music reading apparatus* The bodily components that together function to perform music from score. It comprises the eyes, some of the oculomotor muscles, parts of the skin, some subcutaneous receptors, the musculoskeletal system and parts of the memory system.

music reading process* The functioning of the music reading apparatus in the performance of music from score, including its bodily components and the signals that link them.

musical future See musical present.

musical past See musical present.

musical present* what is being played at the moment, one of the three temporal domains into which the music reading apparatus can be thought of as dividing the score; the others are the musical past (what has already been played during the current performance) and the musical future (what is about to be played).
Glossary of terms

**normal condition** The reading condition in which *tempo* is controlled but minimum span size and leftward refixation are not, as opposed to the *span-stretching condition*.

**number/duration relationship** The relationship between pause number and pause duration when *tempo* is altered in music reading, described by the curve \( nd = \frac{1}{t} \), where \( n \) is the proportional change in pause number, \( d \) the proportional change in mean pause duration, and \( t \) the proportional change in *tempo*.

**number/duration curve** The curve \( nd = \frac{1}{t} \) that describes the relationship between pause number and mean pause duration in music reading.

**oculomotor command** One of the seven signals functioning in the music reading process that directs the movement of the eyes over the score.

**oculomotor feedback** One of the seven signals functioning in the *music reading process* that communicates to the *memory system* the position of the eyes from the level of tension in various oculomotor muscles.

**oculomotor group** One of the two fundamental structural components of the apparatus; 'the reader' as opposed to 'the player', involving the eyes and some of the muscles surrounding them that respond to oculomotor command signals from the memory system, and that generates oculomotor feedback.

**oculomotor response** The physical contribution of the oculomotor muscles to the *music reading process*.

**open-endedness** A characteristic of each of the seven proposed signals in the music reading process such that one side of a signal (either towards the *musical future* or the *musical past*, depending on the signal) consolidates or decays gradually rather than being fixed to either the hand-line or the eye-line. See also fixed-endedness.

**O'Regan effect** The phenomenon whereby the eyes do not fixate on some information on the score during a reading, and apparently combine peripheral perception of the unfixated information with the player's knowledge of the likelihood of its occurrence to process the information into related musculoskeletal commands.

**p-value** A value that indicates the likelihood that the difference between the means of two samples is due to chance, a *p-value* indicates the chance that the difference between two sample means was due to extraneous factors, based on their actual difference, the dispersion of the samples they represent, and the number of values involved.

**pause** A combined variable comprising a *saccade* and the *fixation* it leads to, useful because of the simple mathematical relationship between pause duration and number. See also *number/duration relationship*.

**peak-use area** All *tempos* from the *undercapacity threshold* up to, but not including, the *action-slip threshold*, in which the *music reading apparatus* is
hypothesised to function at or close to capacity without exceeding that capacity. See also action-slip area and undercapacity area.

pedagogy A system or method of training.

perception The uptake and transformation of information from the external environment into a more abstract code.

perceptual span The area of text that is visually perceived, including both central and peripheral perception.

performance instruction* In the context of performing music from score, a unit of information on the score from which part of the performance can be constructed

peripheral area The part of the visual field that surrounds the central area and delivers visual images of low resolution to the memory system.

phonological loop One of the slave systems in working memory postulated by Baddeley et al to store and process certain types of auditory information.

pianoforte style A basic notational and textural style in western tonal music for keyboard in which three parts, soprano, alto and tenor, are notated on a treble stave and share a single stem, and one part, the bass, is notated on a bass stave. See also four-part vocal style.

polyphony One of the fundamental musical textures, comprising two or more distinct melodic lines with a sense of interaction between them such that they compete for the listener's attention. Musicians sometimes use the term polyphony synonymously with the term 'counterpoint'.

posttest A test that measures one or more variables, typically particular aspects of subjects' knowledge or skill, that is conducted after a series of trials; when the results of a posttest are compared with the results of a pretest conducted before the series of trials, the effect of the trials on those variables can be ascertained

pretest A test that measures one or more variables, typically subjects' knowledge or skill, that is conducted before a series of trials; when the results of a pretest are compared with those of a posttest conducted after the series of trials, the effect of the trials on those variables can be ascertained

probability familiarity* A category of familiarity with a musical text that occurs when readers' experience of previous information in the text makes them more sensitive to the likely patterns they will encounter, as opposed to the actual patterns. See also actuality familiarity.

proprioception The uptake and transformation of information from within the body, rather than the external world, into a more abstract code.

refixation A fixation on information that has already been fixated on during the current reading. In music reading there are two forms of refixation: up and down within a chord (vertical refixation), and leftward refixation to a previous chord, as well as the musical equivalent of same-word rightward refixation.

refreshment The fortifying of information in working memory to prevent the decay of that information, either through physical refixation (external refreshment) or imaginative effort (internal refreshment).

rhythm compensator* A device in SightReader for preventing rhythmic activity that is faster than the pulse from distorting the span setting.
Glossary of terms

saccade A rapid sweep of the eyes from one fixation to the next, during which there is no significant uptake of information.

saccade amplitude A measure of the size of a saccade, calibrated in degrees of rotation of the eyeballs. Its equation to distance across the score is mediated by the distance of the score from the eyes.

same-word rightward refixation* A category of refixation whereby information is inspected twice by virtue of overlapping central areas of more two or more fixations during a scanpath.

scanpath The route taken by the eyes over a page.

semibreve A four-beat note in simple time (US = whole note).

sensory buffer The first part of the memory system that information enters, of large capacity but able to hold information for only a fraction of a second.

serialism A style of musical composition that arose in Europe, in particular in Vienna, during the first half of the twentieth century, as a comparatively dissonant alternative to the system of major and minor scales and keys that had predominated for approximately four centuries, referred to as 'tonality'.

short-term memory (also 'short-term store') a term used before the development of the notion of working memory to denote the information the memory system that is being actively used or consciously attended.

SightReader* The computer system invented and developed for this project, to observe and train working memory in the sight reading of keyboard music.

significance (statistical) A quality that arises from samples when there is sufficient difference between them in terms of their means and dispersion to indicate that there is less than a predetermined chance (for example 0.05) that they are drawn from the same population. Significance in this context means 'reliably distinct' rather than 'interesting' or 'important'.

sight reading Traditionally, the performance from score for the first time of a particular musical text. In this study, the term is defined specifically for each reading situation, taking into account the layers of familiarity that mediate the notion of sight reading. See actuality familiarity and probability familiarity.

slave system A part of working memory that is assumed to be dedicated to the storing and processing of information of a particular type or mode.

smooth pursuit The only form of eye movement in which there are no saccades, a behaviour that probably originated in hunting, and is common to many predatory mammals, including humans.

sole-contribution point* The point at which one of two related variables contributes entirely to an adaptation, and the other variable not at all. This study involves four such points: two on the load/latency line (one for load and one for latency), and the other on the number/duration curve (one for pause number and one for mean pause duration). See also equal-contribution point.

span* See eye-hand span.

span size* The distance between the eye-line and the hand-line; in other words, the size of the eye-hand span. In relation to SightReader's span-stretching function, span size is used in this study as an abbreviation for 'minimum span size'.
span setting* One of the two main settings on SightReader for controlling the reading conditions. The span setting determines the minimum span size that the reader will be permitted to maintain during a reading.

span-stretching condition The reading condition in which (minimum) span size and leftward refixation are controlled in addition to tempo, as opposed to the normal condition in which tempo alone is controlled.

splay-threshold setting* A device on SightReader that determines the interval of time after a chord is struck beyond which a subsequent trigger from the musical keyboard will activate a frame change on the monitor.

standard deviation An indication of the dispersion of values around their mean. The larger the standard deviation, the more dispersed the values. Approximately 68% of all values lie within one standard deviation of the mean.

statistics (1) The methodology of collecting, processing, and interpreting quantitative data. (2) The values observed in a sample, as opposed to the true values, or ‘parameters’, of the population from which the sample is drawn.

stimulus Information to which a subject is strategically exposed in a trial, in this study consisting of musical excerpts.

subject A person whose behaviour is measured in a trial.

$t$-test A test of significance designed to indicate the likelihood that the difference between the means of two samples is due to chance. A $t$-value must be evaluated against a table of sampling distributions of $t$ to ascertain significance. A $p$-value, however, is sufficient to indicate whether the data are significant.

tachistoscope A device invented in the 19th century that displays a succession of still images on a screen.

tactile feedback* A signal in the music reading process that transmits information from the tactile system to the memory system.

tactile system* One of the six components of the music reading apparatus, comprising parts of the skin that come into physical contact with the musical instrument, and the related subcutaneous receptors.

tempo The frequency of a musical beat or pulse, normally expressed in beats per minute. See also MM.

tempo setting* One of the two main settings on SightReader for controlling the reading conditions. The tempo setting is in effect a metronome, controlling an electronic pulse to which subjects played in the Stage II trials.

tempo/skill/action-slip fallacy The notion that it is experimentally valid to make comparisons between the eye movement data of skilled and unskilled musicians.

texture The characteristics of a musical passage in terms of its layers; in other words the number, characteristics and interaction of the component individual musical lines.

time-delay calibrator* A function in SightReader that allows the span setting to be calibrated in gradations of beats (for example 2.75 beats) rather than only in whole beats. This is achieved through the timing of the erasure of notation on the display.
Glossary of terms

tonicisation The movement of a passage of music from one key to another, promptly returning to the first key, typically in the space of two or three chords. See also modulation.

top-down An epithet used to describe the processing of visually perceived information by drawing on their context rather than their low-level features. Another way of describing this term is ‘concept-driven’ rather than ‘data-driven’. Concept-driven processes are rather like expectations or plans. See also bottom-up.

triangulation The use of two or more methods of data collection, typically involving quantitative and qualitative data, to study a matter from more than one standpoint in an attempt to explain it more fully.

undercapacity area* The set of tempos at which the capacity of the music reading process is significantly under-used, which can apparently result in the wandering effect. See also action-slip area and peak-use area.

undercapacity threshold* The threshold below which tempo is slow enough such that the capacity of the music reading apparatus is significantly under-used.

variable An aspect of behaviour, or of the conditions in which it takes place, which can take on different values, thus enabling the experimenter to distinguish between one individual and another. See also dependent variable and independent variable.

visuospatial sketchpad (scratchpad) One of the slave systems in working memory postulated by Baddeley et al to store and process visual and spatial information.

voice A musical part or line, whether it is played on an instrument or sung.

voice leading The relationship between two or more musical parts in motion.

wandering effect* A behavioural phenomenon in which the eyes scan a musical score in a comparatively undisciplined manner, apparently associated with undercapacity area.

working memory The system or mechanism underlying the maintenance of task-relevant information during the performance of that task; conscious attention; memory of a limited, and temporary capacity.
Eye movement, memory and tempo in the sight reading of keyboard music