CROSSBOARD: Crossmodal Access of Dense Public Displays

Stephen W. Gilroy, Patrick L. Olivier, Han Cao, Daniel G. Jackson, Christian Kray and Di Lin

Abstract— **Crossmodal displays aim to bridge the gap between ambient display technology and personal mobile human-computer interaction through the exploitation of aspects of crossmodal cognition. We extend the notion of sequential temporal multiplexing, introduced for crossmodal ambient displays, and describe a hierarchical temporal multiplexing approach. We demonstrate this through CROSSBOARD, a prototype public display application that harnesses hierarchical crossmodal cues to support efficient multi-user interaction with dense public information displays. Results of a pilot user study are presented in which the potential of CROSSBOARD for improving the retrieval of unindexed information from dense information displays is clearly demonstrated.**

Index Terms—**public displays, crossmodal attention, human-computer interaction, pervasive computing.**

I. INTRODUCTION AND RELATED WORK

DISPLAY-based systems have been a major topic of interest in ubicomp research from the beginning [1]. Displays are seen less as personal devices on a user's desktop, but as a provider of situated information [2] or as a tool for collaboration [3].

Systems such as that presented by Huang and Mynatt [4] provide a level of interactivity as well as information display, allowing users to leave messages for each other, and obtain information about a specific group's activities. Of more interest are displays that actively support users in finding the information they desire, by tailoring the display according to the groups or users that are present [5], [6], or utilizing the location of the display [7]. Knowing positions and orientation of displays is useful in navigation tasks, as it amounts to tacit information about the user's location.

The CROSSBOARD system aims to pinpoint information relevant to a user by directing the users' attention to the portion of the screen that shows information relevant to them, instead of filtering information according to who is viewing the display. This means the display can cope with a large number of simultaneous users (whose information is all displayed at once) and does not need to attempt to keep track of which user is attending to the display at a particular moment.

An individual user can still find the information that is relevant to them, and all users (including those unknown to the system) can benefit from having a larger amount of information available to them.

Crossmodal cognition and its application to ambient displays are discussed in section 2. A characterization of public display usage is introduced in section 3 and a proposal for multi-user simultaneous access using CROSSBOARD is presented in section 4 and 5. A pilot user study and preliminary results are outlined in section 6 and we conclude with a discussion of outstanding issues and directions of future research.

II. EXPLOITING CROSSMODAL ATTENTION

A. Psychology of Crossmodal Attention

The psychological basis of crossmodal attention is discussed in Olivier et al. [8]. The effectiveness of crossmodal cues is predicated on any bottleneck in a human's information processing [9] being overcome by utilizing the fact that information from unattended sources can also make itself available at the higher levels of processing, bypassing such limitations [10]. There is evidence that demonstrates human ability to utilize valid co-occurrences (simultaneous inputs in more that one sensory input about the same external event) to improve performance [11], a well known example of which is the McGurk effect [12]. The effect of attention in multimodal interfaces has already been documented and put to use in improving the design of such interfaces [13]–[15]. Importantly, humans can integrate spatial cues across a range of modalities (audition, vision, touch and proprioception) [16].

B. Crossmodal Cues in Ambient Displays

Olivier et al. describe a crossmodal ambient display framework that aims to bridge the gap between ambient display technology and personal mobile human-computer interaction through the exploitation of certain aspects of crossmodal cognition [8], [17]. Their prototype, CROSSFLOW, is a crossmodal ambient display prototype for indoor navigation and for which significant improvements in navigation performance was demonstrated.

CROSSFLOW consists of a set of destinations in an environment, to which a user is directed through the projection of a "flow" of objects onto the floor of the environment. The

Patrick Olivier is with the Informatics Research Institute, University of Newcastle Upon Tyne, Newcastle Upon Tyne, NE1 7RU, UK (phone: +441912464939; fax: +441912464905; e-mail: p.l.olivier@ncl.ac.uk).

Han Cao (han.cao@ncl.ac.uk), Stephen Gilroy (steve.gilroy@ncl.ac.uk), Daniel G. Jackson (d.g.jackson@ncl.ac.uk) and Christian Kray (c.kray@ncl.ac.uk) are with the Informatics Research Institute, University of Newcastle Upon Tyne, Newcastle Upon Tyne, NE1 7RU, UK.

direction of the objects in the flow indicates the direction in which a user must move to reach the destination; the objects “point” towards the destination. Figure 1 shows the projection of CROSSFLOW using fish-like flow patterns (to invoke the feeling of the ‘flow’ of a stream).

The possible destinations are cycled through one at a time, the flow pattern displayed changing to point at the currently selected destination. The change of destinations is coordinated with a crossmodal cue (Fig. 2). For example, in time slot 1, directions to destination A are displayed at all locations in the physical space, in time slot 2, directions to destination B are displayed. After all destinations have been displayed, the sequence is repeated.

The destination relevant to the user is identified through the utilization of a crossmodal cue (e.g., a sound or vibration) issued by a personal mobile computing device. When the user receives a cue, the currently displayed flow (matching that cue’s time slot) is that which they should follow. The user is, then, using private information (the cues) that only they can perceive in order to decode the publicly available information displayed in the environment (which in this case is the directions to the destinations).

Fig. 1. Close-up of fish-like flow patterns used in CROSSFLOW (top); projection of the display in the environment (bottom).

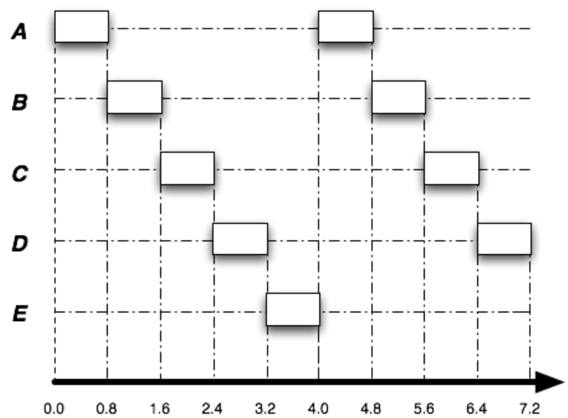

Fig. 2. Sequential temporally multiplexed crossmodal cues in CROSSFLOW.

In CROSSFLOW, appropriate direction information is displayed at every location in the environment, and the information appropriate to users is temporally multiplexed over the cycle of destination. This can be contrasted with traditional, hand-held notions of navigation, where directional information is given privately to the user based on their specific location, that is users’ directional information is spatially multiplexed at any one moment in time.

III. INFORMATION ACCESS ON PUBLIC DISPLAYS

Conventional public displays, as typified by a flight departure board in an international airport, currently support a range of information retrieval tasks. We classify these as follows:

1) Retrieving indexed information
2) Retrieving unindexed information
3) Relational queries (multiple retrievals)
4) Monitoring

Retrieval of indexed and unindexed information relates to the nature for the activity that a user in engaged in. For example, the principal function of an airport display board is to inform users of updates to the departure times of aircraft and the relevant gate from which the departure will occur. Thus the flight time (and destination) is the principal index by which a tabular display of the information is organized. The layout corresponding to this indexing is conventionally (e.g., in international airports), top-down, and left-to-right, both within and between multiple collocated displays. Consequently unindexed information, such as the time and destination corresponding to a particular flight number can only be retrieved through sequential search (or other ad hoc strategies).

Where the quantity of information is particularly large, there are well-understood conventions for limited temporal multiplexing of its presentation on public displays. This can occur within, and between elements of the display. Within-element multiplexing occurs when a field sequentially displays related information, for example, the identifier of the planned departure gate for a flight, and instructions such a "wait in lounge". Between-element multiplexing is used occurs where the amount of information is too large to show at one time on the displays available. Thus displays scroll through pages of information and include subheadings indicating the screen number currently being displayed and the total number of screens.

Relational queries and monitoring both require the stable and indexed display of information. For example, relational queries, such as which of two flights leaves first, involves the identification of the two flights, their relative locations on the display, and a mapping from these locations (via the indexing) to the query. Likewise, monitoring, such as checking for state change in a flight departure display, requires the identification of the spatial location of an element and easy repeated retrieval. Both task require that the elements displayed on the board are relatively stable, that the indexing of the information does not change, and that changes to the location of displayed information occurs in a well understood manner. In case of a departure board the entries move up, and to the left, as time passes.

The stable display of information also has an impact on privacy and collaborative use of displays allowing easier verbal and multimodal coreferencing and shoulder surfing (i.e., identifying what information a user is attending to).

IV. CrossBoard: Crossmodal displays

CrossFlow only supports as many destinations as the sequence of crossmodal cues allows (the sum of the time-slots). There will be a desirable upper limit on the time between cues so that the user does not have to specifically attend to when the cue is issued, breaking the ambience of the system.

In Fig. 2, the time-slot for each cue has a duration of 800ms and if directions to each destination need to be repeated every four seconds this allows the system to provide cues for four destinations.

With CrossBoard, the aim is to again support multi-user access to information, but this time arranged on a large public display, densely arranged, as is found on departure boards in airports and railway stations. Here, large amounts of information, which can run into hundreds of items, are presented on collocated displays, and typically the user is interested in finding the detail for one particular item (the flight/train they are trying to catch). Users regularly encounter difficulties retrieving relevant information in such environments, where the large amount of visually similar information, combined with environmental distractions and time constraints, make the retrieval cognitively tasking.

CrossBoard augments such displays by adding the same crossmodal cues that were utilized in CrossFlow, in order to highlight the location of a user's target information item. The cues are associated with regions of the display that are flashed in sequence. Depending on the number of items displayed, these regions may then be divided into sub-regions, which also have crossmodal cues associated with them. As the cued regions become smaller, it allows a user to rapidly search the region for the required information. This hierarchical cueing does not affect the use of display in a traditional manner, without cues, but allows those users with cues both to narrow down a region of the display that is searchable by some indexed part of the information item (e.g., departure time) and also to quickly locate an item that is not searchable by an index, which would otherwise require systemic scanning of the whole display.

V. Hierarchical crossmodal cues

In order for the user to locate the item of information they are interested in, the display must systematically divide the display up into sections that are highlighted at the appropriate time interval that the user's crossmodal cue is triggered. The number of levels of subdivision applied depends on the granularity to be presented to the user: a block of items, a list of items or the individual item of interest. More levels of subdivision lead the user closer to the item, but require more cycles of highlighting.

In general, the display will highlight all the divided areas in turn, which allows the display to be utilized by several users, whose cues are provided at different time steps. Each user has cues at each level of subdivision that direct them to a single area. All such areas are then divided again, and the user must attend to the highlighting within the relevant division.

As the density of the board increases, the number of time steps required to pick out an individual item in a single level increases. Subdivision trades off the time taken to pick out an area with the cycle time until the user can synchronize with the cues again

A display can be subdivided in a number of different ways at different levels depending on the layout of items. For a simple grid, a row/column division can pinpoint an item in two cycles (of r and c steps respectively), but if the grid is large, it will take a long time to cycles through all of the rows or columns. Instead the display can be divided into fewer columns or rows, made up of more than one item, and then each of these groups is divided on another cycle.

For example, in the CrossBoard prototype, each screen is divided into 16 cells, with 5 items in each cell. Subdivision is used to reveal the cell the item is in, and then a further row division is used to pick the item from a cell.

A. *Area Division*

Area division partitions a display into a number of discrete areas. They can be any size and shape as long as they tile the area completely. CrossBoard commonly uses binary division, or quartering, and highlights the quarters as shown in Fig. 3.

B. *Row/column Division*

Division by row or column is a special case of area division.

A row division has m areas with a width equal to that of the whole display and a height of h/m. A column division has n areas of equal width w/n and a height equal to the height of the section.

Unless the section consists of a single row or column, a pass of rows and a pass of columns are both needed to select a cell in the grid. So, a 2×2 grid can be divided by 1 cycle of 4 square areas (1×1) or 2 cycles of rows (1×2) then columns (2×1). The total cycle time is 4 steps in either case. Selection from within these groups is performed at the next level of subdivision, or left to user scanning. This is shown in Fig. 4.

C. *Multiple Subdivision*

For larger arrangement of cells, there is a choice of how to subdivide them. As an example, consider a 4x4 arrangement of 16 cells. A single cell can be indicated by one level of division, using 4 columns and 4 rows, giving a cycle time of 8 steps. Alternatively it can be divided by two levels of 2×2 area subdivision, of 4 steps each, again a total of 8 steps.

In the one-level arrangement the user has to remember the row until the column is selected too. In the two-level, they only have to remember the current section, which is then further subdivided.

Now consider an asymmetrical arrangement of 8×4 cells. A row/column division will give 8+4 = 12 steps. A subdivision of 4 areas of 2×1 then a row division of the 2×1 areas gives 4+2 = 6 steps—twice as fast.

In general, binary subdivision will use fewer timesteps than two row/column passes, which are more efficient than highlighting every element. Area and row/column divisions result in the same number of steps if a grid is square or consist of 1 row or column. However area divisions imply more subdivision, while row/columns require the user to attend to a previously highlighted area for longer. Subdivision of a display into quarters, then each quarter into 5 rows, is shown in Fig. 5. The cue for the selected item in each cycle is shown underneath.

VI. User Study

A pilot user study was undertaken to investigate the effect of crossmodal cues on information retrieval from a display. This was a within-subjects study of two retrieval tasks for three different board types: plain display, subdivision highlighting without crossmodal cues, and subdivision highlighting with an audio cue. The two retrieval tasks are to find the remaining part of a three-part item given the two other parts. In one task the user is given the information that the item is indexed on, in the other task the user is given the unindexed parts, and must retrieve the indexed part.

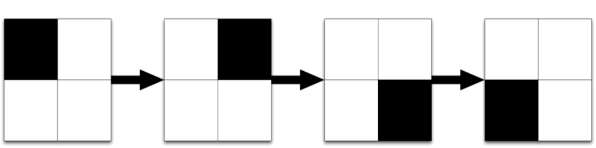

Fig. 3. Binary area subdivision of a display.

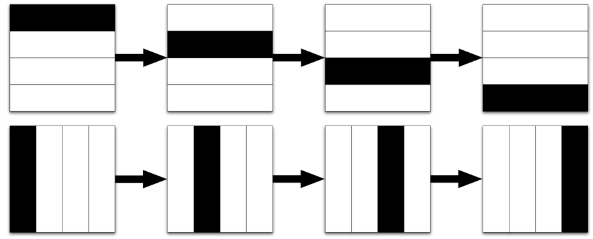

Fig. 4. Row division (top) and column division (bottom) of a display.

A. *Display configuration*

The test display was based on an airport departure board, with each item giving the flight number, destination and departure time. Items were indexed by departure time, as is the norm. The display consisted of 240 items, spread across three large screens. The first cycle of cues selects the relevant screen. A binary area subdivision was used to partition each screen into 16 cells, and then each cell contains 5 items, highlighted by a row subdivision. This setup is shown in Fig. 6, and the sequence of subdivisions in Fig. 7.

The 80 items on each screen are arranged as 20 rows and 4 columns, which can be subdivided in a number of ways ('a', 'r' and 'c' refer to area, row and column respectively):

two levels of quartering plus row division

$$(4a) + (4a) + (5r) = 3 \text{ levels, } 3 \text{ cycles, } 13 \text{ steps} \tag{1}$$

one level of row/column division

$$(20r + 4c) = 1 \text{ level, } 2 \text{ cycles, } 24 \text{ steps} \tag{2}$$

quartering plus row/column division

$$(4a) + (10r + 4c) = 2 \text{ levels, } 3 \text{ cycles, } 18 \text{ steps} \tag{3}$$

row/column plus row division

$$(4c + 4r) + (5r) = 2 \text{ levels, } 3 \text{ cycles, } 13 \text{ steps} \tag{4}$$

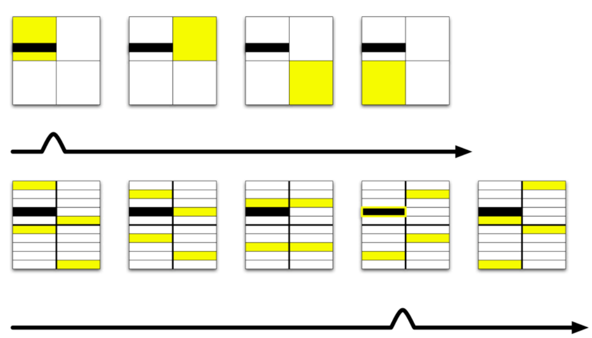

Fig. 5. Hierarchical temporally multiplexed crossmodal cues in CrossBoard.

Fig. 6. Display setup, using a 3-screen CAVE in a "wall" arrangement.

As can be seen, (1) and (4) are equivalent in terms of number of time step and cycles, but (1) utilizes more levels of subdivision. Style (1) was chosen in order to investigate both area subdivision and row/column subdivision while minimizing the total number of steps.

D. Retrieval tasks

Each user performed 10 retrieval tasks for each display condition, in a random order: 5 indexed retrievals and 5 non-indexed retrievals. The dependent variable measured was the time taken to find the required item.

The indexed retrieval task was to find the flight number of a particular flight given the destination and departure time. Non-indexed retrieval required finding the departure time for a given flight number and destination. The flight information was generated randomly, and there were flights to the same destination with different numbers leaving at different times, and also flights leaving at the same time, but to different destinations. A different board of flight times was presented for each task iteration. All users received the same boards in the same order.

E. Expectations

It was expected that without crossmodal cues, users would perform better at indexed task retrieval than non-indexed, as they can utilize the board ordering to effectively scan for the indexed item. Items that were at the beginning of the ordering would be found more quickly than those at the end.

For crossmodal cues, there should be less difference between the two tasks, as the cues drill down to an item irrespective of whether the indexed information is known or not. It was still expected that the indexed task will be completed in less time, as once at a suitable level of subdivision, the user can leverage the existing board ordering to find the item within a division without needing cues.

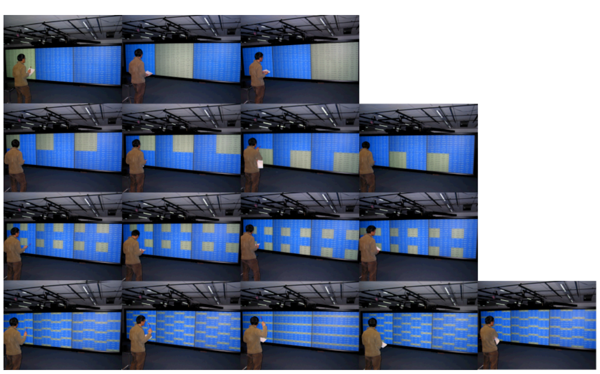

Fig. 7. Subdivision of the display.

Overall, it was expected that using the crossmodal display would be quicker than a static display when accessing non-indexed information.

The highlighted board without audio cues will determine whether the crossmodal display is distracting to users who do not receive the other cues.

F. Results

The dependent variable measured was the time from being shown a display to finding the required item (time taken to read out the requested information was ignored). The experiment was carried on 8 subjects.

Treating each task as a one-way repeated measures ANOVA revealed no significant difference across the three conditions for the indexed information retrieval ($F(2,14)=2.286$; df=2,14; $p=0.138$). However, the mean value and standard deviation of the crossmodal display is the minimum among the three conditions. The performance for unindexed information retrieval revealed a significant difference across conditions ($F(2,14)=24.659$; df=2,14; $p<0.001$). Again, the mean value and standard deviation of the crossmodal display is the minimum.

G. Discussion

These initial results suggest that the crossmodal display is effective at reducing retrieval time for information on a dense display. If the information is arranged in a fashion unordered by the searched value, the crossmodal displaying is significantly faster than a static display. Although the times for a highlighting display without cues were comparable to a non-highlighted display, users did comment that the flashing of the highlighting was distracting. Users were not given an extensive training on using the crossmodal technique (they were shown one example at the start of the experiment).

IV. Conclusions and Further Work

Our initial study of CrossBoard has demonstrated that hierarchical crossmodal cues can be utilized to provide greatly enhanced retrieval performance for unindexed information on dense public information displays without adversely affecting the performance of conventional display usage (for the retrieval of indexed or unindexed information). A number of enhancements are planned including an investigation of the design space for the visual cues, the crossmodal cues, and the integration of crossmodal public and ambient display systems. With a view to increasing the ecological validity of our empirical paradigm we also plan larger scale multi-user and multi-display user studies in natural environments.

More ecologically natural environments will also permit the study of additional properties that are desirable in a public information displays such as *legibility at a distance*. In many spaces and contexts, users access public displays in a dynamic manner, both reading the displays while moving, and/or walking close enough so as to be able to retrieve information. One useful feature of CROSSBOARD is the location of information relevant to users is legible at significantly greater distances than the specific detail of the element. For example, the location of information about a particular flight on a departure board is apparent at distances at which the text is not readable. We anticipate that in dynamic environments users will be able to beneficially coordinate their use of public displays as a result. For example, moving towards sections of the display where their information resides, and only moving close enough to be able to read a single entry rather than a distance where they can comfortably scan all entries.

REFERENCES

[1] M. Weiser, R. Gold, and J. S. Brown. The origins of ubiquitous computing research at PARC in the late 1980s. *IBM Systems Journal*, 38(4):693–696, 1999.

[2] A. D. Craig Wisneski, Hiroshi Ishii, M. G. Gorbet, S. Brave, B. Ullmer, and P. Yarin. Ambient displays: Turning architectural space into an interface between people and digital information. In N. A. Streitz, S. Konomi, and H. J. Brukhardt, editors, *Cooperative Buildings, Integrating* Information, Organization, and Architecture, 1st *International Workshop, CoBuild'98*, pages 22–32, Berlin, Heidelberg, New York, 1998. Springer.

[3] O'Hara, K., Perry, M., and Lewis, S. Social coordination around a situated display appliance. In B. Cockton and P. Korhonen, editors, *Proceedings of the ACM Conference on Human Factors in Computing Systems (CHI 2003)*, pages 65-72, Ft. Lauderdale, Florida, USA, April 05 - 10, ACM Press, 2003.

[4] E. M. Huang and E. D. Mynatt. Semi-public displays for presence and recency information. In B. Cockton and P. Korhonen, editors, *Proceedings of the ACM Conference on Human Factors in Computing Systems (CHI 2003)*, pages 49–56, Fort Lauderdale, Florida, ACM Press, 2003.

[5] J. F. McCarthy, T. J. Costa, and E. S. Liongosari. Unicast, outcast and groupcast: Three steps toward ubiquitous, peripheral displays. In G. D. Abowd, B. Brumitt, and S. A. N. Shafer, editors, *Proceedings of Ubicomp 2001*, pages 332–345, Berlin, Heidelberg, New York, Springer, 2001.

[6] N. A. Streitz, C. Röcker, T. Prante, R. Stenzel, and D. van Alphen. Situated Interaction with Ambient Information: Facilitating Awareness and Communication in Ubiquitous Work Environments. In D. Harris, V. Duffy, and C. Stephanidis, editors, *Human-Centred Computing: Cognitive, Social, and Ergonomic Aspects*, pages 133–137, New Jersey, Lawrence Erlbaum Publishers, 2003.

[7] C. Kray, G. Kortuem, and A. Krüger. Adaptive navigation support with public displays. In R. S. Amant, J. Riedl, and A. Jameson, editors, *Proceedings of the 10th International Conference on Intelligent User Interfaces (IUI 2005)*, pages 326–328, New York, ACM Press, 2005.

[8] P. Olivier, S. W. Gilroy, H. Cao, D. Jackson and C. Kray. Crossmodal Attention in Public-Private Displays. In *Proceedings of the IEEE International Conference on Pervasive Services (ICPS'06)*, Lyon, France, 2006.

[9] H. E. Pashler. The Psychology of Attention. MIT Press (Bradford Books), Cambridge, MA, 1999.

[10] S. J. Luck and S. P. Vecera. Attention. In H. Pashler and S. Yantis, editors, *Stevens' Handbook of Experimental Psychology: Sensation and Perception* (3rd ed.), volume 1, pages 235–286. Wiley, New York, NY, USA, 2002.

[11] P. Bertelson and B. D. Gelder. The Psychology of Multmodal Perception. In C. Spence and J. Driver, editors, *Crossmodal Space and Crossmodal Attention*, pages 141–177. Oxford University Press, Oxford, UK, 2004.

[12] H. McGurk and J. McDonald. Hearing lips and seeing voices. Nature, 264(5588):746748, 1976.

[13] J. Driver and C. Spence. Attention and the crossmodal construction of space. *Trends in Cognitive Sciences*, 7(2):254–262, 1998.

[14] J. Driver and C. Spence, C. Spatial synergies between auditory and visual attention. In: Umilta, C., Moscovitch, M. (Eds.), Attention and Performance XV: Conscious and Nonconscious Information Processing. MIT Press, Cambridge, USA, pages 311-331, 1994

[15] S. Oviatt. Ten myths of multimodal interaction. *Communications of the ACM*, 42(11):74–81, 1999.

[16] J. Driver and C. Spence. Crossmodal attention. *Curr Opin Neurobiol*, 2(8):245–253, 1998.

[17] P. Olivier, H. Cao, S. W. Gilroy, and D. Jackson. Crossmodal Ambient Displays. In *Proceedings of the 20th BCS Annual Human-Computer Interaction Conference (HCI2006: ENGAGE)*, Queen Mary, University of London, UK, 2006.