

# Supporting Cross-Modal Collaboration in the Workplace

Oussama Metatla, Nick Bryan-Kinns  
School of Electronic Engineering and Computer Science  
Queen Mary University of London  
Mile End road, London, E1 4NS, UK  
{oussama, nickbk}@eecs.qmul.ac.uk

Tony Stockman, Fiore Martin  
School of Electronic Engineering and Computer Science  
Queen Mary University of London  
Mile End road, London, E1 4NS, UK  
{oussama, nickbk}@eecs.qmul.ac.uk

**We address the challenge of supporting collaborators who access a shared interactive space through different sets of modalities. This was achieved by designing a cross-modal tool combining a visual diagram editor with auditory and haptic views to allow simultaneous visual and non-visual interaction. The tool was deployed in various workplaces where visually-impaired and sighted coworkers access and edit diagrams as part of their daily jobs. We use our observations and analyses of the recorded interactions to outline preliminary design recommendations for supporting cross-modal collaboration.**

*Collaboration; cross-modal interaction; accessibility; inclusion; haptics; auditory interfaces; workplace.*

## 1. INTRODUCTION

Every day our brains receive and combine information from different senses to understand our environment. For instance when we both see and hear someone speaking we associate the words spoken with the speaker. The process of coordinating information received through multiple senses is fundamental to human perception and is known as cross-modal interaction (Drive and Spencer 1998). In the design of interactive systems, the phrase *cross-modal interaction* has also been used to refer to situations where individuals interact with each other while accessing the same interactive shared space through different senses, e.g. (Winberg and Bowers 2004) and (Metatla *et al.* 2011b)). This is different to typical multimodal collaborations such as audio-video conferencing or shared whiteboards where it is assumed that all collaborators rely on the same set of senses to participate in the shared activity (Cherubini *et al.* 2007). Technological developments mean that it is feasible to support cross-modal interaction in a range of devices and environments, yet there are no practical examples of such systems. This becomes problematic when collaborators have access to differing sets of modalities due to situational or permanent sensory impairment; e.g. Apple's iPhone provides touch, visual, and speech interaction, but there is no easy way for sighted and visually impaired people to collaborate beyond a vocal conversation.

We are particularly interested in exploring the potential of cross-modal interaction to improve the accessibility of collaborative activities involving the use of diagrams. Diagrams are a key form of representation used in all manner of collaborations. Indeed, diagrammatic representations have often become common standards for expressing specialised aspects of a particular discipline, e.g. meteorologists use weather maps, architects use floor plans, and computer scientists make extensive use of nodes-and-links diagrams. However, there is currently no practical way for visually impaired co-workers to view, let alone edit, diagrams. This is a major barrier to workplace collaboration that contributes to the exclusion and disengagement of visually impaired individuals. An RNIB report, for instance, estimates that 66% of blind and partially sighted people in the UK are currently unemployed (RNIB 2009). Addressing the challenge of designing support for cross-modal collaboration in the workplace has thus the potential to significantly improve the working lives and inclusion of perceptually impaired workers.

## 2. BACKGROUND

As technology improves, the inclusion of high fidelity auditory and haptic displays in digital devices is becoming commonplace. Auditory displays make use of speech and non-speech sounds to convey information (Kramer 1994) and are typically used to draw attention to activities outside of the

field of view, or to provide additional information in situations where the eyes are occupied or there is limited screen space. To date, auditory interfaces have been successfully employed in a variety of areas including monitoring applications for complex environments, such as operating rooms and aircraft flight decks, improving accessibility to visually represented information, and supporting data exploration through sonification (Hermann 2002). Haptic and tactile displays on the other hand, are interfaces that convey information through cutaneous or kinesthetic sensation. They allow visually represented objects to be augmented with rich physical properties, such as mass and textures, and can be used to simulate most physical sensations that can be mathematically represented, such as gravitational fields (Kortum 2008). This is usually achieved by using vibrating or robotic devices to convey haptic sensations, allowing a user to perform physical manipulations like pulling, pushing and feeling objects. Research has produced a variety of techniques for conveying information through haptic and tactile feedback. Tactons, for instance, are a form of structured tactile signals that can be used to convey abstract messages non-visually and are equivalent to visual icons and audio earcons (Brewster and Brown 2004).

## 2.1. Non-visual Interaction with Diagrams

Interest in supporting non-visual access to visually represented information grew in parallel with early developments in Auditory Display research (Kramer 1994). A major drive of such endeavours has been and still is the potential to support individuals with temporary or permanent perceptual impairments. For example, (Mansur *et al.* 1985) pioneered a sonification technique to display a line graph in audio by mapping its y-values to the pitch of an acoustic tone and its x-values to time. This approach to using sonification allows visually impaired individuals to examine data presented in line graphs and tables.

Current approaches to supporting non-visual interaction with visual displays employ one or a combination of two distinct models of representation; *Spatial* or *Hierarchical*. The two models differ in the degree to which they maintain the original representation when translating its visual content (Mynatt and Weber 1994), and hence produce dramatically different non-visual interactive displays.

### 2.1.1. Spatial Models

A spatial model allows non-visual access to a visual display by capturing the spatial properties of its content, such as layout, form and arrangements. These are preserved and projected over a virtual or a physical space so that they could be accessed through alternative modalities. Because audio has

limited spatial resolution (Best *et al.* 2003), spatial models typically combine the haptic and audio modalities to support interaction. The GUIB project (Weber 1993) is one of the early prototypes that employed a spatial model of representation to support non-visual interaction with a visual display. The prototype combines braille displays, a touch sensitive tablet and loudspeakers to allow blind users to interact with MS Windows and X Windows graphical environments. More recent solutions adopting the spatial model of representation typically use tablet PC interfaces or tactile pads as a 2D projection space where captured elements of a visual display are laid out in a similar way to their original arrangements. Other solutions use force feedback devices as a controller. In such instances, the components of a visual display are spatially arranged on a virtual rather than a physical plane, and can thus be explored and probed using a haptic device such as a PHANTOM Omni device <sup>1</sup>. The advantage of using a virtual display lies in the ability to add further haptic representational dimensions to the captured information, such as texture and stiffness, which can enhance the representation of data. The virtual haptic display can also be augmented and modulated with auditory cues to further enhance the interactive experience (Yu *et al.* 2003).

### 2.1.2. Hierarchical Models

A hierarchical model, on the other hand, preserves the semantic properties of visual displays and presents them by ordering their contents in terms of groupings and parent-child relationships. Many auditory interfaces are based on such a model as they inherently lend themselves to hierarchical organisation. For instance, phone-based interfaces support interaction by presenting the user with embedded choices (Leplatre and Brewster 2000). Audio is therefore the typical candidate modality for non-visual interaction with visual displays when using hierarchies. One of the early examples that used a hierarchical model to translate visual displays into a non-visually accessible representation is the Mercator project (Mynatt and Weber 1994). Like the GUIB project, the goal of Mercator was to provide non-visual access to X Windows applications by organising the components of a graphical display based on their functional and causal properties rather than their spatial pixel-by-pixel on-screen representations. Other examples have employed a hierarchical model of representation to support non-visual interaction with technical drawings (Horstmann *et al.* 2004), UML (Metatla *et al.* 2008) and molecular diagrams (Brown *et al.* 2004).

<sup>1</sup> Sensable Technologies, <http://www.sensable.com>

## 2.2. Cross-modal Collaboration

Despite significant progress in the use of audio and haptics in multimodal interaction design, research into cross-modal collaboration remains sparse. In particular, very little research has addressed the challenge of supporting collaboration between visually-impaired and sighted users. Nonetheless, initial investigations have identified a number of issues that impacts the efficiency of collaboration in a multimodal interactive environment. An examination of collaboration between sighted and blind individuals on the Tower of Hanoi game (Winberg and Bowers 2004), for instance, highlighted the importance of providing visually-impaired collaborators with a continuous display of the status of the shared game. Providing collaborators with independent views of the shared space, rather than shared cursor control, was also found to improve orientation, engagement and coordination in shared tasks. A multimodal system combining two PHANTOM Omni haptic devices with speech and non-speech auditory output was used to examine collaboration between pairs of visually impaired users (McGookin and Brewster 2007) and showed that the use of haptic mechanisms for monitoring activities and shared audio output improves communication and promotes collaboration. Still, there are currently no studies of collaborations between visually-impaired and sighted coworkers. We therefore know little about the nature of cross-modal collaboration in the workplace and ways to support it through interface design.

## 3. DESIGNING A COLLABORATIVE CROSS-MODAL TOOL

To address the issues identified above we gathered requirements and feedback from potential users to inform the design process. We ran a workshop to engage with representatives from end user groups in order to encourage discussion and sharing of experiences with using diagrams in the workplace. Eight participants attended the workshop including participants from BT and the Royal Bank of Scotland and representatives from the British Computer Association of the Blind and the Royal National Institute for the Blind. Activities ranged from round table discussions exploring how participants encounter diagrams in their workplaces, to hands-on demonstrations of early audio and haptic prototype diagramming systems. The discussions highlighted the diversity of diagrams encountered by the participants in their daily jobs; from design diagrams for databases and networks, to business model diagrams, and organisation and flow charts. Additionally, participants discussed the various means they currently use for accessing diagrams and their limitations. Approaches included using the

help of a human reader, swell paper, transcriptions and stationary-based diagrams, all of which share two main limitations; the inability to create and edit diagrams autonomously, and inefficiency of use when collaborating with sighted colleagues.

We chose to focus on nodes-and-links diagrams because they are frequently encountered in the workplace and we already have evaluated a single user version for audio-only interaction with such diagrams (Metatla *et al.* 2011a). Our cross-modal tool <sup>2</sup> supports autonomous non-visual editing of diagrams as well as real-time collaboration. It allows simultaneous access to a shared diagram by augmenting a graphical display with non-visual auditory and haptic views combining hierarchical and spatial models of representation. The tool supports user-defined diagram templates which allows it to accommodate various types of nodes-and-links diagrams such as organisation and flow charts, UML and database diagrams and transport maps.

### 3.1. Graphical View

Figure 1 shows a screenshot of the graphical view of the tool. This view presents the user with an interface similar to typical diagram editors where a toolbar is provided containing various functions to create and edit diagram content. The user construct diagrams by using the mouse to select the desired editing function and has the ability to access and edit various object parameters such as labels, position, etc.

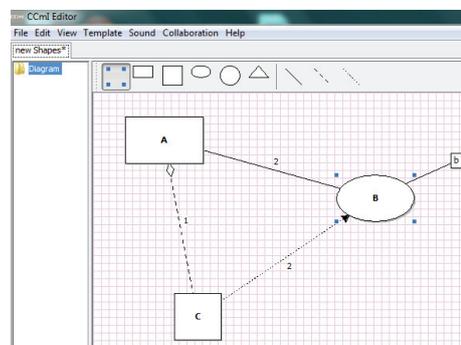
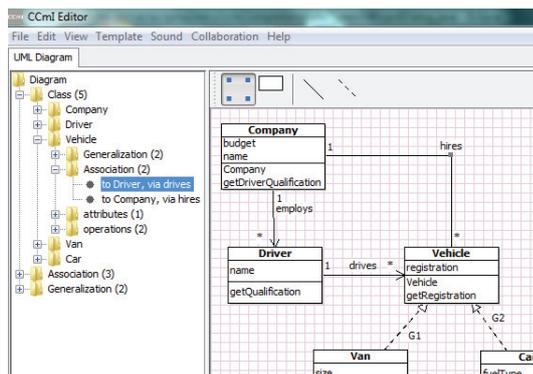


Figure 1: Graphical view of the cross-modal diagram editor.

### 3.2. Hierarchical Auditory View

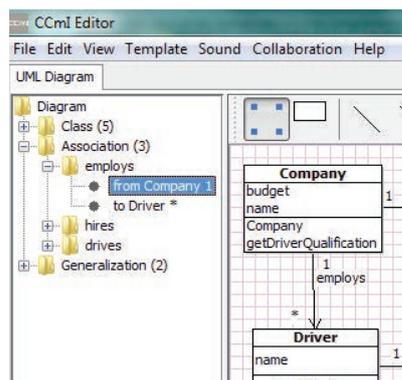
The design of the auditory view is based on the multiple perspective hierarchical approach described in (Metatla *et al.* 2011a). According to this approach, a diagram can be translated from a graphical to an auditory form by extracting and structuring its content in a tree-like form such that items of a similar type are grouped together under a dedicated branch on a hierarchy. This is aimed to ease inspection, search and orientation [ibid.].

<sup>2</sup> An open source release of the tool can be downloaded from: <http://ccmi.eecs.qmul.ac.uk/downloads>



**Figure 2:** Auditory hierarchical view (left) embedded in the editor.

Figure 2 shows how this is achieved for a UML Class diagram. In this case, the diagram’s classes – represented as rectangular shapes – are listed under the “Class” branch of the hierarchy. The information associated with each class, such as its attributes, operations and connections to other classes, is nested inside its tree node and can be accessed individually by expanding and inspecting the appropriate branches. Similarly, the diagram’s associations – represented as solid arrows – are listed under the “Association” branch, and information associated with each connection can be accessed individually by inspecting its branches (see Figure 3). This allows the user to access the information encoded in a diagram from the perspectives of its “Classes”, “Associations” or its “Generalisations”. To inspect the content of a diagram, the user simply explores the hierarchy using the cursor keys, similar to typical file explorers, and receives auditory feedback displaying the content that they encounter.



**Figure 3:** Hierarchical auditory view (left) where a Class diagram is accessed from the perspective of its associations.

We use a combination of speech and non-speech sounds to display encountered content as follows: The successful movement from one node to another is conveyed by displaying the text label of the node in speech together with an earcon in the form of a

single tone with a distinct timbre assigned to each type of item. This is displayed as the sequence (Tone) + “<node name>”. The same technique is used to highlight reaching the end or the top of a list, but in such a case a double beep tone is used instead of a single tone, and is displayed as the sequence (Double beep) + “<node name>”, in which case the user is looped to the other end of the list. The successful expansion or collapse of a branch is displayed using earcons. An *Expand* earcon mixes frequency and amplitude modulation on a basic pulse oscillator to produce a sweep that ends with a bell like sound. A *Collapse* earcon is composed from the reversed sequence of the *Expand* earcon (e.g. “Associations” + (Expand sound) + “Associations” for expanding the Associations branch, and (Collapse sound) + “Associations” for collapsing it). Additionally, when a branch is expanded, a speech output is displayed to describe the number of items it contains (e.g. “Associations” + (Expand sound)+“three” to convey that the diagram contains three associations). In addition to inspecting a given diagram, the hierarchy can also be used to edit its content. To do this, the user first locates the item of interest on the hierarchy before executing a particular editing action that alters its state. For example, to remove a class from the diagram, the user would inspect the appropriate path to locate it on the hierarchy then, once found, issue the command using the keyboard to delete it. The tool then interprets the current position of the user on the hierarchy together with the issued command as one complete editing expression and executes it appropriately. The auditory hierarchical view is thoroughly described and evaluated in (Metatla et al. 2008) and (Metatla et al. 2011a).

### 3.3. Spatial Haptic View

In addition to the auditory hierarchical view, we implemented a spatial model of representation to capture the layout and spatial arrangements of diagrams content. To do this, we use a PHANTOM Omni haptic device (Figure 4) to display the content of a diagram on a virtual vertical plane matching its graphical view on a computer screen (Figure 5). We designed a number of haptic effects to both represent the content of a diagram and support non-visual interaction in this view.

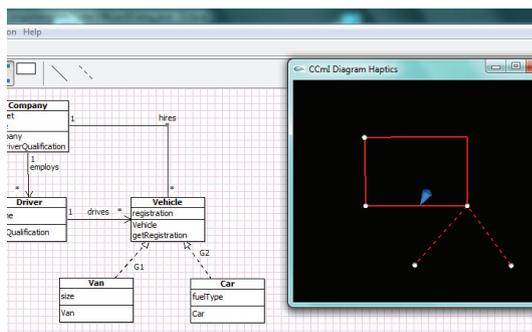
#### 3.3.1. Haptic Representation

The main haptic effect that we use to represent diagrams nodes and links is attraction force. Diagram nodes are rendered as magnetic points on the virtual plane such that a user manipulating the stylus of the PHANTOM device in proximity of a node is attracted to it through a simulated magnetic force. This is augmented with an auditory earcon (of a similar timbre to the single tone earcon used in the auditory view) which is triggered upon contact

with the node. A similar magnetic effect is used for the links with the addition of a friction effect that simulates a different texture for solid, dotted and dashed lines. The user can thus trace the stylus across a line without deviating away to other parts of the plane while feeling the roughness of the line being traced, which increases from smooth for solid lines to medium and very rough for dotted and dashed lines respectively. Contact with links is also accompanied by earcons with distinct timbres, and the labels of encountered nodes and links are also displayed in synthesised speech upon contact.



**Figure 4:** Interacting with the spatial haptic view using the stylus of a PHANTOM Omni haptic device.



**Figure 5:** Spatial haptic view (right) matching the physical layout of the diagram on the graphical view.

### 3.3.2. Haptic Interaction

In addition to representing diagram content using various haptic effects, we implemented two modes of interaction in the spatial haptic view which we refer to as *sticky* and *loose*. In a sticky mode of interaction, the magnetic attraction forces of the diagrams nodes and links are increased to make it harder for the user to snap away from contact with a given item on the diagram. This simulates an impression of being “stuck” to the diagram content and thus one can trace its content by following the connections from point to point. In a loose mode of interaction on the other hand, the magnetic attraction forces are decreased such that a user can freely move around the virtual space without necessarily having to be in contact with any diagram content – in which case the haptic force is set to neutral and no auditory feedback is displayed. Additionally, the user has the ability to move nodes and bend links in space. This can be achieved by locating an item – or a point on a link – on the virtual plane, clicking on the stylus button

to pick it up, dragging the stylus to another point on the plane, then dropping it in a new desired location with a second button click. We designed two extra features to support this drag-and-drop action. First, three distinct auditory icons are used to highlight that an item has been successfully picked up (a short sucking sound), that it is being successfully dragged in space (a continuous chain-like sound) and that it has been successfully dropped in the new location (a dart hitting a dartboard sound). Second, a haptic spring effect is applied, linking the current position of the stylus to the original position of where the item was picked up from. This allows the user to easily relocate the item to its original position without losing orientation on the plane. Once an item is picked up, the user is automatically switched to the loose mode of interaction to allow for free movement while still able to inspect encountered items. Finally, we implemented a synchronisation mechanism to allow the user to switch between the haptic and auditory hierarchical views of the diagrams. The user can locate an item on the hierarchy then issue a command on the keyboard which would cause the PHANTOM arm to move and locate that item on the haptic plane. If the user is holding the stylus, they are then dragged to that location. Similarly, the user can locate an item on the virtual haptic plane then issue a command on the keyboard to locate it on the hierarchy.

### 3.4. Collaborative Interaction

The cross-modal tool runs across-platforms on any computer with a Java Runtime Environment. Simultaneous shared access to a diagram is currently achieved by connecting collaborators’ computers through a local network with one of the computers acting as a server. We have incorporated locking mechanisms which prevents collaborators from concurrently editing the same item on the diagram. Besides these locking mechanisms, the tool does not include any built-in mechanisms to regulate collaboration, such as process controls that enforce a specific order or structure of interaction. This was done to allow users to develop their own collaborative process when constructing diagrams – indeed, there is evidence that imposed structure can increase performance but at the expense of hindering the pace of collaboration and decreasing consensus and satisfaction amongst group members (Olson *et al.* 1993). Thus, the cross-modal tool provides collaborators with independent views and unstructured simultaneous access to shared diagrams.

## 4. EVALUATIONS IN THE WILD

We conducted a study of cross-modal collaboration between visually-impaired and sighted coworkers.

The aim was to explore the nature of cross-modal collaboration in the workplace and assess how well the tool we designed supports it in real-world scenarios. So far, we have deployed the tool to support the work of three professional pairs; these were employees in the head office of a London-based Children and Families Department in local government, an international charity and a private business company.

#### 4.1. Approach & Setup

We first asked pairs to provide us with samples of the type of diagrams that they encounter in their daily jobs. We then created appropriate templates to accommodate these diagrams on the cross-modal tool. Because we wanted to observe the use of the cross-modal tool in real-world scenarios, involving diagrams of real-world complexity, we did not control the type of tasks that the pairs performed nor the way in which they went about performing them. Rather, we deployed the tool in their workplaces and observed their collaborations as they naturally unfolded over a working session. Study sessions lasted for up to two hours, where we introduced the pairs to the features and functionalities of the tool in the first half, then observed them as they used it to access and edit diagrams in the second half.

Visually impaired participants used the audio-haptic views of the diagrams, where audio was displayed through speakers so that their colleagues could hear what they were doing, while the sighted participant used the graphical view of the tool. In all three cases, the pairs sat in a way that prevented the sighted participant from seeing the screen of their colleagues (see Figures 6), and, naturally, the visually-impaired participants did not have access to the graphical view of their partners. We video recorded all sessions and conducted informal interviews with the pairs at the end of the working sessions.



**Figure 6:** An example of the setup used in the workplace of one pair who participated in the study.

#### 4.2. Collaborative Scenarios

We observed two types of collaborative scenarios. In the first scenario, a manager and their assistant, accessed and edited organisation charts to reflect recent changes in managerial structures. In the second and third scenarios, a manager and an

employee assistant and two business partners inspected and edited transportation maps in order to organise a trip.

### 5. OBSERVATIONS & DESIGN LESSONS

All pairs were able to complete the tasks that they chose to undertake using the cross-modal tool. In the following, we focus on aspects of the cross-modal collaborative interaction rather than on the multimodal representation of diagrams. The collaborations that we observed evolved over three distinct phases with differing dynamics of interaction. A first instance is characterised as being driven by the visually-impaired user and includes exploring the diagram, editing its content and altering its spatial arrangements. The sighted coworker in this instance engages in discussions about the diagram and providing general guidance about where things are located and how to get to them. In a second instance of the collaborations, the visually-impaired user continues to drive the interaction with active input from the sighted user who engages in refining the content and spatial arrangements produced by their coworker. In a third instance, both users engage in manipulating the diagram, working independently on different parts of its content while continuing to discuss the task and updating each other about their progress. These dynamics do not necessarily occur in a particular order. For instance, it is likely that the first instance results from the visually impaired desire to establish orientation within the interactive space at the onset of the collaboration, which might be unnecessary for the sighted user, but such reorientation might occur again after a diagram's content has been extensively altered.

Due to the nature of the study – a small number of participants and uncontrolled real world workplace environments – we opted for conducting a qualitative analysis of the recorded interactions rather than attempt to capture quantitative aspects of the collaborations. We present a series of excerpts from the video transcripts to highlight the impact of the cross-modal tool on the collaborations and use these examples to outline a set of design recommendations. Since the constructed diagrams were the property of the organisations that we worked with, we deliberately edited out some content and/or concealed it on the transcripts due to the sensitive nature of the information they contain.

#### 5.1. Exploring and Discussing Diagram Content

In the excerpt shown in Table 1, the pair are editing an itinerary on a transport map. The excerpt starts off with the visually impaired user (VI) locating and deleting a node from the diagram while the sighted

user (S) edits the label of another node. As soon as the node is deleted, S interrupts VI to inform them about the visible changes that resulted from their action: *“you didn’t just delete the node[.]”*. Here the VI user was not aware that deleting a node caused the automatic deletion of the links that were coming in and out of it. The VI user responds with an exclamatory *“yeah?”* while manipulating the haptic device in an attempt to explore the parts of the diagram where the declared changes are said to have occurred. Meanwhile S continues to reason about the outcome of their partner’s action: *“we can recreate the .. part of it needed to be deleted anyway”* while the VI user switches to the audio view to check the diagram, correctly deducing the status of its nodes: *“so it only deleted one node..”*.

What we wish to highlight with this excerpt is the way in which the auditory and haptic views were used in the exchange that occurred between the two colleagues. The VI user was able to seamlessly integrate the discussion about the diagram with their partner with the inspection and exploration of the its content. Here, the cross-modal tool formed an effective part of the collaborative exchange; that is, just as S was able to glance at the diagram while discussing and reasoning about its content, so was the VI user able to access and explore the diagram while actively partaking in the discussion.

Recommendation 1 – Provide explicit representation of the effects produced by a given action to its original author. While the sighted user was able to detect the results of an action as they disappeared from the screen, the original author was completely oblivious of this information. It is therefore recommended to explicitly convey the consequences of an action to its original author. This could also be conveyed in the form of a warning before finalising the executing of an action.

It is important to note that while Recommendation 1 echos well-known usability heuristics, the feedback provided needs to be robust with respect to interaction modes. That is, it needs to convey the result of the action independently of which combination of modalities the user is using, and – due to the fact the user is collaborating – this feedback needs to be provided as close as possible in the same time frame as that in which the result of the action becomes clear to users employing other modes if potential misunderstandings/confusion are to be avoided.

## 5.2. Providing Directional Guidance

There were instances in the collaborations where the sighted user provided directional guidance to their partner while they were executing a given editing action. An example of this is shown in the

excerpt in Table 2. Here, the pair are editing an organisation chart and the visually impaired user attempts to locate a node on the diagram using the Omni haptic device. The excerpt begins with the VI user moving the omni device to locate the node in question, encountering an unexpected node X and announcing: *“I got X”*. The sighted user then uses this information to provide their colleague with relevant directions: *“then go diagonal left”*. The VI user attempts to follow their colleague’s guidance but, failing to go in the specified direction, seeks more clarifications: *“diagonally up or down?”, “from Y or from X?”*. Moving around the haptic plan, the VI user encounters another item on the diagram; a link labelled Z. The sighted user again picks up on the audio triggered by their partner to tailor the guidance they provide them with: *“that’s the right link, follow Z”*. This tailored guidance helps the VI user to locate the node in question. The fact that the audio output was shared amongst the pair helped the sighted user to engage with their partner’s activity. The overlap in presentation modalities in this case created more opportunities for interaction. Information displayed in audio allowed the sighted user to keep track of their partner’s progress and, by referring to the graphical view, they were able to map such information and tailor their own discourse to match such progress.

## 5.3. Transitions Between Collaborative Tasks

The next excerpt, shown in Table 3, shows an example where collaborators executed two dependent actions sequentially. The VI user’s task was to create a link between two nodes on the diagram. To achieve this, the VI user first locates the two nodes in question, selects them, then issues a command to create a connection between them. The sighted user’s task was to arrange the spatial position of the newly created connection. What is noticeable in this excerpt is that the sighted user was able to determine the exact point in the execution where they were required to take action without being explicitly prompted by their partner: *“alright so I’m gonna move that now”*. Here again, having access to their partner’s audio output allowed the sighted user to keep track of their partner’s progress resulting in a seemingly effortless transition between the two dependent actions. Thus, allowing an overlap of presentation modalities helps users to structure sequentially dependent actions.

Recommendation 2 – Allow an overlap of presentation modalities to increase opportunities for users to engage with each other’s actions during the collaboration.

## 5.4. Executing a Spatial Task

A major advantage of using a spatial model of representation to support non-visual interaction with

**Table 1: Extract 1 - Smooth embedding of interaction with device and discussion about content.**

Visually-impaired user	VI actions/audio output	Sighted user	S actions
	<locates node> <deletes node>		<edits node label>
OK, so now I need to			
what?	<moves the omni>	hold on a second	
yeah?	<moves the omni> <moves the omni> <moves the omni>	you didn't just delete the node but also every line that was coming in and out of it we can recreate the ... part of it needed to be deleted anyway but one didn't	
but that segment had to be removed didn't it? let me just .. can i just look for a sec so it only deleted one node..	<explores audio view>  <explores audio view >		
		yeah, but every single line ..	

**Table 2: Extract 2 - Directional guidance.**

Visually-impaired user	VI actions/audio output	Sighted user	S actions
	<moves the omni to locate a node W> <encounters a node X>		
I've got X			
doesn't let me go left it's literally stopping me from going left diagonally up or down? from Y or from X?	<moves the omni to the left > <moves the omni to the left > <moves the omni> <moves the omni> <moves the omni upwards > <moves the omni> <moves omni to relocate X> <system speaks: "Z">	then go diagonal left up left  up  from X	
yeah I'm on ..	<follows Z > <locates node W >	that's the right link, follow Z	

**Table 3: Extract 3 - Smooth transition between actions.**

Visually-impaired user	VI actions/audio output	Sighted user	S actions
	<explores the auditory hierarchy> <locates node X and selects it> <explores the auditory hierarchy> <locates node Y and selects it> <creates a link between X and Y>		
yup	<System confirms the creation of a new link>	alright so I'm gonna move that now	<selects node X and drags it>

**Table 4: Extract 4 - Executing a spatial task.**

Visually-impaired user	VI actions/audio output	Sighted user	S actions
OK, shall we try the others	<moves the omni towards a node>	yup	
yes, X got ya	<locates a node X> <picks up the node> <drags X downwards> <drags X downwards>		
I'm gonna put it down here somewhere What do you reckon? I'm gonna put it here What do you think?	<drops X>	I can't see where you're pointing, drop it first  that is again on the same level as the Y	

**Table 5: Extract 5 - Shared locus.**

VI actions/audio output	S actions
<edits the label of node X> <types new label for X>	<Hovers mouse over node X> <drags X to a new location>
<explores X on the auditory hierarchy> <explores X on the auditory hierarchy>	<drags X to another location>
<synchronises the audio and haptic views to the location of X>	

**Table 6:** Extract 6 - Exchanging updates.

Visually-impaired user	VI actions/audio output	Sighted user	S actions
	<explores the auditory hierarchy> <creates a new node X>		<edits node Y's parameter> <edits node Y's parameter>
	<explores the auditory hierarchy> <selects node X on the hierarchy>	so I'm going though Y and Z just adding their details	<edits node Y's parameter> <edits node Z's parameter>
yeah OK I've created the two ...	<explores the auditory hierarchy>		<edits node Z's parameter>

diagrams is the ability to execute spatial tasks. The visually-impaired users were able to not only add or remove content from the diagram but also engage with their sighted colleagues to alter content's locations on the diagrams. The excerpt in Table 4 shows an example of this. Here, the VI user uses the omni device to locate a node on the diagram, picks up, drags it across the virtual plane and drops it in a new location. Notice how the VI user engages their sighted partner at each step in the execution of this spatial task by supplying cues about what they are doing: "yes, X, got ya", "I'm gonna put it down here somewhere, what do you reckon?". There is therefore a clear attempt by the VI user to use the spatial layout of the diagram as a common reference when negotiating execution steps with their partner. This was indeed a novelty that was well commended by all participants in our study. The sighted user in the excerpt, however, highlights an important point that contributed to his inability to fully engage with their partner to use this common frame of reference: "I can't see where you're pointing, drop it first". Once the VI user drops the node in the new location it appears on the screen of the sighted user, who could then supply the relevant confirmations to their partner: "that is again on the same level as the Y". Because the tool did not provide the users with any explicit representation of their partner's actions – besides final outcomes – it was hard for them to fully engage with each other during execution. In the case of the excerpt on Table 4, the users compensate for this by supplying a continuous stream of updates of what they are about to do.

Recommendation 3 – Provide a continuous representation of partner's actions on the independent view of each user in order to increase their awareness of each other's contributions to the shared space and hence improve the effectiveness of their collaborative exchange.

### 5.5. Shared Locus

The excerpt shown in Table 5 does not involve any conversational exchange. However, the pair's interaction with their independent views of the shared diagrams reveals another way in which the two representations were used as a shared locus. In this excerpt, the VI user has created a new node and

is in the process of editing its label. Meanwhile, the sighted user moves his mouse and hovers over the node that is currently being edited by their partner then drags it to a new location. The interaction in this excerpt enforces recommendation 2. That is, allowing an overlap of presentation between the visual and audio-haptic display modalities allowed the sighted user to identify the part of the diagram being edited by their partner, to follow the editing process, and to seamlessly introduce their own changes to it (in terms of adjusting the location of the node). The VI user in turn, once finished with editing the label of the node, seamlessly synchronises their auditory and haptic views to explore the new location of the node as introduced by their partner. All of this is done smoothly without any need for verbal coordination.

### 5.6. Exchanging Updates

The final excerpt in Table 6 shows a different style of collaborative interaction. Instead of waiting for partners to finish executing an action before proceeding with an another, the pair in this excerpt are working in parallel on two independent actions. The VI user in this case is adding new nodes to the diagram and exploring its content using the auditory hierarchical view, while the sighted user is editing node parameters. The pair are working in parallel and updating each other about the editing actions that they are currently executing: "I'm going through Y and Z just adding their details", "I've created the two..". Each user is therefore engaged with their own task, and unless an update is supplied, the participants remain unaware of each other's progress. Supplying awareness information while both users are jointly engaged with one task is different from supplying it when each one of them is engaged with an independent task. The former, as exemplified in Table 4 was in the form of updates about what the user intends to do, whereas in this excerpt it is in a form of what is currently occurring or what has taken place.

Recommendation 4 – While providing a continuous representation of partner's actions, as outlined in Lesson 3 above, care must be taking to choose the most relevant type of awareness information to provide. This changes in accordance with whether the collaborators are executing independent actions

in parallel, or engaged in the same dependent tasks in sequence.

Although recommendation 4 might appear obvious, how to provide this foregrounding and backgrounding of awareness information is not obvious in a cross-modal context. In the visual modality this might be achieved by changing manipulating the display properties, such as brightness to highlight levels of the awareness situation. In audio one might consider changing amplitude or switching between a normal and a whispering voice to highlight the prominence of the conveyed information.

## 6. CONCLUSION

We presented the design of a collaborative cross-modal tool for editing diagrams which we used to explore the nature of cross-modal collaboration between visually-impaired and sighted users in the workplace. A study that we conducted in the wild with real-world collaborative scenarios allowed us to identify a number of issues related to the impact of cross-modal technology on collaborative work, including coherence of representation, collaborative strategies and support for awareness across modalities. We used our observations to outline a set of preliminary design recommendations aimed at guiding and improving the design of support for cross-modal collaboration.

## 7. REFERENCES

- Best, V., Van Schaik, A., and Carlile, S. (2003) Two-point discrimination in auditory displays. In *Proc. of the 9th Inter. Conf. on Auditory Display*, E. Brazil and B. Shinn-Cunningham, Eds., Boston University Publications Production Department, 17–20.
- Brewster, S., and Brown, L. M. (2004) Tactons: structured tactile messages for non-visual information display. In *Proc. of the fifth conf. on Australasian user interface - Vol28, AUIC'04*, 15–23.
- Brown, A., Pettifer, S., and Stevens, R. (2004) Evaluation of a non-visual molecule browser. In *Assets'04: Proc. of the 6th int. ACM SIGACCESS Conf. on Computers and Accessibility* (Atlanta, GA, USA, 2004), 40–47.
- Cherubini, M., Venolia, G., DeLine, R., and Ko, A. J. (2007) Let's go to the whiteboard: how and why software developers use drawings. In *CHI'07: Proc. of the SIGCHI Conf. on Human factors in computing systems*, 557–566.
- Driver, J., and Spence, C. (1998) Attention and the crossmodal construction of space. *Trends in Cognitive Sciences* 2, 7, 254–262.
- Hermann, T. *Sonification for Exploratory Data Analysis*. PhD thesis, University of Bielefeld, 2002.
- Horstmann M., Lorenz M., Watkowski A., Ioannidis G., Herzog O., King A., Evans D. G., Hagen C., Schlieder C., Burn A. M., King N., Petrie H., Dijkstra S., and Crombie D., (2004) Automated interpretation and accessible presentation of technical diagrams for blind people. In *New Review of Hypermedia and Multimedia*, Vol10, 141–163. Taylor & Francis, ISBN 1361-4568
- Kortum, P. *HCI Beyond the GUI: Design for Haptic, Speech, Olfactory, and Other Nontraditional Interfaces*. Morgan Kaufmann, 2008.
- Kramer, G. *Auditory Display: Sonification, Audification and Auditory Interfaces*. Addison-Wesley Publishing Company, Reading, MA, USA, 1994.
- Leplatre, G., and Brewster, S. (2000) Designing non-speech sounds to support navigation in mobile phone menus. In *Proc. of the 6th Inter. Conf. on Auditory Display*, P. R. Cook, Ed. (Atlanta, GA, USA, 2000).
- Mansur, D. L., Blattner, M. M., and Joy, K. I. (1985) Sound graphs: A numerical data analysis method for the blind. *J. of Medical Systems* 9, 3, 163–174.
- McGookin, D., and Brewster, S. A. (2007) An initial investigation into non-visual computer supported collaboration. In *CHI'07: CHI'07 extended abstracts on Human factors in computing systems*, 2573–2578.
- Metatla, O., Bryan-Kinns, N., Stockman, T. (2008) Constructing relational diagrams in audio: the multiple perspective hierarchical approach. In: *Assets'08: Proc. of the 10th Inter. ACM SIGACCESS Conf. on Computers and Accessibility*, 97–104. Halifax, Nova Scotia, Canada.
- Metatla, O., Bryan-Kinns, N., and Stockman, T. (2011) Interactive hierarchy-based auditory displays for accessing and manipulating relational diagrams. *J. on Multimodal User Interfaces*.
- Metatla, O., Bryan-Kinns, N., Stockman, T., Martin, F. (2011). Designing for Collaborative Cross-modal Interaction. *Proc. of Digital Engagement 2011 the 2nd RCUK Digital Economy All Hands Meeting*, Newcastle, UK.
- Mynatt, E. D., and Weber, G. (1994) Nonvisual presentation of graphical user interfaces: contrasting two approaches. In *CHI'94: Proc. of the SIGCHI conf. on Human factors in computing systems* (Boston, Massachusetts, USA, 1994), 166–172.
- Olson, J. S., Olson, G. M., Storrøsten, M., and Carter, M. (1993) Groupwork close up: a comparison of the group design process with and without a simple group editor. *ACM Trans. Inf. Syst.* 11, 4, 321–348.
- RNIB. *Looking forward to 2014 rnib's strategy to end the isolation of sight loss*, 2009.
- Weber, G. (1993) Adapting direct manipulation for blind users. In *CHI'93: INTERACT'93 and CHI'93 conf. companion on Human factors in computing systems* (Amsterdam, The Netherlands, 1993), 21–22.
- Winberg, F., and Bowers, J. (2004) Assembling the senses: towards the design of cooperative interfaces for visually impaired users. In *CSCW'04: Proc. of the 2004 ACM conf. on Computer supported cooperative work*, 332–341.
- Yu, W., Kangas, K., and Brewster, S. A. (2003) Web-based haptic applications for blind people to create virtual graphs. In *Haptic Interfaces for Virtual Environment and Teleoperator Systems*. HAPTICS 2003, 318–325.