Crossmodal Attention in Public-Private Displays

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

***Abstract*—Striking a balance between the public visibility of a display, and ease by which individuals have access to information, is a key challenge for the developers of interfaces to pervasive services. In this paper we utilize the cognitive phenomenon of crossmodal attention as a means of providing users with personalized cues to content on public displays. We describe two prototype applications that use crossmodal cues to temporally multiplex publicly visible information: CROSSFLOW, an ambient navigation system; and CROSSBOARD, a dense multi-user public information display. We outline the results of pilot preliminary user studies and describe the infrastructure required to support crossmodal displays.**

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

I. INTRODUCTION

Public displays are ubiquitous in our everyday environments, from shopping malls and railway terminals, to hospitals and airports. The very fact that any casual bystander can access this information places significant constraints on the information that can be displayed. Conventional wisdom decrees that information that is anything other than completely public should be provide on personal devices, in current terms, hand held devices such a PDAs and mobile phones.

Shifting information from public displays to personal private displays has a number of consequences. Firstly, it results in a disengagement of users from their environments, as user look away from displays that are embedded in the space encompassing the users to their own devices. Yet the very placement of many public displays is the results of careful design to provide information tailored to the context of a spatial location, the services accessible at this location (whether digital or physical), and tasks that users at these locations are likely to be engaged in.

Spatially contextualized pervasive services can still be achieved within the paradigm of the personal display, but only through the use of interaction heavy user interfaces, whereby users negotiate a set of service choices on their devices, or by tracking the locations users in an environment and adapting the personal hand-held display accordingly. Tracking users in indoor environments presents additional sets of challenges for developers and designer, both technical and ethical.

Although public displays are typically present highly spatially contextualized information, large numbers of display can result in significant clutter in an environment, resulting in user confusion and undermining the effectiveness of individual displays. Furthermore, some classes of information, cannot be identified as naturally residing on any specific location in an environment or as supporting any of the primary tasks of the users in an environment. Such information is ideally presented in an ambient manner.

To address the inherent tension between the need to display information publicly and the individual's desire for an efficient and private display we present two prototype applications that utilizes crossmodal cognition so as to temporally multiplex publicly visible information: CROSSFLOW, an ambient navigation system; and CROSSBOARD, a dense multi-user public information display. We outline the results of pilot preliminary user studies and describe the infrastructure required to support crossmodal displays.

II. PUBLIC DISPLAYS

From the outset of ubiquitous computing research, display-based systems have been a core topic of investigation [1]. There has been a steady interest ever since in display systems supporting people in a variety of tasks (e.g., see [2] for a general discussion and [3] for an overview of systems supporting collaboration). This applies in particular to displays that are not meant to be attended to continuously and exclusively, unlike desktop screens attached to a personal workstation.

Huang and Mynatt [4] presented a projection-based system that was set up in a group area and provided both interactive and information services. Users were able to leave messages for colleagues and obtain information about group activities and people's presence. Some of this information was displayed in an abstract way (e.g., people fading from a group photograph when not being present for a longer period of time) and some in a concrete way (e.g., textual information about group activities).

Like most systems of this kind, the semi-public display did not actively support users in finding the information that was relevant for them (beyond the assignment of certain sections of the display to certain types of information). Another class

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.

of display systems tailors its content to a present user or group. For example, the GroupCast system [5] automatically selects content based on which users it senses in its vicinity. The approach used to build the HelloWall (GossipWall) [6] goes one step beyond that by also adapting the content according to the distance of a user to the display. The navigation system described in [7] further tailors the displayed content based on the physical location and orientation of the actual display.

III. Crossmodal Attention

Psychological research into attention, over many decades, has demonstrated the existence of an information processing bottleneck that implies one-at-a-time processing 18]. When engaged in multiple tasks simultaneously, an inherent limitation in human information processing capability can be observed, resulting in rapidly deteriorating performance when the number of concurrent tasks increases. However, there is also considerable evidence that some information from unattended sources ultimately reaches higher stages of processing [8,9], which allows people to bypass information processing limitations. It also opens up possibilities for people to receive information efficiently in a manner that that does

not require full attention.

Although observations about crossmodal interactions appeared at the very earliest stage of the development of psychological science (e.g. Johannes Mueller refers to ventriloquist effect in his 1839 presentation of his 'law of specific energies of nerves'), in the past three decades considerable empirical research in cognitive neuroscience has given rise to the notion of crossmodal integration, a term used to refer to capacities and effects involved in the process of coordinating (or 'matching') the information received through multiple perceptual modalities. Significant evidence has been accrued that demonstrates our ability to utilize valid co-occurrences (simultaneous inputs in more that one sensory input about the same external event) to improve our performance [10]. Well known examples include the McGurk effect where a video of one phoneme being spoken is dubbed with a sound-recording of a different phoneme – for many viewers the perceived phoneme is a third, different phoneme that is in someway intermediate to the actual and dubbed phoneme [11]. Performance for tasks that rely on spatial attention are also significantly effected by the consistency of the multimodal cues [12]. For example, Driver and Spence identified clear performance deficits for subjects that were required to attend concurrently to separate task-relevant locations [13]. The practical implications of this have already been exploited in the design of conventional multimodal interfaces [14].

Of particular relevance to our goal, that of designing a public interaction technique with low attentional requirements, is the fact that a range of psychological studies have demonstrated that humans spontaneously (and pre-attentively) integrate spatial cues across a range of modalities (audition, vision, touch and proprioception) [15].

Figure 1. CrossFlow uses fish-like flow patterns projected onto the floor of an environment (top). Patterns are designed to be ambient so as to minimize distraction to users of the space. In the user study "fish" were approximately the size of a fist (bottom).

IV. Crossmodal Displays

A. *CrossFlow: Ambient Navigation*

In CrossFlow directions to different locations in the space (including exits) are projected on the floor in the form of directions of flow on fish-like objects on the floor of the environment (see Figure 1). The flow patterns (indicating the direction to the destination) are displayed one-at-a-time on a fixed time cycle in coordination with a crossmodal cue (see Figure 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, and so on until the sequence is repeated. The user identifies (or

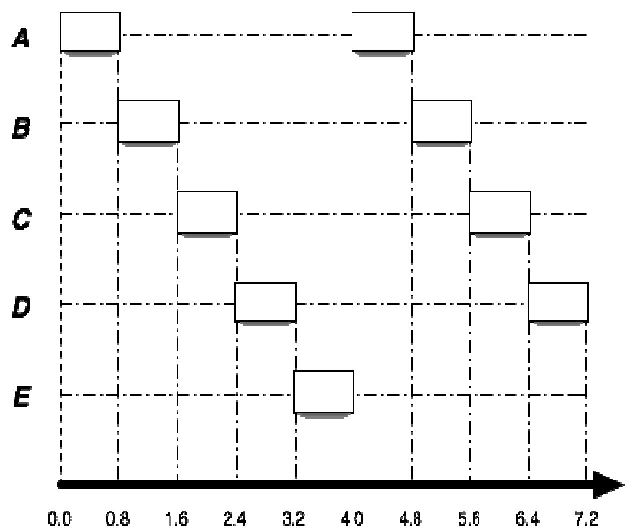

Figure 2. Sequentially time multiplexed crossmodal cues in CROSSFLOW. Crossmodal cues occur on a fixed cycle – in the schedule above each of the five cues has an 800 ms window and repeats every four seconds.

decodes) which time slot in the cycle is relevant to their own request for directions through the utilization of a crossmodal cue (e.g., a sound or vibration) issued by his/her personal mobile computing device. That is, either in response the user's request for directions, or on entry to the physical space, the user's the device communicates with the ambient display infrastructure to establish the schedule of time slots when the different directions will be displayed. In other words, the personal mobile device displays private information cues, that only individual users can perceive, that allow users to decode their personal-public information displayed in the environment (in this case route directions).

Note that the directions displayed at a location depend on the direction of the destination from that location. In traditional hand-held notions of navigation, there is a requirement to track the position of a user and present directions salient to the specific location of the user. We can contrast these two configurations in terms of the multiplexing of information displayed. In traditional mobile device applications, incorporating tracking, information is spatially multiplexed. That is, the position of a user is known and information specialized to the location of the user is displayed on the user's device. In the crossmodal scenario information is temporally multiplexed and information relevant to a location is displayed at all locations (in this case through projection on the floor of the environment) at a specific time.

B. *CROSSBOARD: Access to Dense Public Displays*

CROSSFLOW projects dense information patterns (i.e. the flow-field) indicating the direction to a destination from all points in an environment, but the system only supports as many destinations as the sequence of crossmodal cues allows (the sum of the time-slots). In figure 2, the time-slot for each cue has a duration of 800ms and if directions to each

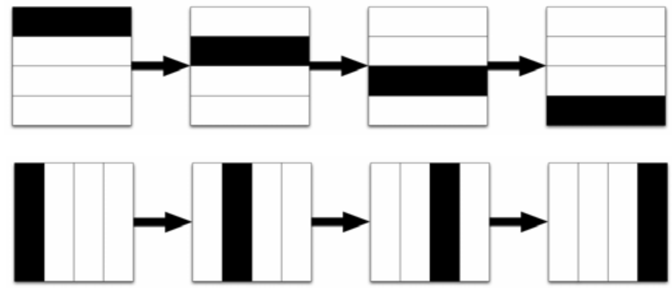

Figure 3. Row division (top) and column division (bottom) of a display. Information displays are typically oriented on a grid and can be divided according to different spatial partitioning schemes.

destination need to be repeated every four seconds this admits the projection of flow patterns for five destinations.

Our aim is to utilize crossmodal cues to support concurrent multi-user access to individual entries that are densely presented on large public displays, such as flight departure boards in international airports. In such contexts, information is typically presented on multiple collocated displays, and the number of elements (where an element is information relating to one user) can run to many hundreds. Users regularly encounter difficulties retrieving relevant information in such environments.

CROSSBOARD is a prototype display that combines a large dense visual information display with hierarchical audio and/or haptic cueing of the location of an individual's information on the display. Structured flashing of regions of the board, coordinated with a sequence of audio and/or haptic crossmodal cues allow CROSSBOARD users to rapidly identify the region of the board that they should search for information (figure 3). For example, in the first two seconds of a display sequence one audio cue will sound in one of four 500ms that four different regions of the display flashes, this indicates a spatial subdivision of the display in which the user's information resides. In the next two seconds a second cue (and coordinated flash) indicates a sub-region of the first subdivision further localizing the information. Depending on the quantity of information (number of elements) on the display, further subdivisions may be required (figure 4).

Hierarchical crossmodal cueing greatly improves the performance of CROSSBOARD users, by highlighting areas of interest, without impacting on the performance of those using

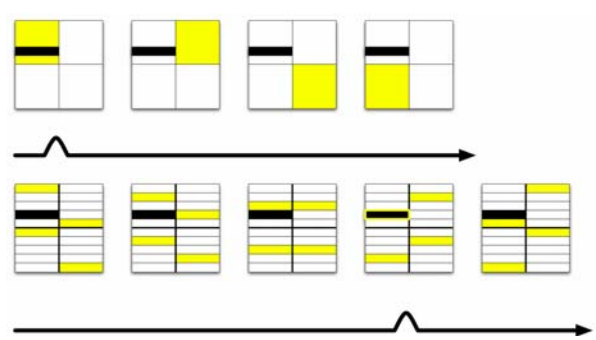

Figure 4. The coordination of cues with multiple subdivision.

the display in a traditional manner. This can be useful for

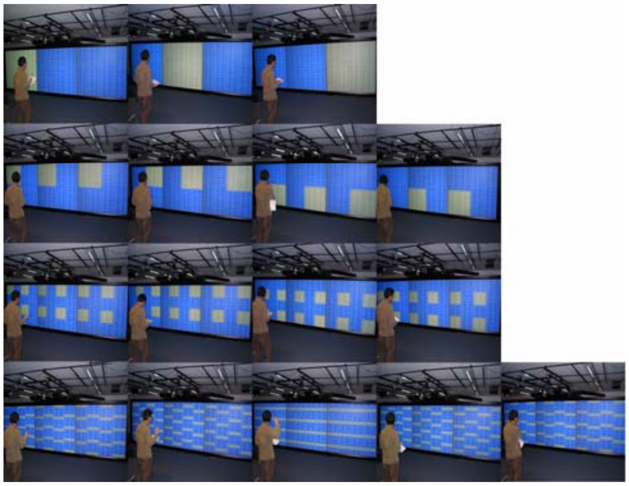

Figure 5. Subdivision of the display.

Figure 6. Display set-up, using a 3-screen CADWall to simulate the physical characteristics of a very large information display comprising details of 240 flight departures.

accessing indexed information (by narrowing the search target) and unindexed information (by indicating an area for direct search). 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.

V. User Studies

A. CrossFlow User Study

Both CrossFlow and CrossBoard were evaluated in pilot studies to assess their potential to support users navigating in indoor environments (CrossFlow) and selectively retrieving information from public displays (CrossBoard). In the navigation case users performance on a small scale navigation tasks was assessed under two conditions: (1) using a traditional map; and (2) using CrossFlow. To examine the peripheral characteristics of the display the study was conducted in a dual task setting in which the primary task involved users answering mental arithmetic questions posed by the experimenter.

Two aspects of the performance on the primary task were considered: (1) the time taken to answer arithmetic questions; and (2) the percentage of correctly answered questions. The average time taken to answer an arithmetic question in the dual-task condition decreased very significantly from using the map to using CrossFlow, with the mean time using CrossFlow being 28% quicker than when using the map. The difference of the accuracy of processing arithmetic questions was only marginally significant between the map and CrossFlow conditions.

The performance on the secondary (navigation) task was compared for the map and CrossFlow condition according to the total time spent finding 5 destinations (total time in the dual-task condition for which it was discovered that the total time spent on the whole experiment in dual-task condition decreased significantly from using the map to using CrossFlow. Finally, the subjective judgments of users was assessed in each of the two conditions using the NASA TLX rating of mental workload. The results show a significant reduction in the perceived mental workload when using CrossFlow as compared to the map

B. CrossBoard User Study

CrossBoard was evaluated in three conditions: (1) visual and crossmodal cues (CrossBoard condition, see figure 5); (2) no visual or crossmodal cues (traditional display board, see figure 6); (3) visual cues only. Users performed two tasks: (1) retrieval of indexed information; and (2) searching for un-indexed information.

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. No significant difference across the three conditions for the indexed information retrieval. However, as anticipated, the performance for unindexed information retrieval was significantly higher in the CrossBoard condition.

VI. Infrastructures for Crossmodal Displays

The two prototypes (CrossFlow and CrossBoard) address two tasks – navigation and information retrieval. They can be viewed as components of a wider crossmodal infrastructure that could be deployed in a space to provided contextualized navigation and information to users within that space. In a basic configuration, these components are deployed in a static manner, with flow-based navigation projection within the space, and crossmodal information displays at appropriate points. A user's interaction is then simply a matter of allocating the correct time slice for notification.

In a scenario where users have networked devices, contextualized information and directions can be sent directly to these devices, but crossmodal techniques provide distinct advantages. Firstly, navigation directions can be directly embedded in the environment, without users having to direct their attention to a separate device in their hand. For information retrieval, using external displays gives a much larger area to display information on, which can be situated contextually in an environment (e.g., times for a specific gate/platform). External displays have the disadvantage that information can be seen by all users, but there are ways to mitigate this. In addition, even on a large display, if enough distinct items of information need to be displayed, more detail per item could be presented on a user-carried device. In some situations, seeing more information than that which is directly relevant to you as a user can be useful, for instance seeing what occurs before and after the event you are interested in, or "the next available train/flight" to a destination.

A. Cue Delivery

Although neither CROSSFLOW nor CROSSBOARD utilize a user's personal computing device to deliver explicit information, it is the means by which the contextual cue is given to the user. This may be a sound, or a vibration, or some other sensory cue. If the device is networked, then it can co-ordinate its time value with the crossmodal system, and be told a specific time to trigger the crossmodal cue. However, one of the aims of the crossmodal system is to deliver cues independent of an external network, to support a wider range of devices, and scenarios where constant network availability is infeasible (e.g., wireless dead spots, lack of GSM signal, restricted environments, cost of connectivity). In this case, the device gives a crossmodal cue at a pre-defined interval, and this interval is synchronized with the rest of the crossmodal system.

This implies two main problems to be solved for a particular crossmodal infrastructure: how to calculate appropriate time intervals, and how to synchronize the cue device with the system. As explained earlier, CROSSFLOW uses linear time slicing, where a set number of destinations are cycled through regularly, implying that for n destinations and a wait time of Δt that the time between cues is $n\Delta t$. There will be a lower limit on Δt by which a user is able to distinguish between different locations; they must be allowed time to react to a cue and associate the cue with the current destination before the destination changes. Long times between cues will mean the user has to interrupt their activity (typically walking) and attend more closely to the cues identify the direction of the destination. In addition, the more destinations there are, the harder it is to distinguish between them. This is analogous to road signs, where a few key destinations and roads are marked on a small number of signs, but as the locations of a particular sign are approached, more destinations and directions are displayed.

CROSSBOARD uses a hierarchical decomposition of a display to narrow down the portion of relevance, so for a decomposition of n levels, the user receives n cues in a particular cycle. The user may not need all cues to identify their information if they have other cross-indexing information. For each level, there are $n(i)$ divisions, and therefore $n(i)$ cues. If the wait between cues is again dt then a single level will last for an amount of time equal to $n(i)\Delta t + \Delta c$ (where Δc is a wait between levels, that may be different to the wait between cues in a single level). So, for a complete cycle, the time is $n(i)\Delta t + \Delta c + C$ (where C is the wait between cycles). For a large number of items, this is less overall cycle time compared to a linear division, but more cues. However, a user does not necessarily need all cues to locate information as all the information is displayed at once, and the user can use cross-referencing information to help location.

Synchronizing the cues with the crossmodal system depends on the network connectivity of the cue-delivery device. As mentioned before, if the device is constantly connected to a network, the system can deliver cues synchronously to the device. However, the usual case is that the device will partially or totally be disconnected from the system's network. Therefore, the device must be able to run asynchronously. This is achieved by having the cycle times programmed into the device, along with the cycle division to be cued. The beginning of the cycle must then by synchronized with the system. Partially-connected devices can retrieve all of this information from the network when they are connected, and update it whenever they re-connect, to keep the cues in synch.

For disconnected devices, the cycles times must be entered directly on the device, and synchronization performed manually. One way to do this is to have a system-generated start-of-cycle cue displayed somewhere, and when this is seen, the device's cycle is started. Care must be taken to factor in user-reaction time as with any other cue. A semi-automatic way of recognizing the cue could be achieved by using the device's microphone or camera to pick up the cue.

B. Co-ordination and Contextualization

Integrating navigation and display boards into a larger infrastructure requires contextual information. The system needs to know which navigation list the user is following or which board they are attending to. If they switch between board and navigation than the cue list must be updated.

By placing displays at the locations in a navigation list, choice of a navigation destination informs the system of the location of the board the user is interested in. Synchronization can be managed across displays and navigation, so all that needs to be done is to tell the system when the user reaches the board. When the user informs the device they are at their destination, the cues are switched to that for the board at that location or the next navigation cues. The device could be programmed with the cues for all appropriate displays, and then user goes through the synchronization step when they reach a display, such as manually entering the display ID or using a camera, microphone or network connection to encode the display details.

Once the system knows which display a user is attending to, information can be further contextualized. That board can add extra details for users it knows are attending. This can be used as a limited form of privacy to ensure information is only displayed when needed. Critically private information can be stored on the user's device which displays the augmented information when attending to a display. As well as other users, consideration must be given to non-users who still are using a display board. The initial CROSSBOARD experiment showed that display cues did not impair a non-cued user's performance, but they anecdotally complained that it was distracting.

The final problem of coordination, is when a display updates timely information. If the information is sorted on the board, this will affect the portion of the display the user's information is in, and hence the cues that apply to that display for that user. Thus, either the cues for that user need to be updated, the board arranged according to cues, not ordered, or better synchronization schemes employed.

VII. CONCLUSION

Crossmodal displays offer an alternative paradigm for the display of public-private information in which the aspects of crossmodal cognition are leverages to provide personalized cues to the occupants of the space in which the displays are situated. We have demonstrated that such displays have significant potential benefits over standard configurations in terms of their situated characteristics, user performances, and the potential to deliver such displays using an infrastructure that maintains user privacy.

ACKNOWLEDGMENT

We would like to thank Andrew Monk (University of York) for kindly providing the mobile phones used in these studies (these were in turn donated by Microsoft Research Cambridge). We would also like to acknowledge Antonio Krüger (University of Münster) with whom we conceived the original notion of using crossmodal displays for indoor navigation.

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

[9] 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.

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

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

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

[13] 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

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

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