RESEARCH ARTICLE

Combining eye and hand in search is suboptimal

Hanneke Liesker · Eli Brenner · Jeroen B. J. Smeets

Abstract When performing everyday tasks, we often move our eyes and hand together: we look where we are reaching in order to better guide the hand. This coordinated pattern with the eye leading the hand is presumably optimal behaviour. But eyes and hands can move to different locations if they are involved in different tasks. To find out whether this leads to optimal performance, we studied the combination of visual and haptic search. We asked ten participants to perform a combined visual and haptic search for a target that was present in both modalities and compared their search times to those on visual only and haptic only search tasks. Without distractors, search times were faster for visual search than for haptic search. With many visual distractors, search times were longer for visual than for haptic search. For the combined search, performance was poorer than the optimal strategy whereby each modality searched a different part of the display. The results are consistent with several alternative accounts, for instance with vision and touch searching independently at the same time.

Keywords Vision · Haptic · Multisensory · Tactile · Search · Eye–hand coordination

Introduction

The question we address in this paper is whether eye and hand can work independently when searching. It is well known that eyes and hands often move in a highly coordinated manner. This happens in simple tasks such as pointing at objects (Neggers and Bekkering 2000, 2002) and drawing ellipses (Reina and Schwartz 2003), as well as in more complicated ones such as manipulating blocks (Johansson et al. 2001) or preparing sandwiches and making tea (Land and Hayhoe 2001). However, eyes and hands can also move independently and perform tasks in parallel. Boucher et al. (2007) studied participant's ability to stop eye and hand movements that had already been initiated. They found that stopping eye movements and stopping hand movements are not completely dependent but also not completely independent processes. Stritzke and Trommershäuser (2007) found that in a rapid pointing task the eye movements are not anchored to the hand movements, but are instead, like in visual search, driven by low-level visual features.

Apart from having to move independently, the eyes and hands would also have to sense independently in order to search independently. Studies on the ability to sense independently with different modalities also presented mixed results. Dalton and Spence (2007) found that irrelevant auditory stimuli interfered with nonspatial visual search (depending on the temporal alignment), leading to interference when they coincided with the appearance of distractors, but to facilitation when they coincided with the appearance of targets. However, Alais et al. (2006) found that, at least in low-level tasks such as auditory pitch and visual contrast discrimination, performance on either the visual or the auditory task is not adversely affected by a concurrent task in the other modality. So when perceiving

H. Liesker · E. Brenner · J. B. J. Smeets (✉)
Research Institute MOVE,
Faculty of Human Movement Sciences, VU University,
Van der Boechorststraat 9, 1081 BT Amsterdam,
The Netherlands
e-mail: j.smeets@fbw.vu.nl

information through two modalities, the two are not always independent. How the modalities affect each other in spatial search tasks has not been investigated.

We compared how participants performed a visual and haptic combined search task with predictions on performance based on their performance in a visual only and a haptic only search task. We designed visual and haptic tasks of comparable difficulty: ones for which the search times were similar. Haptic search for *spatial* properties appears always to be serial, not only when moving the hand from one item to another (Overvliet et al. 2007a), but also even when feeling several objects at the same time (Lederman and Klatzky 1997; Overvliet et al. 2007b). Whether visual search is serial without eye movements depends on how difficult it is to distinguish the target from other (distractor) items. It is definitely serial if one ensures that each item must be fixated with the eyes to see whether it is the target. Such a scanning pattern is critical if we want to study the movement coordination between the eyes and the hand.

In the present experiment, we varied the number of distractors in the visual display (defining the conditions in our experiment) to obtain visual and haptic tasks with comparable search times. In the haptic search task, there was always only one item: the target. Since visual search is obviously faster, when there is only one item, we added distractors in the visual search task to gradually switch from conditions in which visual search is faster to ones in which haptic search is faster. In the combined search task, the visual and haptic stimuli were presented together. The stimuli in the combined task were the same as those used in the visual and haptic tasks, and designed in such a way that the target was at the same position for both modalities.

Performance in the combined search task is unlikely to be worse than for both modalities separately, because participants could only rely on one modality (for instance by not moving their hand or closing their eyes), and if they do consider the other modality, it will always provide consistent information, so doing so will not interfere with the performance based on the original modality. On the other hand, the fact that they can use both their eyes and their hand to find the target might be advantageous: the search times for the combined task may on average be shorter than the search times for the purely visual or haptic task. We will consider three simple search strategies that may speed up the search, and will discuss more complicated strategies after presenting the data.

Many studies suggest that human sensorimotor behaviour is optimal. Optimal behaviour has been reported for planning movements of the hand (Todorov 2004; Trommershäuser et al. 2005; Wolpert 2007) as well as of the eye (Najemnik and Geisler 2005; Munuera et al. 2009). Many recent reports in the sensory domain also favour

optimal combination of information (Ernst and Banks 2002; Faisal and Wolpert 2009; Muller et al. 2009). One might therefore expect that when searching with eye and hand together, the performance would be based on an optimal movement plan combined with optimal sensory processing. We will model the optimal strategy for the present task (Optimal model) as the eyes examining one part of the display and the hand examining the rest of the display. This model assumes that each effector searches a different part of space and that the division of space is made independent of any information about the stimulus. Such a division of the area between hand and eye is not optimal if items are in a limited part of the field, because both modalities could neglect areas in which we more or less instantaneously register (in the visual periphery) that there are no items. Such a strategy could yield even shorter combined search times than our Optimal model predicts.

There are numerous alternative suboptimal strategies for combining manual and visual search. For the purpose of the present paper, we will quantitatively address two of them. In a first alternative model, we assume that the eyes and hand search independently and in parallel until one of them finds the target (Parallel and Independent model). This model is similar to a race model that has been used in other studies of multisensory integration (Hecht et al. 2008). This alternative strategy is clearly suboptimal as time is wasted whenever the eyes and hand examine the same location. A second alternative strategy that can be modelled easily is that subjects concentrate on the fastest modality for each condition (Fastest Modality model).

Methods

Participants

Ten participants, seven male and three female, aged between 25 and 49 years, participated in this experiment. All participants had normal or corrected-to-normal vision. Three of them declared that they were left-handed and the other seven that they were right-handed. Two were authors (EB and JS), the others were unaware of the goals of the experiment.

Apparatus

Participants were seated on a chair in the set-up shown in Fig. 1a. The haptic stimulus (examples in right panels in Fig. 1b) consisted of an A2-sized sheet of paper which was divided into four quadrants, the borders of which were raised so that participants could feel them. This stimulus was made of swell paper (ZY-TEX2, Zychem Ltd) and always contained one item: a raised dot (diameter 0.5 cm).

Fig. 1 a The experimental set-up. The visual stimulus was projected onto a projection surface. Participants saw this stimulus via a mirror, making it appear to coincide in position and in size with the haptic stimulus. **b** The visual stimulus (*left panels*) and the matching haptic stimulus (*right panels*) for the condition with 3 (*upper panels*) and 48 (*lower panels*) items
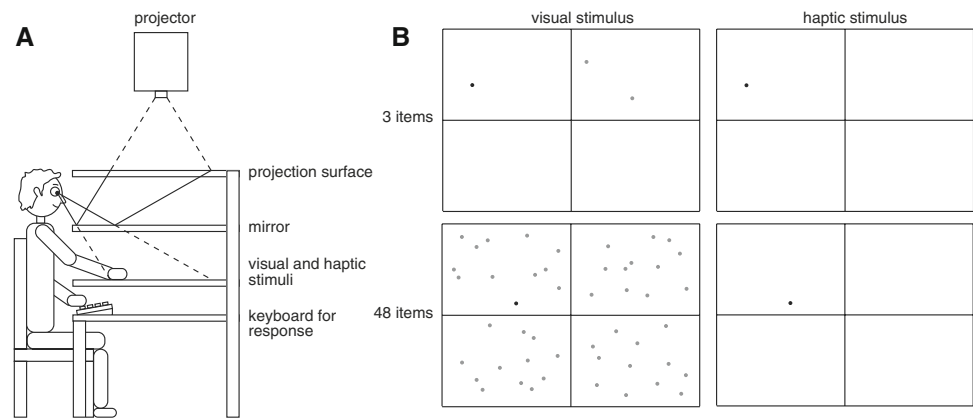
This dot was the target, and could be found by moving the (fingers of the) dominant hand across the paper. The visual stimulus (left panels in Fig. 1b) was generated by an Apple Power Mac G4 and projected by a video projector onto a back projection screen (resolution 1024 × 768 pixels for a 57.5 × 43 cm image; refresh rate 85 Hz). It consisted of a white background divided into four quadrants, separated by black lines, with 3, 6, 12, 24 or 48 items at random positions. The items were dark grey spots (5 pixels diameter), one of which contained a little black dot (1 pixel in size) at its centre. The latter was the target. The luminance of the items was such that in a pilot study the visual search time using 12 items was about the same as the haptic search time.

Participants looked downwards into a mirror where they saw the reflection of the projected image of the visual target stimulus (see Fig. 1a). The image coincided exactly in position and size with the felt surface of the haptic stimulus. Participants adjusted the height of the chair so that they could see the whole image in the mirror and move their dominant hand comfortably across the paper beneath the mirror. The distance from the eyes to the projection of the image was about 55 cm, so that 1 cm corresponds to about 1 degree of visual angle. Participants put their non-dominant hand on the keyboard, which was positioned under the surface containing the haptic stimulus. They indicated that they had found the target by pressing the keyboard's space bar.

Procedure

At the beginning of each trial, the screen was uniformly white. In the haptic and the combined search task, the experimenter put the haptic stimulus in place and then placed the index finger of the participant's dominant hand at the centre of the haptic stimulus, where the four quadrants meet. The participant then pressed the keyboard's space bar and a black fixation cross (10 pixels wide) appeared at the same intersection point (i.e. at the centre of

the image). The participant was instructed to fixate this fixation cross until it disappeared. The fixation cross disappeared after 3 s. In the haptic search task, the image was then white again. In the visual and the combined search task, the visual stimulus then appeared.

As soon as the fixation cross disappeared, the participant started searching for the target. In the haptic search task, this was done by moving (the fingers of) the dominant hand over the haptic stimulus. In the visual search task, it was done by making eye movements. In the combined search task, participants were allowed to search visually, haptically or both together, whichever method they considered to be fastest. Although we did not explicitly instruct participants to use eyes and hand at the same time in the combined search task, we observed that all participants did so. As soon as the participant found the target, he or she gave a response by pressing the keyboard's space bar. In order to ensure that participants had actually found the target, they were required to subsequently report to the experimenter verbally in which of the four quadrants the target was located.

Each of the three tasks (visual, haptic or combined) was performed in a separate session. In order to equate the difficulty across sessions, we used the same set of stimuli with the same target positions in all three sessions (but the participants did not know this). The order of the three sessions was counterbalanced across participants. Each session started with three practice trials to get participants accustomed to the task. This was followed by five blocks of ten trials, with a different random order of target locations for each block and participant. Participants could take a break between blocks. The haptic stimulus always only contained the target (no distractors). For the visual and combined tasks, each block contained trials of a single condition (3, 6, 12, 24 or 48 visible items), presented in a different random order to each participant. Therefore, for each participant, the experiment consisted of 3 sessions of 50 trials: for the visual and combined sessions, the 50 trials were divided into 5 blocks (of 10 trials) with different

numbers of items in the visual display; for the haptic session, all 50 trials were the same except for the target location.

Data analysis

The search time is the time from the moment the fixation cross disappears until the moment the space bar is pressed. Because the search times did not show a normal distribution (Fig. 2), we determined the median search time for each participant in each task and condition. The result figures show averages of these median values with standard errors calculated across participants.

Based on the individual participants' search times on each of the trials in the visual only and haptic only search tasks, three different models were built to predict the search times in the combined search task. The Fastest Modality model assumes that participants in the combined search task will rely on the modality that is fastest for the number of distractors concerned. So when there are few items the combined search task will be similar to that for visual search, but when there are many items, so that haptic search is faster, the combined search task will be as fast as haptic search. The number of items for which (according to this model) a participant would switch from visual to haptic was determined for each participant individually based on that participant's search times in the visual and haptic tasks.

For the Parallel and Independent model, we considered all possible pairs of measured search times in the haptic task (50 trials) and in the relevant condition of the visual task (10 trials), resulting in 500 pairs for each participant and condition. According to this model, participants search with their eyes and hand in parallel and independently. The
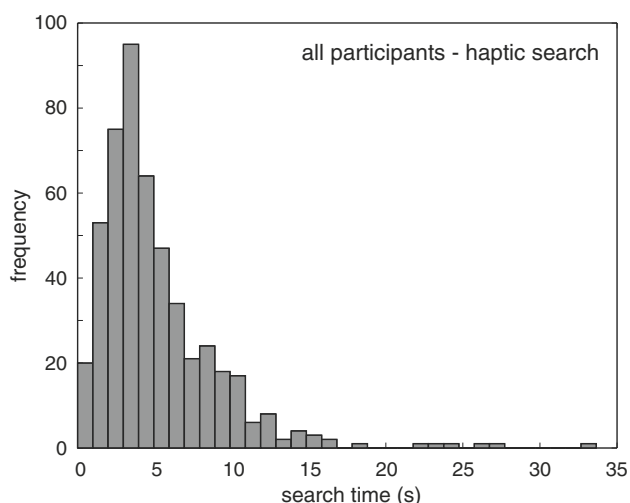
Fig. 2 Distribution of the search times for all participants in the haptic search task

predicted search time of the combined search task is therefore the shortest of each pair of trials.

According to the Optimal model, participants search one part of the display with their eyes and the other part of the display with their hands (in the combined search task). The eyes and hand never search at a location that the other effector has searched (or at least not more than they return to positions when they are the sole effector). Note that this means that if one modality searches faster than the other, it will also process a larger area. In unimodal search, the area processed by the effector will vary from trial to trial. If you are lucky, and encounter the target immediately, you only have to scan a very small area. On the other hand, if you have bad luck, you will have scanned the whole area before you encounter the target. For the trial with the median search time, about half the area will be scanned, independent of the number of distractors. In bimodal search, both modalities have the same time to search. The area A scanned in that time is proportional to the search speed. Therefore, if the search speed is the same in bimodal as in unimodal search, the area A of the workspace processed by each modality in bimodal search is inversely proportional to that effector's search time in unimodal search. The search speed (area per unit of time) with both modalities is the sum of the search speed with each modality, so for any area A (including the area required to find the target, which is on average half the workspace)

$$\frac{A}{t_{\text{bim}}} = \frac{A}{t_{\text{hapt}}} + \frac{A}{t_{\text{vis}}} \Leftrightarrow t_{\text{bim}} = \frac{t_{\text{hapt}} t_{\text{vis}}}{t_{\text{hapt}} + t_{\text{vis}}} \qquad (1)$$

Equation 1 implies that when the visual and haptic search times are equal, the search time for the combined search task is half the visual or haptic search time. We expect this to approximately be the case when there are 12 visual items because the search times were similar for 12 items in the pilot study on which we based our choice of numbers of items. If the unimodal search times differ considerably, the model predicts that the result will be close to the fastest modality, independent of whether the search is easy or difficult. All three models predict the largest advantage of using two modalities for the intermediate number of distractors. Note that this prediction is different from the "inverse effectiveness" of bimodal stimulation. This term is used to describe a reduction of bimodal advantage with the increase in performance of one of the modalities. Such a pattern of results has been found in neurophysiological measures such as the firing rates of bimodal neurones (Meredith and Stein 1986), as well as in behavioural measures such as detection times (Hecht et al. 2008). The reason for this difference is that the increase of search time in our study is not caused by the stimulus being close to threshold, but by a longer sequence of identifications of equal difficulty.

In order to examine whether we can reject one or more of the models, we will test whether the predictions for the three models (all based on the data for the single modalities) differ systematically from the actual data from the combined search task. The difference between the models is the largest if the search times are equal for the two modalities, and negligible if the modalities differ considerably. One could argue that we, therefore, should only analyse the condition with 12 items. To increase the power of our comparisons, and considering that not all participants are expected to perform equally fast for the two modalities when there are 12 items, we will also consider the conditions with 6 and 24 items. We will compare the predictions of each model for these three conditions with the data using a paired t test for the pooled data (three set sizes and ten participants; $\alpha = 0.05$). As the three model predictions for each datapoint in the combined search task are based on exactly the same pairs of datapoints in the unimodal search, we did not introduce additional variability by performing three comparisons. Therefore, we did not correct the significance level for multiple comparisons.

Results

Participants only named a false target quadrant in 3.5% of the 1,000 unimodal trials (15 times for the visual task, 20 times for the haptic task) and only 8 times (1.6%) for the combined task. The combination of modalities thus improved the accuracy of the search (two-sample Z test, $Z = 2.08$, $p < 0.05$).

Figure 3 shows the search times for the visual search task and combined search task for each number of items in the visual display, together with the average search time for the haptic search task. The visual search time depends more or less linearly on the number of items in the display, increasing with about 250 ms per item, as expected for a serial search task involving saccades. The haptic search time is plotted as a horizontal line, because there was always only one haptic item: the target. It is clear from Fig. 3 that when there are 3 visual items visual search is faster than haptic search, whereas with 48 visual items haptic search is faster than visual search. This was true for all ten participants in our experiment. For the other numbers of items, the modality that yielded the shortest search times differed between participants.

Search times for the combined search task were shorter than those for the best modality for each number of visual items. As anticipated, the advantage of using two modalities was smaller for 3 and 48 visual items than for the intermediate number of items. For a display with three items, the search time for the combined search task is about the same as the search time for the visual search task. For a
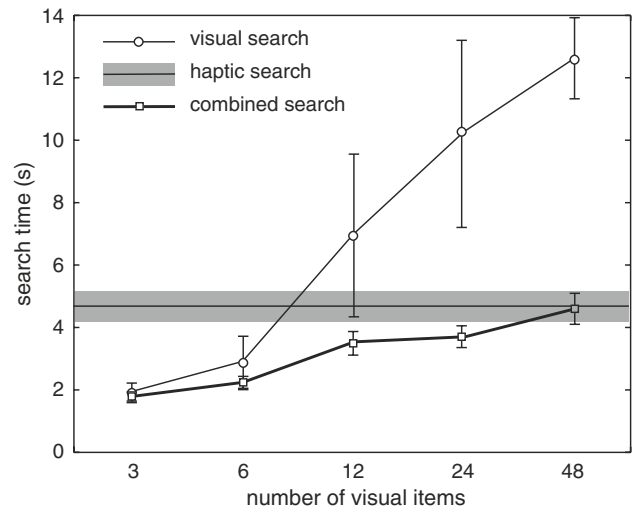
Fig. 3 Median search times for the three tasks (averaged over participants with standard errors)

display with 48 items, the search time for the combined search task is very close to the search time for the haptic search task. For displays with 6, 12 or 24 items the search times for the combined search task are clearly shorter than the search times for either the visual only or the haptic only search tasks.

Figure 4 shows the same data (grey lines) together with the predictions of our three models (black lines) for the combined search task. From a comparison of the data with the models, we conclude that, apart from the fact that for large numbers of items the measured search times are slightly longer than predicted, performance is described very well by the Parallel and Independent model. Based on the paired t test for the conditions with 6, 12 and 24 items, we can reject both the Fastest Modality model ($p = 0.026$) and the Optimal model ($p < 0.0001$). The Parallel and Independent model could not be rejected ($p = 0.93$).

Discussion

We showed that when using both eye and hand, search performance improved compared to searching with one modality: fewer errors and faster search times. The fact that the faster search times are accompanied by a reduction of the number of errors indicates that the reduction in search time is not caused by trading accuracy for speed. From the fact that performance in the combined search task is better than it would have been if participants had relied on the fastest modality, we can conclude that people are able to use both modalities at the same time. It is even more evident that search times were longer in the combined search task than they would have been if participants had searched one part of the display with their eyes and the other part of

Fig. 4 The predictions of our three models for the median search times in combined search (averaged over the 10 participants with standard error). The data from Fig. 3 are shown in *grey* (*without error bars*) for comparison
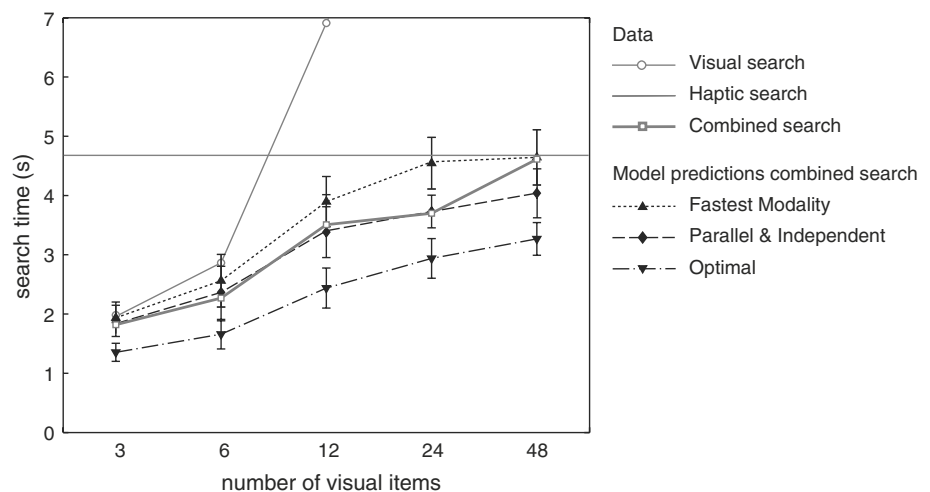
the display with their hands (the Optimal model; assuming that there is no cost in doing both simultaneously).

The search times predicted by the Parallel and Independent model were close to the search times found in the combined search task. The most straightforward explanation for this is that when searching to find a visual and haptic target we use both our eyes and our hand, moving them independently and analysing the sensory input that they provide in parallel. However, the fact that the Parallel and Independent model fits the data so well does not necessarily mean that this model adequately describes the strategy that is used. It might very well be that the participants used a coordinated movement strategy, but that this strategy does not yield the optimal performance that we predict. There may be some cost to searching with two modalities, either in terms of sensory processing or in terms of planning the movements. Moreover, participants might have to search some parts of the space with both the eyes and the hand to be sure that they have not missed any part of the space, because vision and proprioception are not perfectly calibrated (Smeets et al. 2006). They may also use a completely different strategy that leads to better performance than using only one modality, such as moving their hand to the positions at which they see potential targets. They may also increase their search times with respect to optimal performance when using two modalities by checking the target with the other modality after one modality found the target, which would account for the higher accuracy.

The failure to search optimally with two modalities simultaneously could arise because people normally move their eyes and hand together. Thus, participants may have tried to coordinate their movements optimally, but their eyes sometimes made unwanted saccades towards the hand. Alternatively, preventing such unwanted saccades may have slowed the eyes down. Fixation strategies for

visual search in a cluttered environment can be optimal (Najemnik and Geisler 2005). For fixation durations, this has even been demonstrated with stimuli that resemble ours (Over et al. 2007). However, optimality in planning movements has only been demonstrated when determining a single target location at a time (Najemnik and Geisler 2005; Trommershäuser et al. 2008). In order to perform optimally in a combined search task participants have to simultaneously process information about target presence at different locations, and then to pick new locations for both the eyes and the hand, and plan the movements to those locations. Although it is known that the eye can go to a different target than the hand, it has been argued that this is based on low-level features (Stritzke and Trommershäuser 2007). Any cost in planning independent movements for the hand and eyes, or any influence of low-level guidance, would result in performance being suboptimal.

Another possible reason for combined search being suboptimal is that the rate at which information is processed within each modality might be lower when searching with both modalities than when using only one modality. This seems in conflict with many recent experimental results that suggest that multisensory information is combined in a statistically optimal way, but the sensory information in such studies is typically about a single object or body part (van Beers et al. 1999; Ernst and Banks 2002; Niemeier et al. 2003; Alais and Burr 2004). If the information comes from different locations, cue combination is suboptimal (Gepshtein et al. 2005), probably due to violation of the unity assumption (Welch 1986), but it is also possible that spatial proximity is generally necessary for making full use of several streams of information simultaneously.

It may be possible to reject some of the above-mentioned proposals based on the movement patterns of the eye and hand. For instance, if we would see that the eye and

hand never search the same location, we could reject some explanations based on a sub-optimal path. However, we find it too unlikely that performance is suboptimal for only one of the above-mentioned reasons under all conditions for all subjects. Moreover, even if we would for instance find longer fixation times in combined search, we would not be able to tell whether this is because sensory processing or planning the next movement is slower. Similarly, observing overlap between where participants look and touch could indicate that an optimal movement plan is perturbed by unwanted saccades to the hand, but it may also be the consequence of independent control of the effectors or an intentional strategy to improve performance.

The present study primarily demonstrates that performance is suboptimal. It cannot reject the independent model, but also does not provide firm support for it considering all the above-mentioned possible reasons for performance being suboptimal. We conclude that we perform better than we would if we only used the best modality, but worse than we would if we optimally combined search with each of the two modalities on its own.

References

Alais D, Burr D (2004) The ventriloquist effect results from near-optimal bimodal integration. Curr Biol 14:257–262

Alais D, Morrone C, Burr D (2006) Separate attentional resources for vision and audition. Proc R Soc B Biol Sci 273:1339–1345

Boucher L, Stuphorn V, Logan GD, Schall JD, Palmeri TJ (2007) Stopping eye and hand movements: are the processes independent? Percept Psychophys 69:785–801

Dalton P, Spence C (2007) Attentional capture in serial audiovisual search tasks. Percept Psychophys 69:422–438

Ernst MO, Banks MS (2002) Humans integrate visual and haptic information in a statistically optimal fashion. Nature 415:429–433

Faisal AA, Wolpert DM (2009) Near optimal combination of sensory and motor uncertainty in time during a naturalistic perception-action task. J Neurophys 101:1901–1912

Gepshtein S, Burge J, Ernst MO, Banks MS (2005) The combination of vision and touch depends on spatial proximity. J Vis 5:1013–1023

Hecht D, Reiner M, Karni A (2008) Enhancement of response times to bi- and tri-modal sensory stimuli during active movements. Exp Brain Res 185:655–665

Johansson RS, Westling G, Bäckström A, Flanagan JR (2001) Eye-hand coordination in object manipulation. J Neurosci 21:6917–6932

Land MF, Hayhoe M (2001) In what ways do eye movements contribute to everyday activities? Vision Res 41:3559–3565

Lederman SJ, Klatzky RL (1997) Relative availability of surface and object properties during early haptic processing. J Exp Psychol Hum Percept Perform 23:1680–1707

Meredith MA, Stein BE (1986) Visual, auditory, and somatosensory convergence on cells in superior colliculus results in multisensory integration. J Neurophys 56:640–662

Muller CMP, Brenner E, Smeets JBJ (2009) Testing a counter-intuitive prediction of optimal cue combination. Vision Res 49:134–139

Munuera J, Morel P, Duhamel J-R, Deneve S (2009) Optimal sensorimotor control in eye movement sequences. J Neurosci 29:3026–3035

Najemnik J, Geisler WS (2005) Optimal eye movement strategies in visual search. Nature 434:387–391

Neggers SFW, Bekkering H (2000) Ocular gaze is anchored to the target of an ongoing pointing movement. J Neurophys 83:639–651

Neggers SFW, Bekkering H (2002) Coordinated control of eye and hand movements in dynamic reaching. Hum Mov Sci 21:349–376

Niemeier M, Crawford JD, Tweed DB (2003) Optimal transsaccadic integration explains distorted spatial perception. Nature 422:76–80

Over EAB, Hooge ITC, Vlaskamp BNS, Erkelens CJ (2007) Coarse-to-fine eye movement strategy in visual search. Vision Res 47:2272–2280

Overvliet KE, Smeets JBJ, Brenner E (2007a) Haptic search with finger movements: using more fingers does not necessarily reduce search times. Exp Brain Res 182:427–434

Overvliet KE, Smeets JBJ, Brenner E (2007b) Parallel and serial search in haptics. Percept Psychophys 69:1059–1069

Reina GA, Schwartz AB (2003) Eye–hand coupling during closed-loop drawing: evidence of shared motor planning? Hum Mov Sci 22:137–152

Smeets JBJ, van den Dobbelsteen JJ, de Grave DDJ, van Beers RJ, Brenner E (2006) Sensory integration does not lead to sensory calibration. Proc Natl Acad Sci USA 103:18781–18786

Stritzke M, Trommershäuser J (2007) Eye movements during rapid pointing under risk. Vision Res 47:2000–2009

Todorov E (2004) Optimality principles in sensorimotor control. Nat Neurosci 7:907–915

Trommershäuser J, Gepshtein S, Maloney LT, Landy MS, Banks MS (2005) Optimal compensation for changes in task-relevant movement variability. J Neurosci 25:7169–7178

Trommershäuser J, Maloney LT, Landy MS (2008) Decision making, movement planning and statistical decision theory. Trends Cogn Sci 12:291–297

van Beers RJ, Sittig AC, Denier van der Gon JJ (1999) Integration of proprioceptive and visual position-information: an experimentally supported model. J Neurophys 81:1355–1364

Welch RB (1986) Adaptation of space perception. In: Boff KR, Kaufman L, Thomas JR (eds) Handbook of perception and human performance, vol 1, Sensory processes and perception. Wiley, New York, pp 24.21–24.45

Wolpert DM (2007) Probabilistic models in human sensorimotor control. Hum Mov Sci 26:511–524