THE ROYAL SOCIETY

The effects of skill on the eye–hand span during musical sight-reading

S. Furneaux* **and M. F. Land**

Sussex Centre for Neuroscience, University of Sussex, Falmer, Brighton BN1 9QG, UK

The eye–hand span (EHS) is the separation between eye position and hand position when sight-reading music. It can be measured in two ways: in notes (the number of notes between hand and eye; the 'note index'), or in time (the length of time between fixation and performance; the 'time index'). The EHSs of amateur and professional pianists were compared while they sight-read music. The professionals showed significantly larger note indexes than the amateurs (approximately four notes, compared to two notes), and all subjects showed similar variability in the note index. Surprisingly, the different groups of pianists showed almost identical mean time indexes (*ca.* 1 s), with no significant differences between any of the skill levels. However, professionals did show significantly less variation than the amateurs. The time index was significantly affected by the performance tempo: when fast tempos were imposed on performance, all subjects showed a reduction in the time index (to *ca.* 0.7 s), and slow tempos increased the time index (to *ca.* 1.3 s). This means that the length of time that information is stored in the buffer is related to performance tempo rather than ability, but that professionals can fit more information into their buffers.

Keywords: eye–hand span; music reading; sight-reading; saccade; buffer; skill development

1. INTRODUCTION

In musical sight-reading, there is a delay between reading the notes in the score and actually playing them. This lag, the 'eye–hand span' (EHS), can be measured in two ways: either as the time delay from fixation to performance, or as the number of notes between eye position and performance. Both of these values can be positive or negative, but predominantly, as is true for the eye–voice span in reading text aloud (Levin & Addis 1979), the eyes would be expected to be ahead of the hands during performance (Weaver 1943; Sloboda 1974; Truitt *et al.* 1997).

Many things must happen during this separation. Printed material must be recognized, deciphered and processed. This information must then be stored within an internal buffer, and all material to be performed simultaneously must be similarly processed. The concurrent material must then be reassembled, probably within the buffer, before motor output can occur. In addition, the information must be stored until performance has reached the appropriate part of the sequence. Therefore, the length of time from eye fixation to hand performance is a measure of the total time involved in processing and storage, prior to output. If this span is too short, there will not be enough time to fully decipher and reassemble the required information. If this is too long, more information will have to be stored for longer, and as the buffer can only be of a limited capacity, loss or corruption of information is probable. This implies that there is an optimum duration, at least for a particular piece. The number of notes between hand position and concurrent eye position will provide a measure of how much is contained within the buffer at any one time.

There have been very few studies on the eye movements made during the reading of music, no more than a handful since the 1940s (for a review, see Goolsby 1989), and even fewer using EHS data. Weaver (1943) and Van Nuys & Weaver (1943) performed classic research on the characteristics of both eye movements and the EHS while reading dual-staved music, recording how many notes ahead the fixation position of the eye was, compared to the note being played. More recently, Sloboda (1974) ascertained the EHSs of pianists performing single-stave sight-reading, by removing the musical score at some point during performance: the EHS was calculated as the number of notes the pianist could still play. However, he did not measure eye movements, so his results cannot be said to produce a precise measure of the EHS because this method could also be said to include the 'perceptual span'. The perceptual span is defined as the effective field of view about a single fixation point and has been the subject of several text-reading studies (for a review, see Rayner 1998). A recent music-reading study by Truitt *et al.* (1997) measured both perceptual and EHSs, again on single-staved music. It would be more complicated to consider a musician's perceptual span while reading dual-staved scores, because it is possible that sometimes the span extends vertically (between the two staves) during reading, and this would be very difficult to control for experimentally.

All these experiments have shown that, for the majority of performance time, the eyes are ahead of the hands by varying amounts. Weaver (1943) found that subjects looked between one and three notes or chords ahead of performance position on dual-staved music. Truitt *et al.*

*Author for correspondence (s.furneaux@sussex.ac.uk).

Received 9 August 1999 *Accepted* 13 September 1999

(1997), using single-stave melodies, found a mean EHS of only one to two notes, due to a surprisingly large proportion of negative values. A possible explanation for this result is that the experimental music used was too simple. Sloboda (1974) found that skilled subjects looked between six and eight notes ahead but, as mentioned previously, this is likely to be an overestimate.

Surprisingly, no researcher has so far considered the time-delay aspect of the EHS. This measure is just as important as the number of separating notes in considering the reading of music. The time-delay is an indication of how long information is stored and the note separation indicates how much is in the store at any one time: both of these pieces of information are important aspects of the EHS. Equally, both measures can be expected to vary considerably, both within and between performances, so a method of producing continuous data, rather than isolated 'snapshots' during performance, is desirable.

In this study, we have produced continuous EHS results in time and notes, for pianists of several different skill levels. The development of the EHS as skill improves is very interesting. Does the EHS extend in time, capacity, both or neither, as a person becomes more skilled at sight-reading? Or is there a less obvious effect of skill on the EHS, such as a change in variation, rather than a change in mean? Previous studies have suggested that better sight-readers look at a greater number of notes ahead of performance position, although experiments have not been performed on dual-staved music since Weaver's studies in the 1940s. Since so little work has been done on how other variables, such as performance tempo, affect the EHS, we imposed tempos for the performance of each piece of music.

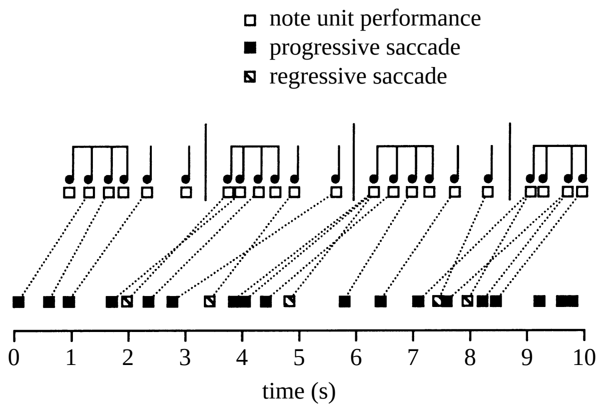

Figure 1. Ten seconds of data from a sight-reading performance by a novice pianist. The upper row of squares indicates the performance of each note unit. The rhythm of the melody from the treble (upper) stave is shown above the corresponding note unit. The lower row of squares indicates the onset of each saccade, and each are linked to the note unit it fixated. Exact time-index data are obtained by subtracting the note performance time from start and end fixation times. Note-index data are obtained by ascertaining which fixation was current at point of note performance, and adding up the number of intervening note units.

2. METHODS

The eye movements of pianists were recorded using a head-mounted video camera system (Land 1993). This non-intrusive system allows full and normal head and body movement during performance. A split-screen video recording was produced via two mirrors: a part-silvered mirror images the scene ahead, and a concave mirror images the subject's left eye (see Land (1993) for more precise details). Subjects were precisely calibrated at the beginning and end of each session, and were also checked between each performance, to ensure no slippage of the headset had occurred.

Eight adult pianists were used as subjects: three 'novices', three 'intermediates' and two professionals. The amateur categories were determined both by the grade standard (as used by the Associated Board of the Royal Schools of Music) of the subject ('novices' were approximately grade 3–4; 'intermediates' were approximately grade 6–7) and by their own estimation of their sight-reading ability. Both professionals were accompanists rather than soloists, and so considered sight-reading their speciality. All subjects performed five different pieces of music.

Subjects were required to sight-read the music as soon as it was presented, and without looking through the piece first, so that differences in the amount of prior exposure could be minimized. Each piece was assigned two tempos for the performances, determined by examining various attributes of each piece (such as time signature and note distribution) and choosing speeds that would be considered 'very slow' and 'very fast' for its performance. Each subject played the piece at both speeds, in random order. The tempo was imposed via a metronome, which was silenced after two full bars of performance.

The pieces were short extracts of musical scores that were already published under a particular grade standard. This allowed the music to be complexity matched to performers (i.e. novice sight-readers were not given very complicated music to attempt to play). The music was presented as single printed sheets of dual-staved score, with an average note-head diameter of approximately 0.4° at a playing distance of about 550 mm, although these dimensions varied to some extent because subjects were free to move their heads.

The position of each fixation on the score was measured from the videos after the recording session, at a time resolution of 50 Hz. This was done via a computer model that uses the shape and position of the iris to determine gaze direction, a process that is accurate to approximately 0.5°. A second generation video was produced which showed the direction of gaze superimposed on the scene ahead. From the computer program, information about fixation durations, saccadic amplitudes and durations, and the timing of these movements could be obtained.

In order to determine the EHS, onset times of each eye movement and note performance are required, and these two sets of information must be time-linked with each other. Time information about the eye movements was acquired directly from the video analysis. Information about the time-course of the auditory signal was produced by a sound-spectrum analyser, which applied a discrete Fourier transform to the input waveform. A logarithmic scale was chosen, so that each frequency band selected was equal to the fundamental frequency of a particular note. The time-window was 20 ms, making it equivalent to the video-frame rate. In this way, onset times for each note performance were obtained, and this information was linked with the eye movements by using an electronic 'clapperboard' (a device that produces an audible tone and a visual light simultaneously) at the beginning of every recording.

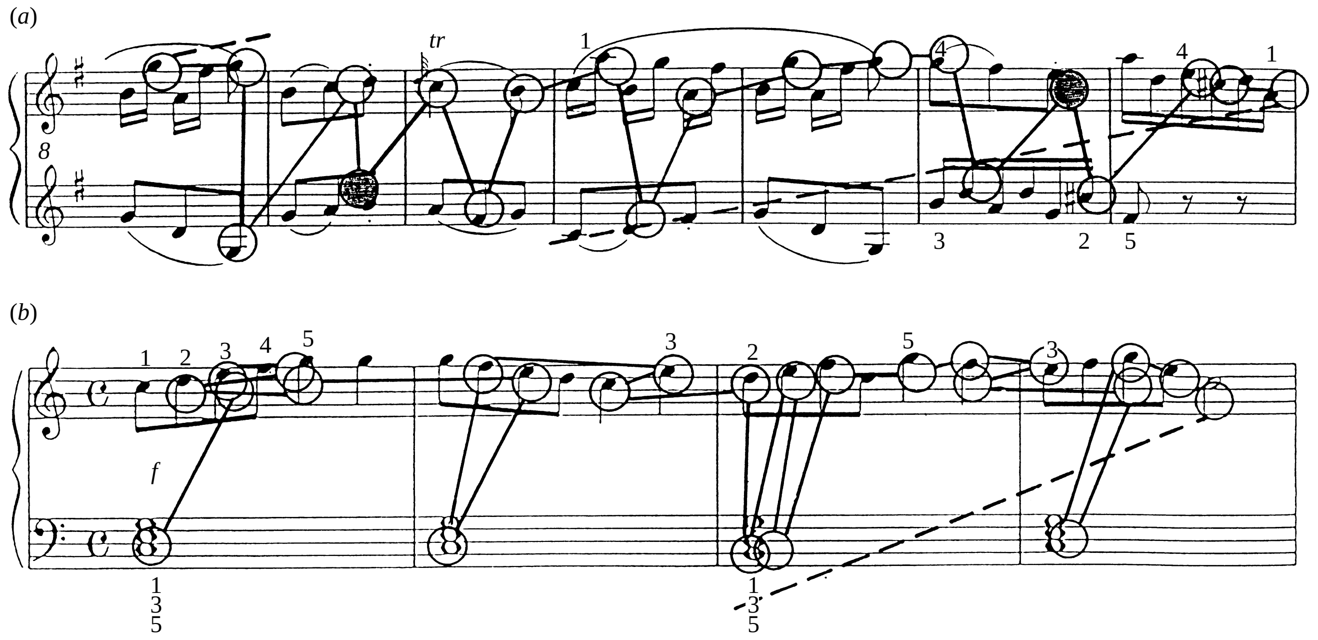

Figure 2. (*a*) One line of a piece by Scarlatti, showing the positions of fixations on the page during sight-reading by a professional subject (D.T.). Circles indicate fixation position and adjoining lines indicate saccades. Filled circles indicate that the next saccade moved the fixation point down to the keys of the piano and are shown linked to the next fixation directed towards the score. The diagonal dashed line indicates the 'new-line' eye movement that returned the gaze to the beginning of the next line. (*b*) As with (*a*), but for a novice sight-reader (P.S.) playing part of a piece by Czerny.

The EHS in time, or 'time index', is the length of time between fixating a scored note and subsequently playing that note. This means that if the time index is calculated for every eye fixation that is directed towards a portion of performable score, a complete and quasi-continuous trace of time-index data will be produced. Such a trace will exhibit a characteristic 'saw-tooth' pattern, because each eye fixation lasts a certain length of time, so the time index will reduce proportionately during that period (see figure 3). Therefore, a maximum time index was obtained at the beginning of each fixation and a minimum at the end of the fixation, immediately prior to the following saccade.

The EHS in notes, or 'note index', is calculated for each performed note or chord, and is the number of notes between the performance position and simultaneous eye position. In other words, at the moment of performing a note from the score, the note index is the number of notes that the fixation point leads the note being performed. Different authors have used different definitions, for example taking note value into account, or assigning different values to chords than to single notes, but for this study, the note index was calculated as a 'note-unit' index, where one note unit was equivalent to all simultaneously performed notes, regardless of duration and stave. This means that our note index can be more accurately described as a 'note-unit performance index'. This is similar to definitions used by other researchers (e.g. Lang 1961), and means that the note index could be calculated each time a piano key was pressed, whether played correctly or not.

Figure 1 shows a short section of a performance by a novice with both the eye and hand movements combined on the same time-scale. Each eye movement has been linked to the note unit it fixated, giving a rough idea of how the time index was calculated. The approximate note index can be acquired by observing the number of notes between performance and saccade. By calculating the time index for every eye movement and the note index for every note performance, continuous and unbroken records of the EHS measures are obtained for each of the pieces performed.

3. RESULTS

Figure 2 shows two short excerpts of music reading with the eye fixations (circles) and saccades (lines) superimposed, so the direction and sequence of each eye movement can be seen, although this gives no indication of the timing of either performance or saccades. Subjects of all skill levels make the large, vertical eye movements down to the keys of the piano (filled circles), and they do not appear to interrupt the pattern of eye fixations on the score. Both figures show a zigzag pattern of reading, with the gaze alternating between the two staves. The amount of alternation between the two staves appears to depend on the relative distribution of notes, with more fixations being directed towards the stave with the greater amount of local information.

These patterns are very similar to those found by other researchers who have recorded the eye movements of pianists as they read dual-staved music (Weaver 1943; Petzold 1995). The gaze follows this zigzag pattern between the two staves, indicating that the two lines of information are acquired separately, in serial form. They must, of course, be performed in parallel, so recombination of the note information must occur later in the system. This also means that there must be a buffer to retain whichever portion of information is acquired first, while the other stave was subsequently read and processed.

Quasi-continuous EHS data were produced for each subject for several different pieces of music. Both time and note indexes were obtained, and it appears that these two measures are affected by different variables.

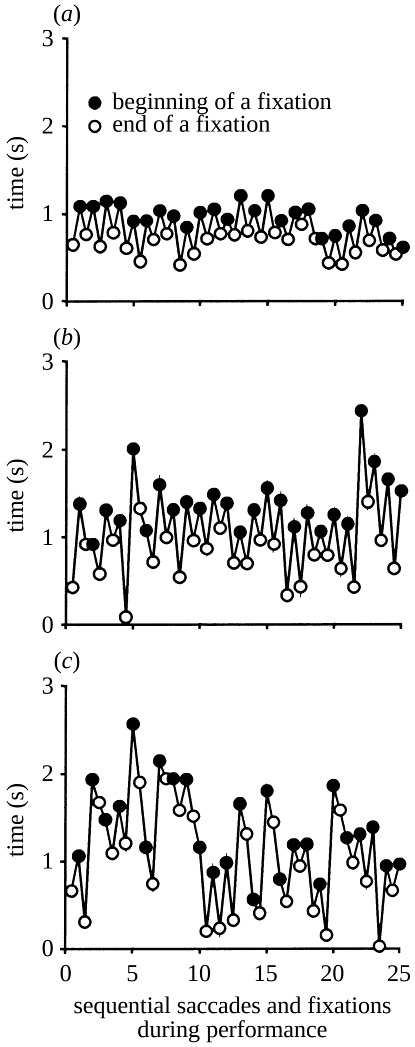

Figure 3. Three graphs showing time-index data (the EHS measured in time) during a sequence of 25 fixations and saccades during sight-reading performances by three different subjects. In each case, a fast tempo was imposed for the performance. The mean time index is approximately equal for each subject, although subjects of lesser ability show a much greater amount of variation than professional pianists. (*a*) Professional; (*b*) intermediate; (*c*) novice.

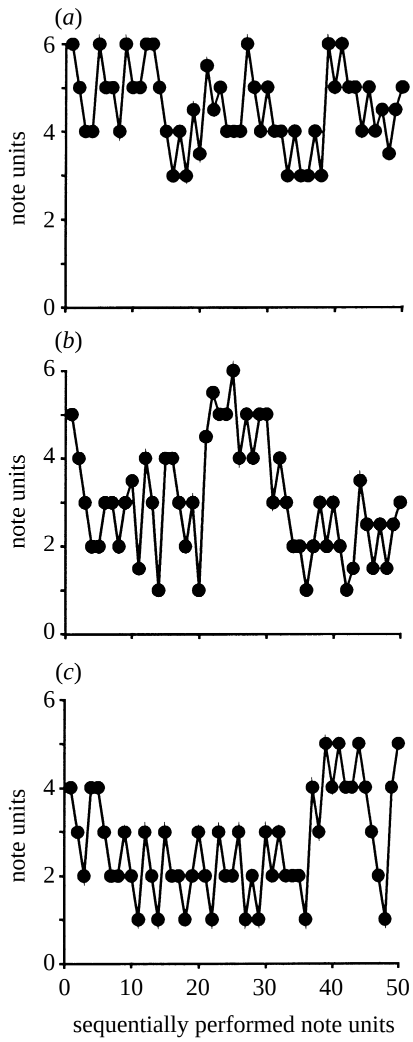

Figure 4. Three graphs showing note-index data during the performance of a sequence of 50 note units from sight-reading performances by three different subjects. The mean note index is higher for subjects of greater skill, and all subjects show a very similar amount of variation during a single performance. (*a*) Professional; (*b*) intermediate; (*c*) novice.

Time-index data were obtained for every fixation centred on performable notation. (A short example of this measure for each ability group is shown in figure 3.) Each subject, regardless of performance ability, produced a mean time index of *ca.* 1 s, and differences between the skill levels do not emerge until the amount of variation about this mean is examined. Professional pianists show significantly less variation about the mean than either the intermediates (*C*-test, d.f. > 120, $p < 0.02$) or novices (*C*-test, d.f. > 120, $p < 0.01$) (Lehner 1996). There was no significant difference between the novices and the intermediates.

Quasi-continuous note-index data were obtained for each performed note unit, including mistakes and corrections (see figure 4 for an example of data for each ability group). Here, there is a relationship between ability and mean note index (Kruskal–Wallis, $H = 133.73$, d.f. $= 2$, $p < 0.01$), with professional musicians looking significantly further ahead than intermediates (Mann–Whitney *U*-test, $p < 0.05$) or novices (Mann–Whitney *U*-test, $p < 0.05$). There was no significant difference between novices and intermediates, nor between the amount of variation shown by any group.

A comparison of the EHSs of subjects when performing at the different tempos indicates that the time index was also affected by the tempo of performance. All subjects showed a reduced time index, of *ca.* 0.7 s, when performing at a fast pace, and an increased time index, of *ca.* 1.3 s when playing at slow tempos (see figure 5*a*). There was a strong significant difference within each skill group between the time indexes obtained from the fast and slow performances (novices, fast versus slow: Mann–Whitney *U*-test, $p < 0.01$; intermediates, fast versus slow: Mann–Whitney *U*-test, $p < 0.01$; professionals, fast versus slow: Mann–Whitney *U*-test, $p < 0.01$) and this was also borne out when subjects of different skill levels are combined (all fast versus all slow: Mann–Whitney *U*-test, $p < 0.01$). The imposed tempos did not affect the note index in any of the skill categories.

The time-index results for the three ability groups and for both tempos are summarized in figure 5*a*, which

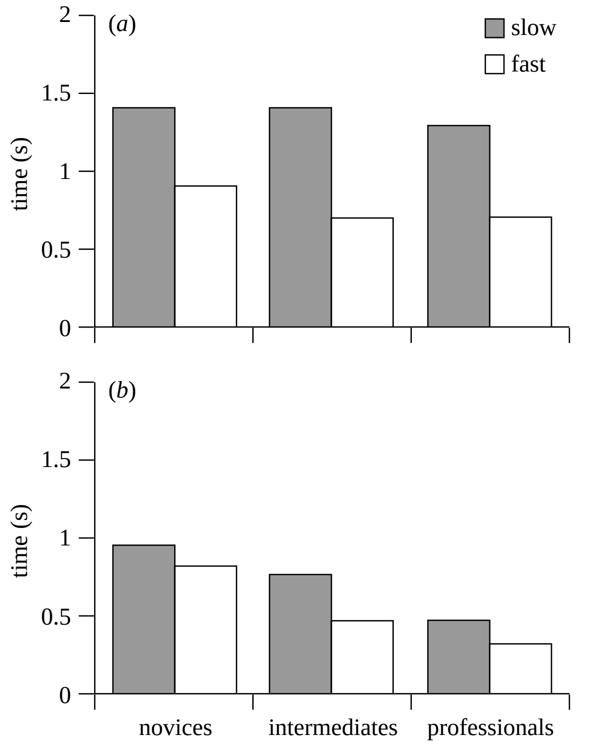

Figure 5. (*a*) Graph showing mean time-index results for each ability group. There is no difference between the different ability groups for mean time index. There is a significant difference in mean time index between the different performance tempos, with slow tempos producing larger time indexes than fast tempos. (*b*) Graph showing the standard deviation of time index for each ability group. There are significant differences in the amount of variation shown by the professional subjects and the two amateur categories, novices showing the greatest amount of variation and professionals showing the least.

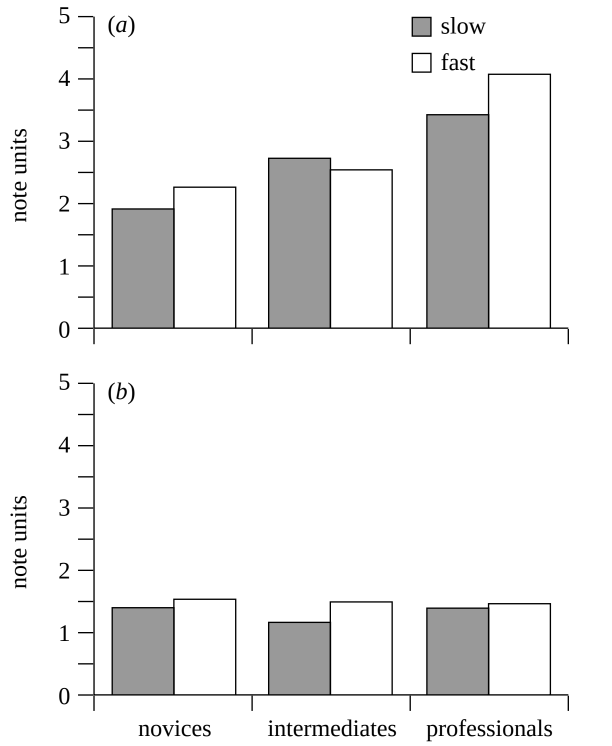

Figure 6. (a) Graph showing mean note-index results for each ability group. There is a significant difference between the mean note index shown by professional subjects and the two amateur categories, with subjects of greater skill producing larger note indexes. There was no effect of tempo. (b) Graph showing the standard deviation of note index for each ability group. No difference was found between the different ability groups, nor between the different tempos.

shows the means, and in figure 5*b*, which shows the standard deviations. Figure 6 shows a similar analysis for the note-index results.

4. DISCUSSION

We found that pianists read the two lines of information on dual-staved music separately. This confirms results from earlier studies on dual-stave music (Weaver 1943; Petzold 1995) where this mostly zigzag pattern of eye movements and fixations was also found. In order for performance of these two staves to occur simultaneously, there has to be a buffer within the system, to allow the reassembly of the two sets of information. The nature of this buffer and what happens within it has not been previously considered, although Kinsler & Carpenter (1995) suggested its existence within a five-component music-reading system. They proposed that a buffer is supplied by a processor (which is itself supplied by an encoder) and emptied by an executive as performance occurs. The processor also activates a saccadic controller, which initiates the next eye movement. The nature of the buffer itself is best investigated through EHS studies, since its capacity must be closely related to the EHS measured in notes (the note index) and the duration that the information is stored within it must be likewise related to the time measure of the EHS (the time index).

The principal findings of this study were that (i) the note index does increase with an increase in skill, (ii) the mean time index does not change with skill, and (iii) the time index reduces with an increase in performance tempo, for all skill levels. The first finding, that the note index is larger for professionals, suggests that the buffer increases in capacity as skill increases, and so more proficient musicians are able to store more information at any instant in performance. This agrees with results from studies on the eye–voice span (EVS) during reading text aloud, where more mature readers show larger spans than younger, poorer readers (Buswell 1920; Levin & Turner 1968). This does not necessarily mean that better readers (of music or text) have simply developed larger storage buffers. An alternative explanation is that rather than reading and processing individual notes, professionals are able to chunk several notes together and process them as a single unit of information, enabling them to store more in a buffer of similar capacity. This idea is reinforced in part by the professional musicians themselves, who report that it is very difficult to perform at faster tempos unless they can see ‘patterns’ in the music, and that these patterns are based on familiar musical forms like scales, arpeggios and chords.

The second finding was that the time index was not different for pianists of different abilities, with all subjects showing an overall mean time index of close to 1 s. When combined with the note-index results, this suggests that if professionals are storing more information for the same length of time, then they ought to be performing at a quicker tempo. This was basically true, although it is difficult to make comparisons between different pieces of music. Professionals did perform more notes per second than the amateurs, due to a combination of actual performance tempo, complexity of the musical score, and a lack of mistakes such as hesitations.

In addition, professionals are more consistent with their time-index results, showing very little variation about the mean, and having very similar maximum and minimum time indexes throughout a performance. Amateurs show a significantly larger amount of variation, partly due to the abnormally large time indexes produced by performance inaccuracies (note errors, repetitions and hesitations), and partly because they were much less skilled at rigidly keeping to a designated tempo, slowing for fast tempos and speeding up for slow ones.

Apart from music reading, there are many other tasks that also have need of a buffer to control the output of processed information. If a sequence needs to be preserved, then motor action must not occur too soon or too late. For example, text reading, typing, and driving all require sequential information uptake and output, and so are likely to involve a buffer. How do the buffers for these different tasks compare? Very few studies have measured the EVS in time, but Geyer (1969) found a temporal span of *ca*. 1 s for mature readers. He also found standard deviations of just over 300 ms, which is quite similar to the variation shown by the professional pianists in this study (350 ms). Hershman & Hillix (1965) and Shaffer & Hardwick (1969) looked at the EHSs of typists and also found a time-lag of *ca*. 1 s, which was about five or six letters for most typists. Also, driving experiments have revealed that drivers look *ca*. 0.8 s ahead along the road (Land 1998), regardless of the speed of travel.

The similarity of these buffer times is intriguing, and suggests either that a single buffer is shared by all these activities, or that the separate buffers have the same temporal properties.

The authors would like to thank C. Plack for his assistance in obtaining the sound spectrograms, R. H. S. Carpenter for comments on an earlier draft, and J. Sloboda for valuable discussion. Supported by a BBSRC grant to the Sussex Centre for Neuroscience (IRC) and a grant from the Gatsby Foundation.

REFERENCES

Buswell, G. T. 1920 An experimental study of the eye–voice span in reading. *Suppl. Edu. Monogr.* no. 17.

Czerny, C. 1986 The five fingers. In *Piano progress studies: book 2* (ed. F. Waterman & M. Harewood), p. 3. London: Faber Music Limited.

Geyer, J. J. 1969 Perceptual systems in reading: the prediction of a temporal eye–voice span. In *Perception and reading* (ed. H. K. Smith), pp. 44–53. Newark, DE: International Reading Association.

Goolsby, T. W. 1989 Computer applications to eye movement research in music reading. *Psychomusicology* **8**, 111–126.

Hershman, R. L. & Hillix, W. A. 1965 Data processing in typing: typing rate as a function of kind of material and amount exposed. *Hum. Factors* **7**, 483–492.

Kinsler, V. & Carpenter, R. H. S. 1995 Saccadic eye movements while reading music. *Vis. Res.* **35**, 1447–1458.

Land, M. F. 1993 Eye head co-ordination during driving. In *Proceedings of the Institute of Electrical and Electronic Engineers Systems Man and Cybernetics Conference*, Le Touquet, vol. 3, pp. 490–494.

Land, M. F. 1998 The visual control of steering. In *Vision and action* (ed. L. R. Harris & M. Jenkin), pp. 163–180. Cambridge University Press.

Lang, M. M. 1961 An investigation of eye-movements involved in the reading of music. In *Transactions of the International Ophthalmic Optical Congress*, pp. 329–354. London: Crosby, Lockwood & Son.

Lehner, P. N. 1996 *Handbook of ethological methods*, 2nd edn. Cambridge University Press.

Levin, H. & Addis, M. B. 1979 *The eye–voice span*. Cambridge, MA and London: MIT Press.

Levin, H. & Turner, E. A. 1968 Sentence structure and the eye–voice span. In *The analysis of reading skill* (ed. H. Levin, E. J. Gibson & J. J. Gibson), pp. 196–220. Final report, project no. 5-1213, Cornell University to US Office of Education. Cambridge, MA and London: MIT Press.

Petzold, L. 1995 Experimental research on the visual perception of the musical score. In *Eighth European Conference on Eye Movements*, Derby, UK.

Rayner, K. 1998 Eye movements in reading and information processing: 20 years of research. *Psychol. Bull.* **124**, 372–422.

Scarlatti, D. 1994 Sonata in G Kp. 2. In *Piano examination pieces*, grade 7 (ed. H. Ferguson). UK: The Associated Board of the Royal Schools of Music.

Shaffer, L. H. & Hardwick, J. 1969 Reading and typing. *J. Exp. Psychol.* **21**, 381–383.

Sloboda, J. A. 1974 The eye–hand span: an approach to the study of sight-reading. *Psychol. Music* **2**, 4–10.

Truitt, F. E., Clifton, C., Pollatsek, A. & Rayner, K. 1997 The perceptual span and the eye–hand span in sight reading music. *Vis. Cogn* **4**, 143–161.

Van Nuys, K. & Weaver, H. E. 1943 Studies of ocular behaviour in music reading. II. Memory span and visual pauses in reading rhythms and melodies. *Psychol. Monogr.* **55**, 1–30.

Weaver, H. E. 1943 Studies of ocular behaviour in music reading. I. A survey of visual processes in reading differently constructed musical selections. *Psychol. Monogr.* **55**, 1–30.