The perceptual cognitive processes underpinning skilled performance in volleyball: Evidence from eye-movements and verbal reports of thinking involving an in situ representative task

José Afonso, Júlio Garganta, Allistair McRobert, Andrew M. Williams and Isabel Mesquita
University of Porto, Faculty of Sport, Portugal, Liverpool John Moores University, School of Sport and Exercise Sciences, UK

Abstract
An extensive body of research exists on perceptual-cognitive expertise. The majority of researchers have used film-based simulations to capture superior performance. We combined eye movement recording and verbal reports of thinking to explore the processes underpinning skilled performance in a complex, dynamic, and externally paced representative volleyball task involving in situ data collection. Altogether, 27 female volleyball players performed as centre backcourt defenders in simulated sessions while wearing an eye-tracking device. After each sequence, athletes were questioned concerning their perception of the situation. The visual search strategies employed by the highly-skilled players were more exploratory than those used by skilled players, involving more fixations to a greater number of locations. Highly-skilled participants spent more time fixating on functional spaces between two or more display areas, while the skilled participants fixated on the ball trajectory and specific players. Moreover, highly-skilled players generated more condition concepts with higher levels of sophistication than their skilled counterparts. Findings highlight the value of using representative task designs to capture performance in situ.

Key words: Perceptual expertise; visual search behaviors; thought processes; in situ testing.

Introduction
An extensive body of research exists on perceptual-cognitive expertise in sport (Dicks et al., 2010; Ericsson and Williams, 2007; McPherson and Kernodle, 2007), as well as in dynamic and time-constrained activities such as driving (Wilson et al., 2008), the military (Janelle and Hatfield, 2008), and medicine (Patel et al., 1990). In the context of sport, athletes are required to closely integrate perceptual, cognitive and motor skills (Gréhaigne et al., 2001). In this vein, attention should be devoted to the processes underpinning expert performance (Ericsson, 2008; McRobert et al., 2009), preferably using tasks that require athletes to combine perceptual, cognitive and motor skills in the same manner as during actual competition (Dicks et al., 2010). It is therefore essential to develop representative task designs that closely resemble the natural performance ecology in order to reveal the full nature of the expert advantage (Mann et al., 2010).

It has been well reported that experts quickly and accurately identify and recognize meaningful patterns in the scenarios (Laurent et al., 2006; Williams et al., 2011). The expert performer is better attuned to the task constraints and more accurate in anticipating the outcome of the action compared to less expert counterparts (McRobert et al., 2011). However, the nature and magnitude of these expertise-related differences vary to a lesser or greater extent depending on a number of factors such as the nature and complexity of the task (McRobert et al., 2011), and the experimental design employed (Button et al., 2011; Dicks et al., 2010). It appears that expertise is increased in complexity and specificity (Ericsson and Ward, 2007). However, as shown by the works of Dicks et al. (2010) and Mann et al. (2010), even simulated environments may not accurately grasp the processes that differentiate skilled from less-skilled individuals. An argument is that wherever possible it is important to try and capture performance in situ using the type of representative design proposed by Brunswik (1955).

Nonetheless, the majority of published work has involved the use of laboratory-based simulations. Although superior performance is typically reported when in film-based simulations are employed (e.g. Vaeyens et al., 2007), it is possible that different processes may be used when viewing film simulations compared to those employed during actual performance. In this vein, Araújo, Davids and Passos (2007) advanced the concern that such oversimplified research designs jeopardize the generalizability and of the findings to ‘real-world’ environments. There have been few empirical attempts to evaluate whether the processes that underpin performance differ when data is collected using film-based simulations compared to in situ (Dicks et al., 2009). Some preliminary evidence exists to suggest that there may be some differences in the processes governing performance across these two types of tasks (Dicks et al., 2010; Mann et al., 2007; Williams et al., 2011).

The two most common methods to scrutinize the processes underpinning perceptual-cognitive expertise are the recording of gaze behaviors and the collection of verbal reports of thinking (Williams and Ericsson, 2005). Although the recording of gaze behavior may provide an indication of how attention is allocated, the relationship between gaze and attention is nonlinear (McPherson and Vickers, 2004). The lack of linearity between point of...
gaze and focus of attention may be particularly evident in experts (Laurent et al., 2006), who are known to rely more so than less expert individuals on peripheral vision and on kinesthetic and haptic information compared to non-experts (Behrmann and Ew ell, 2003). Overall, decision-making implies the integration of sensory data from multiple sources into a composite, meaningful whole (Bogacz, 2007; Lenzen et al., 2009). In this line, the perceptual-cognitive perspective suggests that the expert performance approach should use a manifold of methods, namely combining eye-movement registration with collection of verbal reports of thinking (Ericsson and Williams, 2007; Williams et al., 2004).

According to a perceptual-cognitive account, verbal reports afford a window into the cognitive processes behind perception and action (McPherson and Thomas, 1989; Williams and Ericsson, 2005). Retrospective verbal reports have been applied to collect information concerning the practitioners’ knowledge structures and in-event thoughts in a variety of domains (McPherson and Kendrick, 2007; McRobert et al., 2009; Roc a et al., 2011). However, very few researchers have gathered eye-movement and verbal reports of thinking simultaneously during performance (for exceptions, see McRobert et al., 2011; Roca et al., 2011).

We examine the processes supporting skilled performance in a dynamic, externally paced volleyball task using a representative task design involving an in situ data collection. The task consisted of backcourt defensive scenarios involving 6 vs. 6 simulated offensive situations. The collection of eye-movement registration and verbal reports of thinking were combined in an attempt to provide a more detailed understanding of the processes behind skilled performance in volleyball.

Methods

Participants
A total of 27 female volleyball players were recruited. Using procedures similar to those applied by Baker, Côté and Abernethy (2003), a panel of five expert coaches knowledgeable of the participants stratified them into highly skilled or skilled groups based on an evaluation of their anticipation and decision-making skills. The inter-observer agreement percentages ranged from 88.9 to 96.3%. Highly skilled participants (n = 15; mean age 19.1 ± 8.3 years) presented a mean of 9.2 ± 6.5 years of playing experience as starting players at elite national level teams within their age group. The skilled participants (n = 12; mean age 17.3 ± 4.2 years) had a mean of 5.8 ± 2.3 years of experience practicing at elite national level teams in Portugal. Participants signed an informed consent form, received detailed instruction and rehearsal regarding the tasks and procedures were intended to recreate meaningful play situations. Participants would take turns in the backcourt defensive tasks. The experiment ended after six successful plays had been accomplished. The trials lasted approximately 5 seconds. The endpoint of each play occurred the moment the ball crossed the block, ensuring that initiation times and endpoints were standardized. In volleyball, the impossibility of making more than three contacts with the ball in each play means that no trial lasted more than five seconds. The visual search data were analyzed only until the ball pass to the defender’s side of the net. In total, six trials per participant were considered for analysis. A panel of three expert volleyball coaches established the content and structure of these sequences. The action took place on a standard size volleyball court. The team’s setters received detailed instruction and rehearsal regarding the scenarios that should emerge.

Participants’ eye movements were recorded using the Applied Science Laboratories (ASL) 3000 Mobi-leEye™ registration system (Bedford, MA, USA), which is a video-based, monocular corneal reflection system that records eye point-of-gaze in relation to a head-mounted color scene camera. The system measures the relative position of the pupil and corneal reflection in relation to each other using an infrared light source. These features are then used to compute point-of-gaze by superimposing a crosshair onto the scene image captured by the head.
mounted camera optics. The image is transferred to MiniDV format and later copied to the computer into an .avi file, and has a sampling frequency of 30Hz (30 frames per second). System accuracy was ±0.5º visual angle, with a precision of 0.5º in both the horizontal and vertical fields. The superimposed videos were analyzed twice in a frame-by-frame manner using Avidemux\textsuperscript{9} 2.5.4. for Mac.

Verbal reports were collected using a Sony ICD-UX70 digital audio recorder. The .mp3 files were copied to an Apple MacBook Pro (2.4GHz Intel Core 2 Duo), opened with VLC Media Player version 1.1.111 and copied to a datasheet on Microsoft Excel\textsuperscript{6} 2008 for Mac. Prior to testing, participants were provided with a detailed explanation on how to provide verbal reports of their thoughts (Botelho et al., 2011). They were expected to report cues that were relevant for their action.

Procedure
Before being tested, participants were familiarized with the experimental procedures. For each trial, participants were instructed to take up their ready defensive position and to try to defend the ball. In volleyball, the ready defensive position consists in assuming a small flexion of the ankles, knees and hips, with a slight internal rotation of the hips. The shoulders are slightly in front of the knees and arms are loose (Selinger and Ackermann-Blount, 1986). Participants were positioned in backcourt zone 6. Prior to engaging in the actual trials, the MobileEye\textsuperscript{TM} tracker was fitted to the participant’s head and checked to ensure that it was comfortable and that interference with performance would be kept to a minimum. The eye movement registration system was calibrated using 5 non-linear points in the scene image so that the recorded indication of fixation position corresponded to each participant’s point-of-gaze. An eye calibration was performed for each participant to verify point-of-gaze before the trials and regular calibration controls were conducted during testing. Namely, re-calibration was conducted whenever: a) the participant occasionally made a fall; b) the ball was defended near to the face (implying a vigorous movement of the head); c) the team performing the plays would commit to many fails, prolonging the duration of the testing; d) the participant complained about sweating too much, with drops of sweat in the forehead or eyes’ region, as such drops may impair the functioning of the infrared camera; and e) the participants made arm movements that contacted the goggles and/or the cables. Additionally, random re-calibrations were at times conducted.

Following calibration and instruction, participants stepped into the court and acted as backcourt defenders for as many trials as needed until six trials had been successfully ran. Trials where the team building the attack would fail to make a play or where there was a missed serve were not considered. Interviews were conducted after each trial, and consisted of one question: ‘What were you thinking about while playing that point?’ This recall interview is part of the protocol reported by McPherson (2000), adapted to the requirements of volleyball by Moreno, Moreno, Ureña, Iglesias, and Del Villar (2008) and Araújo, Afonso and Mesquita (2011). Participants were instructed to leave the court after each successful trial and respond as accurately as possible to the question concerning their thoughts during the trial. There was no time limit to respond and additional feedback was provided when necessary. Participants completed six trials and each individual test session was completed in approximately 20-25 minutes.

Data analysis
Visual search data
Search rate included the mean number of fixation locations per trial, the mean number of fixations per trial and the mean fixation duration per trial, measured in milliseconds (Roca et al., 2011). A fixation was defined as the period of time ≥100ms (≥3 video frames) when the eye remained stationary within 3º of movement tolerance (Panchuk and Vickers, 2006). The between-group differences were analyzed using a One-Way ANOVA with Group (skilled vs. less skilled) as the between-participants factor. Partial eta squared values (η\textsuperscript{2}p) effect size measures were calculated.

Percentage viewing time referred to the percentage of time spent in fixation on each area of the display (Dicks et al., 2010). Ten locations were defined: ball trajectories (subdivided into serve trajectory, reception trajectory, and setting trajectory); players performing the action (subdivided into receiver, setter, attacker); players that are not performing an action but may play a role in the action (potential attackers); space (subdivided into space between a potential attacker and the setter, and space between the attacker and the blockers); and unclassified.

The ‘unclassified’ category was incorporated to report all the fixations that fell outside the scope of the other categories (McRobert et al., 2009), although these did not exceed 1% of the occurrences. Data were analyzed using a Two-Way ANOVA with Group (skilled vs. less skilled) as the between-participants factor and Fixation Location as the within-participants factors. Partial eta squared values (η\textsuperscript{2}p) effect size measures were calculated. Significant main effects were followed up using Bonferroni-corrected pairwise comparisons. Interaction effects were followed up using Scheffé post hoc tests.

Verbal reports
The participants’ statements were transcribed verbatim and encoded according to the model of protocol structure for tennis (McPherson, 2000), adapted to volleyball (Araújo et al., 2011; Botelho et al., 2011; Moreno et al., 2008). This adapted protocol includes goal concepts (which refer to the purposes of a chosen action within the context of the game), action concepts (referring to the action selected and its relevance in a specific situation), and condition concepts (specifying under which conditions the action occurred). As reported by Botelho et al. (2011), no goal concepts were articulated. We believe this reflects the nature of the defensive task in volleyball. In fact, the main goal is to play the ball towards the zone where the setter plays, with sufficient height to afford all setting options. As such, participants didn’t verbalize any
goals, presumably because they are implied. Additionally, no action concepts were mentioned, which may relate to the experimental design, since use of the eye-tracker inhibits the action of defense to some extent. Although participants actually moved on the court and tried to defend the ball, they could not, for example, fall to the floor, defend and roll, therefore limiting their action possibilities. Therefore, participants would intercept the ball on the rally, and lasting under 5 seconds, there was only a limited set of information available; hence, the concepts weren’t further divided into sub-concept categories.

These condition concepts were further examined with regard to their hierarchical levels, considering concepts about team members (level 1), and concepts about the opponents (level 2). The hierarchical level 0 (concepts about themselves) did not emerge in our study. The recorded concepts were additionally classified according to their level of sophistication, which reflects the appropriateness and level of detail of the verbal reports. Four levels were considered: inappropriate or weak (quality level 0), appropriate but without any details or features (quality level 1), appropriate with one detail or feature (quality level 2), and appropriate with two or more features (quality level 3). The appropriateness of the concepts was evaluated by comparing the reports with the video images of the corresponding situation, which were available through the eye-tracker’s scene camera. Skill-based differences in the number of condition concepts, levels of sophistication, and hierarchical levels were analyzed using a Mann-Whitney U test with Group as the between-participant factor. Effect size measures were calculated through the formula $r = Z/\sqrt{N}$.

### Reliability of the observation

The reliability of the data was established using the intra-observer and the inter-observer agreement methods. Altogether, 22.2% of the data were randomly selected and re-analyzed to provide agreement figures using the procedures recommended by Tabachnick and Fidell (2007). For search rate and percentage viewing time, Cronbach’s Alpha ranged from 0.954 to 0.977 for intra-observer reliability and from 0.900 to 0.947 for inter-observer testing. Agreement concerning verbal reports’ variables was determined with Cohen’s Kappa. Intra-observer testing showed Kappa values between 0.931 and 1.000. Inter-observer values varied from 0.807 to 1.000.

### Results

#### Visual search data

**Search rate** There were significant skill-based differences in the number of fixations ($F_1 = 4.792, p = 0.030, \eta^2_p = 0.029$) and number of fixation locations ($F_1 = 4.238, p = 0.041, \eta^2_p = 0.026$, see Table 1). The search behaviors of highly-skilled participants involved more fixations to a greater number of different locations compared with the skilled participants. The skilled participants presented superior mean fixation durations than the highly-skilled participants, but with no statistical significance.

**Percentage viewing time** A main effect was found for Fixation Location ($F_9 = 54.559, p \leq 0.001, \eta^2_p = 0.365$), but not for Group. However, a significant Group x Fixation Location interaction emerged ($F_9 = 6.321, p \leq 0.001, \eta^2_p = 0.062$). Post hoc Scheffe tests revealed that highly-skilled participants spent significantly more time fixating the receiver (22.28 ± 6.96%) and the space between the attacker and the blockers (24.92 ± 7.88%) compared to the skilled participants (18.36 ± 6.79 and 18.77 ± 10.27%, respectively). In contrast, the skilled participants spent significantly more time fixating on the attacker (27.15 ± 7.61%) compared with their highly-skilled counterparts (19.40 ± 11.87%).

#### Verbal report data

**Condition Concepts** Significant skill-based differences in the number of condition concepts were observed ($U = 1985.50, z = -4.548, p \leq 0.001, r = 0.357$, see Table 2). The highly-skilled participants generated significantly more condition concepts than their skilled peers. For example, a skilled participant would mention an aspect related to the block, while a highly-skilled participant would mention aspects related to the block, the attacker, and the setter.

**Level of Sophistication** Skill-related differences emerged with respect to the level of sophistication ($U = 1881.00, z = -4.423, p \leq 0.001, r = 0.355$), with the highly-skilled group showing a superior level of sophistication. As an example of a report with a hierarchical level 0, a skilled participant mentioned “the block was open” [meaning that there was space for the ball to pass between two

### Table 1. Differences in search rate per trial across groups. Data are means (±SD).

<table>
<thead>
<tr>
<th></th>
<th>Highly-skilled Mean rank</th>
<th>Skilled Mean rank</th>
<th>df</th>
<th>F</th>
<th>p</th>
<th>$\eta^2_p$</th>
</tr>
</thead>
<tbody>
<tr>
<td>No. Fixations</td>
<td>5.99 (1.33)</td>
<td>5.56 (1.15)</td>
<td>1</td>
<td>4.792</td>
<td>.030</td>
<td>.029</td>
</tr>
<tr>
<td>Mean Fixation Duration (ms)</td>
<td>596.29 (165.70)</td>
<td>627.15 (163.01)</td>
<td>1</td>
<td>1.408</td>
<td>.237</td>
<td>.009</td>
</tr>
<tr>
<td>No. Locations</td>
<td>5.76 (1.12)</td>
<td>5.40 (1.04)</td>
<td>1</td>
<td>4.238</td>
<td>.041</td>
<td>.026</td>
</tr>
</tbody>
</table>

* Significant for the 0.05 level

### Table 2. Differences in verbal reports across groups.

<table>
<thead>
<tr>
<th></th>
<th>Highly-skilled Mean rank</th>
<th>Skilled Mean rank</th>
<th>U</th>
<th>z</th>
<th>P</th>
<th>r</th>
</tr>
</thead>
<tbody>
<tr>
<td>No. Condition Concepts</td>
<td>95.44</td>
<td>64.08</td>
<td>1985.50</td>
<td>-4.55</td>
<td>$\leq 0.001$ *</td>
<td>.357</td>
</tr>
<tr>
<td>Level of Sophistication</td>
<td>89.60</td>
<td>61.94</td>
<td>1881.00</td>
<td>-4.42</td>
<td>$\leq 0.001$ *</td>
<td>.355</td>
</tr>
<tr>
<td>Hierarchical Level 1 – Team Members</td>
<td>85.38</td>
<td>76.65</td>
<td>2891.00</td>
<td>-1.37</td>
<td>.172</td>
<td>.108</td>
</tr>
<tr>
<td>Hierarchical Level 2 - Opponents</td>
<td>91.99</td>
<td>68.38</td>
<td>2295.50</td>
<td>-3.43</td>
<td>.001 *</td>
<td>.269</td>
</tr>
</tbody>
</table>

* Significant for the 0.05 level
blockers], but video data showed that, clearly, it was not. A highly-skilled participant, in a similar situation, mentioned ‘block on time, middle-blocker was late’, a level 3 report, since it presents at least two features and is appropriate to the situation.

Hierarchical Levels There were significant differences across groups for hierarchical level 2 – opponents (U = 2295.50, z = -3.426, p<0.001, r = 0.269), but not for level 1 – team members. The highly-skilled participants produced more condition concepts referring to their opponents (namely the attackers) compared to the skilled participants, with this event explaining the superior number of condition concepts produced by highly-skilled players.

Discussion

We explored the processes supporting skilled performance in a dynamic, externally paced volleyball task using a representative design involving in situ data collection. Participants played in a 6 vs. 6 simulated sequences as centre backcourt defenders, while eye movements and immediate retrospective verbal reports of thinking were collected. We reported the visual search behaviors and verbal reports of thinking underpinning skilled performance in a live-action representative task. Although some movements were limited by the usage of the eye-tracker device (e.g.: to fall and roll), participants could move in the court and they actually tried to intercept the ball.

Our data showed that highly-skilled players employed significantly more fixations to a greater number of locations than skilled players, as previously reported in cricket (McRobert et al., 2009) and soccer (Roca et al., 2011). It has been suggested that more efficient search patterns involve fewer fixations of longer duration and appear to be linked to expert performance (Mann et al., 2007; Piras et al., 2010). However, complex sports may benefit from different visual strategies, making a larger number of shorter fixations to several locations (North et al., 2009). It is apparent that the nature of the task strongly influences the processes underpinning decision-making.

Skill-based differences emerged with respect to percentage viewing time, with highly-skilled participants spending more time fixating on functional spaces, especially just before and during ball contact, as previously reported in soccer (Roca et al., 2011). By fixating on these functional spaces, the participants may be able to retrieve a greater amount of information simultaneously through more effective use of peripheral vision (Behrmann and Ewell, 2003; Laurent and Ripoll, 2009; Williams et al., 2011).

Verbal reports showed that highly-skilled participants generated significantly more condition concepts than the skilled counterparts (cf., McPherson, 2000). As condition concepts reflect the attunement to the surrounding constraints, a superior number of condition concepts may suggest that the highly-skilled participants are better attuned to the task constraints. In support of this idea, highly-skilled participants produced statements that were more sophisticated in comparison to the skilled participants, corroborating previous reports involving baseball (McPherson, 1993), tennis (McPherson and Kernodle, 2007) and volleyball (Botelho et al., 2011). A clear link emerged between visual search behaviors and a more sophisticated knowledge base (expressed through verbal reports of thinking), in the sense that experts attended more to functional spaces, allowing them to capture richer information concerning game problems, which translated into generating more condition concepts of superior sophistication. Furthermore, similar amounts of verbal reports were produced with regard to team-mates, but highly-skilled participants produced more concepts concerning their opponents in comparison to their skilled peers. Hence, highly-skilled participants are better attuned to game constraints created by the opponents.

Conclusion

In conclusion, a representative volleyball task was created involving live-action situations with the purpose of examining skill-related differences in perceptual-cognitive expertise. Although motor responses were not evaluated, they were included in the task, in order to better respect the specific perception-action couplings of real-life situations, therefore capturing more accurately the nature of the skill-based differences (Dicks et al., 2010; Mann et al., 2010). The visual search behaviors of highly-skilled players were more exploratory than those of their skilled counterparts, with highly-skilled players employing more fixations to a greater number of locations. The highly-skilled players spent more time fixating functional spaces (i.e. areas that are intermediate to a number of cues of interest). Furthermore, highly-skilled participants generated significantly more condition concepts than skilled participants, with these statements being of superior quality, potentially reflecting better attunement to the task constraints. Overall, it was established that highly-skilled players present superior ability in detecting relevant information in the visual display (Laurent et al., 2006; McPherson and MacMahon, 2008). The perceptual-cognitive account often distinguishes declarative and procedural knowledge (Thomas and Thomas, 1994). Perhaps verbal reports of thinking provide a measure of the former, with gaze behavior affording a measure of the latter.

References


Key points

- Decision-making in complex sports relies deeply on perceptual-cognitive expertise. In turn, the effect of expertise is highly dependent on the nature and complexity of the task.
- Nonetheless, most researchers use simple tasks in their research designs, risking not capturing performance in a meaningful way. We proposed to use a live action setting with a complex task design, representative of real world situations.
- We combined eye movement registration with collection of immediate retrospective verbal reports. Although the two data sets are not directly comparable, they may be used in a complementary manner, providing a deeper and fuller understanding of the processes underpinning superior performance.
- Highly-skilled players exhibited more exploratory visual search behaviors than their skilled counterparts, performing more fixations into more locations. They also attended more to functional spaces between two or more areas of the display. Furthermore, highly-skilled players produced more condition concepts and with superior level of sophistication, revealing a deeper and more meaningful attunement to the task constraints.