Collaborating Through Sounds:

Audio-Only Interaction With Diagrams

Oussama Metatla

Submitted for the degree of Doctor of Philosophy

Queen Mary, University of London

2010

Collaborating Through Sounds:

Audio-Only Interaction With Diagrams

Oussama Metatla

Abstract

The widening spectrum of interaction contexts and users' needs continues to expose the limitations of the Graphical User Interface. But despite the benefits of sound in everyday activities and considerable progress in Auditory Display research, audio remains under-explored in Human-Computer Interaction (HCI). This thesis seeks to contribute to unveiling the potential of using audio in HCI by building on and extending current research on how we interact with and through the auditory modality. Its central premise is that audio, by itself, can effectively support collaborative interaction with diagrammatically represented information.

Before exploring audio-only collaborative interaction, two preliminary questions are raised; first, how to translate a given diagram to an alternative form that can be accessed in audio; and second, how to support audio-only interaction with diagrams through the resulting form. An analysis of diagrams that emphasises their properties as external representations is used to address the first question. This analysis informs the design of a multiple perspective hierarchy-based model that captures modality-independent features of a diagram when translating it into an audio accessible form. Two user studies then address the second question by examining the feasibility of the developed model to support the activities of inspecting, constructing and editing diagrams in audio.

The developed model is then deployed in a collaborative lab-based context. A third study explores audio-only collaboration by examining pairs of participants who use audio as the sole means to communicate, access and edit shared diagrams. The channels through which audio is delivered to the workspace are controlled, and the effect on the dynamics of the collaborations is investigated. Results show that pairs of participants are able to collaboratively construct diagrams through sounds. Additionally, the presence or absence of audio in the workspace, and the way in which collaborators chose to work with audio were found to impact patterns of collaborative organisation, awareness of contribution to shared tasks and exchange of workspace awareness information. This work contributes to the areas of Auditory Display and HCI by providing empirically grounded evidence of how the auditory modality can be used to support individual and collaborative interaction with diagrams.

Submitted for the degree of Doctor of Philosophy

Queen Mary, University of London

2010

بِسْمِ ٱللَّهِ ٱلرَّحْمَٰنِ ٱلرَّحِيمِ

In the name of God, the most gracious, the most merciful

(*) وَ نَزَعْنَا مَا فِي صُدُورِهِم مِّن غِلٍّ تَجرِي مِن تَحتِهِمُ ٱلأَنهَارُ وَ قَالُوا ٱلحَمدُ لِلَّهِ ٱلَّذِي هَدَانَا لِهَذَا وَ مَا كُنَّا لِنَهتَدِيَ لَولَا أَن هَدَانَا ٱللَّهُ لَقَد جَآءَت رُسُلُ رَبِّنَا بِالحَقِّ وَ نُودُوا أَن تِلكُمُ ٱلجَنَّةُ أُورِثتُمُوهَا بِمَا كُنتُمْ تَعمَلُونَ (*)

- قُرآنٌ كَرِيم ٧:٤٣

And We shall remove from their hearts any lurking sense of injury; beneath them will be rivers flowing; and they shall say: "Praise be to God, Who hath guided us to this: never could we have found guidance, had it not been for the guidance of God: indeed it was the truth that the Messengers of our Lord brought unto us." And they Shall hear the cry: "This is the Paradise which you have inherited for what you used to do."

– Qur'an 7:43

إِلَى جَدَّي ...

الشَيخُ المُجَاهِدُ الإِمَامُ الأُستَاذ عَمَّار ، اَلَّذِي غَرَسَ فِي نَفسِي حُبَّ آلعِلمِ وَ اِحتِرَامَهُ

وَ

المُجَاهِدُ الشَهِيد رَابَح (رَحِمَهُ اَللّه) ، اَلَّذِي عَلَّمَنِي أَثمَنَ دَرسٍ فِي التَضحِيَةِ

To my grandparents ...

Ammar, who instilled me with the love and respect for science

and

Rabah (may God rest his soul), who taught me the ultimate lesson of sacrifice

إِلَى رَائِد النَهضَةِ الفِكرِيَّةِ الإِصلَاحِيَّةِ الجَزَائِرِيَّة

الشَيخ الإِمَام عَبد الحَمِيد ابن بَادِيس (رَحِمَهُ اَللّه)

To the leader of the Algerian intellectual and reformist movement,

Sheikh Abdelhamid Ibd Badis (may God rest his soul).

إِلَى الجَزَائِر.

To Algeria.

Acknowledgements

I would like to thank my supervisors Dr. Nick Bryan-Kinns and Dr. Tony Stockman, first for granting me the freedom to explore my own ideas, then for providing me with timely support, guidance and advice. I am very grateful for the careful reading and constructive criticism that I received from my examiners, Prof. Stephen Brewster and Prof. Anthony Steed. I would also like to thank Prof. Peter McOwen and Dr. Paul Curzon for their feedback on earlier drafts of this thesis, and Hany Azzam and his wife for proof reading the final draft.

My parents have always been there for me in more ways than I can enumerate. There are simply no words that can express how grateful I am for your endless love, trust and sacrifice. My sister Sana and brother Hichem have provided me with much needed companionship and distractions throughout the duration of the PhD. And my dear wife Romaissa encouraged and helped me especially during the hectic period of writing up, and continues to give me constant emotional and practical support, from discussing research ideas, to putting up with my ramblings about diagrams, representation and the human senses, and *always* answering my often annoyingly meticulous questions about human biology and tissue engineering. I also extend my thanks and gratitude to all the Metatla and Boussahel families.

Finally, I thank all my friends and colleagues at Queen Mary University of London, all my friends back home and in the UK, and all those who took part in the experimental studies reported in this thesis. This PhD research was supported by the Algerian Ministry of Higher Education and Scientific Research (MESRS).

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Motivation

Sight is used ubiquitously in everyday activities. From navigating the environment to communicating with others, reading texts and handling pictures and diagrams, the eyes are, for many of us, indispensable to our survival. Not surprisingly, most of our interactions with technology are designed with the assumption that we are able to see controls and feedback. One of the most frequently used interaction paradigms in Human-Computer Interaction (HCI), and arguably the most important (Myers, 1998), is strongly based on this assumption. First introduced to the market in the 1980s, the Windows Icons Menus and Pointers (WIMP) paradigm and the related concepts of Direct Manipulation and Graphical User Interface (GUI) revolutionised the way computers are used. Sophisticated computing capabilities were made more accessible to everyday users who could operate the computer by manipulating spatially arranged and graphically represented items instead of learning and remembering lists of complex text-based commands (Shneiderman, 1981). Within a few years of their inception, GUIs became more popular than command-line interfaces and the overwhelming majority of human-computer interaction now occurs through a graphical user interface.

Today, the prevalence of GUIs extends beyond the desktop computer. With increasingly embedded computing power, cars, TVs, automated teller machines, mobile phones, and even clothing and furniture are being re-engineered to incorporate some form of graphical display and controller. In fact, it has even become difficult to imagine how some of this technology could be operated without one. Indeed, after 25 years chained to the desktop, computers have broken free and are now invading everyday space (Dix, 2010). Advances in Mobile and Ubiquitous Computing, in addition to the rise of the Internet are changing computers from static machines to mobile and social devices that are reshaping fundamental understandings of interaction and communication. Micro-blogging, distributed teamwork and online social communities are but a few of the activities that connect individuals and groups in ways which did not exist just a few years ago. The emergence of such novel contexts of interaction, however, also exposes the limitations of GUI controllers success. As Holmquist (2000) notes, the problem with successful ideas is that they are hard to look beyond, and one of the issues that persists through these innovative times is that most of today's GUI controllers still awkwardly resemble those of the first computers. Buxton (2001) observes:

> *"In April 1981 Xerox introduced the Star 8010 workstation, the first commercial system with a Graphical User Interface (GUI) and the first to use the "desktop" metaphor to organize a user's interactions with the computer. Despite the perception of huge progress, from the perspective of design and usage models, there has been precious little progress in the intervening years. In the tradition of Rip van Winkle, a Macintosh user who awoke today, after having fallen asleep in 1984, would have no more trouble operating a "modern" PC than operating a modern car"* (p.145).

Of particular concern is the largely uni-modal nature of GUIs, which effectively reduce the human user to a set of eyes looking at the graphical display and few fingers manipulating a keypad or computer mouse. This constraint stands in direct contrast to the very nature of the spaces that computing devices are invading. Humans use sight, hearing, touch, smell and taste to make sense of their environment and act in it, but this multimodal experience does not adequately translate to GUI-based interactions with technology. Individuals who are unable to see a device's controls and feedback are thus considerably disadvantaged. This includes not only visually impaired users, who rely predominantly on screen-reader technology[1] to access digital content (Stockman, 2004), but also sighted users interacting with limited visual displays, such as mobile and smart phones, or in contexts where there is poor visibility or high demands on visual attention.

Increasingly, the HCI community is realising that the dominant interaction paradigm no

[1] A screen-reader is a piece of software that extracts text from a screen's display buffer then output the retrieved text using a speech synthesiser or a braille display.

longer adequately accommodates the widening spectrum of users and contexts of interaction (Gentner and Nielsen, 1996; Buxton, 2001; Dix, 2010). As a result, a number of radically different concepts and techniques have been explored to overcome the limitations of GUIs. These techniques include using time rather than space as a metaphor for organising interaction with digital content (Fertig et al., 1996; Rekimoto, 1999), immersing the user in three-dimensional artificial environments (Rheingold, 1992) and using brainwaves as input to control virtual and physical objects (Le, 2009). Of interest to this thesis are approaches that exploit alternative sensory channels to communicate information and support interaction with technology. Hearing and touch are particularly popular alternatives, the former having been explored in the form of auditory displays, which use speech and non-speech sounds to convey information (Kramer, 1994a), and the latter through haptic and tactile displays, which convey information through cutaneous or kinesthetic sensation (Kortum, 2008). As technology improves, the inclusion of such displays in personal digital devices is becoming common place, thus allowing research efforts to focus on issues related to the actual use of such technology rather than its production. In effect, recent research has shown that the auditory and haptic modalities can improve interactive performance over purely visual displays in a variety of interactive settings (Poupyrev and Maruyama, 2003; Chang and O'Sullivan, 2005; Hoggan et al., 2009).

The focus of this thesis is on the auditory modality. Audio accompanied computers since their inception yet its function remained primarily limited to alerting purposes and displaying status information and feedback. These limitations exist despite significant advances in Auditory Display research, a specialised discipline which has emerged in the last few decades to study the use of sound to convey information. Indeed, a recent survey amongst HCI practitioners revealed that audio remains an under-explored medium in interaction design (Frauenberger et al., 2007). Many researchers assert that this is due to the lack of both theoretical support for the design of auditory cues in user interfaces as well as thorough accounts of human interactions with and through the auditory modality (Barrass, 1998; Mitsopoulos, 2000; McGookin, 2004; Murphy, 2007; Frauenberger and Stockman, 2009; Nees and Walker, 2009).

1.2 Aims, Scope and Contributions

The benefits of audio in interaction can only be fully realised when sounds are properly designed and integrated in the user interface. To achieve these benefits, two fundamental aspects of design must be thoroughly explored. First, existing design knowledge must be successfully transferred from researchers to practitioners and from experts to novice designers, and then successfully applied during the design process. Second, existing knowledge must be extended with new understandings of how sound can be used to convey information and support interaction. The aim of this thesis is to contribute to the latter aspect of design; i.e. to unveiling the potential of audio in HCI by building on existing knowledge and contributing new understandings of its practicality as a medium of interaction. This is of course a broad aim and could motivate various lines of research across more than one discipline. Indeed, in their proposed research agenda for the field of Auditory Display, Kramer et al. (1997) write:

> *"[..] development of effective auditory representations of data will require interdisciplinary collaborations using the combined knowledge and efforts of psychologists, computer scientists, engineers, physicists, composers, and musicians, along with the expertise of specialists in the application areas being addressed."*

Thus, to narrow the scope of the research, focus will be directed towards exploring the practicality of the auditory modality in supporting interaction with a subset of graphical displays, namely, with diagrammatically represented information. Diagrams have been chosen as an area of focus for a number of reasons. First, they are used extensively to represent information in both formal and informal human activities. For instance, they can be found in bus stops and train stations and have also been developed into common standards for expressing specialised aspects of various disciplines (e.g meteorologists use weather maps, architects use floor plans, and computer scientists make extensive use of node-and-link diagrams). Second, the dominant WIMP paradigm itself consists of a diagrammatic representation or a collection of such representations (Blackwell, 1998). Thus, knowledge of how diagrams can be modelled for auditory interaction could contribute to overcoming the limitations of GUIs in contexts where vision is not the optimal channel of interaction.

At the same time, the prevalence of GUIs in human-computer interaction means that a variety of GUI-based computer tools have been developed for constructing and manipulating diagrammatic representations, both for industrial and educational purposes. This leads to another important reason for focusing on diagrams: the key role diagrams play in collaborative interaction (Suthers and Hundhausen, 2003; Stahl et al., 2006; Cherubini et al., 2007). It is common for groups of individuals to work on shared documents, reports or lessons which include some form of diagrammatic representations, again both in industry and education. Therefore, understanding how audio can support collaboration may lead to improved designs for supporting individuals involved in distributed teamwork, where team members use devices with varying display capabilities to connect to a shared workspace, or classrooms where sighted and visually impaired members work together with diagrams.

Research questions. The goal of the work presented in this thesis is thus to explore the transformation of a diagrammatic representation into an audio-only accessible form, and to investigate how the resulting translation could be used to support individual and collaborative interaction. The main research question is:

> **“Can audio be a practical medium to support individual and collaborative interaction with diagrammatically represented information?”**

This question implies a number of research questions (RQs) which define the scope of this thesis and implicitly highlight its contributions to auditory research in HCI:

RQ1: How can a given diagram be translated from a graphical form to an alternative form that can be accessed in audio? Human vision and audition are two different perceptual systems. Consequently, the way in which information is conveyed and processed through graphical and auditory displays also differs. These differences must be taken into consideration when translating diagrams from graphical to auditory form. A key step in such a process is grasping exactly how diagrams encode information; this understanding then helps to establish which information about a given diagram should be captured and how it can be modelled and conveyed in audio.

This thesis explores the use of hierarchical structures to model information encoded in a particular type of diagrams, referred to as relational diagrams (see Chapter 3 for more on this). To this end, the thesis addresses the question of how relational diagrams encode information through an analysis of their properties as external representations. This analysis helps to establish a thorough understanding of the functional and structural properties of diagrams that gives them advantages over other forms of representation, thus motivating the design of a hierarchy-based model for supporting audio-only interaction with relational diagrams.

RQ2: How can the activities of inspecting and constructing diagrams be supported through the resulting auditory translation? Once captured and modelled for auditory interaction, the ability to actively inspect, construct and edit diagrammatically represented information should be supported in order to provide a practical means of using audio to interact with diagrams. More importantly, the ability to inspect and edit a shared representation is essential for successful participation in a collaborative task (Cherubini et al., 2007). Thus, investigating the practicality of any approach to translating diagrams from graphics to audio in supporting such activities is essential before examining questions of collaborative use.

This thesis presents two empirical studies that examine the feasibility of the developed model in supporting the above mentioned activities. The first study focuses on the activity of inspecting relational diagrams in audio. To this end, the thesis develops two presentation modes for displaying the hierarchy-based model in audio by varying the extent to which speech and non-speech sounds are combined to represent its content. The first study therefore investigates the feasibility of using the hierarchy-based model to support the activity of inspecting relational diagrams and the impact of the two audio presentation modes on this process. This is done by empirically examining two main hypotheses:

H1 Using a hierarchy-based audio-only model to capture and structure information encoded in relational diagrams allows for successful inspection of such information in audio.

H2 Varying the audio presentation mode of the hierarchy-based audio-only model impacts users' performance on diagram inspection tasks.

The second study focuses on the activity of editing relational diagrams in audio. To this end, the thesis develops two interaction strategies for constructing and editing relational diagrams through the hierarchy-based model. The two strategies differ in the amount of guidance that they provide to the user when editing diagrams in audio. The second study therefore investigates the feasibility of using the hierarchy-based model to support the activity of editing relational diagrams and the impact of the two interaction strategies on this process. This is done by empirically examining two main hypotheses:

H3 Using a hierarchy-based audio-only model to capture and structure information encoded in relational diagrams allows for successful construction and editing of such information in audio.

H4 Varying the interaction strategies with a hierarchy-based audio-only model impacts users' performance on diagram construction tasks.

RQ3: What are the characteristics of collaborative interaction with diagrams when audio is the only medium of interaction? Once a practical model that supports individual inspection and editing of diagrammatically represented information is established, it can be deployed in a collaborative context. The focus in this thesis is on investigating whether audio, by itself, can be a practical means for supporting collaborative interaction. That is, it seeks to establish new understandings of the nature of collaborative interaction when audio is used as the sole means by which collaborators communicate with one another and contribute to a shared workspace.

To do this, the thesis examines collaborative diagram construction in two audio-only workspace settings which differ in terms of the amount of audio that is displayed to or concealed form the collaborators in the workspace. This third study investigates the impact of varying the means for delivering audio to an audio-only workspace on three main dynamics of collaborative interaction: First, the structure and organisation of the collaborative process. Second, the ability of contributors to keep track of their contributions to the shared task and to differentiate it from their partners'. Third, collaborators' exchange of information that contribute towards keeping an awareness of events that occur in the audio-only workspace, and the way this information is used to maintain the fluidity of the collaborations. This is done by empirically examining three main hypotheses:

H5 The means for delivering audio in an audio-only workspace impacts the collaborative process in terms of its structure and organisation.

H6 Displaying the audio output of each collaborator's interactions to all collaborators increases their awareness of self and partners' contributions to shared tasks.

H7 Concealing the audio output of each collaborator's interactions from their partners increases the exchange of workspace awareness information between them.

Context and summary of original work. While there is existing research on the use of sound to access graphical representations (e.g Mynatt, 1995; Alty and Rigas, 1998; Bennett, 1999; Brown, 2008), work in this area has been limited and there is room for improvement, both in terms of the type of graphical representations to address and approaches to supporting audio-only interaction with such representations. Additionally, even less research has addressed the issue of how graphical representations could be actively edited through non-visual means. The few computer-based solutions that have been investigated typically combine the auditory and haptic modalities to support such a task (e.g McGookin and Brewster, 2007a; Rassmus-Gröhn et al., 2007). The practicality of audio as the sole means for supporting such interaction is therefore unknown. Furthermore, studies that investigate the use of audio to support collaborative interaction with diagrammatic representations have also been sparse (e.g Winberg and Bowers, 2004; McGookin and Brewster, 2007b), with no previous work addressing the issue of using audio as the sole means for interaction and communication.

The work presented in this thesis is novel in that it develops a model for translating diagrams into an audio accessible form by building on and extending existing approaches to non-visual interaction with diagrams. Additionally, the research contribution is original in that no previous study has attempted to investigate the audio-only construction of diagrams and collaborative interaction with shared diagrammatic representations where collaborators use audio as the only means for communicating with one another and for accessing and manipulating a shared workspace.

Intended audience. As described above, this thesis aims to deliver empirically grounded evidence of how the auditory modality can be used to support individual and collaborative interaction with diagrams. The theoretical and empirical investigations undertaken in this thesis contribute towards generation knowledge about how to translate relational diagrams from graphical to auditory form, how to use audio to actively support the ability to inspect and edit diagrams, and how to support collaborative interaction in an audio-only workspace. The intended audience for this thesis is therefore the community of Auditory Display researchers and interaction designers who wish to incorporate sound in their designs. By investigating individual and collaborative audio-only interaction, and compiling a set of design lessons throughout the thesis, it is hoped that this work will contribute towards increasing the potential of effectively using audio in the practice of HCI and computer-supported cooperative work (CSCW).

1.3 Thesis Outline

Figure 1.1 outlines the structure of this thesis. It is organised into nine chapters, six of which form the two main working parts of the presented research. Part I focuses on individual audio-only interaction with diagrams, while Part II focuses on audio-only collaborative interaction. First, Chapter 1 presents the motivation for the thesis, defines its scope and contributions and outlines its structure. Chapter 2 then presents a brief account of the benefits and issues surrounding the use of sound in HCI, provides an overview of the field of Auditory Display and a review of current approaches for supporting non-visual interaction with visual displays.

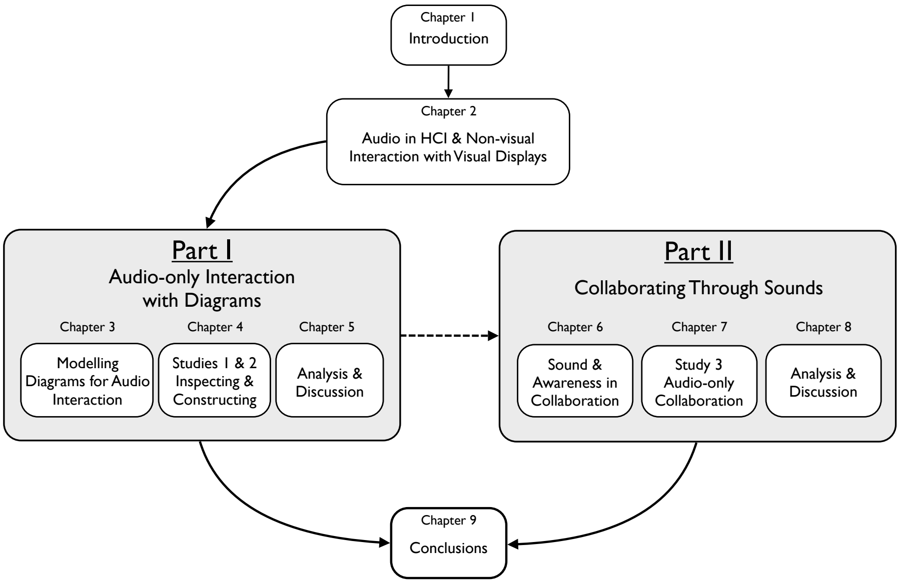


Figure 1.1: Thesis structure.

In Part I, Chapter 3 reviews how diagrams are used to represent information, and includes an analysis of relational diagrams in terms of their properties as external representations. Motivated by these reviews and analysis, Chapter 3 develops a hierarchy-based model that translates relational diagrams for audio-only interaction. Chapter 4 investigates the feasibility of the developed model for supporting audio-only inspection and construction of diagrams through two user studies. Chapter 5, then, presents a detailed analysis and discussion of results obtained from Studies 1 and 2 and a thorough assessment of the feasibility and practicality of the developed hierarchy-based model, together with a set of design lessons for supporting individual audio-only interaction with diagrams. In Part II, Chapter 6 examines the role of audio in both remote and co-located collaborative interaction, and Chapter 7 presents Study 3 where the model developed in Part I is extended and deployed in a lab-based audio-only collaborative study. The results obtained from Study 3 are thoroughly analysed and discussed in Chapter 8 which extends the set of design lessons with recommendations for using audio to support collaborative interaction with and through diagrams. The thesis concludes with Chapter 9, which reviews the research presented in previous chapters and assesses the extent to which the research questions have been answered. Finally, the contributions and implications of the research are discussed and outstanding issues are summarised for future work.

Chapter 2

Audio in HCI and Non-visual Interaction with Visual Displays

2.1 Introduction

This chapter presents a review of the auditory modality in HCI. Particularly, it introduces the field of Auditory Display and the way it addressed the problem of supporting non-visual interaction with visual displays. Understanding the auditory world requires a plethora of reviews to cover the variety of aspects associated with hearing, integrating and interpreting auditory information. The present review is concerned with aspects that are directly relevant to the use of sound as a medium of interaction at the user interface without addressing the physiological, perceptual and cognitive foundations associated with human audition.

It is divided into two main parts. In the first part, Section 2.2 presents a brief history of audio at the user interface and the benefits and issues surrounding its use in HCI. Section 2.3 then introduces the area of Auditory Display. It defines the terminology used throughout the remainder of the thesis by describing various techniques for displaying information in audio and existing methodologies and guidelines for employing such techniques in interface design. In the second part, Section 2.4 reviews current approaches to supporting non-visual interaction with visual displays. In particular, the section explores how auditory display techniques have been used solely or in combination with other modalities to make typical visual representations, such as tables, graphs and diagrams, non-visually accessible. The chapter concludes with a summary in Section 2.5.

2.2 Sound at the User Interface

Human hearing is a sophisticated sense, one that is capable of extracting a wealth of information from the environment and is used ubiquitously in everyday life. Sound communicates linguistic information as well as feedback about events and actions; both of which are often taken for granted. For example, while listening to someone talk, it is possible to infer information about their emotional state from subtle changes in the patterns of stress and intonation in their voice within very few seconds of them starting to talk (Stockman, 2010).

Unlike the visual system, which extracts information from a relatively small field of visual focus, the auditory system gathers information that is emitted from different directions. The ears are also open channels that, unlike the eyes, cannot be voluntarily shut to block sounds, which makes them a powerful instrument for alerting and orientation, often "telling" the eyes where to look (Wenzel, 1992). Indeed, sounds surround the listener with a continuous stream of feedback about ongoing events and activities that help situate perceptions and actions in the environment. The richness of auditory information and the way it is integrated with other senses and used in everyday activities makes audio a naturally attractive modality to exploit in interactive systems.

2.2.1 Brief History

Sound accompanied the personal computer (PC) since its inception. Typing on the computer keyboard, saving or copying a file to its disk drive all make sounds that, although not intentionally designed, allow users to extract information about their interactions with their computers. For example, a user can tell when a saving task was completed by listening to the noise made by the spinning hard drive, and can get on with other activities while peripherally monitoring the progress of such a task.

Audio capabilities. Early PCs were also equipped with hardware capabilities for outputting audio, albeit of limited quality. These were mostly exploited to display alerting messages and convey status information using intentionally designed simple tone sounds displayed through a built-in speaker. Today, most personal computers and digital devices are capable of producing rich audio output ranging from the 16 Bit, 44.1kHz sampling resolution of Compact Disks and mobile devices to the multi-channel input and output of 24 Bit 96 kHz, or 192 kHz of DVDs and desktop sound cards. This means that most interactive devices have the ability to simultaneously output a large number of high quality sounds. Furthermore, almost all audio hardware is nowadays capable of supporting the synthesis of polyphonic and multi-timbral sounds from sampled instruments or Musical Instrument Digital Interface (MIDI)[1] (Brewster, 2002). Ironically, and despite advances in audio input and output capabilities, the use of audio at the computer interface remains mostly limited to the same early functions of alerting and indicating status information.

Speech output. The first speaking computer was the Apple Macintosh 128 when, back in 1984, the company equipped the computer with an 8 Bit mono sound chip with a 22kHz sampling rate that was capable of synthesising speech. In 1986, IBM announced the IBM Screen Reader for DOS; an innovation that allowed visually impaired users to access a PC's terminal. This was followed by the IBM Screen Reader II which, for the first time, was capable of accessing the then-emerging Graphical User Interface (GUI). However, while accessing the dominantly textual, and therefore linear display of command-line interfaces was adequate in speech, screen-readers required a new approach in order to access Graphical User Interfaces (GUIs). Modern screen-reader products, such as JAWS [2], Window-Eyes [3] and Apple's VoiceOver [4] have introduced features that augment speech output with non-speech sounds to increase efficiency of interaction. For example, VoiceOver introduced non-speech sound cues that convey navigational information, while JAWS and Window-Eyes included the ability to link sound files to occurring events on the interface (Stockman, 2004). Despite these advances, generic screen-reader products still use speech as the main means for accessing GUIs.

All of speech, MIDI and non-speech sounds have been combined in various ways to convey information in audio at the user interface. Nonetheless, the role of audio in major operating systems remains marginal, being mostly limited to alerting purposes and to addressing the problem of accessibility.

2.2.2 Benefits of Using Sound in HCI

There are a number of reasons why using sound at the user interface is advantageous, and why its potential should be fully exploited, both when it is combined with a visual display and when it is not. These benefits are interlinked, stemming from the intrinsic characteristics of the human auditory system and of sound as a medium of information presentation. They could be summarised as follows:

[1] `http://www.midi.org/`
[2] `http://www.freedomscientific.com/`
[3] `http://www.gwmicro.com/`
[4] `http://www.apple.com/accessibility/voiceover`

Eyes free interaction. Sound can be received without the user having to focus their ears in a particular direction. If the user is engaged in a task that require them to have visual focus on a particular area of a display, sound can provide an alternative source for conveying information that might be important to the task at hand or for monitoring other ongoing activities. In general, eyes free auditory interaction is beneficial when users are likely to be engaged in activities that leave little room for visual attention (Holland et al., 2002). There are many situations where this is the case. For example, a firefighter whose field of view is impaired by the extensive smoke inside a burning building might require information to track victims locations, or to explore the floor plan of the building. Such information could be effectively conveyed, accessed and manipulated in audio.

Information offload. With increasing reliance on GUIs in modern displays, the visual channel is used intensively and is often overloaded with information. Balancing such load across modalities has been found to increase the efficiency of interaction (Oviatt et al., 2004). Using sound to enhance graphical displays can therefore help reduce information overload and demands on visual attention at any given moment during interaction. This applies both when using large screens or multiple monitors and when using mobile devices that have limited screen capacity for displaying information.

Interaction contexts. The computer is no longer constrained to the desk and users are no longer necessarily holding a mouse and facing a desktop monitor. Computing power and interactivity are embedded in mobile devices and in a variety of everyday objects, including clothes (Mann, 1996). As a result, there is an increasing realisation amongst the HCI community that traditional interaction paradigms, such as WIMP, can no longer adequately accommodate new and emerging contexts of interaction. The incorporation of sound in such contexts has the potential of enriching the interactive experiences that they support.

Assistive technology. While screen-readers can be used to access linearly presented information, such as text or command-line interfaces, they fall short of providing efficient access to GUIs (Barnicle, 2000; Stockman and Metatla, 2008). As a result, visually impaired users are not able to take full advantage of the facilities available through modern technology. As mentioned above, modern screen-readers, such as JAWS and VoiceOver, have only recently started tapping into the potential of using non-speech sounds in combination with speech output to increase the usability and accessibility of graphical content. Exploring and incorporating the full potential of sound at the user interface can help overcome accessibility problems.

Natural interaction. Certain actions or events, such as those containing information that varies over time, maybe be more naturally represented using sound. This is because sound is itself a temporally structured physical phenomenon. Bly (1982), who studied the presentation of multivariate data using sound, suggests that *"in areas in which the number of varying parameters exceeds our visual response to color, rotation and dimension or in areas in which the data does not correspond to our three-dimensional perception, graphical displays are inadequate"* (p.371). Sound could provide an alternative means of presentation that overcomes display limitations in such instances. Sound is also naturally attention grabbing (Brewster, 1994); it is very hard to avoid hearing something as opposed to avoiding looking at something. Indeed, one of the most prominent use of sound at the user interface is for alerting purposes (Kramer, 1994a). The use of sound in the interface to draw the user's attention to the occurrence of particular events or to particular areas on the screen can be a powerful addition to interactive experiences.

Scientific exploration and discovery. Data analysis and exploration have long been supported by a variety of visual displays, such as graphs, tables and diagrams. However, as described above, there are instances where such displays can be inadequate and unnatural means for presenting information. Sonification – which could be considered the auditory equivalent of visualisation – has been successfully employed for exploring data, and even overcoming the limitations of other means of scientific analysis (Hermann, 2002). In addition to providing an extra source of information to a human analyst, presenting information in audio could reveal structures and relationships within data that could remain untapped on if presented through other modalities.

2.2.3 Issues With Using Sound in HCI

There are a number of issues associated with using sound at the user interface. These are also related to the characteristics of sound as a medium of information presentation, but could also be related to the context in which it is used to display information; i.e where the listening takes place. Kramer (1994b) suggests that designers should consider the following issues when using sound to display information:

Low resolution. Many auditory parameters, such as loudness, are not suitable for displaying quantitative information of high resolution (Brewster, 2002). Additionally, the ability of the auditory system to localise sound sources is less accurate than the localisation ability of the visual system. This ability decreases further when two or more sounds need to be separated in space (Best et al., 2003).

Lack of orthogonality. Sound attributes are not independent, which means that changing the value of a given attribute might affect the perception of change in another. For example, the loudness of a sound is dependent on its frequency such that changing the frequency will cause a change in the perception of its loudness (Kaper et al., 1999).

Absolute data. Unless speech is used, the accuracy of interpreting absolute data values in audio is low, hindered by the lack of a reference system – such as the coordinate system in plots. The ability to extract absolute data values from a sound is also largely dependent on the abilities of the listener and the context of where the listening takes place (Moore, 2003).

Transience. Sound is displayed in time and so disappears from the environment immediately after it is presented. A direct consequence of this is that users have to remember what is displayed in sound, which poses issues with lengthy auditory messages. Furthermore, because sound is temporally structured, it is often presented serially, which makes its presentation time consuming and the comparisons of two or more items that occur at different points during an audio message difficult.

Interference and annoyance. Because it is hard to avoid hearing something, sound can be a source of interference and annoyance. Interference can occur in two ways; sounds from an interface might interfere with those in its environment and annoy others; and sounds emitted from the environment might interfere with the audio output of an interface and annoy its user. The loudness of the sounds is a particularly relevant attribute in these situations. Generally speaking, sound is annoying when it does not convey any useful information or pleasure (Buxton, 1989). Using headphones may seem to be a straightforward solution to the problem of annoyance and interference. However, there are complications as it is sometimes desirable that the listener is not isolated from their acoustic environment, particularly if they are working with other people.

Individual differences. The ability to interpret and understand information displayed in audio depends ultimately on the listener. This includes not only perceptual abilities – e.g less than 1% of the population have perfect pitch such that they can identify the pitch of a tone without reference to another tone (Moore, 2003) – but also cognitive abilities, listening skills, cultural biases, training and learning styles (Mauney and Walker, 2007; Stockman, 2010).

Most of the above issues might contribute to why audio is underused in HCI. They are indeed hard to eliminate and, in most cases, should simply be accepted as part of the limitations of the human auditory system and taken into account when designing auditory displays (Brewster, 2002).

2.3 Auditory Display

Auditory Display (AD) is a term employed to describe all uses of sound at the user interface (Kramer, 1994a). Although early research on the use of sound to convey information can be tracked throughout the last few decades (e.g, Pollack and Ficks, 1954; Speeth, 1961; Lunney and Morrison, 1981; Bly, 1982; Mansur et al., 1985), AD is still a relatively young research field. Its origins as an independent discipline could be traced back to the first International Conference on Auditory Display (ICAD) that took place in 1992 [5]. The meeting gathered the then-current research on AD, and the resulting proceedings volume (Kramer, 1994a) is still considered an essential piece of literature and an important foundational reference in the field.

There is ongoing debate about terminology in auditory display research, including the exact definition of what an "auditory display" is. This could be due to the fact that AD is still in its infancy, but also to the interdisciplinary nature of AD research (Hermann, 2008). For the purpose of this thesis, a similar definition to that used by McGookin (2004) will be adopted, which considers an auditory display to be the use of techniques that communicate information about the content or state of an application or computing device to a user in audio. Specifically, the definition excludes the use of auditory techniques for conveying sounds from the user to the device, for example through voice commands.

2.3.1 Techniques

Just as text and graphics are used to convey information in visual displays, a number of techniques are used in auditory displays to encode and communicate data. Seven such techniques are considered in what follows. It should be noted that such a list is not exhaustive and, given the relative recency of the field of Auditory Display, there is still room for new and innovative techniques to address the issues and exploit the advantages described in the previous section. The following were chosen to be discussed because they are commonly used and have been found to be effective for conveying information in sound.

[5] `http://www.icad.org/websiteV2.0/Conferences/ICAD92/about92.html`

2.3.1.1 Speech

Spoken output is an evident and popular means for conveying information in audio at the user interface and is particularly efficient for communicating absolute values. It also has the advantage that, if familiar with the language used, most users would require minimal to no training to be able to use an auditory interface employing speech output. There are two methods for producing speech at the user interface; synthesis or concatenation of recorded speech. The former has the advantage of allowing real-time rendering of text into speech, though it can be of low quality. The latter is less flexible as it requires a pre-stored bank of audio recordings, but allows for high quality speech display (Spiegel and Streeter, 1997). It remains that speech is a relatively slow method to communicate information (Petrie et al., 1998) and can be inadequate for the non-visual presentation of graphs and diagrams (Yu et al., 2003).

2.3.1.2 Auditory Icons

Auditory icons are another popular technique for displaying sound at the user interface. Gaver (1997) defined them as: *"everyday sounds mapped to computer events by analogy with everyday sound-producing events"* (p.18). Auditory icons are thus a means for using sound in computers as they are used in the real world, where familiar sounds are mapped to computer events to which they bear obvious relationship. For instance, the sound of incoming mail can be represented by the sound of a letter being thrown into a mailbox, and that a of a deleted file with the sound of scrunching paper. The SonicFinder (Gaver, 1989) was the first application to use auditory icons to display information in combination with a visual interface. Users could browse file directories and receive information about file types, sizes and locations through variance in the parameters of auditory icons. When enhanced in this way, auditory icons are referred to as *Parameterised Auditory Icons* (Gaver, 1992b).

A disadvantage of using auditory icons is the limited possibilities of representation, as some computer events do not have an obvious relationship with sounds (Brewster, 2002). For instance, the sound of scrunching paper can be directly related to the action of deleting a file, but it is less obvious to determine an appropriate natural sound for more abstract operations, such as renaming a file. It is also difficult to manipulate the parameters of an auditory icon without affecting the characteristics of the natural sound it conveys (Brewster et al., 1996).

2.3.1.3 Earcons

Blattner et al. (1989) defines earcons as: *"non-verbal audio messages that are used in the com-*

puter/user interface to provide information to the user about some computer object, operation or interaction" (p.13). Earcons convey information using musical rather than natural sounds. They are constructed from a basic building block called a *Motive*, which is *"a brief succession of pitches arranged in such a way as to produce a tonal pattern sufficiently distinct to allow it to function as an individual recognisable entity"*(ibid, p.23). Motives can be combined to create families of structured sounds; for example, if an earcon represents a file, and another represents the action "open", then combining the two in a sequence could be used to to represent the "open file" option (Brewster et al., 1994). Familiar examples of earcons include the arrival of a new text message on a mobile phone, the Windows error beep and the Macintosh start-up chord.

Unlike auditory icons, there is no direct relationship between the sound of earcons and the information they represent. Their meaning must therefore be learnt by the listener before earcons can be used efficiently. While this increases the overhead associated with learning, it gives earcons the possibility of representing any event, object or action on the interface.

2.3.1.4 Spearcons

Spearcons are speech-based short audio messages that were developed to represent items on menu entries. A spearcon is created by converting the text output of a menu item into speech then speeding up its presentation to the point where it is no longer comprehensible as speech (Walker et al., 2006). The result is a signature sound that can play the same role of an auditory icon or an earcon. Like auditory icons, spearcons are unique to the particular menu item they represent, but at the same time, the similarities between the initial parts of menu items allows for the creation of families of structures like earcons. For example, when sped up, the initial part of the items "Save", "Save As" and "Save To" would sound the same, yet differ in length and overall acoustic signature making each item unique. Evidence suggests that spearcons are as learnable as speech (Dingler et al., 2008), and that presenting them as hints before fully spoken menu items can improve user speed and accuracy in selection tasks (Walker et al., 2006; Palladino and Walker, 2008).

2.3.1.5 Sonification

Sonification has been defined as:*"the use of non-speech sound to convey information"* (Kramer et al., 1997). Emphasising its function, Scaletti (1994) defines it as:*"a mapping of numerically represented relations in some domain under study to relations in an acoustic domain for the purpose of interpreting, understanding, or communicating the relations in the domain under* *study"*. More recently, Hermann (2008) suggests that a given technique that uses data as an input and generates auditory output maybe called a *sonification* if; the resulting audio reflects objective properties or relations in the data; the transformation is systematic; the sonification is reproducible; and the sonification production system can be used with different datasets.

As mentioned above, sonification could be considered the auditory equivalent of visualisation but uses non-speech sounds instead of graphics for accessing and exploring data. Early examples of the use of sounds for such purposes include the Geiger counter, which maps the level of radiation in the environment to the sound of "clicks"; Sonar, which uses sound for locating and recognising objects under water; and the auditory thermometer, which maps the pitch of a continuous tone to represent temperature. To date, sonification has been employed in a variety of domains including medical (e.g Baier et al., 2006; Pauletto and Hunt, 2006), educational (e.g Upson, 2001; Stockman, 2004) and business (e.g Nesbitt and Barrass, 2002; Janata and Childs, 2004). Sonification is also particularly effective for providing access to graphs data and overviews (e.g Mansur et al., 1985; Nees and Walker, 2006; Harrar and Stockman, 2007).

2.3.1.6 Continuous Ambient Sounds

While the speech and non-speech techniques described above provide short messages to describe elements of an interface, persistent and temporally dynamic objects are better supported with continuous patterns of sounds (Sawhney and Murphy, 1996). Ambient sounds could combine various auditory display techniques while imposing a persistent presentation, such as looping auditory icons, extended melodies of earcons or continuous sonification of data. Such a presentation technique has been found to be particularly effective for monitoring real-time data and background activities and processes (Mynatt et al., 1998; Janata and Childs, 2004; Roginska et al., 2006).

2.3.1.7 Spatial Sounds

Another way in which the above techniques could be displayed is through spatial sounds. Spatialisation is a technique in which a sound is processed so that when displayed it appears to the listener as if it originated from a source placed in three-dimensional (3D) space. 3D sounds can be synthesised and displayed over normal headphones using Head Related Transfer Functions (HRTFs), which model the human head to determine how the frequency and volume of a sound can be manipulated to simulate a 3D effect (Burgess, 1992). This technique has been found to be particularly effective to help separate concurrently presented earcons (McGookin, 2004). A similar effect can also be achieve using stereo output, in which two or more speakers are physically arranged around a listener to display sounds at different levels of intensity and simulate 3D sound effects.

2.3.2 Guidelines and Methodologies

Guidelines and methodologies for designing and incorporating audio at the user interface are considerably lacking in the area of AD. Auditory display design is often based on designers experience, intuition or available resources (Frauenberger et al., 2007; Murphy, 2007). The recency of AD as a research discipline in addition to its highly interdisciplinary nature has led to a slow development of theoretically grounded, comprehensible and generally applicable approaches to understanding and supporting interaction with audio. Nonetheless, a body of theoretical support is emerging as a number of researchers have initiated grounded work for formulating and organising auditory design knowledge.

Design guidelines. Mynatt (1994a) addressed the problem of choosing appropriate sounds to design auditory icons by investigating their usability. She suggested that four factors should be considered to aid this process; the identifiability of a sound; its conceptual mapping, i.e how well it maps to aspects of the interface that it represents; its physical parameters; and users emotional response to it. Mynatt (1994b) derived a design methodology from these factors consisting of the following basic steps; choosing sounds; evaluating their identifiability using free-from answers; evaluating the learnability of sounds that are not readily identified; and running tests to evaluate the cohesiveness of the resulting set of auditory icons.

Brewster (1994) conducted various experiments to examine the understandability of earcons and how they could be designed to effectively communicate information at the user interface. While the choice of earcons to include at an interface depends ultimately on the type of application and the context of its use, Brewster et al. (1995) derived a set of general guidelines for manipulating sound attributes in order to encode information using this technique. Later, McGookin and Brewster (2006a) investigated and and identified guidelines for the concurrent presentation of earcons at the user interface, and Brown and Brewster (2003) presented a set of guidelines for the audio presentation of graphs and tables.

Design methodologies and knowledge organisation. Mynatt (1995) developed a methodology for transforming GUIs into a non-visually accessible format. Her methodology involved two stages; first, an analysis of the design of a system in order to determine the content that needs to be conveyed; then choosing the appropriate method for conveying it in audio. This two-stage approach was later extended by Mitsopoulos (2000) who proposed a methodology for organising the design of auditory displays in three levels. At the first level, *the conceptual level*, the information to be conveyed through an auditory representation should be specified. Mitsopoulos (2000) suggests that such specification should be based on the analysis of the visual artefact to be translated, distinguishing between the conveyed information and the modality of its presentation. At the second level, *the structural level*, the structure of the auditory scene should be determined in terms of interaction possibilities driven by theories of attention and psychoacoustics. At the third level, *the implementation level*, physical dimensions of sounds should be identified to convey the chosen information as captured by the previous two levels.

Barrass (1998) proposed a task and data analysis approach to the design of auditory information in a methodological framework he called *TaDa!* – Task-oriented Data-Analysis. The method consists of four phases; *scenario description* of the supported activity; *requirement analysis* derived from an analysis of the task that the display supports to extract the underlying data that needs to be conveyed; *representation design* in which requirements are used to select potential representations from a database of sound examples; and *realisation* where the sounds of the display are produced. Later, Adcock and Barrass (2004) suggested using the concept of *Design Patterns*, which captures reusable solutions to reoccurring problems in a particular domain, to organise a community effort for collecting and sharing design knowledge. To support such effort, Adcock and Barrass (2004) created a repository of pattern scenarios where auditory solutions are available online for members of the auditory display research community to contribute to. Frauenberger (2008) extended this approach by investigating mode-independent interaction patterns for incorporating sound at the user interface. Frauenberger (2008) developed and evaluated a methodological tool and found that it supports the capture, transfer and application of design knowledge from experts to novices in the form of design patterns (Frauenberger and Stockman, 2009). A potential issue with using design patterns to cultivate and organise auditory design knowledge is its explicit reliance on designers contributions. The body of knowledge could therefore take a significantly long time to build and transfer from academics to commercial practitioners, where the ultimate impact on users lies.

The above is not a complete list of existing guidelines and methodologies for auditory display design. There is, however, a common trend amongst those described in that devising auditory means to display information consists of, firstly, a consideration of the data to be represented without reference to a specific modality of presentation. This is apparent in the first stage of Mynatt (1995)'s methodology, which consists of a system's analysis to determine the content that needs to be communicates; Mitsopoulos (2000)'s conceptual level of design where the visual artefact is analysed to specify the information conveyed through it; the requirement analysis phase of Barrass (1998)'s *TaDa!*; and Frauenberger (2008)'s mode-independent design pattern approach. The author has similarly argued for emphasising modality-independent features of a given visual display when translating it from the graphical to an auditory form (Metatla et al., 2007, 2008); an approach elaborated in Chapter 3 of this thesis. The next section takes a closer look at existing approaches to using audio, either solely or in combination with other modalities, to support non-visual interaction with visual displays.

2.4 Non-visual Interaction with Visual Displays

Interest in supporting non-visual access to visual displays grew in parallel with early developments in AD research (Kramer, 1994a). A major drive of such endeavours has been and still is the potential to support individuals with temporary or permanent perceptual and situational 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. Thus, changes in data values alter the pitch of the tone and allows for an auditory presentation of variance in the data. This approach to using sonification allows visually impaired individuals to examine data presented in graphs and tables. The technique has been thoroughly investigated since it was first introduced; to improve its efficiency and usability, researchers have explored the effects of displaying multiple data series (Brown and Brewster, 2003), adding contextual information (Smith and Walker, 2002), and varying presentation parameters (Stockman, 2004; Harrar and Stockman, 2007) on understanding and interacting with sonified graphs and tables.

Representational models. Of particular interest to the theme of this thesis are approaches to supporting non-visual interaction with graph-based diagrams. Unlike line graphs and tables, which typically convey numerical data, such forms of representation depict concepts that encode other types of information. For example, directed graphs and maps use icons and graphics to show objects, events, relations, movements and so on. It is harder to define a direct mapping between the information represented in this kind of visual displays and acoustic dimensions such as pitch or amplitude. Thus, the aforementioned sonification technique cannot be directly applied to translate such visual displays into sound, and the challenge is to develop intuitive models to support non-visual interaction with the visually encoded information. Current approaches addressing this issue employ one or a combination of two distinct models of representation; *Spatial* or *Hierarchical* models, which differ in the degree to which they maintain the original representation when translating its visual content (Mynatt and Weber, 1994) and produce dramatically different non-visual displays.

2.4.1 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 spatial arrangement. These are preserved and projected over a virtual or a physical space so that they could be accessed through an alternative modality. Because audio has limited spatial resolution, spatial models typically combine the haptic and audio modalities to support non-visual interaction with visual displays.

Example applications. 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. Using a pointing device, such as a stylus or a haptic mouse, and moving it around the projection space or tapping it at specific locations allows for non-visual exploration of otherwise visual content. As the pointing device traverses the plane, the element currently under focus can be displayed through haptic and auditory output. This technique has been found to be particularly efficient in supporting overviewing of tabular data (Kildal and Brewster, 2006) and for exploring other forms of graphs such as pie charts (Wall and Brewster, 2006) and bar charts (Ramloll and Brewster, 2002b).

To support non-visual interaction with node-and-links diagrams, Blenkhorn and Evans (1998) used a spatial model of representation where a diagram's nodes and connections are organised on an N$\times$N grid laid out on a touch pad (where N corresponds to the number of nodes). Users could

(a) Numerical table (b) Pie chart (c) PHANTOM Omni haptic device

Figure 2.1: Example prototypes of non-visual interfaces employing a spatial model of representation for accessing (a) tabular data (Kildal and Brewster, 2006); (b) and pie charts (Wall and Brewster, 2006). (c) A haptic device for virtual interaction with bar graphs.

trace horizontal positions on the grid to locate nodes and vertical positions to locate connections, which were then displayed in audio. In general, the limitation of such approaches when handling complex visual displays makes the spatial model less suited for supporting interaction with this type of diagrams.

Other non-visual solutions employing a spatial model of representation are applications that 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[6] (Figure 2.1(c)). 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 (Avanzini and Crosato, 2006; Yu et al., 2003). This technique has been successfully used to provide access to plots of 3D functions (Fritz and Barner, 1999), line and bar charts (Yu et al., 2003) and for creating bar graphs (McGookin and Brewster, 2007a) and abstract images (Rassmus-Gröhn et al., 2007).

Spatial models and collaboration. One of the main arguments in favour of using spatial models for non-visual interaction with visual displays is support for collaboration. According to this view, visual displays are themselves spatial representations of information, and using a spatial model would allow sighted and visually impaired users to maintain coherent mental models of shared representations (Mynatt and Weber, 1994; Petrie et al., 1995). Consequently, this would provide a common frame of reference and, hence, promote efficient communication between

[6] Sensable Technologies, `http://www.sensable.com`

collaborators when they perform joint tasks.

Although studies of collaboration between sighted and visually impaired users have been sparse, there is evidence that using a spatial model allows for generally smooth execution of joint actions (Winberg and Bowers, 2004; Crossan and Brewster, 2008). However, the claim that supporting coherence of mental models guarantees successful collaboration between sighted and visually impaired users has been contested; Winberg and Bowers (2004) write: *"[Coherence] in and of itself does not guarantee successful collaboration. It is essential to investigate how interfaces are used and how such uses are folded in with the varied things co-participants do (designing gestures, monitoring each other, establishing the state of things and ones orientation in it, reasoning and describing)"*(p.340). Thus, the importance of cooperation during collaborative action in such instances is a far more important factor of successful collaboration than mere coherence of representation (Winberg and Bowers, 2004; Winberg, 2006).

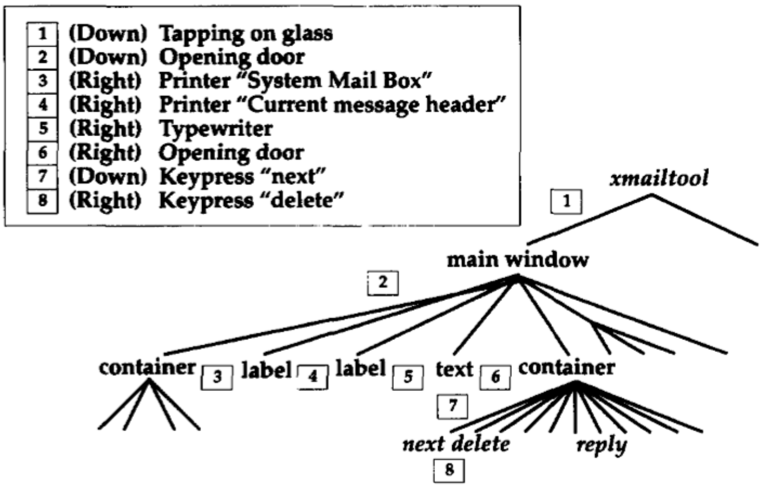
2.4.2 Hierarchical Models

A hierarchical model 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 structuring and organisation. For instance, phone-based interfaces support interaction by presenting the user with embedded choices, and although speech is typically used to display such choices, earcons have been thoroughly investigated to enhance navigation and interaction with hierarchical menus (Brewster, 1997; Leplatre and Brewster, 2000; Leplatre, 2002). Algebra expressions are another form of representation that lend themselves to hierarchical organisation and which have been made accessible in audio using earcons in combination with prosody (Stevens et al., 1997). In their work, Stevens et al. (1997) use a hierarchy-based structure to support non-visual browsing of algebra expressions mediated by a command language. Grouping structures were used whereabout complex objects that includes more than one term in an expression are only referred to by their spoken name at the base level, then, at the user's control, could be displayed in more details at deeper levels of the structure. The user is thus able to move *"into"* and *"out"* of objects to explore information spoken at different levels of complexity. Audio is thus the typical candidate modality for non-visual interaction with visual displays when employing hierarchical models of representation.

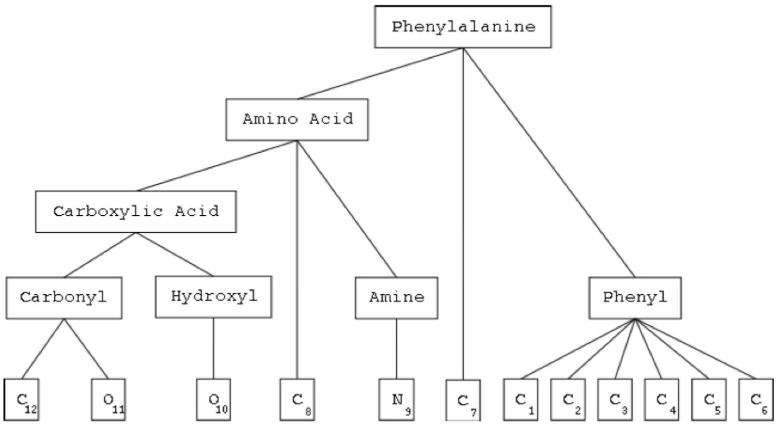
In relation to non-visual interaction with graph-based diagrams, hierarchical models have been used less frequently than spatial models. There are, however, a number of reasons why hierarchical models maybe more suitable for accessing such visual displays in audio. First, theoretical accounts suggest that perceptual representation is hierarchically organised such that visual form is analysed at, at least, three levels of organisation; an overall whole, then moving down to multisegment parts, before considering basic features (Palmer, 1977). According to this view, features of a visual form are selectively grouped together as perceptual units on the basis of connections, importance and contextual relevance. Such top down organisation could thus be captured through hierarchical groupings. In the realm of sounds, these accounts are supported by Auditory Scene Analysis (ASA) (Bregman, 1994), a branch of psychology that studies the perceptual organisation of sounds. ASA describes auditory perception in terms of organisational *Streams*, defined as *"a perceptual unit that represents a physical happening"* (Bregman, 1994, p.10) and used by the auditory perceptual system to construct meaningful elements through grouping principles. Second, a hierarchical organisation supports the notion that obtaining an overview should precede any exploratory interaction with a given dataset; a process expressed by Shneiderman (1996)'s Visual Information Seeking Mantra as: *"overview first, zoom and filter, then details-on-demand"*(p.337). This principle was later extended to the auditory modality when Zhao et al. (2004) proposed the Auditory Information-Seeking Principle of: *"gist, navigate, filter, and details-on-demand"*. By definition, a hierarchical structure could provide such structured organisation by enforcing a top down approach and presenting different levels of details at different levels of hierarchical depth.

Example applications. One of the early examples that used a hierarchical model to translate visual displays into a non-visually accessible representation is Mynatt and Edwards (1992)'s Mercator project. Like the GUIB project, the goal of Mercator was to provide access to X Windows applications for computer users who are blind. Mynatt and Edwards (1992) used a hierarchical structure to organise the components of a graphical display based on their functional and causal properties rather than their spatial pixel-by-pixel on-screen representations. A user could browse such a hierarchy and receive audio feedback that communicates its content using speech and auditory icons (Figure 2.2(a)).

The TeDUB system (Horstmann et al., 2004) combined hierarchical and spatial models of representation to provide non-visual access to technical drawings such as modelling and circuit diagrams and floor plans. User evaluations of TeDUB showed that using hierarchical groupings



(a) Mynatt's hierarchical model of an xmailtool GUI and its auditory icons.



(b) Brown's hierarchical model of a molecule diagram.

Figure 2.2: Example hierarchical models of visual displays; (a) representation of xmailtool interface (Mynatt, 1995); and (b) a molecule diagram (Brown, 2008).

reduced demands on short-term memory and facilitated overviewing, though the way in which information was hierarchically structured was found to cause issues of orientation and made the identification of information about related items on a given diagram difficult. Bennett (2002) extended the TeDUB approach by investigating non-visual strategies specific to node-and-link diagrams and addressing the issue of conveying relational information through a hierarchy. To this end, Bennett (2002) developed and contrasted two alternative navigation strategies for exploring schematic heating systems diagrams in audio; a hierarchical strategy, which emphasised structure, and a connection-based strategy, which emphasised relations. His findings showed that different types of tasks are best supported by a matching navigational model, and that augmenting the hierarchical model with information about spatial positioning – in the form of earcons –

provided no advantage in exploration tasks.

Brown et al. (2004) explicitly cites Palmer (1977)'s proposed model of visual perception as a motivation for developing a hierarchical model for non-visual interaction with diagrams (Figure 2.2(b)). His system, Kekulé, uses hierarchical structures to organise functional groups and structural features encoded in molecule diagrams and allows for exploration of such components in audio. Evaluations of his system showed that grouping of diagram components such that implicit features in data are made explicit reduced memory loads, and that allowing both hierarchical and connection-based browsing of molecular diagrams was most useful. As with the TeDUB system, participants in Brown et al. (2004)'s study found it hard to orient themselves and remember their position on the hierarchy, particularly when identifying previously visited nodes.

The reported orientation and navigation issues could be related to the inconsistencies of the hierarchical structures used for each diagram; in all of the above systems, the way diagram components are hierarchically organised depended primarily on the modelled visual display. That is, every time a new visual display is loaded onto a system, the user has no prior knowledge of how its content will be grouped and where each diagram element will be located on the hierarchy. Only when they browse to each construct of the hierarchy will such organisation be discovered and revealed. The approach developed in this thesis attempts to address this issue by imposing a semi-fixed hierarchical organisation and an indexing of diagram components that allows a user to build an accurate expectation of where diagram components will be located on the hierarchy. This will be described in detail in Chapter 3.

Hierarchical models and collaboration. To date, there has been no published empirical investigation that explored the effectiveness of using hierarchical models to support collaborative interaction. Thus, while the model has been found to successfully convey information in sound, it is not clear whether it is an effective means for supporting non-visual collaboration. Part II of this thesis will address this gap in the field by presenting a detailed study exploring audio-only collaboration using a hierarchical model of representation.

2.5 Summary and Conclusion

This chapter aimed to provide background about the use of audio in human-computer interaction and to address the problem of non-visual interaction with visual displays. Section 2.2 provided a brief historical account of the function of sound at the user interface before listing the potential advantages and disadvantages associated with using the auditory modality to convey information. Section 2.3 then introduced Auditory Display as a specialised field concerned with the use of sound to convey information. It defined existing techniques for displaying audio and subsequently the terminology that will be used throughout the remainder of this thesis.

Section 2.4 reviewed existing approaches to supporting non-visual interaction with visual displays such as GUIs and graph-based diagrams – which form the focus of the work presented in this thesis. In particular, existing approaches were classified with regards to the representational model used for translating visual displays into non-visual forms. A spatial model of representation captures spatial relations, forms and arrangements of the original display and is typically implemented using the haptic and auditory modalities. A hierarchical model, on the other hand, often emphasises the semantics of the original representation over their spatial arrangements and use parent-child grouping and structuring to organise such components. Spatial models have often been claimed to best support collaboration because they are more likely to support coherence of mental models between collaborators who use visual and non-visual displays. However, there isalso evidence in favour of the notion that cooperation is a far more important factor for successful collaboration. Additionally, there is a lack of empirical investigations into the effectiveness of hierarchical models in supporting collaboration; a gap in the field that will be explored in subsequent parts of this thesis.

The next chapter initiates Part I of the thesis, which develops and evaluates an approach for supporting audio-only interaction with relational diagrams using a hierarchical model of representation. The developed model will then form the basis for studying audio-only collaboration, thoroughly addressed in Part II of this thesis.

Part I

Audio-only Interaction with Diagrams

Chapter 3

Modelling Diagrams for Audio-only Interaction

3.1 Introduction

Various approaches to supporting non-visual interaction with visual displays were reviewed in the previous chapter and classified in terms of their representational models. Those employing a spatial model mixed the audio and haptic modalities to support interaction, while those employing a hierarchical model typically relied on audio as the only means of interaction. This chapter addresses Research Question 1 of this thesis by exploring how a hierarchical model of representation could be used to support audio-only interaction with a particular family of diagrams, referred to in this thesis as *relational diagrams*.

In order to translate a relational diagram into a hierarchical form that could be accessed and manipulated in audio, two questions are addressed. First, which information should be captured when translating it into a hierarchy? Second, how should the captured information be hierarchically structured? Section 3.2 reviews the nature of diagrams as a form of information representation in order to examine when and why they are better than other forms of representation. The review aims to understand functional properties that a diagram translation model should support. Section 3.3 then analyses the structural properties of relational diagrams in order to understand how they encode information. The analysis aims to provide a basis for determining which information to capture during a translation process. Section 3.4 uses the presented review and analysis to inform the design of a hierarchy-based model that translates a given relational diagram from the graphical to an audio accessible form. Section 3.5 summarises and concludes the chapter.

3.2 Diagrammatic Representation

Before attempting to translate diagrams into a non-visual form, it is important to understand their characteristics as a means for information representation. In particular, what are the properties of diagrams that give them advantages, if any, over other forms of representations? The following takes a closer look at such characteristics.

3.2.1 Representation

Representation is a key concept in a variety of disciplines. Consequently, the definition of what is exactly meant by *representation* depends on both the context and the domain of its use. For example, in the tradition of Semiotics, which is concerned with the study of signs and symbols, representation has been used to denote various aspects and types of signs and sign processes, while in that of Computer Science and Artificial Intelligence, it is often associated with the representation of knowledge by computational means (Nöth, 1997). A common distinction is also made between representation as a *process* and as the *product* of such a process. The process in this case involves the derivation of a representation from what is being represented, while the product refers to the structural characteristics of the resulting representation (Scaife and Rogers, 1996).

Palmer (1978) describes a representation as *"first and foremost, something that stands for something else [..] some sort of model of the thing (or things) it represents"*(p.262). Thus, according to this broad definition, a representation is essentially a carrier of information rather than the information itself because it captures aspects of perceptions, experiences and thoughts in some medium other than that in which they have occurred (Norman, 1993). Palmer (1978) further describes how such a definition implies the existence of two "worlds"; a represent*ed* and a represent*ing* world, with correspondences between the two. Indeed, a number of views account for representations as part of a system that consists of at least four parts; a represented world; a representing world; a set of relationships that determine how elements in the representing world come to stand for elements in the represented world; and a set of processes or interpreters which operate upon, use and maintain those relationships (Palmer, 1978; Rumelhart and Norman, 1985; Norman, 1991; Zhang, 1991; Markman and Dietrich, 2000).

A representation can take several forms, graphical, tangible or linguistic; it could be a word that refers to an object, an object that refers to a word, a diagram that describes a concept or a thought, or a picture that captures a scene. For example, one could pick up a book, place it on a table and use it to refer to the building one is currently in while instructing a colleague about the shortest route to a desired destination. Then, one could pick up a coffee mug and place it somewhere else on the table while describing how a monument should appear on the right if one's colleague walks for five minutes in the direction that one is gesturing. Then, while gesturing further to the side of the mug, one could describe how one's colleague should take the second right turn to arrive at a destination. The colleague could then retract, pointing to the sides of the book while enquiring about which of the building's many doors should they exit from, highlighting the existence of a construction site outside the eastern entrance of the building. The actual route, building, monument and doors in this scenario are the represent*ed* world; the table, book and coffee mug are the represent*ing* world. Each object on the table is a symbol that stands for a particular feature on the route, accounting for the representing relationships between the two worlds, and the two colleagues, their reasoning and gesturing are the interpreters and processes that act on and maintain these relationships.

Representations are powerful because, in a similar manner to the above scenario, they allow for the capture and manipulation of events and objects which are absent in time and space, including those that never actually existed (Norman, 1993). But representations are interpreted in terms of a particular context and with relation to prior knowledge – such as knowledge about the existence of a construction site in the above scenario. Consequently, representations tend to capture only a subset of the represented information in order to complement what is already known to the processes and interpreters that make use of them. For example, while representing a building and a monument, the sizes of the book and the coffee mug in the above scenario do not reflect the actual sizes of the building and the monument, nor do they capture their scale or distance relative to one another. Such information is actually irrelevant to the task of describing and enquiring about route directions. As shall be described in later sections, a good representation is one that captures aspects of the represented world that are of most relevance to the task at hand and to the knowledge and expertise of the individual or individuals using it.

3.2.2 External Representations

Norman (1988, 1993) refer to *knowledge in the head* and *knowledge in the world* to distinguish between what is known to the individual and what is captured through an artefact. This highlights two further types of representations; that which is *internal* to the mind of an individual and that which is *external* to it. The importance of the role that external representations play in cognitive problem solving, particularly in terms of how they interact with internal representations, has been increasingly stressed (Simon, 1980; Larkin, 1989; Norman, 1991; Vera and Simon, 1993; Zhang and Norman, 1994; Cox and Brna, 1995; Hutchins, 1995). This was considered a progress that marked a shift from the traditional approach to understanding human cognition, which focused on the individual's mental states as the sole frame of reference when analysing cognitive behaviour, towards a view that extends beyond the individual's mind to include references to a broader system of cognition (Scaife and Rogers, 1996). Describing the difference between internal and external representations, Zhang and Norman (1994) write:

> *"Internal representations are in the mind as propositions, productions, schemas, mental images, connectionist networks, or other forms. External representations are in the world, as physical symbols, (e.g. written symbols, beads of abacuses, etc.) or as external rules, constraints or relations embedded in physical configurations (e.g., spatial relations of written digits, visual and spatial layouts of diagrams, physical constraints in abacuses, etc.)"* (p.89).

Norman (1991) uses the term *cognitive artefacts* to refer to information that is externally represented, defining them as *"artificial devices that maintain, display, or operate upon information in order to serve a representational function and that affect human cognitive performance"* (p.17). Norman (1991) argues that the way that information is carried by external cognitive artefacts is as important to the achievements of a task as the information that resides in the mind of the individual. Similarly, Simon (1980) points out that external representation assist the human capabilities by changing the nature of the task so as to make its solution transparent. A simple but effective demonstration capturing this interaction between external and internal representations is presented in (Norman, 1993, p.53-55), and will be described here for illustration.

Consider playing the Game of 15 [1]. In a sample game, player A starts by taking the digit 8, player B then takes 2, then A takes 4, and B takes 3, and then A takes 5. At this point of the game, what should be the next move of player B to prevent their opponent from winning? To answer this question, one must keep track of which digits player A has, which digits player B has, what these sum up to, what digits remain in the list, and calculate which of the remaining possibilities would

[1] In the Game of 15, players take turns in choosing a digit from the set $\{1, 2, 3, 4, 5, 6, 7, 8, 9\}$. Once a digit is taken it cannot be used again, and the first player to collect three digits that sum to 15 wins.

result in a winning move for each player. Most of this would be done internally, and although the arithmetic is simple, keeping track of all the possibilities while doing the calculations could make the game a difficult one.

Now, consider the Naughts and Crosses game [2] shown in Figure 3.1 where player A is **X** and player B is **O**, and it's player B's turn. What should be the next move of player B to prevent their opponent from winning? The answer should be fairly simple; player B should put an **O** on the bottom right corner of the grid to prevent player A from forming a diagonal line of **X**s.

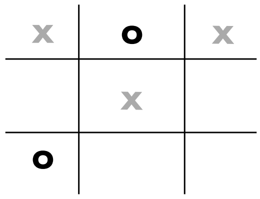


Figure 3.1: External representation of the Naughts and Crosses game.

Norman (1993) describes how the answer to the first question is hard because it requires unaided reflection, while the answer to the second question is easy because it could be answered through perceptual experience. By simply glancing at the tabular grid, a player could see the correct move. The answers to the two questions were therefore achieved through different means; the first was *reflective* the second was *experiential*. Using a different representation, the sequentially presented moves of the Game of 15 could also be displayed so that the task required to solve it changes from a reflective to an experiential one. This can be achieved if, instead of a serial representation of the game moves, the 9 digits are arranged into a 3-by-3 tabular grid so that every three digits on the horizontal, vertical and diagonal lines of the grid sum up to 15 as shown in Figure 3.2 .

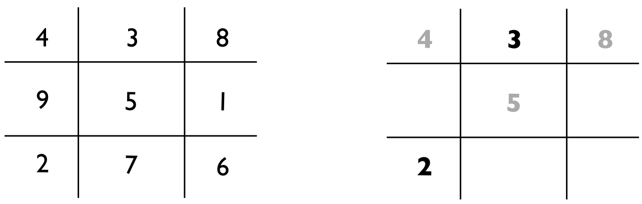


Figure 3.2: External representation of the Game of 15 as a tabular grid.

[2] In a Naughts and Crosses game, two players take turns in placing either an X or an O on a grid. The first player to align three of their symbols in a straight horizontal, vertical or diagonal line wins.

Thus, given this new representation, it is clear that the first and second questions posed above were actually identical, and could be solve by similar experiential means; i.e. looking at the grid and determining the right move. The choice of how the Game of 15 is represented dramatically changes the nature, and thus difficulty of the task required to solve the problem. The first representation required reflection, while the second representation made the solution transparent.

As illustrated by this example, external representations have a number of properties that makes them an indispensable part of problem solving. Zhang and Norman (1994) summarised them as follows:

- *External representations are external memory aids.* For example, instead of having to remember each player's set of digits in the above example, these could be remembered by simply glancing through the grid.
- *External representations can provide information which can be directly perceived and used without being interpreted and formulated explicitly.* For example, the missing digit for completing the sum of the diagonally arranged digits could be directly perceived on the grid.
- *External representations anchor and structure cognitive behaviour.* For example, the grid representation constrained the possibilities of correct moves.
- *External representations change the nature of a task.* For example, a grid representation changed the task of solving the Game of 15 from a reflective task to an experiential one.

3.2.3 Diagrams as External Representation

Diagrams are a form of external representation, and thus inherit the properties described above. But while a representation can take several physical forms, diagrams form the middle part of a continuum between two particular classes of cognitive artefacts; namely, text and pictures (Blackwell, 1998).

The exact definition of what constitutes a diagram is hard to pin down, however. At a fundamental level, a diagrammatic display can be regarded as an arrangement of various graphic elements in space (Cheng et al., 2001). But this blurs the distinctions between elements of the continuum when moving from text towards pictorial graphics. For example, text characters are themselves graphical marks, and a written document that would normally be classified as textual may contain diagrammatic features, such as grouping of paragraphs, indentation, highlighting or other typographical enhancements. Conversely, pictorial and graphical structures are often accompanied by short strings of text such as labels and descriptions. How is it then possible to classify any form of 'mark-making' as diagrammatic?

Fish and Scrivener (1990) suggests that the difference between the textual and the purely graphical could be captured in terms of the trade-offs between *description* and *depiction*, and hence the level of ambiguity in the interpretation of a given representation. According to Fish and Scrivener (1990), text is a descriptive representation that is accessed through a fixed set of vocabularies, its range of expression could thus have a well defined set of interpretation rules. On the other hand, a diagram is a depictive representation, its range of expression is intrinsic to its analog mappings, and so may suffer from ambiguity of interpretation. It is of course possible for text to be ambiguous and for diagrams to have formal definitions of possible interpretations (Gurr et al., 1998; Gurr, 1999). Nevertheless, how information is extracted and interpreted from sentences and diagrams has been thoroughly investigated as a criteria for discriminating between the fundamental characteristics of these two types of representations (e.g Stenning, 2000).

Representations can also be contrasted in terms of informational and computational equivalence. According to Larkin and Simon (1987), two representations are informationally equivalent if *"all the information in the one is also inferable from the other, and vice versa"* (p.67), and computationally equivalent if *" they are informationally equivalent and, in addition, any inference that can be drawn easily and quickly from the information given explicitly in the one can also be drawn easily and quickly from the information given explicitly in the other, and vice versa"* (p.67). The serial and grid representations of the sample Game of 15 described in the previous section are examples of two informationally equivalent but computationally different representations. This is because they contained the same information but differed in terms of the computations required to solve the game.

Larkin and Simon (1987) acknowledge that "easily" and "quickly" are not precise terms on which to base judgment, and suggest that instead focus should be placed on how representations organise and structure data, as well as how they support the processes that operate on them. They thus examined the computations required to solve mathematics and physics problems using informationally equivalent sentential and diagrammatic representations and focused on three main processes; *search*, *recognition* and *inference*. They reported the following conclusions:

Search. Information represented with diagrams is indexed by spatial location and related items are often grouped together, which contributes to reducing the amount of search required during problem solving. Although it is possible to jump to different parts of a sentential representation, searching text is predominantly linear, lacks coordinates, and unless symbolic labels are explicitly added to text, locating related items on written text can be difficult.

Recognition. Recognising features is considerably easier with diagrams than it is with sentential representations. Larkin and Simon (1987) illustrate this by contrasting the diagram in Figure 3.3 with the equivalent sentential description: *"Two transversals intersect two parallel lines and intersect with each other at a point X between the two parallel lines. One of the transversals bisects the segment of the other that is between the two parallel lines"*(p.82). Once drawn, the diagram immediately shows two triangles that form as a result of the intersections (circled on the figure), even though there was no mention of any "triangle" in the textual description. Features which were implicit in the text were thus made explicit through the diagram.

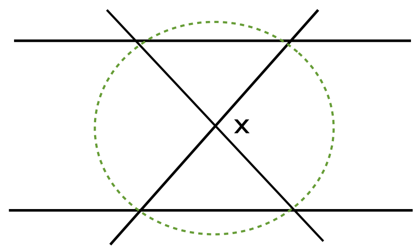


Figure 3.3: A diagram illustrating a geometric problem description (Larkin and Simon, 1987, p.83).

Inference. Larkin and Simon (1987) report no evidence for differences between sentential and diagrammatic representations in terms of easing inference, but this finding is contested by a number of views. For example, Shimojima (1996) uses the term *inferential free-rides* to describe how diagrams automatically include the representation of some conclusions that could be easily inferred when solving syllogisms, and Bauer and Johnson-Laird (1993) demonstrate that solving double disjunction problems, where a reasoner has to bear in mind various possibilities, is significantly quicker when using diagrams. These conclusions are also related to the aforementioned accounts of how external artefacts support problem solving by changing the nature of the task (Simon, 1980; Norman, 1993).

Additionally, Gurr et al. (1998) describes reasoning with diagrams as a two stage process, involving first constructing the required diagram and then reading off the relevant conclusions, and Scaife and Rogers (1996) write that *"diagram production and comprehension are intimately related"* (p.208). According to these views, not only is inference eased with diagrams, but it also linked with the ability to actively construct and manipulate diagrams. In summary, the following are important functional properties that make diagrams a useful form of external representation:

- *Search.* Diagrams make searching for related items easy by grouping them together using locational indexing.
- *Recognition and Inference.* Diagrams make implicit features of the information explicit, rendering certain conclusions more apparent than others.
- *Interactivity.* Diagrams that are interactively created and manipulated support reasoning and inference.

3.3 Analysing Relational Diagrams

Ideally, a model that supports non-visual interaction with diagrams should aim to support the ability to interactively search and locate diagram items, recognise explicit diagram features and interactively access and manipulate such features in order to better support inference. However, as well as supporting such functions, it is important to determine which information should be captured about a given diagram. That is, which information should be interactively searched, recognised and used for inference. This section presents an analysis that provides an approach to answering such a question.

As described in Section 3.2.3, diagrams form the middle part of a continuum between text and pictures. While a wide range of diagrammatic representations exist within this continuum, the work presented in this thesis, and hence the following analysis and the model described in this chapter, will focus on one particular family of diagrams, referred to as *relational diagrams*.

3.3.1 Relational Diagrams

A relational diagram as referred to in this thesis is any diagram that depicts relations between items in a particular domain. In the mathematical sense, relational diagrams could also be classified as part of the family of *Directed Graphs*, which depict sets of nodes or vertices ordered through a set of paired connections called arcs, directed edges or arrows (Harary et al., 1965). This family of diagrams is very common and might be encountered in everyday life as much as in education and the workplace. For example, railway connection networks, underground trains and bus routes are often depicted as connected nodes, and visual programming languages often use box-and-line or nodes-and-links diagrams to represent relations between program constructs.

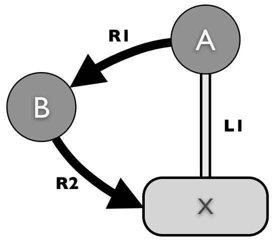


Figure 3.4: An example of a relational diagram.

Figure 3.4 shows a fictitious diagram[3] exemplifying the type of diagrams addressed in this thesis. In this example, items *A* and *B* are connected with a relation labeled *R1*, items *B* and *X* are connected with a relation labeled *R2*, and items *A* and *X* are connected with a relation labelled *L1*. Note that the relations *R1* and *R2* have directions, from *A* to *B* and from *B* to *X* respectively, while relation *L1* does not. Unlike directed graphs, in which directionality is strict, relational diagrams can depict relations between items without necessarily enforcing a particular direction. For instance, *A* and *X* could represent rail stations and *L1* the connecting line, in which case the direction of the relation is insignificant if the train moves in both ways of the line. By the same token, items *A* and *B* could represent structural constructs of a program or a system, and direction may represent relational dependencies, in which case its specification is significant.

3.3.2 Relational Information Displays

Zhang (1996) developed a taxonomy that unifies a variety of external representations under a common form known as *Relational Information Displays* (RIDs). RIDs are displays which represent relations between dimensions, and include various forms of diagrams including line graphs, bar charts, maps and tables. Zhang (1996)'s taxonomy has been used to analyse the structural properties of various displays (Zhang and Norman, 1994, 1995; Zhang, 1996) and can thus serve as a means for analysing the structural properties of relational diagrams in order to determine which information should be captured and modelled for audio-only interaction.

At the core of this taxonomy is the analysis of RIDs in terms of their scale types and dimensional representations.

[3] The terms *diagram* and *relational diagram* will be used interchangeably throughout the reminder of this thesis.

3.3.2.1 Scale Types

According to Zhang (1996), dimensions are the basic structures of RIDs. In general, values along a given dimension are related to each other through a particular mathematical property to form the scale of that dimension. Stevens (1946) identified four major scale types that are commonly used today to measure any physical or abstract property of a dimension; these are *nominal*, *ordinal*, *interval* and *ratio* scales. Table 3.1 summarises the relationships between scale types and their properties.

Table 3.1: Scale types and their formal properties (from (Zhang, 1996)).

	Formal Property			
Scale Type	Category	Magnitude	Equal Interval	Absolute Zero
Nominal	√	×	×	×
Ordinal	√	√	×	×
Interval	√	√	√	×
Ratio	√	√	√	√

Nominal. Nominal scales measure one property, *category*, and are thus used to distinguish items from one another without capturing any information about their magnitudes, intervals or ratios. Colour is an example of a nominal scale as it can be used to categorise a set of items but it does not carry any further properties.

Ordinal. Ordinal scales measure two properties, category and *magnitude* and can thus be used to not only distinguish between different items but also to assign values to them based on their ranking relative to one another. For example, a teacher may use a scale of 1 to 5 to rank students on the basis of their conduct in the classroom from 1 (poor) to 5 (excellent); while a 1 is both different and worse than a 2, there is no indication that 2 is twice as better as a 1, nor whether the improvements in conduct from, say, 2 to 3 is the same as the improvements from 4 to 5.

Interval. In addition to category and magnitude, interval scales have the property of *equal interval*, where the unit distance between any two items is the same regardless of where on a scale the two items are. Temperature is an example of an interval scale since it is possible to distinguish between different temperatures (category); to judge any two temperatures as either greater than, less than or equal to one another (magnitude); and the difference between, say, 10 °C and 15 °C is the same as the difference between 32 °C and 37 °C (equal interval).

Ratio. Ratio scales have four properties, category, magnitude, equal interval and *absolute zero*. Particularly, and unlike the temperature scale, the value “0” on a ratio scale indicates the absence of the measured dimension. This implies that it is possible to make ratio judgements between two items on a given scale. For example, when measuring length, 1cm is different from 4cm (category), 20cm is longer than 5cm (magnitude), the difference between 2cm and 4cm is the same as the difference between 10cm and 12cm (equal interval), and 0 cm implies a non-existing length (absolute zero) hence 4cm is twice as long as 2cm.

Applying a scale types analysis to the relational diagram in Figure 3.4, it is possible to conclude that the diagram represents items using nominal scales. That is, items on the diagram can be distinguished from one another on the basis of at least two categories, *names* or *labels* and *shapes*. The nominal scale of *names* distinguishes between six categories of items – albeit only one item per category – *“A”*, *“B”*, *“X”*, *“R1”*, *“R2”* and *“L1”*. The nominal scale of *shapes* distinguishes between four categories of items or types of shapes; $\{A, B\}$, $\{R1, R2\}$, $\{X\}$ and $\{L1\}$ [4].

3.3.2.2 Dimensional Representations

Dimensional representations refer to the implementation of the scale types using different physical dimensions such as shape, distance, direction and texture (Zhang, 1996). For example, the category type of $\{A, B\}$ on the diagram shown in Figure 3.4 is implemented using textured circular shapes, while the category type of $\{R1, R2\}$ is implemented using a textured geometrical shape that combines a line and a triangle to form arrows and directions. Different colours and shapes on a relational diagram could implement different categories of items, and could also be used to represent other properties, such as size to distinguish between magnitudes, just as different textual markings could distinguish between different items within a particular category.

There are two implications to this analysis. First, the fundamental information conveyed through a given relational diagram resides in its scale types rather than the dimensional representations implementing these scale types. That is to say, it is possible to use different shapes, colours and labels to represent the same categories on a given scale type while still preserving the represented relational information of the diagram. For example, diagrams (1) and (2) on Figure 3.5 use different physical implementations for each category type; star and diamond shapes

[4] If the relational diagram in Figure 3.4 was part of an actual domain, the *shapes* category type would stand for actual items in the represented domain; for example $\{A, B\}$ could stand for train stations or program constructs, and $\{R1, R2\}$ for the connecting lines or exchanged processes.

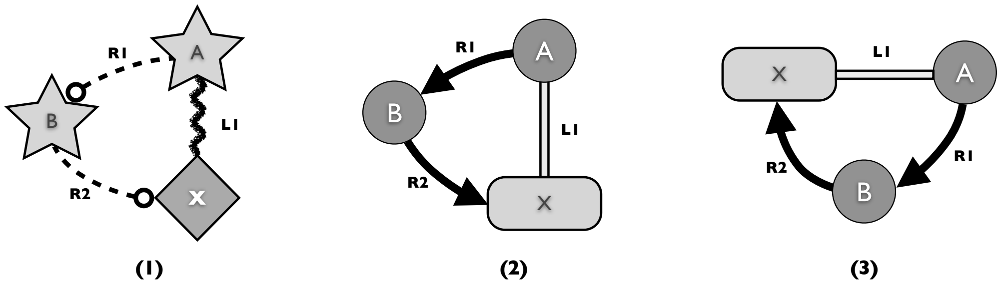


Figure 3.5: Informationally equivalent relational diagrams.

instead of circles and rectangles for the nodes, and dotted or zigzagged lines instead of straight lines for the links or connections, yet the two diagrams remain informationally equivalent. With this in mind, it is possible to consider the information represented through a given relational diagram as two sets of features with the following structural properties:

- *Modality-independent features* that are captured through the scale types without being specific to any particular medium of presentation, and
- *Modality-dependent features* that are captured through the dimensional representations of the scale types and, in the case of visual diagrams, implemented using features that are accessed through visual apparatus.

Second, unless spatial location is itself a represented dimension, varying the spatial distribution of the dimensional representations of a diagram does not affect the represented relational information. For example, diagrams (2) and (3) on Figure 3.5 are informationally equivalent even though diagram items are placed at different locations. As might be expected, while informational equivalence could be preserved when varying dimensional representations and spatial distributions, it is likely that such variance would have an impact on the computational requirements for searching, recognising and inferring information from different relational diagrams[5].

In summary, the above has described the *functional properties* of diagrams, which consist of easing *search* for related information, supporting *recognition* of features that are otherwise implicit, and *inference*, particularly when diagrams are *interactively* constructed and manipulated. Furthermore, the above analysis described the *structural properties* of the information conveyed through relational diagrams as a combination of *modality-independent* and *modality-dependent*

[5] Refer back to the examples described in Section 3.2.2 on page 33 and Section 3.2.3 on page 36 for a discussion on computational versus informational equivalence.

features. Section 3.4 will describe how a hierarchical model could be constructed to support non-visual audio-only interaction with relational diagrams while taking into consideration the particularities of their functional and structural properties.

3.4 Hierarchical Modelling of Relational Diagrams

A relational diagram could be accessed in audio through a spoken description of its content. Indeed, existing guidelines for providing non-visual access to web content emphasise that graphic elements should be accompanied by descriptions of their content, which can be conveyed through the ALT attribute on the IMG element in HTML (Petrie et al., 2005). A user accessing a web page through a screen reader could detect the ALT-text attribute and have the accompanying description read out in speech. As an example, diagram (2) on Figure 3.5 could have the following textual description:

> *" Two circular nodes A and B and one rectangular node X. A is connected to B via a relation R1 that points to B, and B is connected to X via a relation R2 that points to X, R1 and R2 are of the same type. A is connected with X via a relation L1. L1 is of a different type to R1 and R2 and has no direction."*

As highlighted in Section 3.2.3, there are computational differences between using sentential and diagrammatic representations and these differences are likely to translate to spoken text. This is because accessing diagram content by passively listening to its description imposes a linear presentation of the information it contains, which is likely to make it difficult to search for, locate or compare different parts of the represented information. A difficulty that could only increase as the complexity of the diagram increases resulting in lengthier descriptions.

There are other issues that are likely to impact the efficiency of accessing a relational diagram through its spoken description. First, because of the linearity of presentation, the listener is constrained by the order in which information is described, which might not always match the order in which the listener wishes to access the diagram. This leads to a second issue, which is determining an appropriate order for describing diagram content, and, indeed, determining what constitutes an appropriate description at all given that a relational diagram could be described in more than one way. Additionally, any given description has to be produced by an individual or an automated process, both of which require a certain level of domain expertise, which increases the variability of potential descriptions and decreases the autonomy of the listener.

3.4.1 Scale Type Perspectives

This thesis proposes to use the scale types represented by a given diagram as an organising factor for producing descriptions of its content. For instance, the relational diagram (2) on Figure 3.5 represents information using the nominal scales of *names* and *shapes* and could thus be described from the perspectives of the information associated with each value on such scales. Essentially, given a value on a represented scale type, it is possible to determine which relational information could be implied from it and hence produce a description of such information from the perspective of that value.

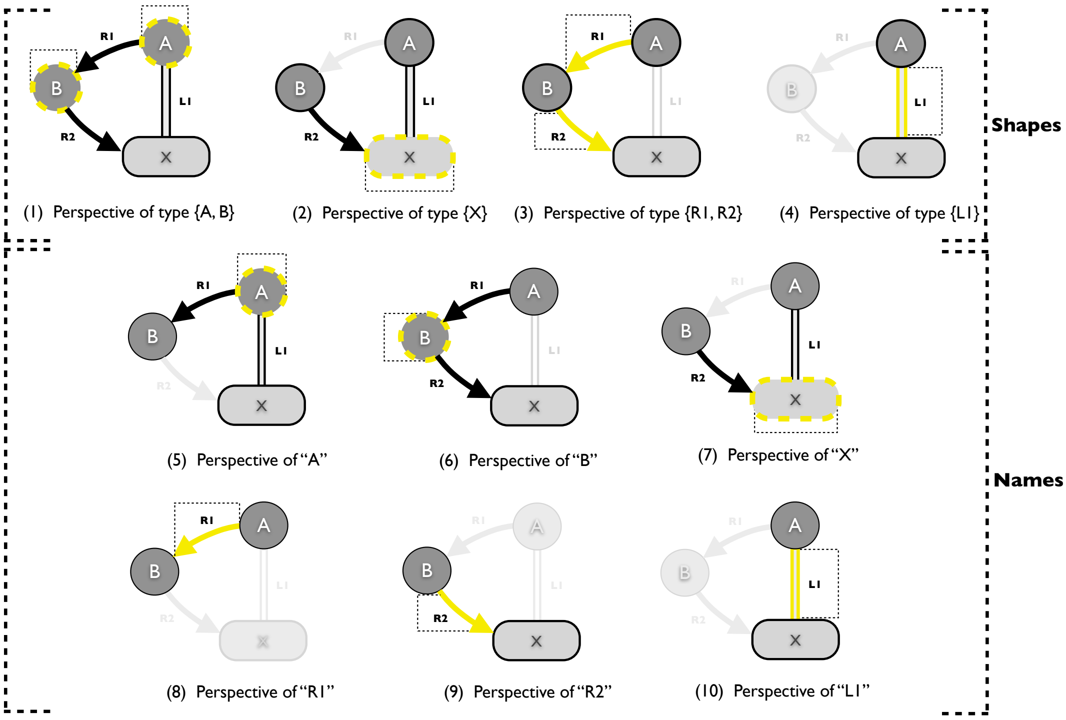


Figure 3.6: Diagram content as captured from the perspective of each value on the *names* and *shapes* nominal scales. Grey shaded elements show items that are overlooked by a given perspective. Bracketed numbers correspond to the numbered rows on Table 3.2.

The following example illustrates this procedure. Consider the *Shapes* nominal scale, which, in the case of the example diagram shown on Figure 3.6, has four values that correspond to the four types of shape categories: 1) value {A, B} represented with circular shape, 2) value {X} represented with rectangular shape, 3) value {R1, R2} represented with directional connections, and 4) value {L1} represented with non-directional connections. It is then possible to produce a description of the relational information that could be implied from each of such values; for instance, starting from the value {X} it is possible to capture two sets of relational information;

Table 3.2: Captured and overlooked relational information as described from the perspective of each value on the represented scale types of relational diagrams. Bracketed numbers correspond to the numbered diagrams on Figure 3.6.

Scale Type	Value	Captured/ Described	Overlooked
Shape	(1) {A, B}	A to B via R1, A linked to X via L1, B to X via R2, B from A via R1	—
	(2) {X}	X linked with A via L1, X from B via R2	A to B via R1
	(3) {R1, R2}	R1 from A to B, R2 from B to X	A linked with X via L1
	(4) {L1}	L1 linking A and X	A to B via R1, B to X via R2
Names	(1) "A"	A to B via R1, A linked with X via L1	B to X via R2
	(6) "B"	B from A via R1, B to X via R2	A linked with X via L1
	(7) "X"	X linked with A via, X from B via R2	A to B via R1
	(8) "R1"	R1 from A to B	B to X via R2, A linked with X via L1
	(9) "R2"	R2 from B to X	A to B via R1, A linked with X via L1
	(10) "L1"	L1 linking A and X	A to B via R1, B to X via R2

that B is connected to X via R2 and that A is connected to X via L1. Figure 3.6 and Table 3.2 show the relational information that is captured from the perspectives of the remaining values on the *Shapes* and *Names* nominal scale types. Organising a diagram's description using its scale types in this way has two main characteristics. First, each instance on a given scale type (as numbered on Figure 3.6 and Table 3.2) emphasises certain aspects of the diagram while overlooking others. For example, the relational information captured from the perspective of the category type $\{R1, R2\}$ – instance (3) – emphasises the the connections between nodes *A* and *B* and between *B* and *X* but overlooks the connection *L1* between nodes *A* and *X*. Second, the same relational information is captured from more than one perspective. For example, the information associated with the connection *L1* is captured from the perspective of the category type $\{A, B\}$ – instance (1) – which includes the following descriptions: *"A pointing to B via R1, A connected with X via L1, B pointing to X via R2"*, as well as from the perspective of the category type $\{L1\}$ – instance (4) – which includes the description: *"connection L1 linking A and X "*.

Since each perspective describes a particular set of values on the represented scale types and overlooks others, all perspectives should be somehow combined within a single model to provide access to a more complete "picture" of the modelled diagram. Sections 3.2.2 and 3.2.3 have already shown how the nature of a task is changed by the perspectives emphasised through the diagrammatic representation that is used to assist such a task. The relationship between task and representation is, however, not exclusive to graphical representations. Evidence from several evaluations of non-visual interfaces reported similar relations between task and artefact when users are provided with the ability to access the same information from more than one perspective (Bennett, 2002; McGookin and Brewster, 2006b). For instance, Bennett (2002) shows that the efficiency of interacting with central heating diagrams through an auditory interface is best supported when the nature of the task matches the perspective of the representation model. Providing access to the same relational information from various perspectives could therefore be potentially useful to support different types of tasks.

Additionally, it is possible to include a description of the positions of diagram items within such perspectives. For example, items' positions could be described in relation to one another as: *"A is at the top of the diagram, B is at the bottom left of A and X is directly below node A"*, although it is easy to imagine that such an approach might be inappropriate for very complex diagrams. Another possibility is to describe the actual *X* and *Y* coordinate values of items' positions on a two-dimensional (2D) plane to convey their positions. However, as mentioned in Section 3.3.2.2 and shown on Figure 3.5, this information is not always important. Unless positional information is itself a represented dimension, it is often a visual convenience rather than an inherent part of the represented information in a relational diagram. Investigating the usefulness of such an approach, Bennett (2002) used earcons to include coordinate information in auditory representations of central heating diagrams and found that such information provided no advantage when users performed navigation tasks. It seems that discarding such information from a translation model does not impact its informational equivalence with the original diagram, but impact of discarding it on their computational equivalence is unclear.

3.4.2 A Multiple Perspective Hierarchy-Based Model

If using the perspectives of scale type values to produce descriptions provides a potential approach to modelling diagrams for audio-only interaction, it is only half a solution. The other half should address the problem of how to navigate through such descriptions in a way that facilitates searching for and locating information. As described above, this is problematic with a spoken diagram description due the linearity of presentation. This section describes how a hierarchical model could be developed to address such a problem.

What is noticeable from an examination of the modality-independent features of the relational diagram shown in Figure 3.6 is that they lend themselves to hierarchical organisation. Specifically, the nominal scale of *shapes* could be used to group together similar values on the nominal scale of *names*. For instance, the values *"A"* and *"B"* on the *names* scale form the category type $\{A, B\}$ on the *shapes* scale because both values are implemented using the same dimensional representation; a circular shape. Similarly, the values *"R1"* and *"R2"* form the category type $\{R1, R2\}$ because they are both implemented using directional arrows, and so on.

This relationship between the two nominal scales could be exploited to create hierarchical groupings as shown in Figure 3.7. In this structure, values of a given nominal scale occupy a unique branch on a hierarchy to group together information about individual values within corresponding scale types. Thus, the more categories there are in a relational diagram (i.e on the *shapes* nominal scale) the more branches there would be at level 1 of such a hierarchy, and the more items there are within a particular category type (i.e on the *names* nominal scale) the more branches there would be at level 2 of the hierarchy.

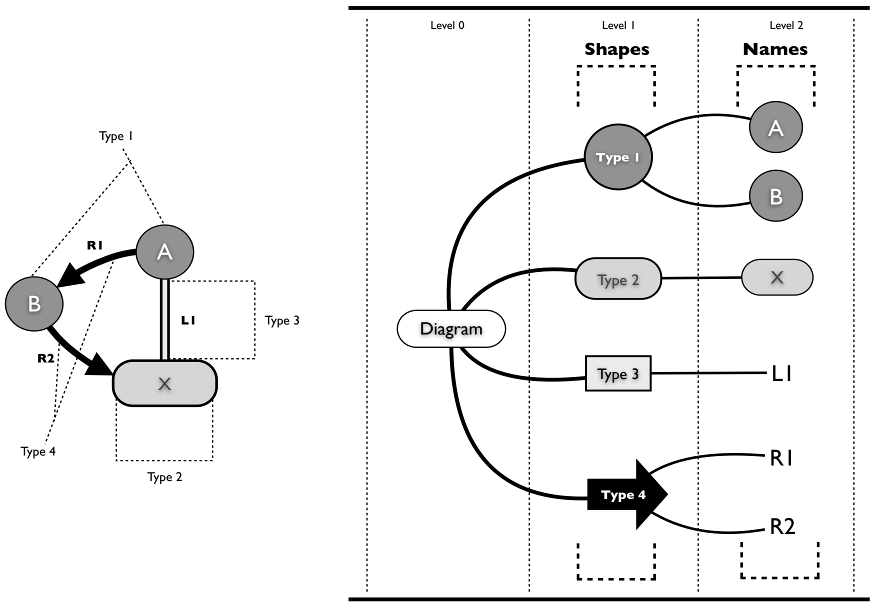


Figure 3.7: Hierarchical grouping of represented scale types.

Using these levels of hierarchical groupings, diagram items are indexed on the basis of their scale types and could be searched and explored by allowing a user to navigate through such a structure. Once located on the hierarchy, the relational information associated with each value of the grouped scale types is then attached to it using the same principle of grouping by scale type. This is shown in Figure 3.8. For instance, it is possible to locate the category type $\{A, B\}$, then

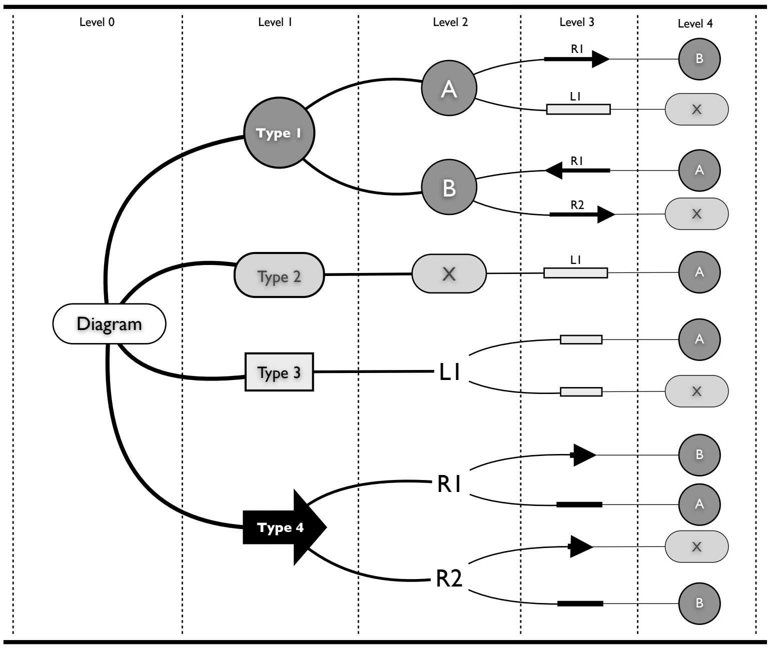


Figure 3.8: Multiple perspective hierarchy.

proceed to the value *"A"* within that group, then explore the relational information associated with *"A"*, which in this case includes two connection types *"R1"* of type $\{R1\}$ and *"L1"* of type $\{L1\}$, and are thus represented with two distinct sub-branches on Level 3 of the hierarchy, and each connecting the value *"A"* to two distinct shape types, also represented with distinct sub-branches on Level 4 of the hierarchy: *"B"* and *"X"* respectively. Thus, the diagram descriptions associated with each perspective, as listed on Table 3.2, are *spread* across the various levels of this strucrure to facilitate searching for and locating the information they contain; e.g *"a circular node (level 1) "A" (level 2) connected via a relation "L1" (level 3) to a circular node "B" (level4)"*. The result is a multiple perspective hierarchy that organises the information encoded in a given relational diagram through a grouping structure using the represented scale types as an organisational index.

The proposed model can support auditory access to the content of a relational diagram by allowing a user to navigate through such a structure and displaying auditory feedback that convey information about encountered items. Using this model, a user should be able to interactively explore diagram content as captured through the perspectives of the represented scale types. That is, rather than being constrained by a linear presentation of a diagram's description, the user can be in control over which part of the description they wish to be displayed as they move through each level of the multiple perspective hierarchy.

3.4.3 Reflections on the Proposed Model

As mentioned in previous sections of this chapter, a good representation is one that captures aspects of the represent*ed* world that are of most relevance to the task it is designed to support. The multiple perspective hierarchical model is thus a representation of a representation as it attempts to capture relevant information about a given relational diagram and present it in way that could potentially be accessed and manipulated using an alternative modality. To achieve this, the representational analysis which described the structural properties of a relational diagram is used to capture its modality-independent features. That is, features that form the underlying information represented by a diagram rather than those related to its implementation using visual means.

The model then attempts to address the problem of supporting the functional properties of the original visual diagram that it translates; i.e searching, recognising and inferring knowledge through interaction with a representation. As indicated by (Simon, 1980) a representation assists the human capabilities by changing the nature of the task to match them. The proposed model follows this notion by changing the task of searching for and exploring diagram content from a process that relies on locational indexing on a 2D plane to one that is based on browsing lists of grouped items. The use of a scale type indexing in this context is an attempt to impose a grouping structure that facilitates the process of locating information of interest and orientation around the captured information. In this sense, the model follows Larkin and Simon (1987)'s approach of focusing on how a representation organises and structures the data it conveys in order to support processes of searching, recognition and inference.

While the scale types grouping attempts to support ease of searching for and locating information, the processes of recognising and drawing inference from such information are not explicitly supported through the proposed model. The process of recognising information would ultimately depend on the auditory display techniques used and how effectively they convey the hierarchically structure content of a diagram. In relation to inference, such a process is only implicitly supported in that an implementation of the proposed model would provide capabilities for interactively accessing and manipulating its content, i.e capabilities which were described to be intimately related to the ability to draw inferences from diagrammatically represented information (Gurr et al., 1998; Scaife and Rogers, 1996). The question of how effective is the proposed model in supporting the abilities to search through, recognise and interactively access and manipulate auditorally displayed information forms the core of Research Question 2 of this thesis, which will be empirically addressed in the remaining chapters of Part I.

3.5 Summary and Conclusion

This chapter motivated and presented an approach to modelling relational diagrams for audio-only interaction. The approach is based on understanding the functional and structural properties of diagrams as a form of external representation. In terms of their functional properties, diagrams facilitate searching for and recognising features within represented information using locational indexing. They help inference processes by making some conclusions more apparent than others, particularly when they are interactively constructed and manipulated.

An analysis of diagrams as a RIDs concluded that the structural properties of any given relational diagram are a combination of two sets of features; 1) modality-independent features that are captured through the represented scale types without being specific to any particular medium of presentation, and 2) modality-dependent features that are captured through the dimensional representations of the scale types and made accessible through visual apparatus. It was concluded that the former set of features should be the focus of the developed hierarchical model because they could be implemented using alternative structural tokens that are appropriate to the desired modality of translation.

To model such functional and structural properties, diagram descriptions were developed from the perspective of each value on the scale types represented by a relational diagram. Each of the produced descriptions captured aspects of the diagram and overlooked others, with similar relational information being captured from more than one perspective. Such descriptions were then combined into a hierarchical structure that uses a scale type rather than locational indexing to organise relational information. The result was a multiple perspective hierarchy-based model that could be interactively searched and explored to support non-visual access and manipulation of diagrammatically represented information. The question remains as to how such a model could be displayed in audio and used for accessing and manipulating actual relational diagrams. The next chapter will present two studies that empirically investigate answers to this question.

Chapter 4

Inspecting and Constructing Diagrams in Audio – Preliminary Studies 1 and 2

4.1 Introduction

This chapter presents two studies of audio-only interaction with relational diagrams. The studies aim is to assess whether the multiple perspective hierarchical modelling approach developed in Chapter 3 could be used to support audio-only inspection of existing relational diagrams and construction of new ones; i.e. answering Research Question 2 of this thesis.

Section 4.2 describes the design of two audio presentation modes that were developed to display the content of the hierarchy-based model using a mixture of speech and non-speech sounds. The two modes differ in the amount of spoken output used to display the hierarchy. The section then presents Study 1, which explores the question of whether the auditorally displayed hierarchy could be used to inspect and extract the relational information encoded in relational diagrams. The study compares user performances on inspection tasks as the verbosity in each presentation mode is varied. Section 4.3 then describes the design of two interaction strategies that were developed to augment the audio-only inspection of relational diagrams with features to support construction and editing. The section presents Study 2, which explores the question of whether the auditorally displayed hierarchy could be used to construct and edit relational diagrams. The study compares the usability of the two interaction strategies when supporting construction tasks. Section 4.4 concludes the chapter by summarising the results obtained from the two studies, which will be thoroughly analysed and discussed in Chapter 5.

4.2 Study 1 – Inspecting Diagrams in Audio

In order to explore how the hierarchy-based model developed in Chapter 3 could be used to inspect relational diagrams, it needed to be applied to an actual diagramming domain. Class diagrams of the Unified Modelling Language™ (UML®)[1] were chosen as a test case.

4.2.1 Test Case - UML Class Diagrams

UML is a set of visual modelling techniques used to capture and communicate the structure, function and interrelations between data structures, information systems and business processes (Bennett et al., 2005). In particular, Class diagrams are used in the software engineering discipline to model the relations between different components of a software system or a computer program. The choice of UML Class diagrams was based on a number of reasons. First, they represent programs as a set of connected objects, which fits our definition of relational diagrams. Second, the UML diagrammatic notation has been developed into an established standard, which provides a rich but at the same time constrained language that, in turn, helps keep the present investigation at a manageable level; for example, the language only includes a limited number of types of constructs. Finally, the UML modelling language is popularly used in the workplace, which increases the potential of Study 1 to contribute towards a practical solution to the accessibility of these diagrams, and is widely used in education, which provides a ready pool for recruiting potential participants in the study.

A Class diagram consists of nodes (the *Classes*) and connections between them (the *Relations*) and contains a wealth of graphically encoded information, such as the types of connections and the characteristics of objects. In order to keep the problem of translating such diagrams into a hierarchy at a further manageable level, Study 1 focuses on two main types of connections only; namely, *Associations* and *Generalisations*. Classes' characteristics, such as attributes and operations were also omitted from the translation process. Figure 4.1 shows the simplified UML Class diagram notation and its translation into a hierarchy. Any Class diagram – simplified by the above constraints – represents three types of constructs; classes, associations and generalisations. According to the approach outlined in the previous chapter, these three categorical types form distinct branches at the first level of a hierarchical structure (level 1 on Figure 4.1). The second level of such a hierarchy (level 2 on Figure 4.1) lists the items within each category type; i.e. the

[1] `http://www.uml.org/`

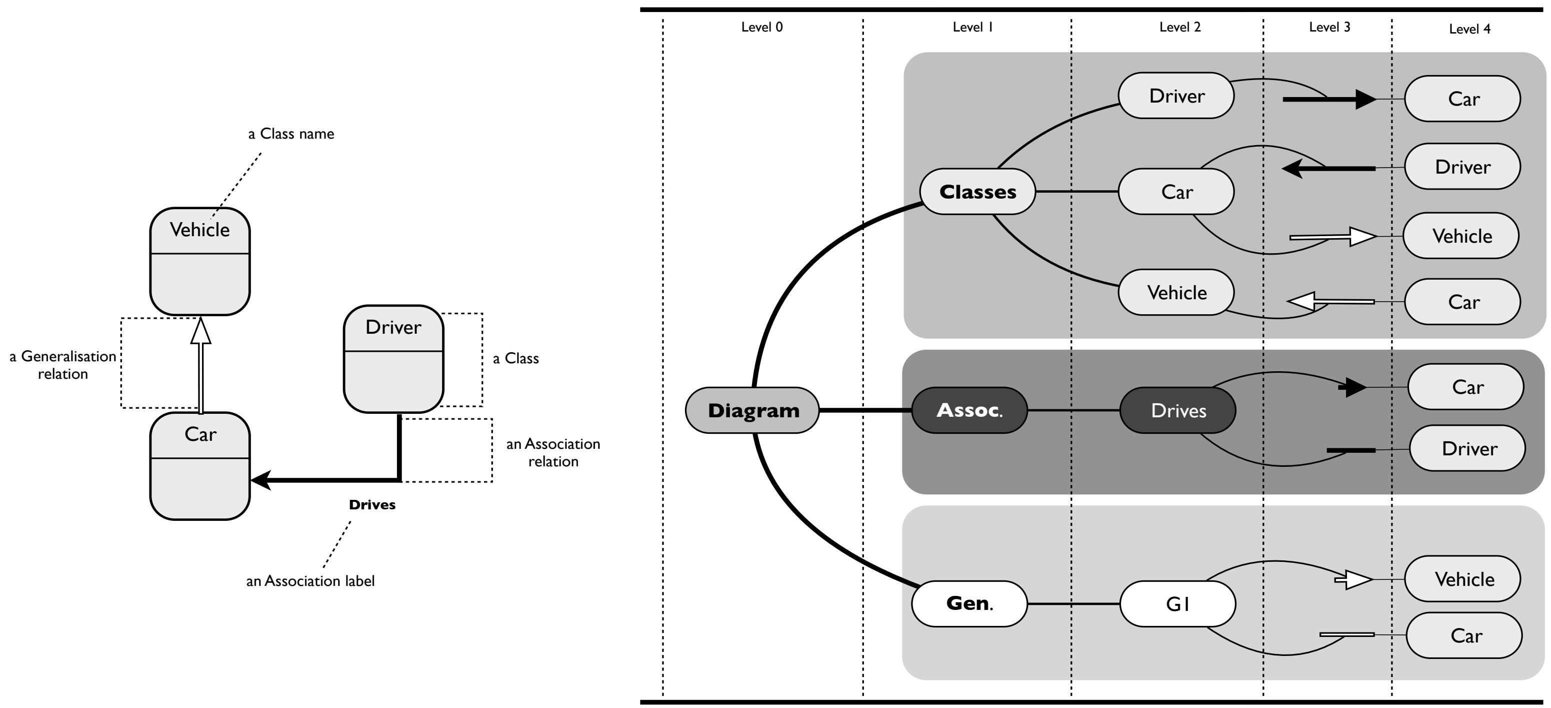


Figure 4.1: A simple UML Class diagram (left) modelled as a multiple perspective hierarchy (right). The three main categories of items; Classes, Associations and Generalisations are translated into distinct branches on a 5-levels hierarchy, each branch listing the items of the corresponding type together with the details of associated relational information. The hierarchy captures and presents the same relational information from different perspectives. A user can browse the hierarchy using a four-way control, such as the keyboard's cursor keys, to move between the nodes of each branch and discover its content.

diagram classes under the *Classes* branch, the association relations under the *Associations* branch, and the generalisation relations under the *Generalisations* branch. The third and fourth levels then list the relational information associated with each item in a given category type, such as the connection types of each class. For example, traversing the path *Diagram* > *Classes* > *Driver* reaches the level of the hierarchy from which relational information that is specific to the class *Driver* can be explored, including its connection types (in this case an association), directions (pointing outwards of *Driver*), and the classes connected through them (linking *Driver* to the class *Car*). The relation between the *Driver* and the *Car* classes can also be inspected through the *Associations* branch of the hierarchy; traversing the path *Diagram* > *Associations* > *Drives* allows a user to discover the label of an association relation, together with its decomposed direction (i.e. the arrow head at the class *Driver* and the arrow tail at the class *Car*). Thus, the same relational information can be inspected from more than one perspective and at different levels of details.

Prototype tool. Modern software engineering programming and modelling tools include features for drawing technical diagrams and exporting such drawings into textual markup formats. The XML Metadata Interchange (XMI)[2] is a standard format for capturing and describing UML diagram elements through cross-referenced lists of classes and their relations. In order to support user interaction with multiple perspective hierarchical models, a prototype tool was developed whereabout XMI formats of Class diagrams are read as input, parsed and processed to extract the information they encode, which is then structured into a multiple perspective hierarchy. The prototype implementation employs the computer keyboard to allow a user to navigate through the hierarchy using the cursor keys to expand and collapse its branches and move up and down its lists of items. Encountered items on the hierarchy are then displayed in audio to support audio-only inspection of captured diagram content.

4.2.2 Auditory Presentation Modes

Visually accessible tree structures, such as the ones found on computer file explorers and application menus, convey a wealth of information that assist users when navigating and inspecting the information they represent. For example, the menu structure in Figure 4.2 shows not only the currently selected menu item, but also its depth position within the list of sibling items, the

[2] `http://www.omg.org/technology/documents/formal/xmi.htm`

size of such a list, the parents menu items' positions and depth, and so on. All such information would be lost if the highlighted menu item is simply spoken.

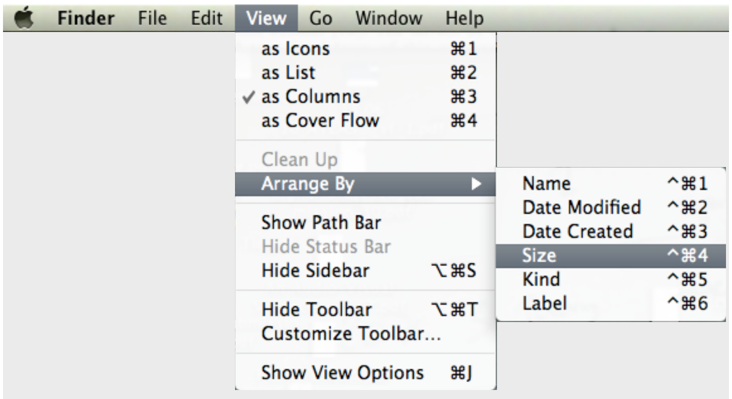


Figure 4.2: Hierarchical menu structure of the Macintosh Finder application showing expanded options.

The interactive functionalities of the prototype system were developed in accordance with a number of guidelines suggested by Smith et al. (2004) for designing non-visually accessible tree structures. These were derived from usability studies that compared the functional needs for navigating both visually and non-visually presented menu-based hierarchies. The adopted guidelines can be summarised as follows:

- Provide a "where am I" option to convey the current position on the hierarchy.
- Present the list of a node's children as a circular list.
- Draw users' attention to the occurrence of illegal moves.
- Allow direct movement to cousin nodes.
- Provide information about the distance of the current node from the root node.

These guidelines were used to drive the design of the audio presentation techniques that display the hierarchy-based model in sound and support interaction with it. The design of such techniques aims to capture the wealth of information needed for interacting with hierarchies. They are described in more details in the following sections.

4.2.2.1 *Information Types*

In order to drive the audio design process, the information contained in a hierarchy was divided into three types; *navigational*, *content* and *contextual* information, summarised in Table 4.1.

Table 4.1: Types of information conveyed through a hierarchical structure.

Information type	Description
Navigational	Information communicating the outcome of navigational actions, such as the successful expansion of a branch on the hierarchy and the listing of the items it contains.
Content	Information describing the actual content on the nodes of a hierarchy, such as the textual label of a node.
Contextual	Information communicating the context of the current position on the hierarchy with relation to the rest of the structure, such as the parent(s) of the current node.

Navigational information communicates cues about the outcome of a user's navigation actions on the hierarchy. Expanding a node on a visual interactive hierarchical structure, such as the one in Figure 4.2, is typically followed by the visual display of its children branching out of the expanded node. This display communicates the fact that the expansion of a branch was successful. Similarly, successful movements between nodes on a visual hierarchy is typically conveyed by alternating the highlighting of the nodes in accordance with a user's movement commands to indicate the displacement of position from one node to the next. This information is lost if there is not enough screen space to properly display the hierarchy or when using non-visual means of interaction and no equivalent of highlighting is provided. *Content* information communicates cues about the actual content of the current node on a hierarchy. On a visual hierarchy, this is typically the textual label of a node, but could also be an iconic image or a combination of text and images. *Contextual* information communicates cues about the context of the current node in relation to the rest of the hierarchy. As shown in Figure 4.2, it is possible to infer information about the parent(s) of the current node, its depth within the hierarchy as well as its position within the current list of children by simply looking at the expanded position of the node of interest. Navigational, content and contextual information are also not necessarily mutually exclusive. For example, it is sometimes possible to infer the context of a currently selected node from its content.

4.2.2.2 Sound Design

It is possible to convey all of the three types of information in audio using speech. Synthetic speech has in fact been used to support inspection in most of the hierarchical structures described in Chapter 2, particularly to convey content information (Mynatt and Weber, 1994; Bennett, 2002; Brown et al., 2004), but also to convey navigational information through parameterised speech output that maps an acoustic dimension such as pitch to the depth of the hierarchy (Shajahan and Irani, 2004). Non-speech sounds have also been used to convey navigational information, particularly in the form auditory icons (Mynatt, 1995) and hierarchical earcons (Brewster, 1997, 1998; Bennett, 1999). The prototype tool developed as part of this thesis uses a mixture of speech and non-speech sounds to communicate the three types of information outlined above. The collection of these techniques were used to design two distinct modes of audio presentation; a *low-verbosity* and a *high-verbosity* mode, such that it is possible to switch between the two during interaction. The auditory display techniques used in each mode are summarised in Table 4.2.

Navigational Sounds. The prototype tool supports four main navigation actions which were identified as requiring explicit auditory feedback to communicate cues about their occurrence during an interaction; *Expanding* a node, *Collapsing* a node, *Browsing* between nodes and *Switching* perspectives on the hierarchy. Additionally, two events were further identified as requiring explicit auditory feedback; reaching the end of a list, and issuing an illegal move.

In the high-verbosity mode, the successful expansion or collapse of a node on the hierarchy is conveyed in speech by proceeding a user's action with the words *"OPENED"* or *"CLOSED"* combined with the corresponding node label (e.g. *"Associations Opened"*). The same message is conveyed using a non-speech sound in the low-verbosity mode. Two earcons were design for this purpose; an *Expand* earcon uses a mixture of frequency and amplitude modulation on a basic pulse oscillator to produce a sweep that ends with a bell like sound, and a *Collapse* sound uses the reversed sequence of the sounds used in the Expand sound (e.g. *"Associations" (Expand sound)* for expanding the Associations branch, and *(Collapse sound)"Associations"* for collapsing it). The two sounds are intended to produce an analogy for a successful expansion and closure of a branch on the hierarchy. Additionally, when a node is expanded, a speech output is displayed to describe the number of items it contains (e.g. *"Associations" (Expand sound) "two"* to convey that the diagram contains two associations).

The successful movement from one node to another is conveyed by displaying the content of a node in either speech or non-speech sounds (described below). Additionally, an earcon is used to augment the speech output of node content with a *Browse* sound in the form of a single beep. This is displayed as the sequence *(Beep)"<node name>"*. The same technique is used to highlight reaching the end of a list, but in such a case a double beep is used instead of a single beep, and is displayed as the sequence *(Double beep)"<node name>"*. These techniques are used in both the high-verbosity and low-verbosity modes of presentation. The pitch of the browsing beeps are mapped to the depth of the hierarchy, so that the deeper the node being browsed the higher the pitch of the accompanying beep.

Table 4.2: Three types of information conveyed through a hierarchical structure and their speech and non-speech audio presentation as used in Study 1. Bolded text highlights the major differences between the two audio presentations modes. Quoted text refers to speech output. Bracketed text refers to non-speech output where (e:) = earcon, (pai:) = parameterised auditory icon.

Information Type	**High-Verbosity**	**Low-Verbosity**
Navigational Information		
1) **Expand a branch**	***"<node name> Opened"***	***"<node name>"*** **(e:Expand Sound)**
2) **Collapse a branch**	***"Closed <node name>"***	**(e:Collapse Sound)** ***"<node name>"***
3) Browse nodes	(e:Browse Sound) *"<node name>"*	(e:Browse Sound) *"<node name>"*
4) Reaching end of list	(e:End of List Sound)*"<node name>"*	(e:End of List Sound) *"<node name>"*
5) **Switch perspective**	***"<description> <node name>"***	**(e:Switch Sound)** ***"<node name>"***
6) Illegal moves	(e:Error Sound)	(e:Error Sound)
Content Information		
7) Class name	*"<class name>"*	*"<class name>"*
8) Relation name	*"<relation name>"*	*"<relation name>"*
9) **Arrow type**	***"<Description of direction>"***	**(pai:Arrow)**
10) **Arrow head**	***"<Description of type>"***	**(pai:Arrow Head)**
11) **Arrow tail**	***"<Description of type>"***	**(pai:Arrow Tail)**
Contextual Information		
12) **Branch**	**"<Description of context>"**	**(Continuous ambient sound)**

The prototype tool allows a user to switch from one perspective on the hierarchy to another; essentially rapidly transporting to the top level of a given branch type from anywhere on the hierarchy using a single keystroke. Conveying a successful switch in the high-verbosity mode is done through a spoken description: *"switched to <node name>"*. The same information is conveyed in the low-verbosity mode using an earcon combined with the spoken description of the destination node (e.g. *(Switch sound) "Associations"*). There are four main nodes that the user can take a shortcut to; the root node, and the three main nodes Classes, Associations and Generalisations. Finally, an earcon is used in both the high-verbosity and low-verbosity presentation modes to highlight the occurrence of illegal moves. This is referred to as the *Error* sound and designed as a low pitched version of browse sound. An example of an illegal move is attempting to expand an already expanded branch, or attempting to browse beyond level 4 of the hierarchy.

Content Sounds. In most cases, speech is used to display content information in both presentation modes, particularly when the content of a node is a textual label (see Figure 4.1). Combined with the navigation Browse sound, this is displayed as *(Beep) "<node name>"*. Level 3 of the hierarchy contains information about connection types and directions. In the high-verbosity mode, this is conveyed using speech messages based on the terminology of the UML notation where association relations are described as *"Association From"* when the arrow points towards a class and *"Association To"* when the arrow points outwards of a class. Generalisation relations are described as *Super Class Of* when the arrow points towards a class and *Inherits From* when it points outwards.

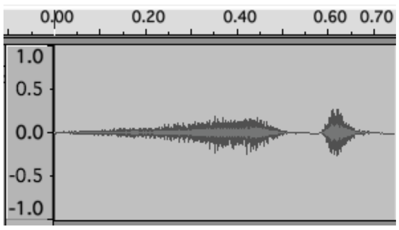


Figure 4.3: A waveform of the "Auditory Arrow", a parameterised auditory icon used to convey arrow type and direction. The length of the parameterised auditory icon was 0.7 seconds.

Parameterised auditory icons are used as alternative non-speech auditory display of connection types and directions in the low-verbosity mode. The auditory icon mimics the sound of drawing an arrow on a chalkboard. Connection types are conveyed by using different timbres for association and generalisation relations, while direction is conveyed by combining a short and a long sound of the same timbre where the short sound represents the arrow head and the long sound represent its tail (see Figure 4.3). Thus, manipulating the order of the sequence in which these two sounds are displayed convey the direction of an arrow; where a long sound followed by a short sound represent an arrow pointing outwards, and the short followed by the long sound represent and arrow pointing inwards. Amplitude modulation on the tail part of the "auditory arrow" is used to enforce the effect of direction.

Contextual Sounds. Non-speech sounds are used to convey one particular type of contextual information in the low-verbosity mode; feedback about the parent branch. Continuous ambient sounds are triggered whenever a user expands a particular branch on the hierarchy and continuously displays while the user remains within that branch. Since hierarchical branching represent a nominal scale, different timbres are used to reflect each of the three main category types/ branches of the hierarchy; i.e. Classes, Associations and Generalisations. The same timbre used to convey the type of a connection is used as a continuous ambient sound in the Association and Generalisation branches. The ambient sound increased in pitch the further a user travels down a particular branch. Shaded areas on Figure 4.1 highlight the parts of the hierarchy where each continuous sound is active. Alternatively, a spoken message describing the current node and its immediate parent node is used in the high-verbosity mode. Rather than continuously displayed, the context description can be requested by the user at any moment during interaction to obtain contextual information.

4.2.3 Study Overview

The aim of the Study 1 is twofold. First, to examine whether the hierarchy-based model can be used to support inspection of a relational diagram. Secondly, to assess the difference between the high-verbosity and low-verbosity presentation modes described in Section 4.2.2 when supporting such interaction. The study therefore examined the following hypotheses:

H1 Using a multiple perspective hierarchy-based model to capture and structure the information encoded in a relational diagram allows for successful inspection of such information in audio.

H2 Varying the presentation modes of the hierarchy-based audio-only model has an effect on users' performance on diagram inspection tasks as follows:

- **H2a** Completing diagram inspection tasks using a high-verbosity presentation mode takes longer than when using a low-verbosity mode.
- **H2b** Answers to diagram inspection tasks are more accurate when using a high-verbosity presentation mode than when using a low-verbosity mode.
- **H2c** Developing expertise in using the hierarchy-based audio-only model is faster when using a high-verbosity presentation mode than when using a low-verbosity mode.

To test these hypotheses, verbosity was manipulated as an independent variable in a between-subjects experimental design factor of presentation mode. In a *Verbose* condition participants used a high-verbosity presentation mode, where speech sounds are the dominant means for the audio presentation of diagram content. In a *Terse* condition participants used a low-verbosity presentation mode, where non-speech sounds are the dominant means for the audio presentation of diagram content.

4.2.4 Method

4.2.4.1 Participants

Twenty sighted individuals were recruited to take part in this study; Seven undergraduates and thirteen postgraduates, nine were female and eleven were male. All participants were from the computer science or electronic engineering departments at Queen Mary University of London and had varying knowledge of UML modelling ranging from low to intermediate. Participants were asked to choose a convenient time to participate in a two hour session that took place at the Usability Lab in the Department of Computer Science within the School of Electronic Engineering and Computer Science at Queen Mary University of London. All participants received a cash incentive of £15 for their participation.

4.2.4.2 Setup and Procedure

A pilot study was conducted with two sighted participants to fine tune the procedure of the training and testing, and to ensure that the tasks and the means for capturing data were adequate. Study sessions were made up of three parts; introductions, training and testing. An informal interview was conducted with each participant at the end of their designated session.

Introductions. Participants were randomly assigned to one of the two conditions. When they arrived at the lab, they were briefed that they were taking part in an evaluation study testing the usability of an audio-only UML Class diagram browser. They were asked to read through information sheets that described the purpose of study and how they are to be involved in it (see Appendix A.1). A short oral summary of such information was also given to make sure the participants understood the procedure. Participants were then asked to sign consent forms for subsequent anonymous use of study materials and to fill a pre-test questionnaire, which gathered basic information about their familiarity with UML Class diagrams and screen-reader technology (see Appendix A.2). Introductions typically lasted for five to ten minutes.

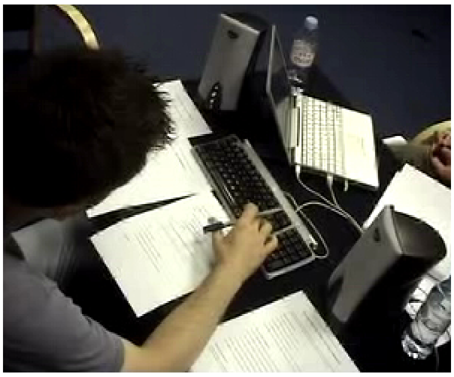


Figure 4.4: Experiment setup during the testing part in Study 1.

Training. All participants were introduced to the basics of UML Class diagram notation used in this stud (see Figure 4.1) and to the audio-only prototype tool using either a high-verbosity or a low-verbosity mode of presentation depending the condition that they were assigned to. Once familiar with the tool, an example diagram was loaded onto the system and participants were given a set of training tasks similar to those used in the testing. With assistance from the experimenter, they were shown how to solve each task to ensure that they had a good understanding of the audio-only tool and the commands to use to navigate and extract information about the diagram from it. During the training, participants could refer to a visual diagram, but this was not allowed during the actual testing. The training typically lasted for twenty to thirty minutes.

Testing. In the actual testing, participants were asked to complete three diagram inspection tasks, without any assistance, in a total of four scenarios each involving a different UML Class diagram. Figure 4.4 shows the experimental setup during this part of the study. Participants were not blindfolded, they controlled the audio-only interface using a computer keyboard and sat facing two standard computer speakers which displayed the audio output of their interactions. The complexity of the diagrams increased from one scenario to the next, and the order of scenarios was kept constant for all participants. Diagram complexity was defined in terms of the number of items on a given diagram as a tuple: {<classes>,<associations>,<generalisations>}. The training diagram for instance was of a {5, 3, 2} complexity because it was made up of five classes three associations and two generalisations; a relatively medium complexity in comparison with the diagrams used in the testing, shown on Figure 4.5.

During the testing, participants were allowed to ask the experimenter to clarify the spoken output of the system if they did not understand it. The experimenter would answer their requests

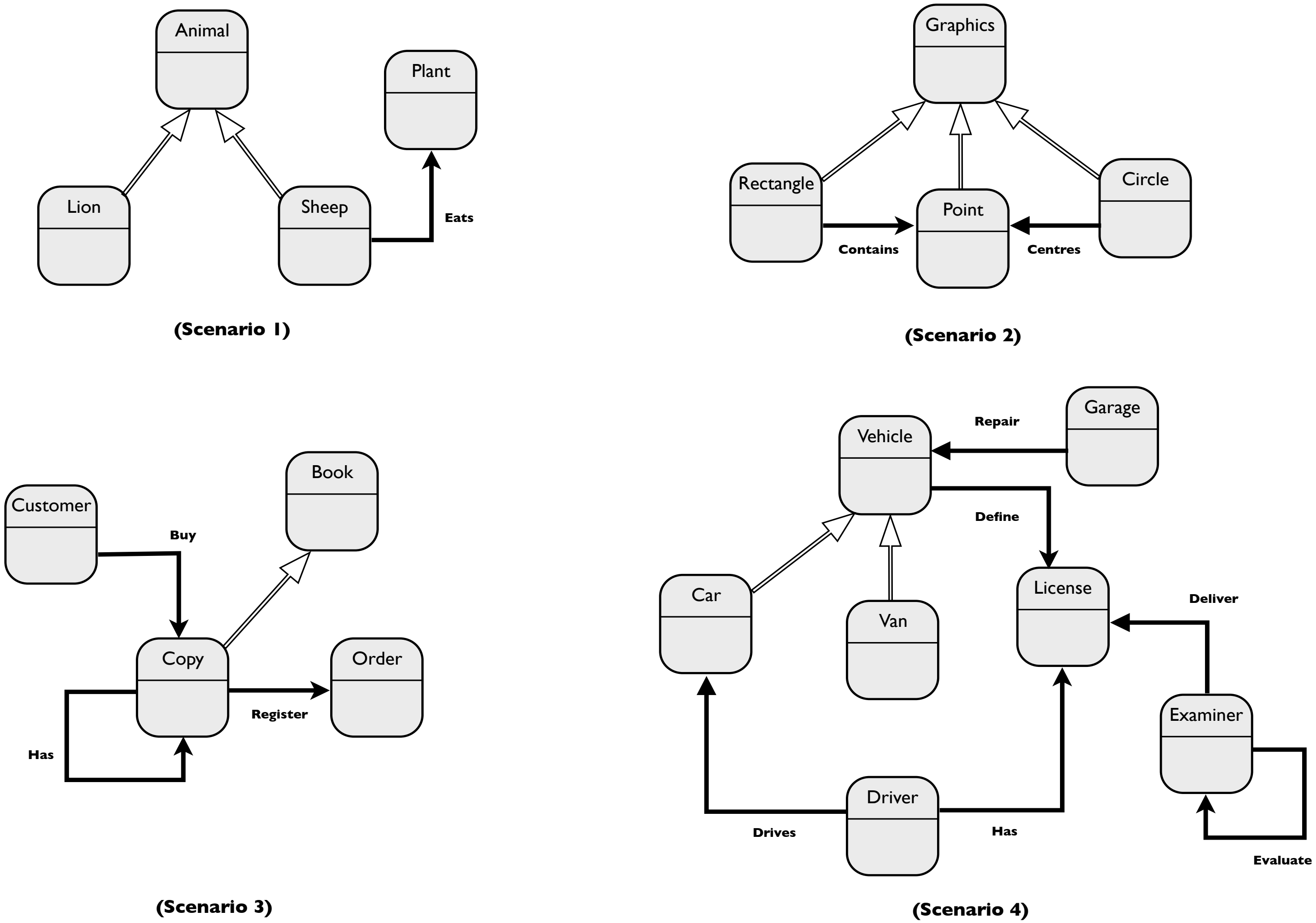


Figure 4.5: UML Class diagrams used in Study 1 with complexities {3, 1, 2}, {4, 2, 3}, {4, 3, 1} and {7, 6, 2} for the 1st, 2nd, 3rd and 4th scenario respectively. Note that the diagrams in the 2nd and 3rd scenarios varied mainly in terms of representational complexity rather than number of items where an involuted relation (i.e. where an object points to itself) was added. The diagram in the 4th scenario included both an involuted relation and an increased number of items.

by repeating what was spoken more articulately. Participants were asked to let the experimenter know when they finished a task and are ready to move on to the next. No time limit was set for completing the tasks, and participants were made aware that they could give up a task or a scenario and move on to the next if they felt stuck, or withdraw from the whole experiment at any point without loosing their cash incentive. At the end of each scenario, the experimenter asked each participant to briefly describe any difficulties experienced while completing the tasks. The testing typically lasted for up to one hour and thirty minutes.

4.2.4.3 Tasks

Participants completed a set of diagram inspection tasks similar to those described in (Bennett, 2002) in that they required an ability to inspect a given relational diagram and to extract information about its nodes and its connections. There were a total of three tasks to complete per scenario and participants were provided with pre-formatted sheets to write their answers on (see Appendix A.3), and the other of the tasks was kept constant across all scenarios:

- Task 1 required a participant to locate a class on a diagram, to explore its relations and find out how it links to other classes on the diagram.
- Task 2 required a participant to locate a relation on a diagram, explore which classes are linked through it and determine its direction.
- Task 3 required a participant to find out the total number of classes and relations on the diagram.

4.2.4.4 Data Gathering and Measures

A variety of means were used to gather data for analysis; the testing part of every session was video recorded; participants' interactions with the audio-only tool were automatically logged and timestamped; and informal interviews with the participants were conducted at the end of each session to gather personal reflections on the use of the audio-only tool.

Participants' ability to use the hierarchy-based model to inspect relational diagrams (H1) is measured in terms of their performance on diagram inspection tasks. Hypothesis H2a is tested by measuring the times it takes the participants to complete the diagram inspection tasks. Hypothesis H2b is tested by measuring the scores achieved when assessing the correctness of the retrieved information. Hypothesis H2c is tested by measuring Participants' expertise in using the audio-only interface in terms the efficiency of their interactions and how this develops across scenarios, where efficiency is reflected in the quality of executing strategies for completing each diagram inspection task. Specifically, the concept of *interaction traps* (Blandford et al., 2003) was used as a model for assessing the efficiency of an interaction. The concept has previously been used to evaluate and analyse complex interactions with graphical user interfaces where users have multiple objectives, shifting objectives, or interleaving tasks (Blandford et al., 2001, 2003), all of which are typical activities that occur when interacting with diagrammatice representations (Scaife and Rogers, 1996).

Interaction Traps. Interaction traps are associated with a subset of negative experiences that occur when using interactive systems. Blandford et al. (2003) defined an interaction trap as a *"situation where, due to the way features of the system design interact with the users understanding of the system and their objectives, the user forms an incorrect understanding about the achievability of an objective, or about how to achieve an objective"*(p.57)[3]. Blandford et al. (2003) described two instances where interaction traps can potentially occur during an interaction. In a first instance, a user believes that an objective is no longer achievable given the current system state when in fact it is – the user then either ceases to try to achieve their objective or tries to achieve it through a different strategy. In a second instance, a user believes that an objective is achievable given the current system state when in fact it is not – after trying for some time, the user might realise that they hit a barrier and eventually stop trying to achieve their objective.

The concept of interaction traps is therefore based on an analysis of how well a user understands the achievability of an objective; if there is a mismatch between such an understanding and the actual achievability of an objective, the user is likely to fall in an interaction trap. Thus, fewer occurrences of traps during an interaction can be a manifestation of an improvement in a user's expertise in using a given system to achieve the objectives they set for themselves, including an improved ability to develop workarounds system features.

Assessing Interaction Efficiency. The retrieval of any particular information about a given diagram using the audio-only tool can be represented as the series of the interactive steps needed to traverse a hierarchical path and reach such information. It is therefore possible to define the optimal traversal path for retrieving each piece of information from the hierarchy. In the context of this study, an occurrence of an interaction trap was identified if a user deviated from such

[3] The definition of interaction traps bears similarities to that of *action slips* (Norman, 1981), where a mismatch between the misinterpretation of a system's state in relation to a user's goals and intentions results in failures to carry out correctly planned actions.

an optimal path when completing the diagram inspection tasks. Three categories of interaction efficiencies were defined to assess the efficiency of execution of a traversal strategy, these are:

- *Inefficient*, if the execution of a traversal strategy includes an occurrence of one or more interaction traps.
- *Less Efficient*: if the execution of a traversal strategy does not include an occurrence of an interaction trap, but involves interaction errors; where a user issues illegal moves, triggering the system to display the error sound.
- *Efficient*: if the execution of a traversal strategy includes neither an occurrence of an interaction trap nor interaction errors.

4.2.5 Results

All participants completed all diagram inspection tasks in each scenario. All data was adequately captured, allowing for two-samples unrelated Student's T tests to be performed to test the study hypotheses.

4.2.5.1 Task Completion Times

Figure 4.6 shows the mean task completion times averaged across the four scenarios for the ten participants in each of the Verbose and Terse conditions. Participants spent an average of 66.3 sec (SD=18.2), 42.9 sec (SD=21.8) and 32.6 sec (SD=7.18) on Task 1, 2 and 3 respectively in the Verbose condition, and 65.4 sec(SD=17.7), 43 sec(SD=16.4) and 23.1 sec (SD=8.16) on Task 1, 2 and 3 respectively in the Terse condition.

An unrelated Student's T test revealed that the differences between the times captured from two conditions were not significant for the first ($t=0.117$, $p=0.91$) and second ($t=-0.12$, $p=0.991$) tasks, but significant for the third tasks ($t=2.761$, $p=0.013$). Thus, participants in the Terse condition spent significantly shorter time executing the third task than did participants in the Verbose condition but not the first and second tasks. These results only partly support hypothesis H2a.

4.2.5.2 Scores on Diagram Inspection Tasks

Figure 4.7 show participants' mean scores on the diagram inspection tasks averaged across scenarios. In the Verbose condition, participants scored an average of 94.2% (SD=8.6) on Task 1, 98.8% (SD=3.9) on Task 2, and 98.8% (SD=3.9) on Task 3. In the Terse condition, participants scored an average of 99.3% (SD=1.7) on Task 1, 96.6% (SD=6.1) on Task 2, and 99.2% (SD=1.9)

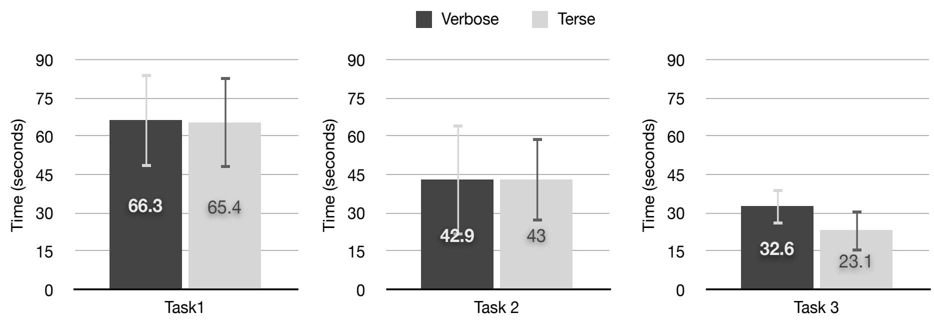


Figure 4.6: Average time spent on completing each diagram inspection task in the Verbose and Terse conditions, averaged across the four scenarios – Error bars show the standard deviations.

on Task 3. An unrelated Student's T test revealed no significant difference between the scores in the two conditions (t=-1.814, p=0.086 for Task 1; t=0.894, p=0.383 for Task 2; and t=-0.376, p=0.711 for Task 3). These results do not support hypothesis H2b.

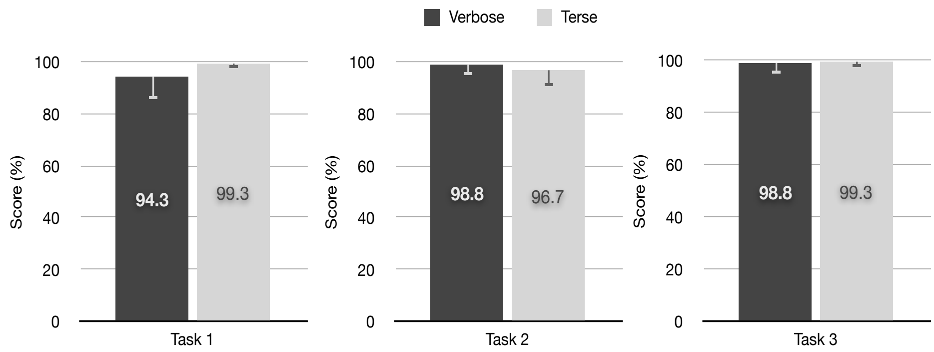


Figure 4.7: Participants' average scores on each diagram inspection task in the Verbose and Terse conditions, averaged across scenarios – Error bars show the standard deviations.

4.2.5.3 *Interaction Efficiencies*

Figures 4.8 shows the proportions of interaction efficiencies in each scenario as averaged across the three tasks in each condition. In the Verbose condition, the proportions of *Inefficient* interactions decreased from 26% in the first scenario to 6%, 0% and finally 3% in the second, third and fourth scenarios respectively. The proportions of *Less Efficient* interactions also decreased from 40% in the first scenario to 36% and 23% in the second and third, then increased slightly to 26% in the fourth scenario. In contrast, the proportions of *Efficient* interactions increased from 33% in the first scenario to 56%, 76% in the second and third, then decreased slightly to 70% in the fourth scenario.

In the Terse condition, the proportions of *Inefficient* interaction were similar in the first and

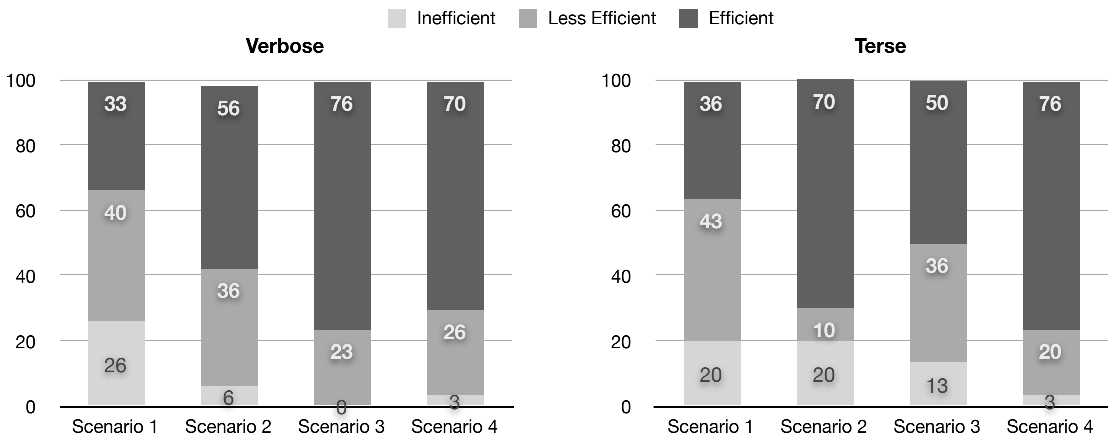


Figure 4.8: Participants' interaction efficiencies in each scenario as averaged across the three tasks.

second scenarios, comprising 20% of the total interactions. Inefficient interaction then decreased from 13% in the third scenario to only 3% in the fourth. The proportions of *Less Efficient* and *Efficient* strategies fluctuated from one scenario to the next. Less Efficient interactions went from 43% in the first scenario to 10% in the second, then up to 36% in the third scenario then down again to 20% in the fourth. There was thus an overall decreasing rate of Less Efficient interactions as scenarios progressed. Efficient interactions went from 36%in the first scenario to 70% in the second, then down to 50% in the third scenario then up again to 76% in the fourth. There was thus an overall increasing rate of Efficient interactions as the scenarios progressed.

Figure 4.9 shows the rate of interaction efficiencies occurrences in each scenario as compared across the two conditions. A Chi-Square test was used to compare these rates and found no significant differences between the two conditions in the first ($c^2(2, N{=}60) = 0.37$, p=0.83) and fourth scenario ($c^2(2, N{=}60) = 0.29$, p=0.86). There was, however, a significant difference between the two conditions in the second ($c^2(2, N{=}60) = 6.99$ p=0.03) and third scenario ($c^2(2, N{=}60) = 6.57$, p=0.037). Thus, participants produced similar rates of interaction efficiencies in the first and last scenarios but not in the second and third. In the second scenario, participants in the Verbose condition produced lower rates of efficient interaction; higher rates of less efficient interactions; and lower rates of inefficient interactions than expected. Participants in Terse condition produced higher rates of efficient interaction; lower rates of less efficient interactions; and higher rates of inefficient interactions than expected. In the third scenario, participants in the Verbose conditions produced higher rates of efficient interactions; and lower rates of less efficient inefficient interactions than expected. Participants in the Terse condition on the other hand produced lower rates of efficient interactions, and higher rates of less efficient and inefficient interactions than expected.

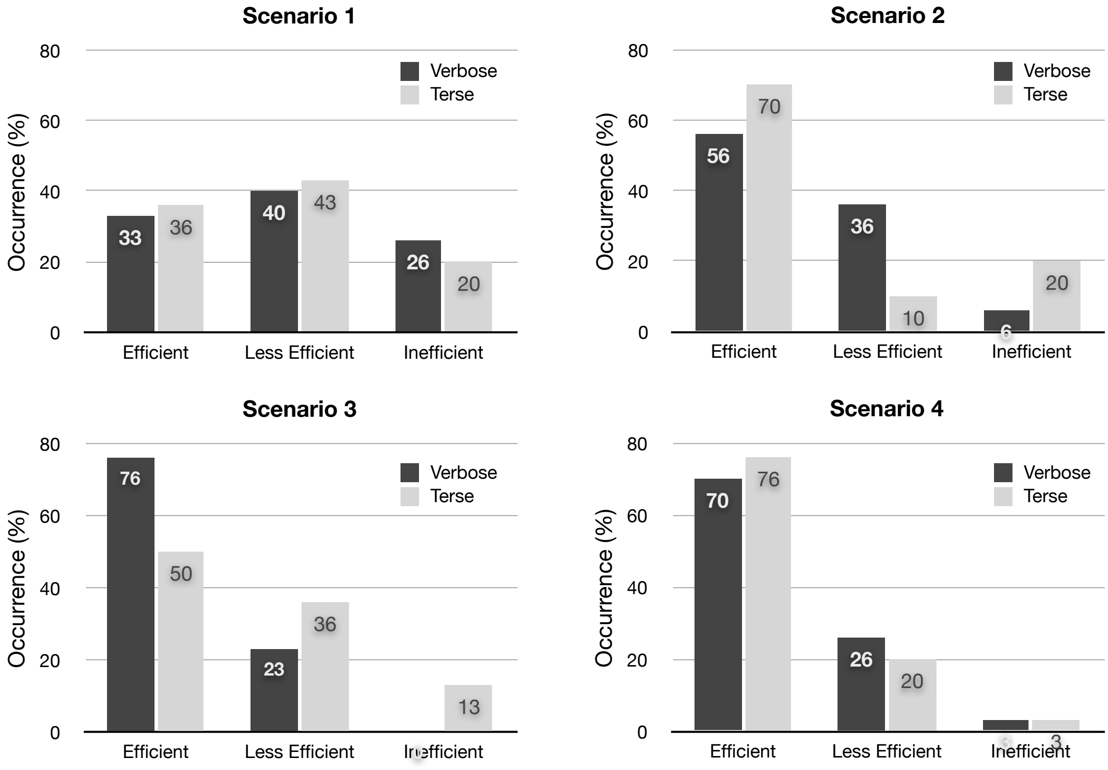


Figure 4.9: Participants' interaction efficiencies in each scenario compared across conditions.

Overall, participants in both conditions showed an increasing rate of improvement in executing interaction strategies for completing diagram inspection tasks as they progressed through the four the scenarios, both reaching equal to or over 70% of efficient interactions and only 3% of inefficient interactions by the fourth scenario. Even though the progress of developing expertise seems to be more steady in the Verbose condition, the Chi-Square test results show that it cannot be considered to be overall faster in this condition than it is in the Terse condition, which does not support hypothesis H2c.

Detailed discussion of the above results will be presented in Section 5.2 which will explore the impact of the hierarchy-based model on participants performance on diagram inspection tasks; the impact of the audio presentation techniques on such performance; and participants learning patterns when developing expertise in inspecting UML Class diagrams in audio.

4.3 Study 2 – Constructing Diagrams in Audio

In order to explore how the hierarchy-based model developed in Chapter 3 could be used to construct and edit relational diagrams, it needed to be applied to an actual diagramming domain. Entity-Relationship (ER) diagrams were chosen as a test case.

4.3.1 Test Case - Entity-Relationship Diagrams

The choice of using ER diagrams was driven by similar motivations to those of choosing UML Class diagrams. There are in fact numerous similarities between ER and Class diagrams; both being a standardised visual modelling technique that represent relational information through rich yet constrained notational language. Additionally however, ER was chosen in order to examine the applicability of the developed hierarchical modelling approach to a different diagrammatic notation. At an abstract level, the main differences between the two notations are in the terminology and graphical tokens used to describe diagram items, but ER application areas are also often different to those of Class diagrams. This will therefore increase the potential of Study 2 to contribute towards a practical solution to the accessibility of a diagrammatic notation that is used in other domains.

ER diagrams are used by system analysts and software engineers to model the conceptual structure of a system prior to its development and are particularly popular for modelling database systems (Connolly and Begg, 2001). Notations for drawing ER diagrams are varied and they also differ in terms of the graphical tokens used to represent diagram elements[4]. An ER diagram represents two main categories of items; *Entities*, drawn as labeled boxes, and *Relations*, drawn as labeled diamond shapes. An entity is associated with a number of *Attributes* where one or more attributes can be used as its unique identifier, in which case it is referred to as a *Primary Key*. A relation connects one or more entities, and each entity in a given relation is associated with a *Cardinality*[5] For the purpose of Study 2, a simplified version of ER diagrams is used whrere only one relation type and its cardinalities are considered. Such relations do not encode direction. Figure 4.10 shows a simple ER diagram and how it translates into a hierarchy. According to the approach outlined in the Chapter 3, the two categorical types of Entities and Relations form dist-

[4] Common ER notations include Chen's notations – used in Study 2 – and Crow's Foot notation. An example in the difference between the two notations is the use of a diamond shape to represent relations and textual labels to represent cardinalities in the former, in contrast to textual labels for relations and iconic representations for cardinalities in latter.

[5] A cardinality in an ER diagram is the maximum participation of an entity in a given relation. For instance, the ER diagram in Figure 4.10 on page 72 represents a *Design* relation which involves a maximum of *One* entity of type *Doctor* and *Many* entities of type *Therapy*.

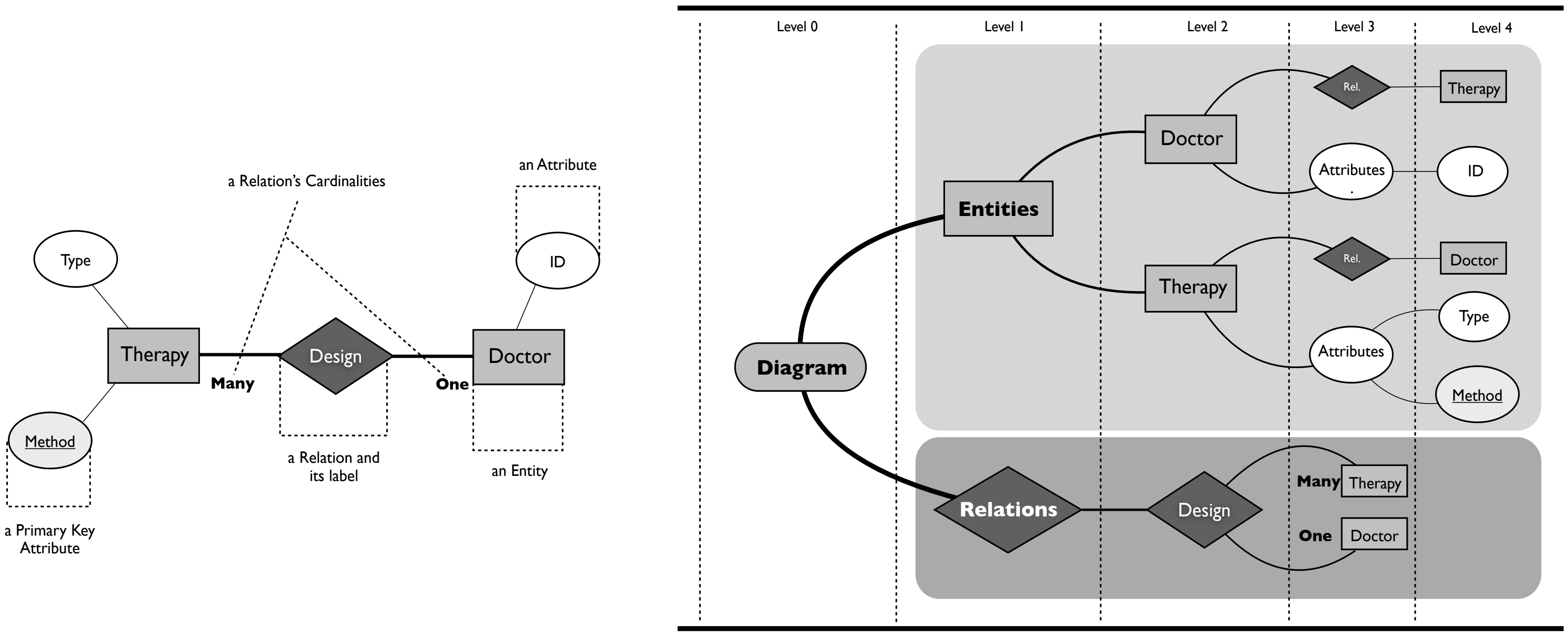


Figure 4.10: An Entity-Relationship diagram (left) modelled as a multiple perspective hierarchy (right). The two main categories of items; Entities and Relations are translated into distinct branches of 5-levels hierarchy, each branch listing the items of the corresponding type together with their specific details.

inct branches at the first level of a hierarchy (Level 1 on figure 4.10). Individual entities and relations are then listed under their corresponding branch on the second level of the hierarchy (Level 2 on figure 4.10). The third and fourth levels of the hierarchy list information specific to each item within a given category type. For example, traversing the path *Diagram* > *Entities* > *Doctor* allows for exploring the attributes of entity *Doctor* as well as its connections to other entities on the ER diagram. Traversing the path *Diagram* > *Relations* > *Design* allows for exploring details of a given relation, including the entities related through it and their associated cardinalities. Here too, the hierarchy allows for the same relational information to be inspected from different perspectives at different levels of details.

4.3.2 Interactive Audio-only Construction Strategies

To support audio-only construction of ER diagrams through a hierarchy-based model, two interaction strategies were designed to augment the prototype audio-only tool to support construction and editing. These are referred to as *Guided* and *Non-Guided* interactions. There is a lack of research into audio-only construction of visual displays such as relational diagrams, and guidelines for designing such interaction were virtually non-existent at the time of writing this thesis. The designs of the two strategies outlined in the follow sections are mainly based on existing human-computer interaction techniques and inspired by Hutchins (1989)'s analysis of interaction metaphors in interface design.

4.3.2.1 Guided Interaction Strategy

One way to support the execution of diagram editing actions through the hierarchy-based model is to define the set of steps necessary to complete every possible editing action on a diagram, then use each set as a template to guide a user when executing corresponding actions. For example, to connect two entities with a relation, generic steps can be defined whereabout the entities to be related must first be identified, a relation between them created, then a unique label added to distinguish the created relation.

The *Guided* strategy was developed around this principle of template matching. In a guided interaction strategy, the system plays the role of an agent that assists the user when executing an editing action. The user invokes the system's assistance by expressing the editing action they wish to execute (using the computer keyboard in the case of the developed prototype tool), then follows the system's instructions to complete its execution. Thus, the system and the user engage in a conversation-like style of interaction, where they collaborate by exchanging information to complete the desired task. When the user expresses a desired action, a series of system prompts are triggered, each requiring the user to supply more details about what they wish to achieve. By responding to each prompt, the user is essentially guided through the necessary steps required to accomplish an editing action.

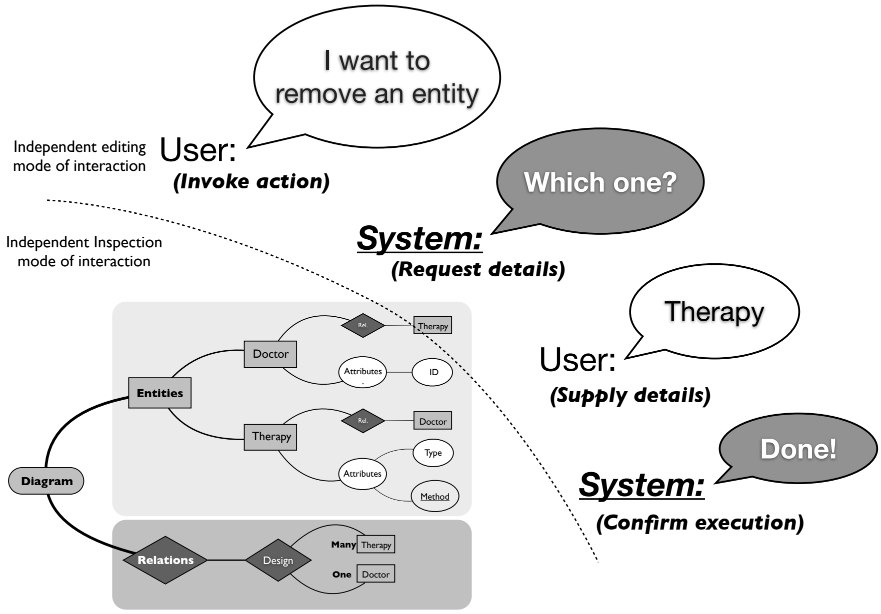


Figure 4.11: In a Guided interaction strategy the user responds to systems prompts away from the hierarchy. Inspecting and editing the diagram are thus isolated into two independent modes of itneraction

Figure 4.11 exemplifies this concept; if the user wishes to remove an entity from a diagram, they specify their desire to do so by issuing the appropriate command using the keyboard. The system then requests any necessary details related to the issued command, in this case; a label to identify the entity in question. The user supplies the relevant details by typing in the label of the entity, and the system uses this information to execute and confirm the successful accomplishment of the requested editing action.

The series of system prompts that the user responds to depend primarily on the type of action that initiates an editing task. This strategy therefore allows the execution of any action on any given item on the diagram without requiring the user to locate it within the hierarchy. That is, when editing an item on a diagram, the user is doing so away from and independently of the hierarchy. The tasks of inspecting and editing the diagram are therefore isolated into two distinct modes of interaction; an inspection mode, and an editing mode, and the user will essentially be switching between these two modes when constructing a diagram. Thus, while engaged in an editing action, the system acts as an implied intermediary between the user and the world in which actions are taken (Hutchins, 1989). Once actions are completed, the user can observe their effects by inspecting the part of the hierarchy where they were applied. The hierarchy in this case represents the model-world where the outcomes of user actions can be perceived.

4.3.2.2 Non-Guided Interaction Strategy

The system in a *Non-Guided* interaction strategy on the other hand does not provide any explicit assistance to the user. To construct or edit an item in a diagram, the user must first locate it within the hierarchy before executing a particular editing action that alters its state. Figure 4.12 exemplifies this concept; to remove an entity from a diagram, the user must first inspect the appropriate path to locate it on the hierarchy then, once found, issue the desired editing command to delete it. The system then combines the current position of the user together with the issued command and interprets the two as one complete editing command expression before executing it appropriately.

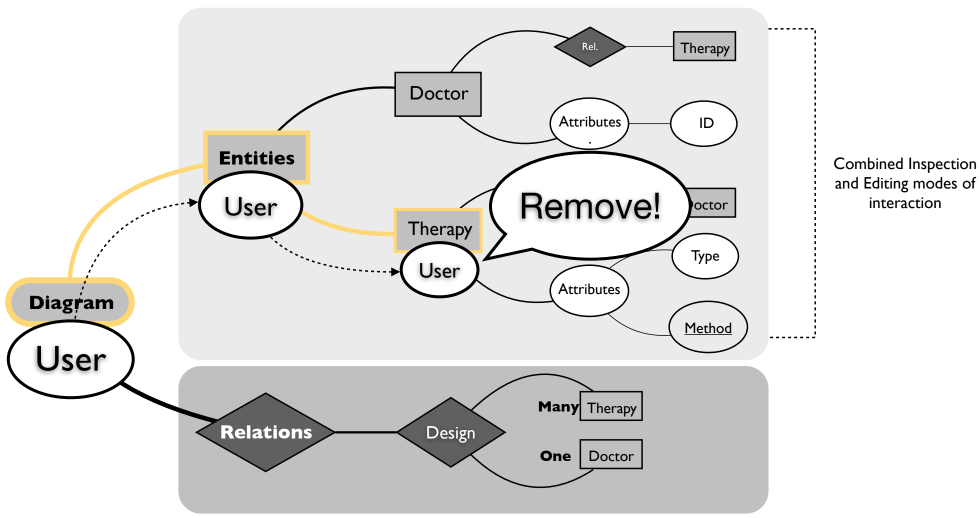


Figure 4.12: In a Non-Guided interaction strategy the user locates the item on the hierarchy to edit it. Inspecting and editing the diagram are thus combined into a single mode of interaction.

The completion of an editing action therefore depends primarily on the status of the hierarchy at the moment it is executed, and the user can directly perceive the changes in the state of the hierarchy as soon as the action is completed. Thus, the strategy puts an emphasis on the user as the main actor within a model-world where interactive expressions can be realised (Hutchins, 1989). The hierarchy in this case represents both the model-world where actions are executed, and part of the language which expresses the editing actions. In contrast to the guided interaction strategy, there is no implied intermediary between the user and the hierarchy since the user is directly engaged with the hierarchy throughout the execution of any editing action. In the Non-Guided interaction strategy the tasks of inspecting and editing the diagram complement one another, and the strategy eliminates multiple modes by combining them into one complementary and inter-dependent mode of interaction.

In both the Guided and Non-Guided strategies, content added onto the diagram is automatically structured into a hierarchy. That is, the user issues commands to create entities, attributes, relations and cardinalities rather than nodes and branches, and the system dynamically updates corresponding parts of the hierarchy to reflect the user's edits.

4.3.2.3 Sound Design

The audio display used in both construction strategies were similar to those described in Section 4.2.2 with the following exceptions. First, all audio output was displayed using a low-verbosity mode in the first instance. High-verbosity alternatives were available at the user's request. This is done so in order to avoid audio clutter and improve the pace of interaction[6]. Second, since relations in ER diagrams do not encode direction, the "auditory arrow" was discarded and relations were simply displayed using speech describing their labels. Third, only two continuous ambient sounds were used to match the two main branches of the hierarchy; an Entities and a Relation sound. Shaded areas on Figure 4.10 highlight the parts of the hierarchy where each continuous ambient sound is active. In addition however, a third background sound was used to distinguish between editing and inspection modes of interaction in the Non-Guided construction strategy. This distinction was further emphasised with the use of two voice genders for the speech output that occur during each mode; a male voice was used for speech output during the inspection mode, and a female voice was used in the editing mode.

Finally, the cardinalities of a relation were displayed using two alternative techniques; an auditory icon in a first instance, then a spoken description in a second. That is, when a user encounters a cardinality in the *Relations* branch, an auditory icon is displayed to convey its

[6] Although results from Study 1 are analysed and discussed in Chapter 5, they were in fact used to motivate and improve on the audio design of the prototype tool in Study 2. For example the choice of using a low-verbosity presentation mode in the first instance was based on the improved performance observed in task completion times which did not compromise levels of performances on inspection tasks.

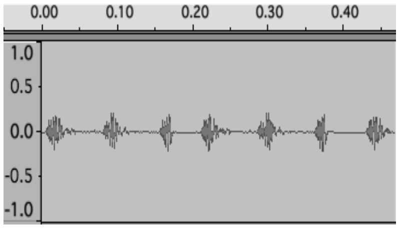


Figure 4.13: A waveform of the "Auditory Cardinality" used to convey a relation's MANY cardinality. The length of the parameterised auditory icon was just under 0.47 seconds.

meaning, the user can then request a spoken description of the encountered auditory icon if they so wish. The simplified ER diagrams used in the Study 2 included two type of a cardinalities; a *Many* cardinality and a *One* cardinality. The speech output used simply described what each cardinality is; i.e. using the spoken words *"MANY"* or *"ONE"*. Cardinality auditory icons were designed using a percussion metaphor as a non-speech alternative. One percussion tap sound was used to correspond to the *One* cardinality, and a sequence of seven percussion taps were used to correspond to the *Many* cardinality (see Figure 4.13).

4.3.3 Study Overview

The aim of Study 2 is twofold. First, to examine whether the hierarchy-based model can be used to support construction and editing of a relational diagram. Second, to examine the usability of each interaction strategy by assessing whether and how differences between the way in which interaction is structured in each strategy affects users' ability to build and retain a coherent understanding of the constructed diagrams. The study therefore examined the following hypotheses:

H3 Using a multiple perspective hierarchy-based audio-only model to capture and structure information encoded in relational diagrams allows for successful construction and editing of such information in audio.

H4 Varying the interaction strategies with a hierarchy-based audio-only model has an effect on users' performance on diagram construction tasks and comprehension of the constructed diagrams as follows:

- **H4a** Constructing a relational diagram using a guided interaction strategy takes longer than when constructing it using a non-guided strategy.

H4b Constructing a relational diagram using a guided interaction strategy supports more coherent understanding of the diagram than when constructing it using the non-guided strategy.

To test these hypotheses, the combination of inspecting and editing modes was manipulated as an independent variable in a within-subjects experimental design factor of interaction strategy. In a *Guided* condition, participants construct relational diagrams using the guided interaction strategy where inspecting or editing a diagram are isolated into two distinct modes of interaction. In a *Non-Guided* condition, participants construct relational diagrams using the non-guided interaction strategy where the two modes are combined. A third condition, where participants list diagram elements on paper rather than through the audio-only tool was introduced as a control condition. The three conditions are described in more details in what follows.

4.3.4 Method

4.3.4.1 Participants

Twenty four sighted participants, different from those who took part in Study 1, were recruited to take part in this study. Participants were a mixture of undergraduate and postgraduate students from different disciplines including computer science, engineering and business and studying at various University of London institutions. They had varying knowledge of ER diagrams ranging from none to advanced expertise. Participants were asked to attend a two hour session at the Usability Lab of the Computer Science department at Queen Mary University of London. All participant received a £15 cash incentive for their participation.

4.3.4.2 Setup and Procedure

A pilot study was conducted with two sighted participants to fine tune the procedure of the training and testing, and to ensure that the tasks and the means for capturing data were adequate. Study 2 sessions were also made up of three parts, introductions, training and the actual testing. Unlike the procedure in Study 1 however, training and testing in Study 2 were intermixed such that each participant would receive a training on a given interaction strategy then be tested on that particular strategy before receiving training on the second strategy. Post-test preference questionnaires and informal interviews were conducted with all participants at the end of each session.

Introductions. Participants read through information sheets describing the purpose of the study and the nature of their involvement in it (see Appendix B.1). They were briefed that they were taking part in an evaluation study to test the usability of an audio-only tool for constructing ER diagrams, and a short oral summary of the information sheet was given to make sure participants understood the procedure. They were then asked to sign consent forms for subsequent anonymous use of study materials. Introductions lasted for five to ten minutes.

Training. All participants were introduced to the basics of ER modelling as used in this study (see Figure 4.10) and to the hierarchy-based audio-only tool. Once familiar with the tool, participants were trained on how to construct ER diagrams using the guided and non-guided strategies. Here participants were presented with a diagram description similar to those used in the testing part, and were assisted in constructing various parts of an ER diagram until they felt comfortable using the tool. During the training, participants could refer to a visual version of the diagram they were constructing, but this was not allowed in the actual testing. Training typically lasted for up to forty minutes.

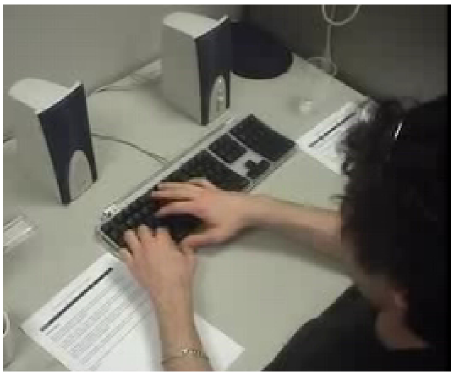


Figure 4.14: Experiment setup during the testing part in Study 2.

Testing. The testing part of the study was made up of three phases. In each phase, participants were provided with a textual description of an application design and were asked to construct an ER diagram to model each system (see Appendix B.3). All systems were of similar complexity; they could all be modelled by constructing ER diagrams of six to eight entities and five to eight relationships, each entity having one, two or three attributes (see Figure 4.15 for an example diagram complexity). No time limit was imposed and participants were informed that they could take as long as they wished to construct their diagrams. Figure 4.14 shows the experimental setup during this part of the session. Participants were not blindfolded, they used a computer keyboard to control the audio-only tool and sat facing two standard computer speakers which presented the audio output of their interactions with the diagrams. The testing part typically lasted for up to one hour and thirty minutes.

4.3.4.3 Tasks

In the first phase – the control condition – participants were asked to go through the textual description, identify the list of potential diagram elements (i.e. entities and their attributes, potential relations, the entities these relate and their cardinalities) and to write these down on a provided pre-formatted sheet (see Appendix B.4). They were explicitly asked not to use any annotation or sketching as they extract the items from the written description, and to simply write them down on the provided sheets. This task was chosen as a control for two reasons. First, it is often used in the analysis phase of ER modelling, where the analyst identifies and lists the details of all elements necessary for constructing an ER model before drawing the actual ER diagram (Connolly and Begg, 2001). Second, it is similar to constructing the diagram using the audio-only tool in that both tasks involve going through a written description and writing/or typing the diagram elements into a browsable list. Thus, the task acted as a control against which performances using the audio-only tool could be compared. In the second and third phases – the test conditions –

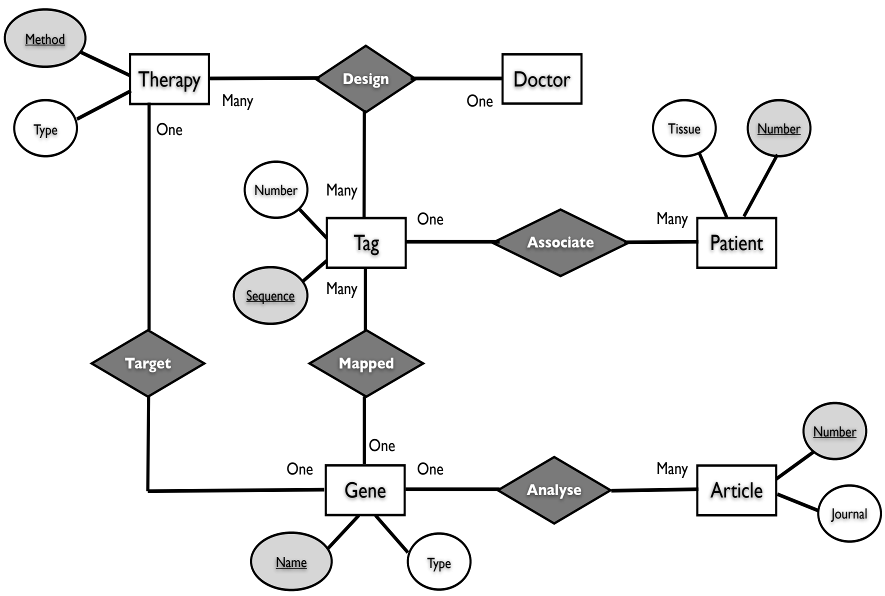


Figure 4.15: Example of the complexity of ER Diagrams that model the textual descriptions.

participants were asked to construct two ER diagrams, one at a time, from the provided textual descriptions, using the audio-only tool set to support one of the two interaction strategies each time around. The order in which the three textual descriptions were presented to different participants, as well as the order in which they used the interaction strategies were randomised across all participants to minimise learning effects.

After constructing each diagram in a given condition, participants were asked to answer a series of questions about it. These post-construction questions were designed to test participants' understanding of information about the diagrams which was not explicitly mentioned in the written descriptions – for example, the answer to the question of *"Which entity had the largest number of relations?"* could only be explicitly revealed through the construction process (see Appendix B.5 for the full list of questions). The questions were not related to the technicalities of modelling systems using entity-relationship diagrams, and so did not required expertise in using the notation. Rather, they were related to the particular features of the constructed diagrams as well as the process of their construction, and hence addressed each participant specifically to test their knowledge of what they have done; i.e. each participant against their own standard and abilities.

4.3.4.4 Data Gathering and Measures

The testing parts of the study were video recorded, and all participants' interactions with the audio-only tool were automatically logged and timestamped. Questionnaires and informal interviews were used to capture participants' performance and personal experience and preferences. Hypothesis H3 is tested by measuring Participants' ability to use the hierarchy-based model to construct relational diagrams in terms of their performance on diagram construction tasks. Further, task completion times and scores were measured as dependent variables. Specifically, Hypothesis H4a is tested by measuring participants performance in terms of the times it took them to construct the diagrams in the Guided and Non-Guided conditions. Hypothesis H4b is tested by measuring participants' comprehension of the constructed diagrams in terms of their scores on the post-construction questions.

4.3.5 Results

All participants completed the construction tasks using pen and paper in the Control condition and the hierarchy-based audio-only tool in the Guided and Non-Guided conditions. Construction times were only captured for the Guided and Non-guided conditions, a Related Student's T test was thus used to test hypothesis H4a. Scores on post-construction tasks were captured from all three conditions, allowing for a within-groups one-factor ANOVA test to be performed to test hypothesis H4b.

4.3.5.1 Construction Completion Times

Figure 4.16 shows the mean construction times in the Guided and Non-guided conditions. Participants spent an average of 23.41 minutes (SD=5.7) to complete constructing the diagrams in the Guided condition and an average of 19.2 minutes (SD=6.4) to complete constructing the diagrams in the Non-guided condition. A Student's T test revealed that these differences were significant (t=3.039, p<0.005). This result supports hypothesis H4a.

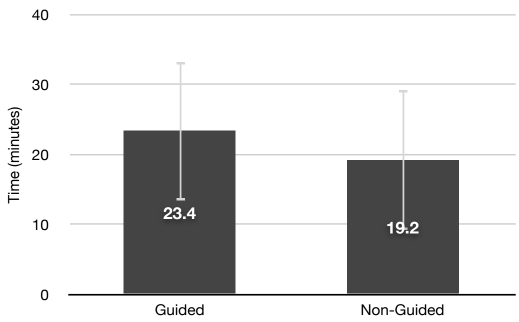


Figure 4.16: Average construction times for all participants in the Guided and Non-guided conditions – error bars show the standard deviations.

4.3.5.2 Scores on Post-Construction Tasks

Figure 4.17 shows the mean scores on post-construction tasks in the Guided and Non-guided conditions and in the control condition. Participants scored an average of 80.3% (SD=10.8) in the Control condition, 76.3% (SD=8.4) in the Guided condition and 77.45% (SD=11.3) in the Non-guided condition. The results of a within-group ANOVA test showed no significant difference between these scores (F(2)=1.711, p=0.192). This result does not support hypothesis H4b.

4.3.5.3 User Preferences

Out of the twenty four participants, sixteen preferred the non-guided interaction strategy to the guided strategy, while only six preferred the guided strategy; two participants had no preference. Among the six who preferred the guided strategy, three rated their expertise in ER modelling as intermediate, one expert, one beginner, and one had no previous experience. Among the sixteen who preferred the non-guided strategy, two rated themselves as experts, seven as intermediate,

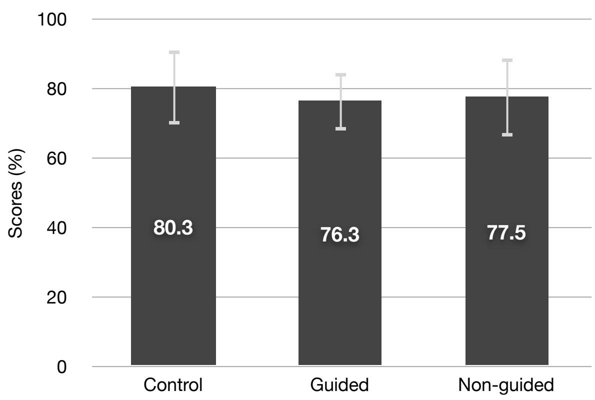


Figure 4.17: Average participants' scores on post-construction tasks in the Control, Guided and Non-guided conditions – error bars show the standard deviations.

and seven as beginners. Further within-group ANOVA tests were conducted on each preference group in order to establish whether the difference between the scores within these groups were significant. The results showed no effect (F(2)=0.263, p=0.774 for N=6 participants preferring the Guided; and F(2)=2.111, p=0.139 for N=16 participants preferring the Non-guided).

Detailed discussion of the above results will be presented in Section 5.3 which will explore the impact of the hierarchy-based model on participants performance on diagram construction tasks; the impact of the Guided and Non-guided interaction strategies on such performance in terms of their usability.

4.4 Summary and Conclusion

This chapter presented two studies that addressed Research Question 2 of this thesis. The aim of the two studies was thus to assess whether the hierarchy-based approach developed in Chapter 3 could be used to support audio-only interaction with relational diagrams.

The first study focused on the activity of inspecting existing relational diagrams in audio through the hierarchy-based model. Section 4.2.2 described two audio presentation modes that differed in the amount of speech and non-speech sounds used to communicate navigational, content and contextual information of the hierarchy. A high-verbosity mode used speech as the main means for conveying information, whereas a low-verbosity mode combined speech and non-speech sounds. Section 4.2.3 presented Study 1, which evaluated and compared the two presentation modes when used for solving diagram inspection tasks. The results were presented in

Section 4.2.5 and showed that completing some diagram inspection tasks using a high-verbosity presentation mode took longer than when using a low-verbosity mode, but scores on all tasks were equivalent irrespective of presentation mode. The results also showed that the progress of developing expertise in using a hierarchy-based audio-only tool to solve diagram inspection tasks was equivalent irrespective of the audio presentation mode used to display the hierarchy.

The second study focused on the activity of constructing and editing relational diagrams in audio through the hierarchy-based model. Section 4.3.2 described two construction strategies for supporting such interaction. The two strategies differed in the way they combine or isolate inspecting and editing modes of interaction. The Guided interaction strategies isolates the two modes and assists the user when executing diagram editing actions by guiding them through the steps necessary for completing such actions. The Non-Guided strategy cedes control to the user to act directly on the hierarchy, combining inspection and editing into one mode of interaction to support the execution of diagram editing actions. Section 4.3.3 presented Study 2, which evaluated and compared the two construction strategies when used for diagram construction tasks. The results were presented in Section 4.3.5 and showed that constructing relational diagrams using the Guided interaction strategy took longer to complete than when using the Non-Guided strategy, and that both strategies supported similar levels of diagram comprehension when users were tested on post-construction tasks. The next chapter presents a thorough analysis and discussion of these results highlighting their implications for the design of audio-only interaction with relational diagrams.

Chapter 5

Analysing Audio-only Interaction with Diagrams

5.1 Introduction

The previous chapter presented two studies that examined audio-only interaction with relational diagrams. Results from Study 1 showed that sighted participants were able to use a hierarchy-based audio-only tool to inspect and extract information from existing relational diagrams. The results also showed that varying the level of verbosity in such a tool affects the time it takes to complete some diagram inspection tasks but not the outcome of, nor the progress of developing expertise in using the audio-only tool to complete such tasks. Results from Study 2 showed that sighted participants were also able to construct and edit relational diagrams using a hierarchy-based audio-only tool, and that combining or isolating inspection and construction modes of interaction affects construction times but not users' comprehension of the constructed diagrams.

This chapter presents a detailed analysis and discussion of these findings and compiles a set of design lessons learnt from them, which are presented in pop-out boxes throughout the chapter and will be summarised later in Chapter 9. Section 5.2 discusses aspects of using a hierarchy-based audio-only tool as a means for inspecting and extracting meaningful information about relational diagrams, particularly in terms of affordances for navigation and orientation, users' interpretation of auditorally displayed diagram content and development of expertise during such interaction. Section 5.3 then discusses participants' performance on diagram construction tasks in terms of construction times and comprehension of constructed diagrams, and addresses issues related to the usability of the developed interaction strategies. Section 5.4 concludes the chapter.

5.2 Inspecting Diagrams in Audio

To recap, the first study tested two main hypotheses; first, that a hierarchy-based audio-only tool allows for inspection of a relational diagram; and second, that varying the audio presentation mode of such an interface, specifically in terms of its verbosity, affects users' performance when completing diagram inspection tasks. Participants' ability to inspect relational diagrams using the developed hierarchy-based tool was assessed in terms of the scores they achieved when solving diagram inspection tasks. High scores were achieved (an overall mean of 97.3% and 98.4% in the high-verbosity and low-verbosity conditions respectively) and the differences between scores in the two conditions were not statistically significant. This result is thus considered a reflection of participants' ability to use a hierarchy-based audio-only tool to inspect relational diagrams, effectively supporting the first hypothesis (H1) tested in Study 1.

Completing diagram inspection tasks required a participant to *navigate* through the hierarchy in order to locate diagram content, to *extract* meaningful information about the encountered content by interpreting their audio descriptions, and to *develop expertise* in doing so in order to be more effective at inspecting relational diagrams using the provided hierarchy-based audio-only tool. The following sections discuss how these activities were supported through such a tool.

5.2.1 Performance on Diagram Inspection Tasks

5.2.1.1 Navigation and Orientation

In addition to some shortcut keys, which allowed rapid transportation between hierarchical perspectives, participants used the keyboard arrow keys (as per Figure 5.1) to navigate around the hierarchy. The layout of the keys represented the basic layout of the hierarchical structure so that directional movements of the keys matched the directional movements on the hierarchy.

To find their way around, participants relied on a number of features of the hierarchy as they navigated through it. The first is its *semi-fixed structure*. Diagram content was organised in terms of categorical types, where each branch of the hierarchy was exclusively reserved for one type of information, and included details specific to each item within that particular category of items. Using this approach, any UML class diagram was structured in terms of classes, associations and generalisations, and any item within each of these three higher level types had a fixed number of possibilities for representing its relational information. A class, for instance, would always be located on the third level of the *Classes* branch, and would always have up to four types of relations (*association from/ to* another class, and be a *super class of*, or *inherits from* another class). Similarly, an association relation will always be located on the third level of the *Associations* branch, and always link no more than two classes. Thus, the basic structure of the hierarchy is fixed, what changes is the details of its content depending on the details of the diagram being inspected. This imposed semi-fixed structuring of diagram content allowed participants to anticipate where specific information will be located on the hierarchy at the onset of navigation, decreasing the potential for disorientation. This was then reinforced the more experience participants gained as they used the tool. This approach to hierarchical structuring is different to previous work on using hierarchies to organise graphically represented information (e.g Mynatt and Weber, 1994; Bennett, 2002; Brown et al., 2004) where different diagrams typically yield different hierarchical structures and organisation.

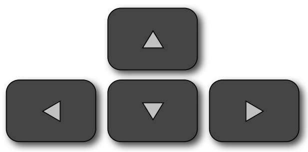


Figure 5.1: The layout of the arrow keys as found on most standard computer keyboards.

The second feature that facilitated navigation of the hierarchy was its *fixed depth*. Because the structure of diagram content was semi-fixed, a complex diagram would only expand the hierarchy in terms of its breadth but not its depth[1]. In this case too, correctly interpreting the displayed content of the current position on the hierarchy, and correctly matching this interpretation to the fixed number of possible locations on the hierarchy where such content could potentially be found allows for orientation within the structure. Observations of participants as well as data collected from the interviews revealed that the fixed depth assisted such judgements by constraining the number possible interpretations. Additionally, the interactive shortcut keys were particularly helpful in supporting this process. When participants felt lost during navigation, they would switch to one of the main top nodes on the hierarchy and start their navigation over. The shortcut keys thus served as a safety measure for re-establishing orientation.

[1] Refer back to Section 3.4.2 on page 47 for a detailed outline of how hierarchical depth is maintained while diagram complexity increases.

Design Lesson 1 – Use grouping by type to structure information. When possible, enforce a semi-fixed structure of the hierarchy by grouping items together by their categorical types, with the parent node labelled to reflect its content. Also, shortcut commands should be provided to allow users to jump to the main parent types from anywhere on the hierarchy.

Another feature that assisted orientation within the hierarchy was its auditory display, specifically, the *Error* sound and the continuous ambient sounds (in the low-verbosity condition). Observations and interviews also revealed that, while designed to convey illegal moves on the hierarchy, the *Error* sound was also used as a reference sound since it conveyed feedback about navigation possibilities. Some participants also reported trying to listen out for the ambient sounds at points during the interaction where they felt disoriented or felt they needed a confirmation about their current position on the hierarchy. But not all participants relied on this particular feature of the hierarchy and, as shall be described below, some simply ignored the audio display conveying navigational information, and relied solely on the semi-fixed structure and depth of the hierarchy for navigation and orientation.

5.2.1.2 Control Over the Display

When visually inspecting a diagram, one can control which information one attends to by gazing at a particular part of the spatially organised information. Sound is an inherently transient phenomenon, however, and any information encoded in audio disappears from the environment as soon as it is displayed or heard – unless it is continuously replayed. As reviewed in Chapter 2, this is considered one of the issues associated with using sound to convey information. The interactivity of the hierarchy-based audio-only tool provided a way to overcome such a shortcoming by allowing a user to have explicit control over which parts of the hierarchy is displayed in audio. One could imagine an audio display of the hierarchy where a user controls the sounds by simply pressing a play button, then passively listening to the displayed audio as it scans through the full hierarchy or parts of its branches.

A number of systems previously reviewed relied on a similarly passive approach to support non-visual access to graphically represented information (e.g Kennel, 1996; Blenkhorn and Evans, 1998). Instead of this, participants in Study 1 actively listened to the hierarchy by interactively triggering the audio display of different parts of its branches; they were effectively able to play and replay a given sound as many times as they wished while attempting to extract the meaning it encodes. The interactivity of the hierarchy in this case affords control over the flow of information that is received, and thus, in a somewhat similar way to that of controlling gazing around a space, allowed a user to choose both which part of the diagram to inspect and the number of times to inspect it. Examples of systems where active exploration was similarly supported include Mathtalk programming for reading and browsing algebra expression in audio (Stevens et al., 1997), the TeDUB system for accessing technical drawings (Petrie et al., 2002), and Kekulé (Brown et al., 2004) for audio-only interaction with molecule diagrams.

Design Lesson 2 – Allow for active control of the display. The user should be allowed to control which information is auditorally displayed by actively browsing the hierarchy. Content information should thus only be displayed when the hierarchy is interrogated by the user.

Design Lesson 3 – Hierarchical path selection. To further support orientation within the hierarchy, the selection of a hierarchical path should only be changed in response to a user's action or when accompanied by a detailed message that explicitly highlights the occurrence of a change and the new position on the hierarchy.

5.2.1.3 Task Completion Time

There was no statistically significant difference between task completion times in the high-verbosity and low-verbosity conditions except on Task 3. Task 1 took the longest in both conditions; over 60 seconds (compared to 40 seconds and 30 seconds for Tasks 2 and 3). This could be due to the nature of Task 1, which required a participant to locate a class on the hierarchy and discover its connections to other classes on a diagram. A class can have up to four types of connections each involving one or more classes, which potentially increases the number of items to inspect in order to complete such a task. In contrast, Task 2 required a participant to locate a relation and identify the classes related through it. Relations in UML Class diagrams typically involve no more than two classes, and so inspecting a relation is often likely to be a shorter task to complete than that of inspecting the relational information associated with a given class.

Task 3 took the least amount of time to complete in both conditions, but using the low-verbosity presentation mode yielded significantly shorter task completion times than using the high-verbosity mode. Task 3 required participants to enumerate the diagrams items; that is, establishing how many classes, associations and generalisation there are on a diagram. An examination of the interaction logs identified two distinct strategies employed by participants to retrieve such information. In the first strategy, participants established the number of items on the diagram by browsing through and counting the list of items in each of the main branches of the hierarchy. In the second strategy, participants relied on the audio display of the hierarchy, which automatically speaks the number of children of a node when it is expanded. Clearly, the former strategy would take longer to execute than the latter, and it seems that more participants in the high-verbosity condition relied on the former, slower, strategy to complete Task 3 than did the participants in the low-verbosity condition.

Thus, while the overall results partly supported hypothesis H2a, it seems unlikely that the significant difference in the completion times for Task 3 was an effect of the auditory presentation mode, but could rather be attributed to participants' interaction styles. Given the above analysis, hypothesis H2a could in fact be fully rejected; varying the presentation mode did not affect the time it takes to complete diagram inspection tasks.

5.2.1.4 Task Structure

While capturing the relational information encoded in UML Class diagrams, the hierarchy-based audio-only tool used in Study 1 did not capture any spatial information related to the layout of the classes or their relations on the original visual diagrams. All relational information was decomposed and structured across several levels of the hierarchy. For instance, a user encounters a particular class on level 2 of the hierarchy, its connection types on level 3 and the classes linked to it via each relation on level 4[2]. It was therefore the user's task to traverse the path containing the relational information of interest, to retrieve it one bit of information at a time, then to integrate these bits of information and form a coherent understanding of the represented relational information.

The tasks used to assess participant ability to inspect relational diagrams in audio required them to retrieve such relational information from the hierarchy; a task that does not involve nor necessitate knowledge about the spatial distribution of the retrieved information on the original diagram. In fact, spatial arrangements do not encode any information about a UML Class diagrams[3] – though, visually, the spatial arrangements of diagram elements on a two-dimensional plane can affect reading and interpretation (Purchase et al., 2001). There was thus a close match

[2] Refer back to Figure 4.1 on page 54 for an illustration of a Class diagram modelled as a multiple perspective hierarchy.

[3] Refer back to Section 3.3 on page 39 for a thorough discussion about how a relational diagram encodes information.

between the captured information, the way this information was hierarchically structured and the structure of the diagram inspection tasks.

Not capturing spatial information, combined with the match between the structure of the tasks and the structure of the hierarchy changed the way participants perceived their interactions with the relational diagrams. When interviewed at the end of the their sessions, almost all participants reported that, while completing the diagram inspection tasks, they thought about the information in terms of a hierarchy rather than a Class diagram. That is, they tended to have a picture of a hierarchy in their minds as they retrieved information from it, rather than that of a diagram with boxes and arrows. Only when a subset of relational information is retrieved, for example that a class A has an association from a class B, did the participants picture such information as two labelled boxes connected with an arrow, but this "picture" was discarded when the corresponding task was completed. This change in perception is likely to be caused by the decomposition of relational information across a number of hierarchical levels, requiring a participant to experience the relations one bit of information at a time as they traverse the hierarchy.

Since relational information was discarded from memory as soon as retrieved and used to complete a task, participants ability to keep and overall picture of the diagram as a whole; i.e. to build an overview of its structure, was hindered. Indeed, retaining all relational information encoded in the hierarchy would be necessary for acquiring a complete overview of the represented diagram, yet such a task would put too much mental demand on memory, which is fundamentally characterised by a limited capacity for holding information in the short-term, typically not exceeding 7 ± 2 items at a time (Miller, 1956). The problem of supporting users constructing an accurate overview of an auditorally represented set of information is well known (e.g Zhao et al., 2004; Kildal and Brewster, 2006) and the need for developing a means for presenting diagrams in audio so that a user can integrate information from various sources into a coherent whole has previously been acknowledged (Bennett, 2002). Augmenting the hierarchy with a further auditory representation to convey an impression of the topology of the relational diagram it represents, such as the sonification technique developed by Brown et al. (2006), could be a potential solution to such a problem. However, higher complexity diagrams may still cause problems as their sonification is likely to increase both in display time and complexity of composition. Providing overviews in audio is indeed an important and challenging problem, but is beyond the scope of the work presented in this thesis.

5.2.2 The Audio Presentation

In addition to navigating through the hierarchy, participants needed to interpret the audio display triggered by their interactions in order to effectively inspect and extract information about the diagrams from the audio-only tool. As discussed in Section 4.2.2, three types of information about the hierarchy were identified and displayed in audio; navigational, content and contextual information, all of which were found to be useful to varying extents. Both video logs and interview data were used to collect and analyse participants' reactions to the sounds used to convey each type of information in the two presentation modes.

> ***Design Lesson 4 – Convey three types of information.*** Three types of information should be communicated when a hierarchy-based model is presented in audio; *content*, *navigational*, and *contextual* to capture the richness of information that is contained in a hierarchical structure.

5.2.2.1 Navigational Sounds

Participants showed three distinct reactions to the designed navigational sounds; they listened out for some; appreciated others for the added aesthetics; and ignored the rest. The first reaction was mainly towards two non-speech sounds; the *End of List* and the *Error* sounds. When interviewed, all participants mentioned these two sounds as being informative and helpful during navigation. As described above, the *Error* sound did more than flag illegal moves and participants used it as a reference sound to assist orientation within the hierarchy. For instance, hearing the *Error* sound when pressing the expand key could confirm that the deepest level of the hierarchy has been reached or that the current node is already expanded and so movement should be directed elsewhere. Similarly, the *End of List* was used to recognise the event of reaching the last item within the list, but this information was also found to be important for completing other actions, such as counting the number of elements contained in a list.

The second kind of reaction was noted towards sounds that conveyed the successful expansion and collapsing of hierarchical branches and the switching between different hierarchical perspectives. These included both the non-speech *Expand*, *Collapse* and *Switch* sounds used in the low-verbosity modes and their equivalent spoken descriptions in the high-verbosity mode. Participants noted that such sounds helped them confirm that their actions were successful, but did not find them necessary because they were already aware that they pressed a command key corresponding to executing one of such actions. When asked whether they preferred to do without these sounds, however, most participants noted that they actually prefer hearing confirmations of their actions and that they appreciated such additions to the audio-only experience. There is a contrast between the type of information conveyed by the sounds corresponding to participants first and second reactions that can explain such reactions. On the one hand, the *Error* and *End of List* sounds conveyed information that is somewhat out of a user's control, drawing participants attention to unexpected events that occurred as a result of their interactions with the hierarchy. On the other hand, the expanding, collapsing and switching sounds conveyed information that is directly linked to a user's control, flagging events that are essentially expected to occur.

The *Browse* sound, which was used in both presentation modes and which can also be considered as a confirmation sound flagging expected outcomes of interaction, was ignored by most participants. When asked about the *Browse* sound, most participants noted that they simply did not recall hearing it while they browsed the hierarchy. This also meant that mapping the pitch of such a sound to indicate the depth of the hierarchy was not picked up by any participant and, as described above, participants relied on other means for keeping track of their position on the hierarchy to navigate it. The fact that the *Browse* sound accompanied sounds describing nodes' content could explain why it was ignored; as the participants inspected the hierarchy to complete diagram inspection tasks, they were more likely to focus on the displayed content of the nodes to extract needed information rather than on the accompanying non-speech sound. This observation should not be considered a recommendation against the use of the depth to pitch mapping, however. It is possible for example to switch the application of such a mapping to the content sounds rather than the navigational sounds; for example by varying the pitch of the spoken labels of the nodes. Such technique has in fact been shown to improve navigation within menu-based telephone interfaces (Shajahan and Irani, 2004; Irani et al., 2006) and could further improve navigation within the hierarchy-based audio-only tool used for diagram inspection.

Design Lesson 5 – Emphasise the occurrence of unexpected events. The occurrence of unexpected events should be explicitly highlighted; for example users in Study 1 found the sounds used to highlight reaching the end of a list and the occurrence of an illegal move particularly useful for orientation within the hierarchy. Sounds communicating feedback about expected events (e.g. moving between nodes, expanding and collapsing branches) should not be excluded unless they interfere with other sounds in the interface.

5.2.2.2 Content Sounds

Nodes' content was displayed using either speech or non-speech sounds. When encountered on the hierarchy, classes and relations names were spoken in both modes of presentation. Relations types and directions on the other hand were displayed using a parameterised auditory icon in the low-verbosity mode and using a spoken description in the high-verbosity mode. The non-speech sounds were well received by the participants who used the low-verbosity condition; as one participant commented *"[the relations] sounded like how they would have been drawn"*. Indeed, there was no statistically significant difference between participants's scores on the first and second diagram inspection tasks in the two conditions, which required accurate interpretation of the relations types and directions. Participants who heard parameterised auditory icons scored just as well as those who heard explicit spoken description of the relations types and directions.

An observation that is in line with these comments is that most participants who heard the parameterised auditory icons *drew* their answers on the tasks answers sheets, in contrast to participants in the high-verbosity condition who wrote the description of the relations types and directions as spoken by the system. Furthermore, when interviewed, participants in the high-verbosity condition described that they sometimes found it difficult to interpret the direction of an association relations, a complaint not mentioned by participants in the low-verbosity condition. Again, there was no significant differences between participants scores on the diagram inspection tasks, and it therefore seems that extracting direction from a spoken description required more effort than extracting it from a non-speech representation. For example, determining that the spoken description *"Association From"* is in fact referring to an arrow that is pointing outwards rather than towards a class – which would be described as *"Association To"* – was a more difficult task than hearing a long followed by a short sound forming the non-speech "auditory arrow".

The difficulty in extracting a relation's direction was not mentioned for the generalisation relations – spoken as *"Inherits From"* for a relation pointing outwards, and *"Super Class Of"* for a relation pointing towards a class. It seems that the terminology of UML Class diagrams notation for associations relations was somewhat less intuitive in this case. Of course the difficulty in interpreting direction from a spoken description would be affected by a participant's expertise and familiarity, both with UML Class diagrams and with using the provided audio presentation mode. Further, while the parameterised auditory icons representing relations types and directions were somewhat superior to the equivalent spoken description – which can testify to the design of such non-speech sounds – it is not possible to generalise these observations to any non-speech versus a speech description of iconic graphical representations.

> ***Design Lesson 6 – Use non-speech sounds to display iconic content.*** Using timbre to display relation types and varying the order of the short and long sounds to display a relation's direction was more intuitive than using speech to display iconic content. Where possible, describing iconic content of relational diagrams in speech should be avoided and replaced with equivalent non-speech sounds.

5.2.2.3 Contextual Sounds

Audio was used to convey one specific type of contextual information; feedback about the parent branch of the current node. In the high-verbosity condition, a spoken description of such context information described the immediate parent of the current node. Participants could request such information at any moment during their interaction, but this feature was seldom used. Instead, participants often "physically" moved between levels of the hierarchy, back and forth, in order to determine their current context; for instance by pressing the collapse arrow key to return to the current node's parent and infer context from the displayed content audio, then going back to relocate the node of interest.

As previously mentioned, not all participants in the low-verbosity condition listened out for the continuous ambient sounds. As was the case with the *Browse* sound, the designed mapping, which increased the pitch of the continuous ambient sounds the further a user traveled down a particular branch, was not picked up by our participants. Participants noted that they were more aware of the ambient sounds when they switched to different perspectives rather than when they were navigating within a single perspective. Mapping the timbre of such sounds to match the hierarchical branch was thus an effective means for conveying contextual information, but only at points where participants switched between contexts. Thus, the continuous ambient sounds seem to effectively communicate transitional information. Parente (2008) uses a similar approach to providing contextual information to support auditory interacting with GUI applications through "audio streams", which are continuous ambient audio that vary in timbre and rhythm to reflect the context of interaction. Parente (2008)'s results show that these streams benefit usability of audio-only interfaces, particularly in terms of increasing users awareness of peripheral information. When interviewed, participants in our study reported that they did not find the continuous ambient sounds annoying or irritating, which makes their unexpected function of conveying information about transitions between contexts worth including in similar designs of audio-only interaction.

> ***Design Lesson 7 – Avoid displaying context information using spoken descriptions.*** Contextual information should be conveyed through less intrusive means than spoken descriptions, particularly when lengthy messages need to be displayed. In the case of our evaluations, lengthy context messages pushed users to avoid requesting such information all together.

> ***Design Lesson 8 – Use ambient sounds to convey hierarchical perspectives.*** Users in our evaluation were more aware of the continuous ambient sounds at points of the interaction where they switched perspectives. Mapping the timbre of such sounds to match the hierarchical branch was thus an effective means to convey context, but the ambient sounds were more effective at communicating transitional information and should therefore be gradually faded out to a minimum amplitude when user movements are limited within a single branch.

5.2.3 Developing Expertise

To effectively inspect a given relational diagram, participants also needed to develop expertise in navigating the hierarchy and interpreting the audio display triggered by their interactions. Participants' scores were consistently high across the four scenarios while the diagrams complexity increased, which could be considered an indication that they were indeed getting better at using the provided audio-only tool to complete the diagram inspection tasks.

The concept of *interaction traps* was used to motivate an approach for capturing participants' progress and development of expertise in a systematic way. An interaction trap identified instances where a user deviated from the optimal hierarchical path for retrieving information about a diagram, which typically resulted in the user retrieving incorrect information, or having to retract back to the optimal path. Participants' strategies for traversing the hierarchy and retrieving information necessary for completing diagram inspection tasks were classified as either *inefficient*, when they included an occurrence of one or more interaction traps; *less efficient*, when they included no interaction traps but involved one or more illegal moves; or *efficient* when they included no interaction traps and no illegal moves[4].

[4] Refer back to Section 4.2.4.4 on page 65 for a detail outline of the devised approach for capturing and measuring interaction efficiencies.

5.2.3.1 Interaction Efficiencies

Across the two conditions, 50% of inefficient interactions occurred in the first scenario, where participants were still unfamiliar with the hierarchy-based audio-only tool, their only experience having been gained from the training that preceded the testing part of the study. Because inefficient interactions were measured in terms of interaction traps, they were typically a result of a mismatch between participants' understanding of how to achieve an objective and how it could actually be achieved using the provided tool. For example, when asked to locate the class *Sheep* and identify its connections to other classes on the diagram, seven out of twenty participants initiated their interactions by taking a shortcut to the *Classes* branch of the hierarchy, then browsed through the list of classes to locate the *Sheep* class (see Figures 4.1 and 4.5). However, as soon as they encountered a class labelled *Animal* on such a list, they decided to explore the children nodes of this class with the assumption that *Sheep* would be a child of *Animal*. This incorrect move was based on the incorrect assumption that all "animals" will be listed under the *Animal* node, and is thus a result of misunderstanding the basis for grouping classes on the hierarchical structure. Both the *Sheep* and *Animal* in this case represented distinct classes on the diagram rather then lists of classes. Participants had thus selected the wrong strategy to achieve a correctly interpreted objective and fell into an interaction trap.

The interaction trap in this example occurred as a direct consequence of the modelled domain. That is, participants confusion can be said to be a consequence of the names used to label the diagram classes, which influenced participants' interpretation of grouping structures. There were other instances of inefficient interactions that were a direct consequence of the way in which the hierarchy-based model captured and presented relational information. Specifically, the decomposition and organisation of relational information across several levels of the hierarchy meant that class names were duplicated at various levels[5]. This was particularly the case in the *Classes* branch of the hierarchy, where classes were nested within one another. As a result of this duplication, some participants misinterpreted their current position on the hierarchy – for example, thinking they are on level 2 while in fact they were on level 4, or vice versa. This caused them momentary confusion as they issued browsing commands that did not trigger sounds corresponding to what they expected to hear, and they fell into an interaction trap. They typically remained trapped until they reestablished orientation by, for example, relying on the *Error* sound

[5] Refer back to Figure 4.1 on page 54 which shows a simple Class diagram translated into a multiple perspective hierarchy.

to determine navigation possibilities – as described above – or using a shortcut key to switch back to the top node of the desired perspective and start navigation over.

Interaction errors, which were used to classify participants' interactions as less efficient, occurred when participants attempted to issue illegal moves that triggered the system to display the *Error* sound. These were incorrect navigation actions of two main types. The first is issuing a command to execute an already executed action; for example, attempting to expand a node that is already expanded. The second is issuing a command to execute an action that, given the current state of the hierarchy is not possible to execute; for example, attempting to expand a leaf node (i.e. a node not nesting any other nodes, typically found on level 4 of the hierarchy). In both instances, less efficient interactions were typically a result of participants misinterpreting the available possibilities for navigation. Thus, both inefficient and less efficient interactions were a result of misinterpreting auditory cues that conveyed the state of the hierarchy and navigation possibilities. Improving the auditory display of the hierarchy could therefore improve the support for more accurate interpretations of such information. For instance using hierarchical earcons (Brewster, 1998) or varying the parameters of the speech display of hierarchical node (Shajahan and Irani, 2004) could be used to reflect nodes nesting positions.

Design Lesson 9 – Constrain navigational possibilities to reflect context. Movements on the hierarchy should be constrained depending on where the user is located on the hierarchy. In particular, movements between cousin nodes should be disabled such that a user loops to the first child of a list rather than move to the next cousin node when reaching the end of such a list. Movement to parent nodes should also be constrained; to avoid confusion, the interface controller used to move to parent nodes should be different to the controller used to move to sibling nodes. For example, if a 4-way navigation controller is used to navigate the hierarchy such as a joystick or the keyboard cursor keys, then the functions of the keys for moving within the hierarchy should be mapped to match the layout of the hierarchy.

5.2.3.2 Learning Patterns

Overall, and in both conditions, the rate of inefficient and less efficient interactions decreased, while the rates of efficient interactions increased from one scenario to the next. Decreasing and increasing rates of less efficient and efficient interactions fluctuated as the scenarios progressed, but the fluctuations differed in each condition. In the high-verbosity condition, the highest rate of efficient interactions occurred in the third scenario before decreasing slightly in the fourth scenario, while the rate of less efficient interactions decreased in the third scenario then increased again in the fourth scenario. In contrast, the highest rate of efficient interactions in the low-verbosity condition occurred in the fourth scenario after having drastically decreased in the third scenario, while the rate of less efficient interaction increased drastically from the second to the third scenario, before decreasing again in the fourth scenario.

There was no statistically significant differences between the two conditions in the first and fourth scenario, demonstrating that participants' initial and final levels of expertise were equivalent. The statistical differences in the second and third scenario, however, demonstrate that the patterns of attaining such levels were different in the two conditions. Participants in the low-verbosity condition were overall more efficient at executing diagram inspection tasks in the second scenario, while participants in the high-verbosity condition were overall more efficient in the third. Thus participants' expertise development in using a high-verbosity presentation mode was slow between the first and second scenarios, then steadily progressed to the better throughout the remaining scenarios. In contrast, participants' expertise development in using a low-verbosity presentation mode progressed well between the first and second scenarios, was hindered between the second and third scenarios, then reached an equivalently high level in the final scenario.

The main variable between the scenarios was the diagrams complexity, which increased in terms of the number of diagram elements between the first and second scenarios – from a {3, 1, 2} to a {4, 2, 3} complexity – and in terms of representational complexity between the second and third scenarios – from a {4, 2, 3} to a {4, 3, 1} complexity[6]. Diagrams complexity increased from a {4, 3, 1} complexity in the third scenario to a {7, 6, 2} in the fourth, differing in terms of both the number of items and representational complexity, yet participants in both conditions produced most of the highest rates of efficient interactions and lowest rates of inefficient interactions in this scenario. However, there are a number of reasons that make it hard to interpret the differences in the fluctuations of the captured learning patterns in the second and third scenarios as an effect of the presentation modes. First, the differences between diagram complexities in these scenarios were not prominent (from a {4, 2, 3} to a {4, 3, 1}). Second, the observed effect was not consistently associated with one presentation mode; participants in the low-verbosity condition outperformed those in the high-verbosity condition in the second scenario, then participants in the high-verbosity condition outperformed those in the low-verbosity condition in the third scenario.

[6] Refer back to Figure 4.5 on page 64 which illustrates the complexities of the Class diagrams used in each scenario.

Finally, and most importantly, there were no explicit auditory presentation techniques used to highlight the differences in the representational complexity between the two scenarios (e.g the addition of involuted relations). Such complexity was equally captured in the high-verbosity and low-verbosity conditions and was reflected in the hierarchical structuring of involuted relations rather than its auditory display. The achieved results, therefore, only allow for the conclusion that participants in both conditions learned to solve diagram inspection task at overall similar rates, therefore rejecting hypothesis H2c.

5.3 Constructing Diagrams in Audio

Study 2 tested two main hypotheses; first, that a hierarchy-based audio-only tool allows for construction of relational diagrams; and secondly, that varying the interaction strategies for supporting diagram construction in terms of isolating or combining inspection and editing modes affects users' performance on diagram construction tasks.

All participants were able to construct relational diagrams using both interaction strategies. High scores were achieved when participants answered questions about relational information that was constructed using the audio-only tool (a mean of 76% and 77% in the Guided and Non-guided conditions respectively). There was also no statistically significant differences between these scores and those achieved in a visual control condition. This result is thus considered a reflection of participants ability to use a hierarchy-based audio-only tool to construct relational diagrams, effectively supporting hypothesis (H3) tested in this study. The following sections examine aspects of participants' interactions that affected their performance on diagram construction tasks.

5.3.1 Performance on Diagram Construction Tasks

5.3.1.1 Construction Times

The reported results have shown that participants spent significantly more time constructing diagrams when using the Guided strategy, effectively supporting part of hypothesis (H4a) tested in Study 2. Constructing a diagram using the hierarchy-based audio-only tool involved both inspecting it and executing editing actions to alter its content. Inspecting the diagram was supported through equivalent means in both the Guided and Non-guided strategies, while the execution of editing actions differed. Participants using the Guided strategy could execute any editing action from anywhere on the hierarchy and were guided through the steps required to complete such actions. Their inspection of the hierarchy was therefore independent of their editing, and vice versa. On the other hand, participants using the Non-guided strategy could only execute editing actions that were relevant to their current position on the hierarchy. Editing actions were therefore dependent on inspection. However, in both conditions, editing actions were equivalent in terms of the number of operational steps required to execute them.

Operational steps are the building blocks of the interactive expressions issued by a user to the system in order to execute an editing action. In both strategies, an editing expression consisted of three main parts; a command (*C*), an argument (*A*) and a parameter (*P*). A user expresses their desire to execute an editing action by issuing the appropriate command, specifying the corresponding arguments, and providing the desired parameters. For example, to add an attribute to an entity on the diagram, a user uses the computer keyboard to issue the command *C* = <*cmd* + *T*>, specifies which entity to add the new attribute to *A* = <*a specific entity*>, and supplies a label to be used as a name for the new attribute *P* = <*new attribute name*>. The difference between executing editing actions in the Guided and Non-guided strategies was in the way the operational step of specifying an argument (*A*) within an editing expression was realised. To specify which entity should the new attribute be added to in the above example, a user in the Guided strategy was provided with a list of entities to choose from after they issued the command for adding an attribute, while a user in the Non-guided strategy had to locate the desired entity on the hierarchy before issuing such command. An editing expression therefore takes the form <*Command* + *Argument* + *Parameter*> in the Guided strategy, and the form <*Argument* + *Command* + *Parameter*> in the Non-guided strategy. That is, the argument (*A*) in the Guided strategy is postfixed, always proceeding the editing command *(C)* and specified by following the system suggestions, but prefixed in the Non-guided strategy, preceding the editing command *(C)* and specified using the hierarchy to select the desired item at the moment of issuing the desired editing command.

Because the command (*C*) and parameter (*P*) parts of editing expressions are equivalent in the two strategies, it is likely that the time it takes to specify the argument part (*A*) affected the speed by which a given editing action was executed, and hence impacted the overall diagram construction times. Engaging in an exchange with the guiding system to specify (*A*) increased execution times in the Guided strategy, significantly more than did the inspection of the hierarchy to search for and locate an item to specify (*A*) in the Non-guided strategy. Hypothesis H4a anticipated this outcome, because the main audio output used to guide the user during the argument specification exchange in the Guided strategy was speech, which takes time to be uttered. It is of course possible to decrease interaction times when using the Guided strategy by increasing the speed at which speech is presented, but such an approach should be applied with care as processing rapidly spoken information is not a trivial task and can depend on both user experience and the complexity of the spoken words or sentences (Asakawa et al., 2003). A possible alternative way to increasing the speed of interaction while maintaining efficiency in the Guided strategy would be to provide more concise speech output by using spearcons, which have been found to speed up performance on menu navigation while preserving the intelligibility of an interface (Walker et al., 2006). The significant difference in construction times was not only a consequence of using speech when guiding users, however. The way in which the two strategies supported participants' when completing certain editing actions seems to have also contributed to speeding up performance times. This is discussed further in Section 5.3.2 below.

5.3.1.2 Diagram Comprehension

Hypothesis H4b predicted that using the Guided strategy would promote more coherent understanding of the relational information encoded in the constructed diagrams. This was measured in terms of participants' scores on post-construction tasks, which involved answering questions that probed their understanding of implicitly embedded relational information that could be explicitly revealed through the construction process [7]. The results did not support this hypothesis; no statistically significant difference was found between participants' scores on post-construction tasks in the Guided and Non-guided conditions as compared to a visual control condition.

Hypothesis H4b was based on the assumption that isolating inspection and editing actions into two distinct modes of interaction would enable the user to focus on one rather than two activities at a time while editing the diagram. A user could thus focus on the content they are constructing without being distracted by other content that could be displayed as a result of inspection actions. When interviewed about their experience in using the Guided strategy to edit diagrams, most participants had in fact described feeling detached from the hierarchy, and hence from inspection while editing diagram content. However, participants reacted negatively rather than positively to such detachment because they found it harder to relate newly constructed

[7] Refer back to Section 4.3.4.3 on page 80 for a detailed description of these tasks and Appendix B.5 for the full list of post-construction questions.

content to existing content when using the Guided strategy. Participants still appreciated the ability to execute editing actions from anywhere on the hierarchy instead of being bounded by their locations on it. This was in contrast to their experience of using the Non-guided strategy, which they felt allowed them to situate themselves within the existing diagram content as they edited it in new content, but constrained their movements to the part of the hierarchy being edited.

Thus, the Guided strategy afforded a freedom from the hierarchy that made editing the diagram more flexible, but at the cost of increasing the difficulty of integrating new and existing content. The Non-guided strategy on the other hand afforded such integration, but at the cost of constraining movement within the hierarchy. As mentioned above, regardless of participants' impressions of each strategy, these differences did not have a significant impact on participants ability to retain a coherent understanding of the constructed diagrams. Since both strategies promoted similar levels of diagram comprehension, their distinctive features, as well as participants' preferences should be further explored in order to identify how interaction with each strategy could be improved in other terms, such as their usability. This is explored further in the following sections.

5.3.2 Editing Categories

Participants were interviewed at the end of study sessions and asked to specify which of the two strategies they preferred to use to construct diagrams in audio. Additionally, participants were asked to specify which of the two strategies they preferred to use for executing each single editing action supported by the provided system. Two categories of editing activities were identified along which participants's preferences could be classified. These are referred to as *Global* and *Local* editing categories, shown on Figure 5.2. Table 5.1 summarises the editing actions that make up each editing category.

5.3.2.1 Global Editing

Global editing actions alter the overall state of each categorical type, affecting higher level branches on the hierarchy, i.e. the *Entities* and *Relations* perspectives, mainly through the addition or deletion of entities or relations from the diagram. When a global editing action is executed, the audio output of level 1 and 2 of the hierarchy changes in terms of what is spoken when the list is browsed as well as their enumeration when the list is expanded (see Figure 5.2 and Table 5.1). There was no significant majority in favour of one strategy over another when executing global editing actions; some participants preferred to be guided through the process or editing the diagrams entities or relations since this meant that they would do less browsing, while others found locating entities or relations within the hierarchy much more intuitive and saves them time that would be wasted on short exchanges with the guiding system.

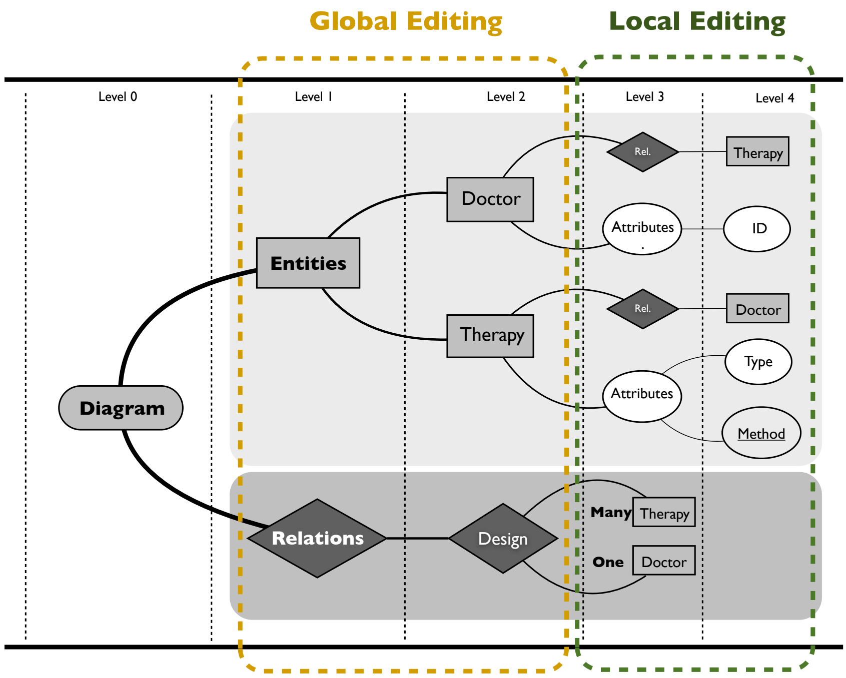


Figure 5.2: A multiple perspective hierarchy of an ER diagram highlighting the areas affected by each category of editing actions (see also Table 5.1).

5.3.2.2 *Local Editing*

Local editing actions alter the state of individual items within each categorical type, thus affecting the deeper levels of the hierarchy (see Figure 5.2 and Table 5.1). This include actions that add or delete an entity's attributes or a relation's cardinalities, consequently altering the audio output of levels 3 and 4 of the hierarchy also in terms of displayed content and enumeration. The majority of participants preferred using the Non-guided strategy, having found the Guided strategy to be too cumbersome when executing such actions.

In the Guided strategy, the deeper a diagram element is nested in the hierarchy the more system prompts the user has to respond to in order to complete the editing of that element. For example, the user/system exchange for adding the cardinality *Many* to the entity end *Therapy* of the relation *Design* is as shown on Table 5.2. The execution of the same action using the Non-guided strategy can be achieved by traversing the path *Diagram* > *Relations* > *Design* >

Table 5.1: Editing actions in each editing category (see also Figure 5.2)

Global Editing Actions	**Local Editing Actions**
Adding a new entity	Adding a new attribute
Deleting an entity	Deleting an attribute
Renaming an entity	Renaming an attribute
Adding a new relation	Adding a new cardinality
Deleting a relation	Deleting a cardinality
Renaming a relation	Changing a cardinality

Therapy then issuing the command <*cmd Y*> and choosing the desired cardinality.

Table 5.2: An example of User/System exchange in the Guided strategy. <command>, *"speech output"*, (comments).

Command: *C* = <*cmd + Y*>	
1) User:	<cmd Y> (issues command to add a cardinality)
Argument: *A* = <*Therapy*>	
2) *System:*	*"new cardinality. choose relation?"* (requests details)
3) User:	<browse> <select> (selects *Design* from the provided list)
4) *System:*	*"relation Design selected. choose entity end?"* (requests details)
5) User:	<browse> <select> (selects entity end *Therapy* from the provided list)
Parameter: *P* = <*Many*>	
6) *System:*	*"entity end Therapy selected. choose cardinality?"* (requests details)
7) User:	<browse> <select> (selects cardinality *Many* from the provided list)
8) *System:*	*"cardinality Many added to entity end Y"* (confirms action)

When performing local editing using the Guided strategy, most participants found themselves forgetting their original selections (steps 3 or 5 in the above example) before they could finalise the editing action (i.e. reaching step 7). This was because, while editing the diagram, participants often switched their attention between manipulating the diagram and reading through the written descriptions to extract relevant information about diagram content. Forgetting where they left the process of editing at when switching back to the interface, participants found themselves having to cancel the editing action and start over again. Many participants found this to be a frustrating process. Furthermore, participants tended to execute local editing actions systematically. That is, they preferred to complete all local editing related to a particular item within a categorical type before moving on to the next (e.g. adding all cardinalities to a relation before moving on to editing other content). The Guided strategy did not support such systematic editing in a way that the participants found useful. Progressing through the same set of user/system exchange more than once for each local editing action, in addition to the possibility of losing track of the order of the steps, affected performance times and further increased users' frustration with the tool. The Non-guided strategy on the other hand allowed for the systematic execution of related local editing actions. A user would have to locate the desired item on the hierarchy only once, then, while situated within that part of the hierarchy, the user could systematically issue the necessary local editing commands. The Non-guided strategy therefore matched how the participants structured and planned the construction, and was thus preferred by the majority of participants to execute local editing actions. A possible solution to improve local editing in the Guided strategy is therefore to allow the user to specify multiple edits to the same local area of the hierarchy. This way, the Guided strategy will function in a similar way to the Non-guided strategy with the added advantage of allowing flexibility in the execution of such actions from anywhere on the hierarchy.

Design Lesson 10 – Combine aspects from the Guided and Non-guided strategies to support editing diagrams through a hierarchy. This could be achieved by tracking users editing actions and position on the hierarchy, the sequence of guiding steps could then be triggered if the action and location bare no relevance to one another; the non-guided process should proceed otherwise. Additionally, allow users to specify multiple edits to the same local area of the hierarchy when using the Guided strategy.

5.3.3 Interaction Modes

In contrast to the Non-guided strategy, the Guided strategy structured the interaction in terms of two independent modes; an inspection mode where a user can navigate through the hierarchy and discover diagram content; and an editing mode where they can alter and manipulate such content. However, because participants had to switch between these two modes of interaction, they were prone to execute mode errors.

Mode errors are a common class of errors in human-computer interaction and, as defined by Norman (1981), result from a false classification of a situation leading to performing actions that are appropriate for the analysis of the situation but inappropriate for the actual situation. Mode errors are also said to occur when certain system features behave differently under different circumstances, and where interpretation of user actions are shifted depending on the current context or state of the system (Thimbleby, 1982; Johnson and Engelbeck, 1989). A common example of this can be found in text editing, where characters typed into an editor are interpreted as text in one mode and as commands in another. In order to produce the correct input, a user must keep track of which mode the system is in, if they misinterpret one mode for the other then they would be likely to confuse erroneous for correct interaction. The cause of mode errors is therefore typically associated with the user's inability to correctly interpret system states, either because cues conveying such information are inappropriately communicated to the user, or are not noticed because the user is busy with other tasks. These situations were observed when participants in our study used the Guided strategy to construct diagrams. Some functionalities in the inspection mode behaved differently in the editing mode, and vice versa, while the availability of other functionalities depended on the current mode of interaction. For instance, the keyboard arrow keys, which allowed for four navigation movements in the inspection mode (matching the basic structure of the hierarchy) were reduced to provide only two navigation movements in the editing mode (allowing a user to browse up and down a list of system's suggestions); the shortcut keys, which allowed a user to jump to the top level nodes of the hierarchy in the inspection mode, were completely disabled in the editing mode; and the *Enter* key, which was used to replay sounds in the inspection mode, was used to confirm selections in the editing mode.

Examining the interaction logs and the data collected from interviewing participants, it was possible to identify instances where participants switched between the inspection and editing modes and encountered problems related to mode confusion. As is the case with mode errors, these problems occurred as a result of either one, or a combination of two reasons; 1) participants drawing their attention away from manipulating the system and into other tasks – mainly reading through the diagram textual descriptions – and 2) the system's failure to properly convey information about its state.

5.3.3.1 Mode Confusion

As described above, switching back and forth between manipulating the tool and reading through its textual descriptions caused participants to be frustrated because they could not keep track of which step they were at in a given editing process. A further source of frustration in such situations was mode confusion, where upon switching from reading the textual description back to manipulating the tool led participants to also forget whether they were editing or inspecting the diagram. Table 5.3 shows an extract from an interaction log illustrating this situation.

Table 5.3: An example of mode confusion.

Command: ***C* = <*cmd + R*>**	
1) User:	<cmd R> (issues command to create a relation)
Argument 1: ***A1* = <*2*>**	
2) *System:*	*"new relation. how many entities?"* (requests details)
3) User:	<browse> <select> (selects the number *2* from the provided list)
Argument 2: ***A2* = <*Therapy*>**	
4) *System:*	*"two selected. choose first entity"* (requests details)
5) User:	<browse> <select> (selects an entity *Therapy* from the provided list)
Argument 3: ***A3* = <*Doctor*>**	
6) *System:*	*"Therapy selected. choose second entity"* (requests details)
7) User:	**(interrupts the editing process and reads through the textual description)**
8) User:	<expand> (attempts to browse by pressing the right arrow key)
9) *System:*	*(error)* (actions unavailable in the current mode)
10) User:	<switch> (attempts to jump to *Entities* by pressing a shortcut key)
11) *System:*	*(error)* (actions unavailable in the current mode)
12) User:	<browse> (attempts to browse by pressing the down arrow key)
13) *System:*	*"Tag"* (displays the next item on the suggestion list)

Here, a participant aims to create a relation *Design* between two entities *Therapy* and *Doctor*. The process of editing this relation requires a command (<*cmd+R*>), three arguments (a specification of how many entities should be related *A1* = <*2*>, and of which entities to relate *A2* = <*Therapy*> and *A3* = <*Doctor*>) and one parameter (a label for the new relation *P* = <*Design*>). The participant in this extract interrupts the construction of the relation at step 7 to turn back to the textual description of the diagram in order to retrieve more information about the relation. However, upon coming back to manipulating the tool in step 8, the participant starts interacting with the tool as if it were in the inspection mode. They first attempt to expand what they consider to be the currently selected node on the hierarchy, then to take a shortcut to the *Entities* branch. These actions are in fact unavailable in the editing mode, and the system displays the *Error* sound in response to both keystrokes. The participant then presses the down arrow key, and the system displays the next item on the current suggestion list *"Tag"*; in this case, the down movement was available for execution but it is likely that the participant still thought they were browsing the hierarchy rather than the list of system's suggestions.

In this extract, there were two audio cues which conveyed information about the fact that the current mode of interaction was the editing and not the inspection mode. The first was the *Error* sound displayed in response to the *switch* command in step 10. While this sound could be displayed in response to an illegal expand command in the inspection mode (for example, if the user attempts to expand an already expanded node), the shortcut key only triggers the *Error* sound in the editing mode because it could be invoked anytime and from anywhere on the hierarchy in the inspection mode. The participant could therefore have inferred that they were editing rather inspecting the diagram. The second cue was the gender of the voice used to display the speech output of the system. A male voice was used in the inspection mode and a female voice was used in the editing mode[8]; because the speech output *"Tag"* in step 13 of the extract was spoken using a female voice, the participant could have inferred that they were in the editing mode of interaction. Both cues were not picked up by the participant who fell into a mode confusion and continued to interact with the system as if it switched modes, when in fact it did not.

There were other instances where participants interacted with the system as if it did not switch modes, when in fact it did. In such instances too, participants fell into a mode confusion because they did not manage to pick up on auditory cues that conveyed mode information. Examples of this second instance of mode confusion occurred when a participant initiated then cancelled an editing action. Cancelling an editing action causes the system to automatically switch back to inspection mode, but some participants often continued interacting with the system as if it was still in the editing mode.

5.3.3.2 Preventing Mode-Confusion

The existing audio cues were not effective in communicating mode information. The most explicit of these; using different gender voices for speech output in each interaction mode, was not used by the participants to overcome mode confusions. When interviewed, most participants stated that they could recall hearing two different voices during their interactions but the majority had failed to rely on such cues to manage their interactions across the editing and inspection modes.

Another feature of the audio design which might have further contributed to mode confusions is the continuous ambient sounds, which were designed to distinguish between hierarchical perspectives. As described in Section 4.2.2; two distinct sounds with different timbres are contin-

[8] Note that the function of these voices was explicitly highlighted during the training part of the study.

uously displayed when a user browses a particular branch of the hierarchy. Although participants did not explicitly highlight this as an issue, when an editing action is initiated, these ambient sound remained audible. For example, if a participant was inspecting the *Entities* branch of the hierarchy and issued a command to create a relation, the *Entities* ambient sound continued to be displayed while the new relation was being edited. The original intention behind this design feature was to help users retain an awareness of which part of the hierarchy they left the system at as they detached from it to go into the editing mode. Ideally, users would be able to tell that they were in the *Entities* branch and that they will go back to this branch when the current editing action is completed. Then, because the audible continuous ambient sound would be accompanied by either a female or a male voice, it was expected that users will be able to interpret their current mode of interaction without confusion, but as described above, this was not the case and the opposite was actually true.

There are a number of auditory cues that could be used to communicate the system states more effectively and hence prevent or decrease the potential of users falling into mode confusions. The first possibility is to display more explicit speech messages that describe information about the current mode of interaction. For example, the single non-speech *Error* sound that was displayed when a participant issued a command that was not available in the editing mode (see steps 9 to 11 on Table 5.3) could be replaced with an explicit spoken description such as *"This command is not available because you are in the editing mode"*. Thus, instead of expecting the user to infer that they are in the wrong mode from the non-speech sound, the spoken description conveys the message more explicitly. There was in fact a similar feature that was used to provide a description of context on the hierarchy using speech; the user could request a description of the immediate parents of their current position[9].

Another possibility is to design audio cues that are contingent to each mode, such that a user hears an accompanying sound with every keystroke that correspond to the current mode of interaction. Monk (1986) successfully demonstrates that such a method can be used to alert users to potential errors when manipulating a "moded" visual interface, effectively reducing mode errors. Sellen et al. (1992b) described how Monk (1986)'s approach provided *reactive auditory feedback* where upon a user can determine the current mode *after* executing and action, and how *proactive feedback* could provide a superior alternative by allowing a user to determine current

[9] Recall that, in Study 1, this feature was seldom used for requesting context information (see the discussion on *Contextual Sounds* on page 95)

mode *before* initiating a possibly erroneous action (Sellen et al., 1992b). Adapting the continuous ambient sounds used in Study 1 to produce such proactive feedback could also be another possibility to increase a user's awareness of the current mode of interaction. In fact, as discussed in Section 5.2.2.3, ambient sounds effectively communicated transitional information, conveying information about context switch[10]. Parente (2008) has also shown that varying background sounds increased users awareness of their context of interaction. An additional continuous ambient sound could therefore be designed to distinguish between inspection and editing modes such that, when an editing action is initiated, existing ambient sounds from the inspection mode are switched off or faded away to the background while the editing mode sound is foregrounded to dominate the auditory display and highlight the successful switch of context.

Design Lesson 11 – Manage transitions between interaction modes explicitly. The Guided strategy breaks the interaction into two independent modes; an editing mode and an inspection mode. Care must therefore be taking to prevent the user from falling into mode confusion. Explicit auditory cues should be designed to convey mode information; for instance, using a distinctive continuous ambient sound that conveys mode status, or other cues that are contingent to each mode, such that a user hears an accompanying sound with every keystroke to correspond to the current mode of interaction.

5.4 Summary and Conclusion

This chapter presented a detailed analysis and discussion of the results obtained from two studies that examined audio-only interaction with diagrams. Overall, sighted participants were able to use the developed hierarchy-based audio-only model to both inspect and construct relational diagrams, which supports both hypothesis H1 and H3.

The hierarchy-based audio-only tool provided a semi-fixed structure and depth that allowed participants to anticipate where specific diagram content would be located prior to commencing navigation, and this helped orientation within the hierarchy. A fixed depth kept the hierarchy at a manageable level regardless of the complexity of the diagram, which also decreased the potential for disorientation. These features, coupled with the interactivity of the audio-only tool, allowed participants to control which part of a diagram's content is inspected and when. However, the afforded control matched the structure of the inspection tasks of searching for and retrieving

[10] Refer back to page 95 for a discussion about the use of the continuous ambient sounds to convey transitional events.

information, and since the hierarchy did not encode spatial information, it may not be effective for supporting inspective tasks that require interpretation of the spatial arrangements of diagram elements.

The analysis also showed that varying the verbosity of the audio presentation mode did not impact task completion times as the results seemed to show, but the strategies for completing diagram inspection tasks did, thus fully rejecting hypothesis H2a. Further, sounds which alerted participants to the occurrence of unexpected events were found to be most informative and were actively used for orientation within the hierarchy. Sounds communicating feedback about expected events were mostly appreciated for the added aesthetics. Navigational sounds that accompanied content sounds were completely ignored, which meant that mapping their pitch to reflect the depth of the hierarchy was ineffective. The parameterised auditory icons used to convey diagram content were well received by the participants, and seemed to be more intuitive in communicating relations directions than the equivalent spoken descriptions, and the continuous ambient sounds were found to be effective in conveying transitional information. Participants efficiency in completing such tasks increased from one scenario to the next, but was equivalent across the two audio presentation modes therefore rejecting hypothesis H2c. More longitudinal studies would be needed to establish whether long term use would change these learning patterns. The causes for inefficient and less efficient execution of inspection strategies were found to be related to the users' ability to interpret their position on the hierarchy and to issue the correct navigational actions in order to move around it.

In relation to constructing and editing diagrams, the analysis showed that the significant difference supporting hypothesis H4a was a consequence of the difference between the means for specifying editing actions in the Guided and Non-guided strategies. Particularly, the deeper the location of item on the hierarchy the longer it will take to edit it using the Guided strategy. Other aspects affecting performance when editing diagrams using each strategy were also identified. For instance, because the Guided strategy isolates inspection and editing into two distinct modes, it afforded a freedom from the hierarchy that allowed flexibility of executing editing actions, but made it difficult to integrate newly added items with existing content. The Non-guided strategy situated participants within the content of the diagram, affording integration of new and existing diagram content, but at the cost of restricting movements within the hierarchy.

Participants preferences when using each interaction strategy to execute diagram editing actions highlighted two categories of editing actions; *Global* and *Local* editing actions. The Non-guided strategy was preferred for *Local editing* actions, which affected deeply nested nodes on the hierarchy. This was because this strategy matched the way participants planned the execution of editing actions targeting those parts of the hierarchy. The Guided strategy was found to be cumbersome in these instances because it took longer to execute such actions, involving more steps which in turn made it hard for participants to keep track of the editing process. Frustration was further increased because participants fell into mode confusion when using the Guided strategy. Mode errors were due to participants' distraction from the construction task and to the inadequacy of the audio display to convey mode information. The section then discussed possible improvements to the sound design, such as developing more explicit speech messages, or adapting the ambient sounds to convey transitions between modes.

The presented analysis was used to compile a set of lessons for designing support for audio-only interaction with relational diagrams through hierarch-based models of representation. This concludes Part I of this thesis, Part II will focus on the use of the hierarchy-based audio-only model to support audio-only collaborative interaction with relational diagrams.

Part II

Collaborating Through Sounds

Chapter 6

Sound and Awareness in Collaboration

6.1 Introduction

Part I developed and evaluated a hierarchy-based audio-only model for supporting individual inspection and construction of relational diagrams. It also complied a set of lessons for designing support for audio-only interaction with relational diagrams. The ability to inspect a shared representation and to contribute to its construction and editing is essential for successful participation in collaborative design activities (Cherubini et al., 2007). The previous studies were thus an important and necessary step to complete before arriving at a stage where a study into collaborative audio-only interaction with diagrams could be conducted.

This chapter reviews the various ways in which the auditory modality is used in computer-supported cooperative work (CSCW) in order to motivate the focus of the audio-only collaborative study reported in Chapter 7. In general, audio is used to augment both remote and co-located collaborative shared spaces with means for mediating collaborators' conversational exchange, communicating incidental sounds emitted from shared environments, and representing aspects of the shared workspace or artefacts of collaboration. Sections 6.2 and 6.3 present the function of audio in each of these cases, and highlight a common trend amongst existing research that point to the important role that audio plays in communicating awareness information in remote and co-located collaborative workspaces. Section 6.4 then takes a closer look at the concept of awareness in collaboration, and reviews a framework for understanding and developing support for awareness in collaboration. Section 6.5 summarises and concludes the chapter.

6.2 Audio in Remote Collaboration

As a research field, CSCW is concerned with studying *"how collaborative activities and their coordination can be supported by means of computer systems"* (Carstensen and Schmidt, 2003). With computers being increasingly used to connect geographically separated locations, one of the aims of CSCW research is to develop applications that support interaction between remotely located collaborators. Due to the primary role that audio plays in communication (Chapanis, 2002), such support typically include means for allowing synchronous and asynchronous verbal exchange.

6.2.1 Mediating Conversational Sounds

Videoconferencing and Media Spaces are examples of applications that capture and transmit audio between geographically separated locations. They allow for several people to converse at the same time using video and computing equipments that are augmented with open audio channels (Fish et al., 1990; Gaver et al., 1992; Bly et al., 1993). A media space is defined as:

> *"An electronic setting in which groups of people can work together, even when they are not resident in the same place or present at the same time. In a media space, people can create real-time visual and acoustic environments that span physically separate areas. They can also control the recording, accessing and replaying of images and sounds from those environments."* (Stults, 1986)

> *"a computer-controlled teleconferencing or videoconferencing system in which audio and video communications are used to overcome the barriers of physical separation"*. (Baecker, 1994)

The two definitions emphasise the technology connecting remotely located people and the modalities that mediate their communication; a "space" is created across remote locations through the sharing of visual and auditory information, transmitted by incorporating cameras, microphones, monitors and speakers at each location. One of the early examples of such spaces is the Xerox PARC Media Space (Bly et al., 1993), in which continuous video and audio connections were used to mediate informal interaction between two remote research laboratories. The RAVE media space at Xerox EuroPARC was a sister system that was also developed to overcome the constraints of working together when physically separated (Gaver et al., 1992). Offices and common areas acted as nodes within a linked network, and users were able to display views from different video cameras around the building and set up two-way audio-video connections with any node in the network. Other early examples of media spaces include the VideoWindow (Fish et al., 1990), Cruiser (Root, 1988), and the Hydra system (Sellen et al., 1992a), which simulated face-to-face meetings by combining a camera and monitors with directional microphones to enhance the sense of presence and proximity. With the advancements in technology, media spaces have also extended beyond office space and research labs to include home environments (Hindus et al., 2001; Neustaedter and Greenberg, 2003).

Even though media spaces and videoconferencing developments initially aimed to mimic face-to-face interaction, it was soon realised that remotely shared media spaces offer different affordances than shared physical space, and hence different opportunities and challenges for collaborative action and the design of the tools that support it (Gaver, 1992a; Dourish et al., 1996). The potential of such affordances for communication was thus the focus of many investigations (e.g. Sellen, 1995; Olson et al., 1995) but a full discussion of what these affordances are and how they compare to face-to-face interaction in the current boost of technological development is beyond the scope of this review. What is of specific interest however, is the role of the auditory modality in supporting remote collaboration in such settings.

Audio in Media Spaces. Sellen (1995) compares remote collaborative interaction in audio-only, audio-video and face-to-face settings and reports that users manage their interactions well in the audio-only condition, that turn-taking is not affected when visual cues are absent, and that indicators of problems in coordination, such as interruptions and simultaneous starts of conversation, are more or just as likely to occur when collaborators share the same physical space. The main significant difference, concludes Sellen (1995), lies in whether interaction is mediated by technology or not, rather than whether a mix of audio and video or audio-only is used. In a similar study, Olson et al. (1995) compare face-to-face, video and high-quality audio conditions, also reporting no difference between the quality of work performed in face-to-face meetings and in settings with high-quality audio and video. However, in contrast to Sellen (1995)'s findings, Olson et al. (1995) show that the task quality is lower in the audio-only condition, with participants in their study preferring to interact with video rather than without.

Tang (1992) reports similar results to those of Sellen (1995) and, in addition, concludes that problems in managing turn-taking and coordinating eye gazes occur if the audio transmission is of low quality. A similar stance is made by Pagani and Mackay (1993), where users of media spaces in real world scenarios are reported to often switch off their connecting links and turn to other forms of exchange when audio quality degrades. When discussing the affordances of media spaces, both Gaver (1992a) and Dourish et al. (1996) also refer to the impact of the quality of audio on the perceived quality of the interaction. However, according to Tang (1992), audio is relatively more important than video in supporting remote collaboration and so recommends that, if network constraints require trade-offs to conserve bandwidth, video quality should be degraded before audio quality is in order to ensure more usable experiences of video conferencing. This recommendation is supported by findings from Hindus et al. (1996) and Ackerman et al. (1997) who present field studies of an audio-only media space called Thunderwire, and conclude that audio can be sufficient for a usable media space. Ackerman et al. (1997) highlight: *"conversation characteristics, such as turn-taking and overlapping speech [in Thunderwire] were in notable contrast to low-quality audio use, such as one finds with the telephone"*(p.62).

In general, the bulk of the studies on the impact of the type of communication media on collaborative interaction described above point to similar conclusions; that using audio is important in remote collaboration; that using audio alone to create a shared space results in performances that are either as effective or almost as effective as face-to-face or video communication; and that the quality of the transmitted media has a direct impact on collaborators' perceived quality of their interactions.

6.2.2 Mediating Incidental Sounds

In addition to mediating spoken conversation, audio captured by microphones at remote locations include sounds that are emitted from the environment at each end of a connection. Ackerman et al. (1997), for example, describe how users of their audio-only media space reported hearing the sounds of mouse clicks, paper rustling, background conversation, phone ringings and people moving through the building while they interacted through Thunderwire. Here, these types of sounds are referred to as "incidental" sounds to distinguish them from "engineered" sounds, which are intentionally designed into an environment to, for instance, communicate some aspects of its state. The ensemble of incidental sounds create an ambient atmosphere rich with "noise" that could be a source of annoyance, but it has often been reported to benefit the overall experience of remote collaboration. Users of Thunderwire, for instance, found low-level background noise to be not only acceptable but useful and sociable (Hindus et al., 1996). An example of this is an audible "click" sound that was triggered whenever a Thunderwire user switched their audio connection on or off. Such sound was actively used as a resource for determining when a participant joined or left the shared space. Ackerman et al. (1997) field observations report that a social norm evolved around this particular incidental sound: it was usual for someone to announce their presence as soon as they "clicked" in, and, if a click was heard without an announcement, someone asked who it was. Social sanctions, such as derisive comments, were applied by group members to those who were caught evading this norm.

In general, some of the most useful aspects of media space environments are their support for informal, awareness-based facilities because they provide means for communicating information about ongoing activities across locations (Fish et al., 1990; Gaver, 1992a; Bly et al., 1993; Olson et al., 1995). Incidental sounds contribute to this process, and are more prominent than video data because the information they communicate can be attended to without the need for constant monitoring of video screens. Indeed, Dourish et al. (1996) report that continuous audio access is an invaluable source of information that allows for the *"flexible management of collaboration and interaction through peripheral monitoring of ongoing activity and the assessment of anothers availability for interaction"*(p.49). Similarly, Bly et al. (1993) highlight that one of the most powerful uses of open channels in the PARC media space is the transmission of ambient sounds because they communicate peripheral awareness information that convey overviews of who is present and what they are doing.

Audio and Awareness. In describing shared work, Gaver (1991) makes the important distinction between focussed collaboration and divided labour: *"just as most people don't work alone at all times, nor do they always work together. Often people are merely aware of each other - aware of others' presence, perhaps their activities and progress [and] shift from working alone to working together, even when joined on a shared task"* (p.294). Gaver (1991) goes on to assert the potential of using auditory cues to communicate serendipitous information that increase peripheral awareness of activities and events. According to Gaver et al. (1991), it is in the appropriate provision of such information that the tension between the support for focused collaboration and individual work in shared spaces is potentially eased.

This potential was demonstrated in the ARKola simulation (Gaver et al., 1991), in which pairs of users manipulated a shared workstation displaying a number of machines in a virtual factory plant for handling the production of a virtual soft drink. Each user had a limited view of the plant and controlled its machines using the computer mouse. Feedback about users actions on the machines was displayed graphically and using auditory cues, which included a mixture of continuous ambient and short sounds. Users were remotely located and could communicate through an audio-video link. Because a user had a limited view of the workspace, coordination and communication with their partner was necessary for the smooth running of the simulated factory. Gaver et al. (1991)'s results show that users notice warning sounds and that confirmation sounds provide an awareness of partners actions. Gaver (1991) describes that, when used in this way, audio provides a new dimension of reference that is absent from the graphics-only condition. He writes:*" [sound] smoothed the transition between division of labour and focussed collaboration. Being able to hear the status of offscreen machines allowed a dissociation of focussed visual attention and more general awareness, so that each participant could have an area of primary responsibility and still join together to solve problem"* (p.302).

Gaver's work on auditory icons (Gaver, 1986) and the affordances of media spaces (Gaver, 1992a) was incorporated in the RAVE media space to expand the naturally occurring incidental sounds beyond the "earshot" with intentional "engineered" sounds. The EAR system augments the media space by transmitting short auditory cues communicating information about upcoming or ongoing events to people's offices (Gaver, 1991). For instance, the sound of a door opening or closing is used to indicate that a connection between two nodes is established or terminated, and the sound of a knock is displayed to signify an incoming video call. In addition to announcing formal events, EAR was found to support casual awareness of colleagues activities and to increase opportunities for informal communication. Cohen (1993) describes ShareMon, a file-sharing application which also employs auditory icons to notify users of ongoing group activities in an unobtrusive manner. Cohen (1993) used auditory icons to indicate such events as user's logging-in and out, and file transfer progress using knocking sounds and slamming doors, chair creaks and footsteps.

The Audio Aura is another example where physical actions in the environment are captured and translated into background auditory cues that are incorporated back in the workplace. Mynatt et al. (1998) describe the design of the Audio Aura as a pattern-detecting system in which people's activities in the workplace are detected and translated into speech and non-speech auditory clues. Combined with wireless headphones and online data sources, such as calendars and emails, electronic tags were placed around the workplace to detect physical actions and trigger the delivery of relevant auditory cues to a user. For instance, when a person walks up to a colleague' office, their proximity is detected and audio cues are displayed to communicate information about whether the colleague have been in, whether they have been gone for long or whether they have just left their office. Mynatt et al. (1998) describe such auditory cues as *"[delivering] a sense akin to seeing [the colleague's] light on and their brief case against the desk or hearing a passing colleague report that the person was just seen walking toward a conference room"* (p.568). The Audio Aura also included an audio representation of co-located groups, referred to as a "group pulse", which displays a continuous sound of waves that change in intensity to reflect group's dynamics, such as access to shared artefacts and presence, though some users found the meaning of sounds sometimes difficult to remember (Mynatt et al., 1998).

The sounds incorporated into the media spaces described above are a mixture of verbal, musical and auditory icons. These sounds seem to serve two main functions. The first is to emphasise what would otherwise have been environmental incidental sounds, and the second is to communicate information that would be out of visual reach. Whether incidental or intentionally designed, these sounds draw users attention to important events that they otherwise would have been unaware of. Sound in these cases communicates peripheral background information that could be attended to without being disturbing, that is important for keeping track of ongoing activities and that contribute to creating a sense of shared social space.

6.3 Audio in Co-located Collaboration

Scenarios of co-located collaborations in human activity are vast, ranging from formal meetings to game play, and from training to teaching activities. In accordance with the theme of this thesis, co-located collaborations that involve working on shared representations to complete a common task are of particular interest. Tools that assist such activities have been referred to as Single Display Groupware (SDG) (Stewart et al., 1999), which emphasise the use of a common shared display or set of equipment to support collaborative work between people that are physically close to each other. Verbal exchange in such contexts is facilitated by the properties of the shared physical environment. For instance, in their studies of collaboration in the London Underground control rooms, Heath and Luff (1992) describe how operators sharing the same physical space of a control room rely on the ability to listen to each other's talk to fine-grain interaction and coordinate actions. Research investigating technological support for co-located collaboration therefore tend to explore the auditory modality for other purposes than the capture and transmission of spoken conversation.

6.3.1 Augmenting Shared Displays

Augmenting collaborative systems with sounds to convey information about the actions that occur in the shared space is one such purpose. This has been explored by using audio as an additional information channel in collaborative interaction with groupware such as vertical wall displays and horizontal tabletops. For example, Müeller-Tomfelde and Steiner (2001) describe the DynaWall, an interactive board that combines a large touch sensitive display with gesture recognition and auditory and haptic feedback to support co-located collaboration. The interactive board enables traditional actions such as writing and sketching, in addition to displaying multimedia objects and files. Müeller-Tomfelde and Steiner (2001) use the auditory modality to augment users' gestures with sounds that communicate the type of gestures and the progress of the actions that are executed through them. For example, an earcon accompanies the creation or deletion of an object to indicate whether the system has correctly recognised the captured gesture. Spatial audio is also used to augment animated motions of objects on the board. Because sound is used to represent user actions, DynaWall designers highlight its potential to increase peripheral awareness amongst team members, though no formal evaluation is reported to confirm such a claim.

Hancock et al. (2005) address the question of awareness by examining the impact of adding auditory feedback to an interactive multi-touch multi-user tabletop display. The type of feedback provided by audio is manipulated, together with the means for delivering audio to users (localised: using a speaker for each user, or coded: one speaker for the group and assigning a different timbre to each user) to assess the impact on group awareness of activities and individual performances. Their results show that adding audio feedback to a multi-user tabletop environment increases group awareness but at the cost of decreasing individual performance. In this study, group awareness is assessed on the basis of participants' scores on a post-test recall task where participants are asked to distinguish their contributions to the product of a shared task from the group's. However, participants in their study reported being confused as to which sounds were triggered by which group member during the actual execution of the shared task. Thus, while the group might have shown greater awareness of the overall activity, it seems unlikely that such awareness translates to the moment by moment knowledge of who is contributing what onto the shared space during the actual collaboration.

Morris et al. (2004) compare the use of headphones versus shared speakers to deliver auditory feedback to a group of multi-touch tabletop users. Their system allows users to browse and manipulate movie scenes and music files to create soundtracks for movies. The audio feedback is thus the actual sounds of movie scenes and music files. Their results show that groups' task strategies change when users wear headphones, and that using headphones does not impede group communication (Morris et al., 2004). These results contrast those reported by Blaine and Perkis (2000), who compare three different sound setups for delivering audio to a group of users collaborating to create music on a shared tabletop. Blaine and Perkis (2000) report that using a shared speaker to display a mix of all collaborators audio makes it difficult for users to identify their own sounds. Using a headphones mix that places each players own sounds in the centre foreground, with a spatial mix of the remaining players surrounding it to match the relative physical position showed improvements, but only for users with musical experience, and at the cost of impeding group communication. Using individual speakers for each player is reported to help users identify sounds sources without impeding group communication (Blaine and Perkis, 2000).

The contrast between these two results can be attributed to both the type of collaborative task studied and type of headphones used in each study. The task in Blaine and Perkis (2000)'s study is purely musical, whereas that of Morris et al. (2004) involves the manipulation and organisation of shared objects. Unlike creating music, the audio output of movies and music files is not displayed continuously during collaboration, which leaves room for verbal exchange and hence group communication is not impeded. Also, participants in Morris et al. (2004) study use earbuds on a single ear, whereas those in Blaine and Perkis (2000) use headphones that cover both ears - albeit with a spatial mix. Based on a long-term study of the use of media spaces, Dourish et al. (1996) advise against the use of headphones that cover both ears in order to maintain coherence between the physical and virtual space, they write: *"headsets may remove problems from the interactional perspective, [but] the effects can be less positive when viewed from the communal or societal perspectives [..] headsets distance the wearer from the local environment, separating the media space from the physical space."*(p.49). Both Blaine and Perkis (2000) and Morris et al. (2004) acknowledge the potential impact of using speakers and headphones on public and individual awareness of actions, and consequently on the dynamics of the collaboration, but neither explore the details of such impact in their studies. This is explicitly explored in the study described in Chapter 7 of this thesis, which examines the effect of using individual speakers and "single ear" headphones on collaborative organisation and individual and group awareness.

6.3.2 Supporting Accessibility in Shared Displays

As described in Chapter 2, many researchers have explored the use audio to support non-visual interaction with visual displays such as GUIs, graphs, tables and diagrams. However, relatively fewer researchers have looked at supporting accessibility to such visual displays in collaborative contexts of interaction.

Plimmer et al. (2008) explore the use of sound to augment drawing gestures in a multi-modal tool for collaborative teaching and learning of handwriting. Their system, which builds on Crossan and Brewster (2008)'s work on audio-haptic trajectories, combines a PHANTOM Omni haptic device and audio output to capture teachers' writing gestures and guide visually-impaired students along the captured trajectories. While force feedback constrains the movements of the haptic device, Plimmer et al. (2008) use a sonification that maps vertical strokes to the pitch of a sine tone, and horizontal strokes to a right or left stereo pan depending on the direction of the gesture. They also used two additional and distinct sounds to indicate the beginning and end of a gesture. Evaluating this collaborative technique with sighted teachers and visually impaired children, Plimmer et al. (2008) demonstrate that it can significantly improve children handwriting skills as well as make the learning experience more enjoyable and motivating.

Winberg and Hellström (2001) also describe the sonification of gestures made with a mouse or a pen stylus, but in terms of manipulating spatially distributed objects to simulate the concept of Direct Manipulation in audio. This simulated concept is used to examine collaboration between pairs of visually-impaired and sighted individuals on an implementation of an audio-only and a graphical interface for the Tower of Hanoi game[1] (Winberg and Bowers, 2004), and a drag and drop task (Winberg, 2006). In these studies, visually impaired participants use headphones to access the auditory interface, while sighted participants use a computer screen to access the graphical interface. Participants in a pair share a single controlling cursor, and so do not have

[1] In the Towers of Hanoi game, a number of disks of differing sizes are placed on one of three rods in an ascending order, the object of the game is to move the stack of disks to another rod while obeying a number of constraining rules.

simultaneous access to the shared representation.

Winberg and Bowers (2004) report that together with the partner's talk, the continuous display of audio allow the visually-impaired participants to monitor their partner's actions, but their disengagement from the active manipulation of objects hinder their understanding of the state of the game. They write: *"Blind participants have to re-establish their understanding of the game and their orientation in it at the beginning of game turns. If a transition between participants has been inelegantly dealt with (e.g. if both mice are momentarily moved), it may be problematic to recover a sense of the game even leading to an extended disengagement from taking a turn in it"*(p.339). The movements of the sighted participant might be executed at a faster pace than the pair's talk, which makes them hard to follow. This, combined with the fact that the audio-only interface used continuous sonification to display the progress of the game might have contributed to visually-impaired participants' feeling of disorientation. Indeed, Winberg and Bowers (2004) state that: *"For our blind users, cursor-sharing sometimes disrupted not just the fluency of turn-taking but the delicate linkages between gesture, sound, talk and game-play needed to maintain an understanding of the state of things."*(p.339). Such disorientation could thus be potentially avoided by providing more explicit cues about system states and making these available at the request of the visually impaired participant, or by providing each participants with independent control over their view of the shared representation.

In his study of collaborative drag-and-drop, Winberg (2006) enables independent views of the shared space in the form of divided private workspaces. Winberg (2006) also uses additional non-speech sounds to communicate users' actions, such as the grabbing and dragging of objects to support awareness of partner's activity. Results from this study show improved orientation and engagement when participants execute coordination and sorting tasks. Describing such improvements, Winberg (2006) writes: *"the auditory interface made it possible for the blind subject to take part in the problem solving, both by active inquiries in the interface (exploration of the auditory space) as well as repairing breakdowns (realising when an error has been made and taking the necessary steps to correct this)*" (p.108). Winberg and Bowers (2004) and Winberg (2006) focus on addressing the issue of accessibility as a cooperative activity between visually-impaired and sighted individuals with a particular focus on the importance of the context of collaborative interaction. The question of awareness is only implicitly implied within this perspective; for example, the authors write *"monitoring the other persons actions was indirectly supported by [..]*

the manipulation in the interface and the social interaction" (Winberg, 2006, p.108). The audio display therefore contributed information that is necessary to the collaborative interaction, but no details is given as to how such information was gathered from the displayed audio and used in the collaboration. Furthermore, these studies involved a small number of participants (three pairs in the Tower of Hanoi study, and two pairs in the drag and drop study) and more thorough studies need to be conducted in order to confirm the reported findings.

McGookin and Brewster (2007b) describe explicit mechanisms for supporting monitoring partners activities in non-visual collaboration and explicitly address the potential of the auditory and haptic modalities in supporting awareness of interactions. Their collaborative system, which is based on the GraphBuilder application (McGookin and Brewster, 2007a), combines two PHANTOM omni haptic devices with speech and non-speech auditory output to support collaborative access and manipulation of bar graphs. The system employs different voices and timbres to differentiate between the audio output of each user, which is delivered through stereo speakers. McGookin and Brewster (2007b) describe two haptic mechanisms that allow users to locate each other on the shared workspace; a "come to me" mechanism, which causes a partner to be dragged towards one's location; and a "go to you" feature, which causes a user to be dragged towards their partner's location.

In their study, McGookin and Brewster (2007b) observed pairs of visually-impaired individuals as they collaborated to complete graph validation tasks. Their results show that the use of the haptic mechanisms for monitoring activities is dependent on the collaborative strategy, with the "go to you" mechanism being more popular than the "come to me" mechanism. Both mechanisms are reported to be used to locate partners on the shared space and to provide them with assistance when they are lost in the workspace. In relation to sharing audio over speakers, McGookin and Brewster (2007b) report that their participants found it useful but at times distracting, and that this was also related to the collaborative strategy employed by each pair. Pairs who chose to divide the task and separately work on different parts of the graphs found the audio to be sometimes distracting and confusing, particularly when the partner's audio output overlapped with theirs. Pairs who took turns in completing the task on the other hand found the audio, particularly the speech output, to be useful as it informed them of their partner's progress and status. The shared audio is also reported to encourage and mediate communication and to serve as a reference point during the collaborations. To reconcile the reported advantages and disadvantages of sharing audio McGookin and Brewster (2007b) suggest that *"the amount and type of shared audio should be altered dependent on the strategy that participants adopt"* (p.2577). Exactly which type of awareness information and how much of it should be altered remains an open question. Because McGookin and Brewster (2007b) study is also of a small scale (four pairs), more thorough studies are needed to confirm these initial findings.

Ramloll and Brewster (2002a) explore the use of spatial models of representation to support collaborative access to sonified line graphs. Their AudioCave system allows up to four participants to share access to an auditory workspace using their index finger, which is mounted by motion sensors, to trace a virtual soundscape on a 2D plane and receive spatialise auditory cues in a WYHIWIH (What You Hear Is What I Hear) configuration – reminiscent of the WYSIWIS (What You See Is What I See) configuration typically deployed in visually dominant multi-user interfaces (Stefik et al., 1987). Ramloll and Brewster (2002a) studied how pairs of participants used the AudioCave to solve graph reading tasks in two conditions. In one condition, collaborators wore headphones to access a shared soundscape but heard each other's speech as collocated with the position of the finger carrying the sensor. In a second condition, collaborators also wore headphones but heard each other's speech coming from a fixed point at the centre of the workspace. Their results show that participants used the spatial position of their peer's speech to maintain awareness of their interactions with the shared space. Ramloll and Brewster (2002a) do not report any specific details about which awareness information is extracted from the localised speech nor how this information is used to maintain the collaborations.

There is a common trend amongst the above reported research in that they all point to the close relation between augmenting collaborative settings with auditory information channels and the function that this plays in conveying awareness information. In general, *awareness* is highlighted as a crucial element of successful collaboration (Dourish and Bellotti, 1992) and the omnidirectional and attention grabbing characters of sound seem to contribute to its usefulness in providing awareness information in an unobtrusive and useful manner. Many issues seem to impact the efficiency of sound to communicate awareness information in ways that collaborators find useful, and answers to these issues in audio-only interaction remain unclear. These include; the impact of the audio setup; the specification of what is meant by awareness; which awareness information is actually important; and how it is extracted and used in collaboration. These issues are explicitly addressed in Chapter 7. The next section takes a closer look at the concept of awareness in collaboration in order to understand and leverage its impact in collaborative interaction.

6.4 Awareness in Collaboration

In simple terms, awareness is *"knowing what is going on"* (Endsley, 1995, p.36). The line of research on awareness originated from studies of civil, commercial and military aircrafts, in which pilots interact with highly dynamic, highly complex and information-rich environments. In such situations, information demand exceeds human ability to attend to relevant events and act appropriately, and *situation awareness* emerged as a crucial construct that affects decision making and performance under such circumstances. Situation awareness (SA) is defined as *"the up- to-the minute cognizance required to operate or maintain a system"* (Adams et al., 1995, p.85) and *"the perception of the elements in the environment within a volume of time and space, the comprehension of their meaning and the projection of their status in the near future"* (Endsley, 1995, p.36). Endsley (1995) suggested that three different stages are required in achieving SA:

1. *Perceiving elements in the environment* – this refers to the ability to selectively attend to and perceive events on the environment that are relevant to one's current task.
2. *Comprehending these elements* – then to extract meaning from those events using prior knowledge and integrating those into new knowledge in light of the current situation.
3. *Projecting their status in the near future* – then to anticipate and predict changes in incoming information from the environment.

The emphasis of SA is on an individual's ability to operate efficiently in an informationally complex workspace. In situations involving multiple actors, however, an individual's SA must extend beyond their workspace to include knowledge that has consequence for the collaboration. For a group of individuals to be able to collaborate efficiently, they need to not only be able to communicate with one another, but also to pick up clues from their shared environment to establish an understanding of who is around and how they are contributing to the shared activity, to notice what others are referring to and what changes they are making to shared resources (Hutchins, 1995). While gaining and using this knowledge in face-to-face interaction is often taken for granted, it is considered a significant challenge for both users and designers of computer-supported collaborative systems (Dourish and Bellotti, 1992; Gutwin and Greenberg, 1996). A major research area concerned with this problem has thus emerged in CSCW, and the term *awareness* was adopted to refer to collaborators' ability to construct knowledge about various aspects of their joint activity, and to use it to efficiently integrate interdependent activities. Gutwin and Greenberg (2002) write:

> *"maintaining this awareness has proven to be difficult in real-time distributed systems where information resources are poor and interaction mechanisms are foreign. As a result, working together through a groupware system often seems inefficient and clumsy compared with face-to-face work [..] It is becoming more and more apparent that being able to stay aware of others plays an important role in the fluidity and naturalness of collaboration"*(p.411).

There does not seem to be a consensus within the CSCW literature as to the exact definition, and hence the exact use of the term *awareness*, however. Schmidt (2002) notes that *"CSCW researchers are obviously far from confident with using the term [awareness] and thus often use [it] in combination with different adjectives"*(p.286). For instance, both Gaver (1991) and Bly et al. (1993) use the terms *General Awareness* and *Casual Awareness* to refer to the importance of knowing who is around and what they are doing. Bly et al. (1993) assert that such knowledge supports informal interaction between distributed groups and increases opportunities for communication and interaction, and Gaver (1991) describes its importance for managing transitions between individual work and focussed collaboration.

As described in Section 6.2.2 above, the notion of *Peripheral Awareness* and *Passive Awareness* grew out of research into media spaces to also refer to knowledge about collaborators presence and ongoing actions (Gaver, 1992a; Bly et al., 1993; Dourish et al., 1996; Dourish and Bellotti, 1992). According to Dourish and Bellotti (1992), passive awareness mechanisms in collaborative systems allow users to move smoothly between close and loose collaboration. They describe awareness as an *"understanding of the activities of others, which provides a context for your own activity"*(p.107). Peripheral awareness in such cases is thus a set of information that is passively gathered from the environment, and that envelops a user without distracting them from their ongoing activity. Schmidt (2002) note that in highly cooperative activities, participants actively seek to retrieve information about their colleagues' ongoing activities by monitoring the shared environment for clues and indicators before adapting their actions accordingly. Schmidt

(2002) supports the concept of awareness as knowledge of colleagues' activities that is used to adjust how one goes about their own work, but opposes the notion of passivity in capturing and constructing such knowledge. Results from ethnographic studies conducted by Heath and Luff (1992) within a Line Control Room on the London Underground show a number of examples where coworkers actively monitor, as well as intentionally display aspects of each other's conduct in order to efficiently coordinate collaborative tasks. Thus, according to such observations, gathering awareness information is a very active and deliberate act. Gutwin and Greenberg (1996) use the term *Workspace Awareness* to refer to the collection of up-to-the minute knowledge a person uses to capture another's interaction with the workspace. Their description explicitly bounds the kind of knowledge that a person constructs to information about colleagues' interactions inside the immediate and synchronously shared workspace. Neuwirth et al. (1998), on the other hand, emphasise the importance of supporting awareness information in collaborations that occur asynchronously throughout long-term collaboration to support collaborators establish understanding of a situation after time away from a shared task.

Although diverse, the descriptions stated above have a common denominator in that in describing the notion of awareness they all refer to a person's awareness *of something* (Schmidt, 2002). This common dominator is at the same time what makes each description unique and distinct from the others; the type of knowledge that forms the concern of an awareness and the way in which it is used in a collaboration yield great distinctions between the different descriptions and hence approaches to supporting awareness in collaborative systems. Carroll et al. (2003) use theses distinctions to categorise three main types of awareness; *Social*; *Action*; and *Activity*. These are summarised Table 6.1.

Social awareness. Social awareness refers to the individual's understanding of the social context of their environment. This usually boils down to knowledge about presence of collaborators, their activity level and availability for interaction (Carroll et al., 2003). Media spaces, for instance, were shown to provide such background information to remote collaborators through continuous video and open audio links, simulating the kind of information that could be obtained in co-located office environments. Adequate levels of social awareness can thus increase one's understanding of potential opportunities for interaction and communication. Other examples of technologies supporting social awareness include the use of avatars, icons, social proxies and other forms of visual features to indicate presence, contributions and general activity levels of

Table 6.1: Three types of awareness as suggested by Carroll et al. (2003)

Awareness concern	**Information needed to address concern**
Social: Who is present?	Presence of collaborators; features of a collaborator that convey motivational state or attitude; timing, frequency, or intensity of individual or group activity or communication
Action: What is happening?	Timing, type, or frequency of collaborators interactions with a shared resource; location and focus of collaborators current activity
Activity: How are things going?	Creation or changes to shared plans, evaluations, or rationale; assignment or modifications of project roles; task dependencies based on roles, timing, resources, etc.; exception handling

group members in remote collaboration (Benford et al., 1994; Erickson et al., 1999; Bryan-Kinns and Hamilton, 2009).

Action awareness. Action awareness goes beyond knowing just who is around and whether they are available to interact. It also comprises knowledge about which actions are carried out, by whom and on which part of the shared resources. Thus, action awareness applies specifically to instances where collaborators need to coordinate manipulations of shared resources, such as documents, images or databases, and is thus particularly important in synchronous collaboration.

Activity awareness. Activity awareness encompasses both social and action awareness and goes further beyond to include knowledge that affects long term coordination of activities. Carroll et al. (2003) describe *Activity* awareness, as *"an awareness of people's plans and understandings"* (p.614) and therefore applies to an individual's needs to develop appropriate understandings of how their contributions fit in the overall goal of the team. This includes knowledge of interdependencies among tasks, team members' responsibilities and roles with the group, and modifications to the overall goals and plans. Activity awareness therefore pertains to group activity that takes place over an extended period of time, implying extended asynchronous collaboration process.

In general, awareness information refers to users need to establish and maintain background information that may not be directly related to the task currently in focus (Carroll et al., 2003).

The awareness supported through media spaces for instance, is collected through passive observations of available monitors and speakers, and rarely is such information used on a focussed task that require coordination. In synchronous collaborations through shared workspaces, awareness information might be actively sought from the environment to support joint actions. In both cases, the importance of such information is in its potential to influence how an individual goes about conducting their actions for the benefit of the overall group performance. This is therefore an important dynamic to observe when studying the nature of collaborative interaction in an audio-only workspace. How is such information mediated in an audio-only environment? how do collaborators attend to such information, extract it from an audio-only environment? and use it in their collaborations? are questions that will thus be thoroughly explored in the study presented in Chapter 7. But first, the following sections will take a closer look at theses activities as reported in the literature on visually dominant collaborative workspaces.

6.4.1 Monitoring and Displaying Activities

Whether passively collected or actively acquired, collaborators manage to extract awareness information from their environment and make practical use of it in their social, action or activity contexts by monitoring colleagues' activities and displaying one's own activities to colleagues.

Monitoring Activities. Monitoring colleagues' activities is achieved by observing or listening for cues that occur in the shared environment. Heath and Luff (1992) report that operators in a control room carefully monitor both shared resources and colleagues' actions to assess the implications of certain activities for their own conduct. They note that more often than not, operators do this surreptitiously while they are engaged in unrelated and independent actions. For instance, to produce timely relevant information for passengers, an Informational Assistant systematically monitors the available information about the service and the actions produced by other operators present in the control room, before transforming the collected information into tailored announcements (Heath and Luff, 1992). Monitoring is thus done to establish an understanding of the state and progress of occurring events, and to adjust one's own work to complement what is being monitored. The number of collaborators in a team, in this case, can impact individuals' abilities to efficiently monitor each others work (Beaudouin-Lafon and Karsenty, 1992)

An individual is also able to pick up clues about the state and progress of colleagues' activities by scanning the shared environment to detect changes to shared resources. That is, even if unable to see or monitor aspects of a colleague's actions as these are being executed, it is possible for a competent collaborator to infer the state and progress, as well as the plans and intentions of a colleague by monitoring the state of the shared workspaces itself (Schmidt and Simone, 2000). The following characteristics are thus important aspect of the process of collecting awareness information through monitoring:

- The act of monitoring does not necessitate a response from those colleagues being monitored.
- The act of monitoring is done in a deliberately unobtrusive manner so as to not interrupt the focus of both one's own current actions and those of colleagues.
- The level of obtrusiveness in monitoring colleagues is made appropriate to the situation or task at hand.

Displaying Activities. The ability to monitor the state and progress of colleagues' activities is in turn dependent on how available and easily accessible are clues about them in the shared environment. Knowing that one's actions are likely to be monitored by one's colleagues, a competent collaborator might explicitly conduct their work so as to provide appropriate clues about the state and progress of their activities (Schmidt and Simone, 2000). Heath and Luff (1992) describe how operators in control rooms intentionally make their activities "publicly visible" to display clues about aspects of their work which they think might be relevant to their colleagues. Knowledge of how certain aspects of one's work affects others is thus important in shaping the way in which these aspects are displayed to ensure that they are appropriately perceived and, hence, used in the collaboration. Schmidt (2002) writes:*"actors may display their work in ways that are designed to attract the attention of colleagues to the activity or certain features of it, by gazing at certain objects, humming, thinking aloud, placing artefacts in certain locations or orientations, leaving traces in the setting, etc"*(p.292). In this sense, the act of monitoring can in itself become a form of displaying one's activities to colleagues; an individual might be very explicit in the way they monitor a colleague's action during a critical point of a coordination. The following characteristics are thus important aspect of the process of facilitating the collection of awareness information through displaying one's own activities:

- The act of displaying one's actions is itself driven by an awareness of who is monitoring one's actions.

- The act of displaying one's actions is done in a deliberately explicit manner to draw colleagues attention.
- The level of explicitness in displaying one's activities is made appropriate to the situation or task at hand.

Monitoring and displaying activities are therefore interrelated acts, that dynamically change in terms of explicitness of display and obtrusiveness of monitoring to appropriately match the needs of a given situation or stage of a collaboration.

6.4.2 Workspace Awareness

Workspace awareness (WA) is the *"the up-to-the-moment understanding of another person's interaction with a shared workspace"* (Gutwin and Greenberg, 1996, p.208). WA is distinct from other types of awareness previously described in its focus on people's interactions with a synchronously shared workspace, rather than on the workspace itself. WA is thus a special case of *situation awareness* because it extends knowledge about a given situation and a domain task with knowledge about collaboration with others, and overlaps with the notion of *action awareness* because it focuses on real-time synchronous collaboration.

A number of frameworks have been suggested to understand and design support for awareness. From an implementation perspective, the Big Watch (Kirsch-Pinheiro et al., 2003) and the Atmosphere (Rittenbruch, 2002) frameworks support developers when incorporating mechanisms for providing awareness information, and to address the problem of workload when providing contextual awareness in asynchronous collaboration. Gutwin and Greenberg (2002) devised a three-part descriptive framework which sets out elements of knowledge that make up WA, outlines the perceptual mechanisms used to maintain it, and describes ways that people use such knowledge in collaboration. According to Gutwin and Greenberg (2002), the framework is applicable to collaborations that have the following characteristics:

- Collaborators interact with each other in real-time but from different locations and within a shared workspace.
- Collaborators can see and manipulate shared artefacts that are related to the shared activity.
- The focus of the collaboration is on visible and manipulable shared artefacts.

- Collaborators engage in the creation, exploration and physical manipulation of new and existing artefacts within the share workspace.
- Small groups of collaborators (two to five) are involved in carrying out such tasks, where they shift between individual and shared activities during a work session.

These characteristics overlap with those of the collaborative setting addressed in this thesis, where pairs of individuals engage in the creation and manipulation of diagrams using an an audio-only model that acts as a shared workspace. Because of this overlap, the following describes aspects of Gutwin and Greenberg (2002)'s WA framework in further details, as these are used to inform the analysis of the audio-only collaborative study presented in Chapter 7.

6.4.2.1 WA Knowledge Elements

According to Gutwin and Greenberg (2002), when designing support for WA in the type of collaborative settings listed above, designers of groupware systems should take into account *what* information should be captured about collaborators' interactions within the workspace, and *how* to present such information to other participants. The first part of Gutwin and Greenberg (2002)'s framework provides guidance for deciding what information should be captured. It sets out elements of knowledge that make up WA around answers to the questions of 1) who is present, 2) what they are doing, 3) where they are working, 4) when various events occur in the workspace, and 5) how do these events take place. These elements are summarised in Tables 6.2 and 6.3. WA knowledge is concerned with both actions that are currently taking place and those that have been performed. In relation to current activities, answers to the "Who" category of questions constitute knowledge about whether there are others in the workspace, who they are, and which actor is performing which action. Answers to the "Where" questions constitute knowledge about where the person is working, where they are looking, what they can see, and the area of the workspace that they can act on. Finally, answers to the "What" questions constitute knowledge about what another person is doing and what objects they are working on. In relation to past activities, workspace awareness includes knowledge about how changes to shared objects came about and who was behind such changes.

Gutwin and Greenberg (2002) do not include elements about future events and actions in their framework, arguing that designers of collaborative systems are unlikely to support maintenance of such information since future events and actions cannot be determined from raw perceptual information. It could be argued, however, that raw perceptual data could provide a basis for determining future events. Schmidt and Simone (2000), for example, describe how the ability to perceive the state of a shared work field and changes to it enables competent collaborators to infer the plans and intentions of colleagues. This is particularly the case if colleagues' actions are part of a known work process; perceiving cues of performing the initial steps of such a process can thus be used to infer the performing of the remaining steps.

Table 6.2: Elements of workspace awareness related to the present

Category	**Element**	Specific Questions
Who	Presence	Is anyone in the workspace?
	Identity	Who is participating? Who is that?
	Authorship	Who is doing that?
What	Action	What are they doing?
	Intention	What goal is that action part of?
	Artefact	What object are they working on?
Where	Location	Where are they working?
	Gaze	Where are they looking?
	View	Where can they see?
	Reach	Where can they reach?

Table 6.3: Elements of workspace awareness related to the past

Category	**Element**	Specific Questions
How	Action history	How did that operation happen?
	Artefact history	How did this artefact come to be in this state?
When	Event history	When did that event happen?
Who (past)	Presence history	Who was here, and when?
Where (past)	Location history	Where has a person been?
What (past)	Action history	What has a person been doing?

The past and present elements therefore constitute basic knowledge pertaining to workspace awareness that should be gathered and presented to users of a collaborative system. Gutwin and Greenberg (2002) highlight that designers of groupware system should not simply support all such elements equally, however, but instead treat each element with relation to the context of the collaboration and the type of collaborative task. They emphasise two critical factors for handling each element: *"First, the degree of interaction between the participants in the activity indicates how specific or general the information in the interface should be. Second, the dynamism of the element – how often the information changes – indicates how often the interface will need to be updated."* (p.421). For example, if there are only two individuals collaborating in the workspace, then there is little or no need to support the collection and distribution of information about presence or identity. This is similar to the settings used in Study 3 presented in Chapter 7 and will therefore be taken into consideration when analysing the data obtained from that study.

6.4.2.2 Extracting WA Knowledge from the Environment

While Tables 6.2 and 6.3 list the information needed for maintaining WA, the second part of the framework addresses the question of how such information is retrieved from the workspace. The second part is thus essentially concerned with supporting people in the processes through which they manage to establish answers to the questions listed above. Gutwin and Greenberg (2002) suggest three mechanisms for gathering WA (also shown on Figure 6.1):

1. Intentional Communication,
2. Consequential Communication, and
3. Feedthroughs

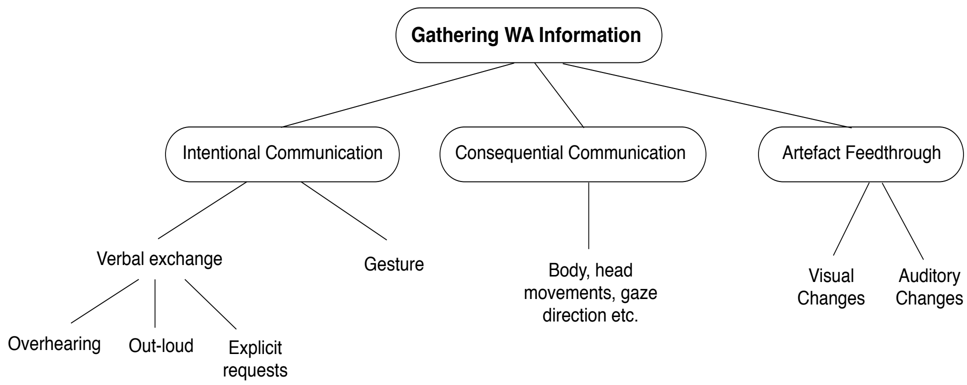


Figure 6.1: Mechanisms for gathering Workspace Awareness information from a shared environment

Intentional Communication. This refers to the use of verbal and gestural exchange in communication. Naturally, awareness information can be gathered from verbal descriptions of awareness elements, particularly when colleagues *explicitly request* such information from each other – for example asking a colleague where they are or what they are doing. Awareness information can also be gathered from *overhearing* others' conversations in the workspace or when colleagues verbally describe what they are doing as they are doing it. The latter has been referred to as verbal shadowing (Gutwin and Greenberg, 2002) or *outlouds* (Heath and Luff, 1992), which is *"the running commentary that people commonly produce alongside their actions"* (Gutwin and Greenberg, 2002, p.424). Gestures can also communicate workspace awareness information, particular when they are used for *illustration* or in the form *emblems* (Gutwin and Greenberg, 2002).

Consequential Communication. This refers to those actions that are carried out as part of performing a given task without being specifically intended for communication. If actions are publicly visible, they provide clues about one's activity and can therefore be a source of workspace awareness. Body movements, posture and position, in addition to head and gaze directions are all a consequence of a person's activity, and thus send signals about such activity. Consequential communication is thus related to the notion of *passive awareness* (Dourish and Bellotti, 1992; Bly et al., 1993) where an individual picks up clues from the environment that were not intended to address them specifically, and use these clues to build an awareness of what is going on around them.

Feedthrough. This is information communicated through artefacts. The state of an artefact usually has indications about the sort of changes that occurred to it. For example, a physical object can be moved around, annotated or transformed, and detecting such changes can give clues as to which actions were performed to incur such changes. This includes both "live" actions – those that can be perceived as they are performed – as well as past actions, and involve visual as well as auditory clues; for example the sound of a pen on a board indicates that a colleague is actively writing on it.

6.4.2.3 Benefits of WA in Collaboration

The third part of Gutwin and Greenberg (2002)'s framework describes the collaborative activities that can benefit from WA information. These are: managing coupling; simplification of communication; coordination of action; anticipation; and assistance (see Table 6.4).

Managing Coupling. Collaboration involves transitions between focused individual activities and joint work (Gaver, 1991; Dourish and Bellotti, 1992). Gutwin and Greenberg (2002) used the term *coupling* to refer to the amount of work an individual can do before requiring interaction with another person. Movement between tight and loose coupling can occur as a result of certain task requirements, or due to the occurrence of opportunities for collaboration during a work session. Because WA can help a user keep track of their colleagues' activities during loosely coupled activities, they can make accurate judgements to determine when opportunities for tighter coupling are available.

Table 6.4: Summary of the activities in which workspace awareness is used.

Activty	**Benefit of WA**
Management of coupling	Assist people in noticing and managing transitions between individual and shared work.
Simplification of communication	Allows people to use the workspace and artefacts as conversational props, including mechanisms of deixis, demonstrations, and visual evidence.
Coordination of action	Assists people in planning and executing low-level workspace actions to mesh seamlessly with others.
Anticipation	Allows people to predict others actions and activity at several time scales.
Assistance	Assists people in understanding the context where help is to be provided.

Simplification of Communication. Knowing where a person is working and what parts of the workspace they are manipulating allows collaborators to use the shared workspace as an external representation (Gutwin and Greenberg, 2002). This means that parts of the shared workspace become a part of the language used for communication. Verbal communications is thus simplified by using the shared workspace as a context for deictic reference (pointing to and highlighting objects on the workspace); as a means for demonstrating meaning (e.g. tracing a path on the workspace); to manifest actions (replacing an entire verbal utterance with a manipulation inside the workspace); and to provide back-channel feedback for confirmations of understandings (e.g. visual manipulations as evidence for understanding instructions).

Coordinating Actions. Coordinating actions in collaboration refers to the ability to complete joint actions in the right order, at the right time and within the constraint of the task at hand (Gutwin and Greenberg, 2002). When coordination is not achieved through explicit communication, WA conveys the necessary information for maintaining temporal and spatial boundaries of joint actions. That is, the knowledge of what others have done, what they will do next and when, enables collaborators to seamlessly integrate their actions, and is particularly important in continuous action where people are sharing the same objects.

Anticipation. Because WA provides up-to-the-moment knowledge of others' interactions with the workspace, it is possible for an individual to predict what their colleagues will do next based on where they are currently acting and what their current action is. In turn, the ability to properly anticipate the action of others allows one to prepare for one's own actions, and thus avoid conflicts, or provide one's colleagues with relevant resources when they are needed. WA knowledge also provides information that assist one in determining whether the actions performed by others match existing expected patterns of actions.

Assistance. WA allows a person to detect whether their colleagues require assistance on certain tasks. Knowing what one's colleagues are doing, what they intend to do and how their work is progressing allows one to provide assistance even when not explicitly requested to do so.

According to Gutwin and Greenberg (2002), the three parts framework can be used both to determine where awareness support is most appropriate and to analyse the degree of support needed in a given collaborative context. Following this recommendation, the framework is used as a tool for analysing workspace awareness in audio-only collaboration, which forms the focus of the remaining chapters in Part II of this thesis. What is noticeable from the descriptions of these three parts is the strong emphasis on visual features of the workspace and of the collaboration in contributing to WA. The mechanisms for gathering awareness information in the workspace described the use of gestures for intentional communication, of body movements, posture and position as well as eye gaze and head directions in consequential communication, in addition to visual features of physical artefacts as spatial organisation and manipulation in the form of feedthroughs. Furthermore, the potential benefits that WA brings to the collaboration are described in terms of collaborators ability to extract visual clues for the workspace (e.g. tracing a path, or visual manipulations as evidence for understanding instruction). Thus, visual information forms an integral part of Gutwin and Greenberg (2002)'s framework.

With the exception of explicit verbal exchanges and some auditory feedthrough (e.g. the sound of writing on the board) there were relatively fewer references to the potential of auditory information in communicating or contributing to maintaining workspace awareness information. Applying this framework for analysing audio-only collaborative interaction should therefore provide a means for understanding such potential.

6.5 Summary and Conclusion

This chapter aimed to provide background about the role of sound in computer-supported collaboration and to highlight the link between the auditory modality and the concept of awareness in collaboration. Sections 6.2 and 6.3 reviewed the use of audio in remote and co-located collaborative interaction. Sound has a primary role in communication, mediating not only verbal exchange, which is essential in collaboration, but also a variety of incidental sounds that contribute to enriching collaborators awareness of the context of their interactions. Audio in remote collaboration has been exploited to deliver both functions, and has been shown to provide a usable and sociable space for interaction even in the absence of other modalities (e.g. Hindus et al., 1996). Researchers have learnt from the way in which people use incidental sounds, and used this knowledge to design and integrate auditory cues back in shared spaces. Such engineered sounds were found to support awareness and enrich the experience of remote collaboration (e.g. Gaver, 1991; Cohen, 1993, 1994; Mynatt et al., 1998).

In co-located collaboration, audio has been used to augment visually-dominant collaborative displays (such as interactive boards, wall displays and tabletops) with auditory feedback that communicate information about users' actions and represent aspects of the shared workspace itself. Relatively fewer studies have examined the use of audio to support accessibility in collaboration (e.g. Winberg and Bowers, 2004; McGookin and Brewster, 2007b; Tanhua-Piiroinen, Pasto, Raisamo, and Sallnäs, 2008). As a result of conveying information about actions, the use of audio was shown to increase individual and group awareness of the activities that take place during a collaboration (e.g. Müeller-Tomfelde and Steiner, 2001), but the means for delivering audio in such multimodal contexts – through headphones, individual or shared speakers – can impact levels of awareness as well as strategies for completing shared tasks (Morris et al., 2004; McGookin and Brewster, 2007b), though it remains unclear whether this is also the case in audio-only contexts of collaboration.

Because of the evident link between sound and awareness in group work, a closer examination of the concept of awareness was presented in Section 6.4. Generally speaking, gaining an awareness *of something* in a collaboration involves acts of monitoring both the shared space and co-present individuals, as well as displaying one's own activities to others. Competent collaborators typically adjust the levels of obtrusiveness in their monitoring of co-workers and explicitness in displaying their own actions to match the demands of the current task or state of the collaboration (Heath and Luff, 1992). *Workspace awareness* is one particular type of awareness that refers to the ability to keep track of collaborators' interactions within a synchronously shared workspace. Details of Gutwin and Greenberg's framework for workspace awareness were presented because of similarities between the applicability of such framework and the work explored in this thesis. The framework describes what knowledge constitutes workspace awareness, which perceptual mechanisms are used to extra such knowledge from a shared workspace and how it benefits collaboration. One important characteristic of workspace awareness is its focus on collaborators' interaction within a defined space, in realtime, and with shared resources, and thus involves knowledge of which parts of the workspaces others are working (*location*), what they are currently doing (*actions*) and what they will do next (*intentions*) (Gutwin and Greenberg, 1996). The potential of audio to deliver and maintain workspace awareness in an audio-only context of collaboration remains unclear, and parts of this framework are used in the following chapters as a basis for capturing and exploring this potential.

Chapter 7

Audio-only Collaboration With and Through Diagrams – Study 3

7.1 Introduction

The previous chapter reviewed the use of audio in computer-supported collaboration and highlighted a common trend amongst existing research which points to the important role that audio plays in supporting awareness in collaboration (e.g. Hindus et al., 1996; Gaver, 1991). The means for delivering audio to shared workspaces is also reported to impact such support (e.g. Morris et al., 2004; McGookin and Brewster, 2007b). This evidence is based on studies of collaboration where sound is used as an additional modality in otherwise multimodal shared spaces. It is therefore not clear if such support and impact extend to situations where collaboration takes place in an audio-only workspace[1]. Earlier, Chapter 2 reviewed existing approaches to supporting non-visual interaction with visual displays and highlighted the lack of research exploring the use of hierarchical models in non-visual collaboration.

This chapter presents Study 3, which examines audio-only collaboration. The study focuses on a subset of collaborative situations in which pairs of collaborators share access to the same workspace in realtime to produce and manipulate diagrams, all of which occur in an audio-only setting. The study goes further than previous research to establish:

[1] An audio-only workspace in the context of this thesis refers to situations where collaborators use verbal exchange and the shared workspace is accessed through audio alone – as apposed to other modalities, such as graphics or haptics. Specifically, audio-only collaboration in the in this context excludes the use of verbal commands for inputing instructions to any shared artefacts in the workspace.

- How an audio-only tool that is based on a hierarchical model of representation is used as a shared workspace to support collaboration.
- How the means for delivering audio to such a workspace impacts collaborative processes and organisation and collaborators' awareness of their interactions.

Section 7.2 specifies the hypotheses of the study. Section 7.3 describes the collaborative tool used as a shared audio-only workspace, and Sections 7.4 and 7.5 present the details of the methodology and results. The chapter concludes with a summary of the results in Section 7.6.

7.2 Hypotheses

The overall research question of this thesis, stated in Section 1.2, asked whether audio can be a practical medium to support individual and collaborative interaction with diagrams. Study 3 contributes to answering this question by examining collaborative construction of diagrams using the hierarchy-based audio-only model developed and evaluated in Part I as a shared audio-only workspace. In particular, the study examines how the means for delivering audio to such a workspace impacts 1) the collaborative process in terms of its structure and organisation, and 2) collaborators' exchange and use of workspace awareness information in the collaboration; i.e. addressing Research Question 3 of this thesis, which translates to the following hypotheses:

H5 The means for delivering audio in an audio-only workspace impacts the collaborative process in terms of its structure and organisation.

H6 Displaying the audio output of each participant's interactions to both participants in a pair increases their awareness of self and partner's contributions to shared tasks.

H7 Concealing the audio output of each participant's interactions from their partner increases the exchange of workspace awareness information between them.

To evaluate these hypotheses audio output in an audio-only collaborative workspace is manipulated in a within-subjects experimental design factor of means of audio delivery. In a *Shared* condition, pairs of participants collaboratively construct diagrams using the hierarchy-based audio-only model while the audio output of each participant's interactions is delivered through speakers, rendering the audio present in the collaborative workspace. In a *Non-Shared* condition, the audio output of each participant's interactions is delivered through headphones, rendering the audio absent from the collaborative workspace.

7.3 Overview of the Audio-only Collaborative Tool

The hierarchy-based audio-only tool that was developed to support single user interaction with ER diagrams was adapted to allow multiple users to simultaneously access and manipulate shared diagrams. As described in Chapter 5, the Non-Guided interaction strategy was preferred by most participants in Study 2 and was found to support faster interaction times, and was therefore chosen to be adapted for multi-user interaction.

The adaptation did not include any built-in mechanisms to regulate collaboration through the audio-only tool, 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 too much 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 adapted collaborative tool provides both users in a pair with an equivalent and independent "view" and unstructured interaction with a shared diagram. Collaborators could each control how they access and manipulate a diagram, therefore triggering differing audio output, with one exception: a "protection" mechanism that prevents users from simultaneously editing the same item on a given diagram. The *Error* sound in such instances is displayed to the user attempting to edit the protected item.

In the single user version, objects that are to be related must be selected from the hierarchy before a relation can be created between them. Two features for relating objects on a relational diagram were added in the adaptation process. First, the multi-user version provides two possible types of selection: a *public* and a *private* selection. A "publicly" selected object can be used by any user in the workspace, while a "privately" selected object can only be used by the user who issued the selection. Second, the two types of selection are accompanied by two alternative commands for creating relations. Issuing a command to create a private relation would relate all privately selected objects, while issuing a command to create a public relation would relate all publicly selected objects. Once created, the produced relation can be accessed by all users. A high pitched voice was used in the single user version to highlight a selected object when it is encountered on the hierarchy. This was extended in the multi-user version by using a high pitched voice to highlight publicly selected objects, and a low pitched voice to highlight privately selected objects. The low pitched voice is only displayed to the user who issues the private selection. Besides these audio display techniques, no modifications were implemented to distinguish the audio output of each user in the workspace; i.e. the same voices and non-speech sounds were used for both users in a pair. The remaining features of inspecting and constructing diagrams were as described in the previous studies.

7.4 Method

7.4.1 Participants

Thirty-two sighted individuals, different from those who took part in Studies 1 and 2, were recruited for this study. They had replied to an email that was sent to departmental mailing lists of various University of London institutions, and which asked those interested to bring a partner a long to evaluate a novel system for constructing diagrams using audio. Twelve participants were undergraduates, twelve were studying for a Masters, seven at various PhD levels and one participant at a post-doctoral level. Twenty-four were male and eight were female. This made up a total of sixteen pairs who knew each other. All participants were from the computer science or electronic engineering disciplines and had varying knowledge of ER modelling ranging from very low to very high expertise. Each pair was asked to choose a convenient time to participate in a two hour session that took place at the Usability Lab in the department of Computer Science within the School of Electronic Engineering and Computer Science at Queen Mary University of London. Each participant received a £15 cash incentive for their participation.

7.4.2 Setup and Procedure

A pilot study was conducted with two pairs of sighted participants to fine tune the experimental setup, procedure of the training and testing, and to ensure that the tasks and the means for capturing data were adequate. Any given session consisted of four parts; an introduction, a training, a test and a post-test part.

Introductions. Upon arrival to the lab, pairs were asked to read through information sheets that described the purpose of the study and how they are to be involved in it (See Appendix C.1). A short oral summary of the information sheet was also given to the participants to make sure they understood what to expect from the study. Participants were then asked to sign consent forms for anonymous subsequent use of study materials, and to fill a pre-test questionnaire (see Appendix C.2), which gathered basic information about each participant's familiarity with ER diagram modelling, screen-reader technology and musical training. The introduction part typically lasted for five to ten minutes.

Figure 7.1: Experimental setup in the Shared condition

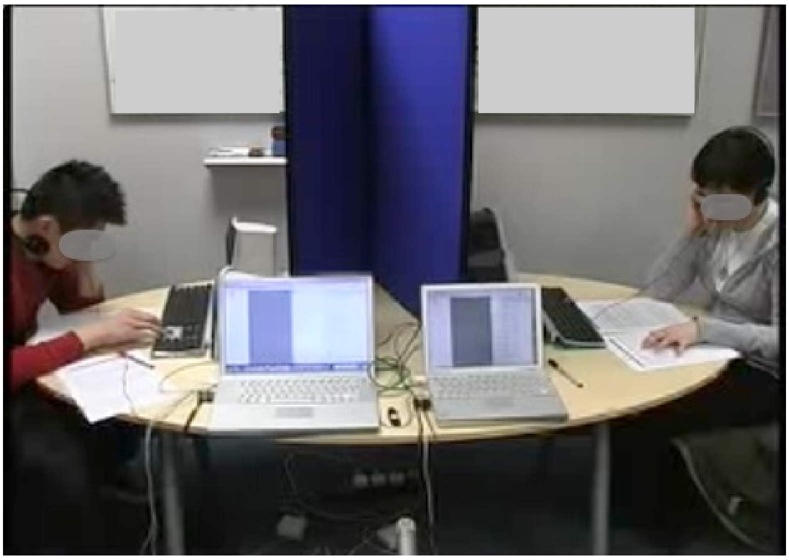


Figure 7.2: Experimental setup in the Non-Shared condition

Training. A training session then followed where all participants were introduced to the basics of ER modelling and to the hierarchy-based audio-only collaborative tool. Features of the tool and its auditory display were thoroughly demonstrated until the participants felt comfortable using it. They were then presented with a sample visual diagram and a textual description – similar to those in the testing part (see Appendix C.3) – and were given time to use the tool to construct various parts of an ER diagram while being closely assisted by the experimenter. They were able to refer to the visual diagram as they constructed it but this was not allowed during the testing part of the study. The training part lasted for up to thirty minutes.

Testing. Once familiar with the tasks and how to perform them using the collaborative tool, the next part was the actual testing. Figures 7.1 and 7.2 show the experimental setups during this part of the study (participants faces are intentionally concealed on the figures). Pairs sat facing each other and had each a keyboard to manipulate the diagram and a set of computer speakers or headphones to use in the Shared and Non-shared conditions respectively. The keyboard, speakers and headphones were connected to two computers, one for each participant with one of those acting as a server and linking the pair to a shared diagram. An opaque board was placed between the two participants to eliminate any form of visual communication (body language, facial expressions, etc.). Participants could hear each other's audio output (in the Shared condition) as well as converse comfortably. The pair then collaboratively constructed two diagrams, one at a time, under each of the two experimental conditions; the order of the conditions was randomised across the sixteen pairs to cancel out learning effects. While they constructed the diagrams, pairs were allowed to ask the experimenter to clarify the audio output if, for example, they found it hard to understand the synthesised speech output, or to be reminded of any of the system commands if they forgotten what they were. The testing part lasted for up to an hour.

Post-tests. After constructing each diagram, participants were asked to complete an annotation task where they highlighted visual versions of the diagrams that they constructed to indicate their contributions and their partner's.

7.4.3 Collaborative Task

At the start of each test, an initial diagram, such as the one in Figure 7.3, was loaded onto the tool and participants were given a textual description containing information about how the diagram could be completed. The descriptions contained complementary information (see Appendix C.3); for instance, one participant might have had information about the name of an entity, while their partner had information about its attributes and primary keys. This was done to ensure participants talk to each other during the construction process. Participants were instructed to consult the textual descriptions and to complete the diagram as they see fit. They were informed that they had complementary information on each description and therefore needed to consult with one another. They were given no time limit to complete the diagram, and were free to decide which information to include from the description and which to omit or delete from the provided initial diagram; it was therefore entirely up to the participants to decide when a diagram was complete. Figure 7.4 shows an example of the typical complexity that the finished diagrams reached.

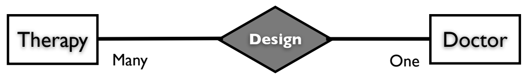


Figure 7.3: Initial ER diagram

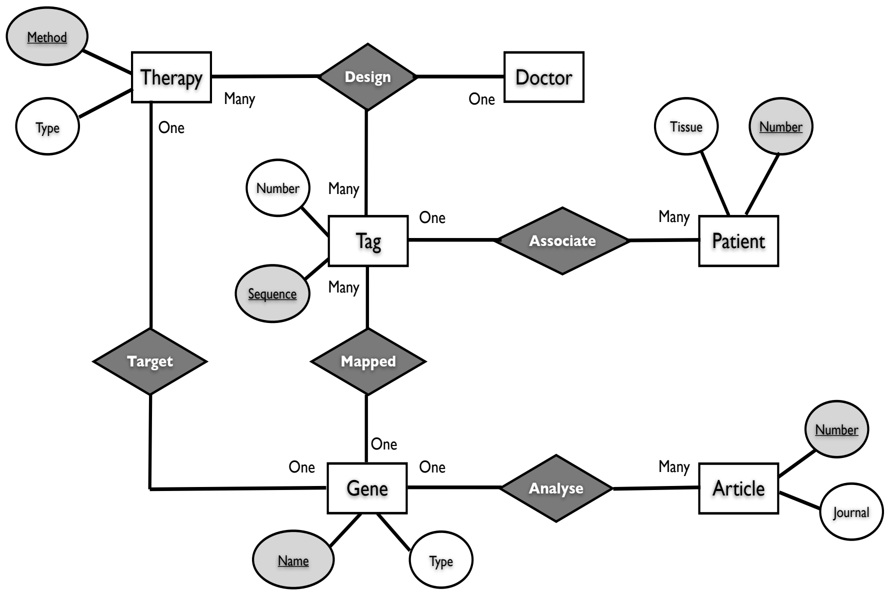


Figure 7.4: Example of ER diagram complexity

7.4.4 Data Gathering and Measures

The testing part of the sessions were video recorded; all interactions with the tool were automatically logged and timestamped; and informal interviews were conducted with the participants at the end of each session to gather personal reflections on the collaborative experience.

The aim of the study was to examine how sharing access to an audio-only workspace under different conditions affected participants' 1) collaborative process in terms of structure and organisation, and 2) workspace awareness information exchange, including awareness of their contributions to the shared tasks. A coding scheme was developed to capture these aspects of the collaborations from the video data (describe below). To test hypothesis H5, the coding scheme was used to categorise the activities that took place during the sessions in order to assist the examining of how the collaborative process was structured and organised in each condition. To tests hypothesis H6, scores on the annotation tasks were used to measure participants' awareness of each other's contributions to the constructed diagrams. To tests hypothesis H7, the framework for workspace awareness described in Section 6.4.2, was used to identify and measure the occurrences of requested and supplied workspace awareness information in the coded video transcripts.

7.4.4.1 Video Data Transcription and Coding

All videos were transcribed by the author and a coding scheme was developed to capture two main aspects of the collaborative interaction; 1) the activities that the pairs engaged in as they constructed the diagrams, and 2) the exchange of workspace awareness information between participants.

Activity Categories. The choices of which categories to use to describe and catalogue participants' activities emerged from the coding process itself while drawing on previous analysis of group design meetings (Olson et al., 1992) and shared document editing (Bryan-Kinns et al., 2007). Essentially, the general topic of discussion that a pair engaged in at any given moment was used to tag chunks of conversations. Participants discussed four main topics when constructing diagrams which were tagged with the labels *"Strategy"*, *"Content"*, *"Labour"*, and *"Execution"*. A different approach was used to account for activities which did not involve discussion and which still formed part of the collaboration. For such instances, the type of activities that the participants were performing was used to tag the extra categories, these were labeled as *"Other–Interacting"*, *"Other–Reading"* and *"Other–Interacting/Reading/Etc."*. Table 7.1 summarises the descriptions of the chosen categories.

The *Strategy* category refers to instances where pairs discussed how they should go about constructing a diagram or part of it, as well as how they should organise episodes of construction, such as moving between different parts of the textual descriptions. An example of this is deciding whether all entities of a diagram should be created first before moving on to its relations, or whether individual entities and relations should be finalised interchangeably. Discussing *Content* typically involved deciding which values to assign to different elements that are to be included on the diagram, such as deciding how many attributes an entity should have, what should their

Table 7.1: Activity categories used in the developed coding scheme to capture the structure and organisation of the audio-only collaborations from the video transcripts – WI is short for "While Interacting", NI is short for "Not Interacting".

Activity	**Description**
Strategy-WI	Discussing how to go about constructing a part of the diagram while one or both participants are manipulating the diagram.
Content-WI Related	Discussing which values to use when constructing a part of the diagram while one or both participants are manipulating the diagram and at least one participant's interaction matches the discussed content.
Content-WI Unrelated	Discussing which values to use when constructing a part of the diagram while one or both participants are manipulating the diagram and both participants' interactions do not match the discussed content.
Labour-WI	Discussing the division of labour (who should do what) while one or both participants are manipulating the diagram.
Execution-WI	Discussing how to use the collaborative tool (which commands to use and in which order) to construct a part of the diagram while one or both participants are manipulating the diagram.
Strategy-NI	Discussing how to go about constructing a part of the diagram while no participant is manipulating the diagram.
Content-NI	Discussing which values to use when constructing a part of the diagram while no participant is manipulating the diagram.
Labour-NI	Discussing division of labour while no participant is manipulating the diagram.
Execution-NI	Discussion how to use the collaborative tool to construct a part of the diagram while no participant is manipulating the diagram.
Other–I	Both participants are manipulating the diagram without conversing with one another.
Other–R	Both participants are reading through the textual descriptions without conversing with one another.
Other I/R/Etc.	One participant is manipulating the diagram while their partner is reading the description or waiting for their partner without conversing with one another.

labels be, and so on. *Labour* refers to discussions where a pair decided who should do what on the diagram; i.e. division of labour. Finally, *Execution* refers to discussions about issues related to the technicalities of using the collaborative tool, such as discussing specific commands for executing a given action or the order in which certain commands should be issued to the tool.

The *Other–Interacting* (Other-I) category refers to instances in the collaborations where both participants were engaged in manipulating the diagram without conversing with one another. The *Other–Reading* (Other-R) category refers to instances where both participants were neither using the tool nor talking to each other (e.g. both reading through the textual descriptions). The *Other–Interacting/Reading/etc* (Other-I/R/etc) refers to instances where one participant was manipulating the diagram while their partner was idle or engaged in another activity (e.g. reading through the description, or simply waiting for their partner while they finish their manipulations).

Discussions about *Strategy*, *Content*, *Labour* and *Execution* occurred when either one or both participants were manipulating the diagram, or when none of them was. The labels *"While Interacting"* (WI), and *"Not Interacting"* (NI) were thus associated with each of the these activity categories to distinguish between these two cases. Additionally, two subcategories were used to capture a further characteristic of the Content-WI activity; a *"Related"* tag was used when the content being discussed matched at least one participant's manipulations of the diagram, and a *"Unrelated"* tag was used when it did not. There were thus eight categories based on discussion topics, and three categories based on other activities not involving discussion. Taking into consideration the two further modes of interaction (WI and NI), this made up a total of twelve categories that were used as a coding scheme for capturing the structure and organisation of the audio-only collaborative process from the video transcripts.

Workspace Awareness. According Gutwin and Greenberg (2002)'s framework, WA knowledge is based around eleven elements[2]:

1. *Presence* – Who is participating in the activity?
2. *Location* – Where are they working?
3. *Activity Level* – How active are they in the workspace?
4. *Actions* – What are they doing? What are their current activities and tasks?
5. *Intentions* – What will they do next? Where will they be?
6. *Changes* – What changes are they making, and where?

[2] Refer back to Section 6.4.2 on page 134 for a more thorough description of the elements that constitute workspace awareness knowledge.

7. *Objects* – What objects are they using?
8. *Extents* – What can they see? How far can they reach?
9. *Abilities* – What can they do?
10. *Sphere of Influence* – Where can they make changes?
11. *Expectations* – What do they need me to do next?

This thesis uses a refined version of Gutwin and Greenberg (2002)'s framework to identify occurrences of its elements in the video transcripts. To adapt the above elements, the developed coding scheme focussed on *Actions*; *Intentions*; and *Locations* and refined them as follows: knowledge about the level of activity (element 3), current activity (element 4), changes made to the workspace (part of element 6) and the objects of change (element 7) were combined to reflect knowledge about *actions* in the workspace and extended to account for past and current actions. Part of the knowledge about *intentions* (element 5) was preserved to reflect knowledge about future actions. Knowledge about future locations (part of element 5), locations of changes (part of element 6), and element 2 were combined to reflect knowledge about *location* in the workspace. Knowledge about presence (element 1), extent (element 8), abilities (elements 9) and sphere of influence (element 10) were discarded from the coding scheme since these are enforced by the experimental settings. For example, because collaborating groups in Study 3 were made up of two participants, it was obvious who was present in the activity. Also, both participants were made aware that they had equivalent access to the diagrams, and thus knowledge about extent, abilities and sphere of influence were also given. knowledge about expectations (element 11) was also left out as such knowledge could be explicitly negotiated amongst participants or inferred from the other elements.

Thus, in relation to *Actions* and *Intentions*, the coding scheme captured indicators of past, current and future activities as well as indicators of action completion statuses. That is, all explicit references made by participants where they supplied or requested information from their partner about what actions they had undertaken, what actions they are currently performing, what actions they plan to perform, and when they completed an action. In relation to *Location*, the coding scheme captured supplied and requested information about position on the diagram or within the hierarchy. WA information exchange was thus divided into *Supplied* and *Requested* types. The supplied type refers to information provided by a participant to their partner without the latter having asked for it, and the requested type refers to instances where a participant explicitly asks

Table 7.2: The refined version of Gutwin and Greenberg (2002)'s workspace awareness elements and indicators used in the developed coding scheme to capture participants exchange of WA information from the video transcripts.

Type	Element	Description
Supplied	What I Did	A participant describes the editing actions that they have just completed without explicit request from their partner; this information is supplied after the action is completed. *E.g."ok, I made three entities ok gene tag and article"*.
	What I'm Doing	A participant describes the actions that they are currently performing without explicit request from their partner; this information is supplied after the action is initiated and while it is is being performed. *E.g."I'm making that the primary key."*
	What I will Do	A participant describes what action they are about to perform without explicit request from their partner; this information is supplied before the action is initiated. *E.g."ok I'm gonna go and select Bus"*.
	Where I Am	A participant describes their current position on the diagram or on the hierarchy without explicit request from their partner. *E.g."So I'm inside the Therapy"*.
	Completion Status	A participant inform their partner that an action has been completed without explicit request from their partner and without explicit reference to which action has been completed. *E.g."OK, I'm done."*
Requested	What Did You Do?	A participant requests information from their partner about what actions they have completed. *E.g."Which relation did you add?"*
	What Are You Doing?	A participant requests information from their partner about what action they are currently performing. *E.g."are you creating the relation?"*
	What Will You Do?	A participant requests information from their partner about what action they intend to perform. *E.g."which entity are you going to do next?"*
	Where Are You?	A participant requests information from their partner about where they are on the diagram or on the hierarchy. *E.g."Which entity are you on?"*.
	Completion Status	A participant requests information from their partner about whether they have finished performing an action but without explicit reference to a particular editing action. *E.g."done?"*

their partner for information regarding their actions, intentions or location. Each of the supplied and requested type had five elements, these are summaries in Table 7.2.

Intercoder reliability. To verify that the rest of the videos were coded reliably, two independent coders used the developed coding scheme to code video transcripts from two pairs' collaborations. A Cohen's Kappa measure revealed high levels of reliability; 0.93 and 1.0 for the activity categories and the refined WA framework respectively.

7.5 Results

All sixteen pairs successfully completed the construction of two ER diagrams under each experimental condition using the provided hierarchy-based audio-only collaborative tool. All data was adequately captured with the exception of data from one pair, which was partly lost due to a system failure and was therefore excluded from the reported results. The average time spent constructing diagrams was 33.18 minutes (SD = 11.99) in the Shared conditions and 32.68 minutes (SD = 8.19) in the Non-Shared condition, a related t-test revealed that the difference was not statistically significant (t=0.151, p=0.881).

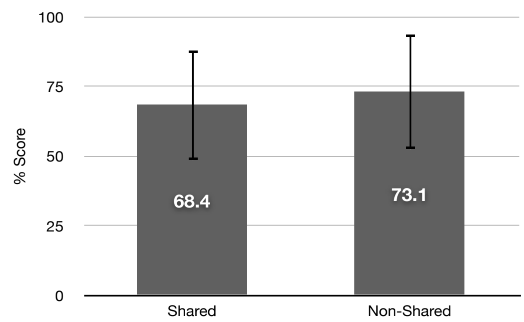


Figure 7.5: Means of participants scores on annotation tasks in the Shared and Non-Shared conditions – error bars show the standard deviation.

7.5.1 Annotation Task

Participants' ability to distinguish their contributions to the constructed diagrams from their partner's was measured by counting the proportions of their annotations that matched the information captured in the interaction logs. As shown on Figure 7.5, the mean proportion of correct annotations was 68.4% (SD=19.92) in the Shared condition and 73.1% (SD=20.67) in the Non-Shared condition. Although participants scored higher in the Non-Shared condition, the difference was not statistically significant ($t=-1.064$, $p=0.296$). This result does not support hypothesis H6.

7.5.2 Workspace Awareness Information Exchange

A related-sample Wilcoxon Sign Ranks test confirmed that overall participants exchanged significantly more WA information in the Non-Shared condition than they did in the Shared condition ($W=25.5$, $p<0.005$). This result supports hypothesis H7. As shown in Figure 7.6, participants supplied significantly more WA information to each other than they requested from one another in both conditions (70% supplied versus 30% requested in the Shared conditions; $W=79.5$, $p<0.0$, and 75% supplied versus 25% requested in the Non-Shared condition; $W=5$, $p<0.01$).

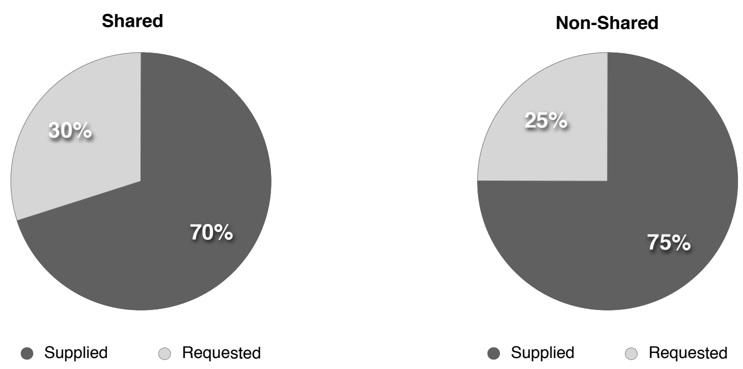


Figure 7.6: Proportions of supplied versus requested types of awareness information exchange in the Shared and Non-shared conditions.

A separate comparison of the supplied and requested types of WA information across the two conditions revealed that the supplied type of exchanges were significantly higher in the Non-Shared condition ($W=23$, $p<0.005$), but differences between the requested type of exchanges across the two conditions were not statistically significant ($W=106.5$). Comparing the occurrences of each of the five elements of the supplied type across the two conditions revealed that exchanges of three out of the five elements were significantly higher in the Non-Shared condition; pairs supplied significantly more information of type *"What I Did"* ($W=25.5$, $p<0.01$), *"What I Am Doing"* ($W=81$, $p<0.05$) and *"Supplied Completion Status"* ($W=15$, $p<0.01$) when audio was delivered through headphones. Figure 7.7 shows the exchanged proportions of all the elements of WA information.

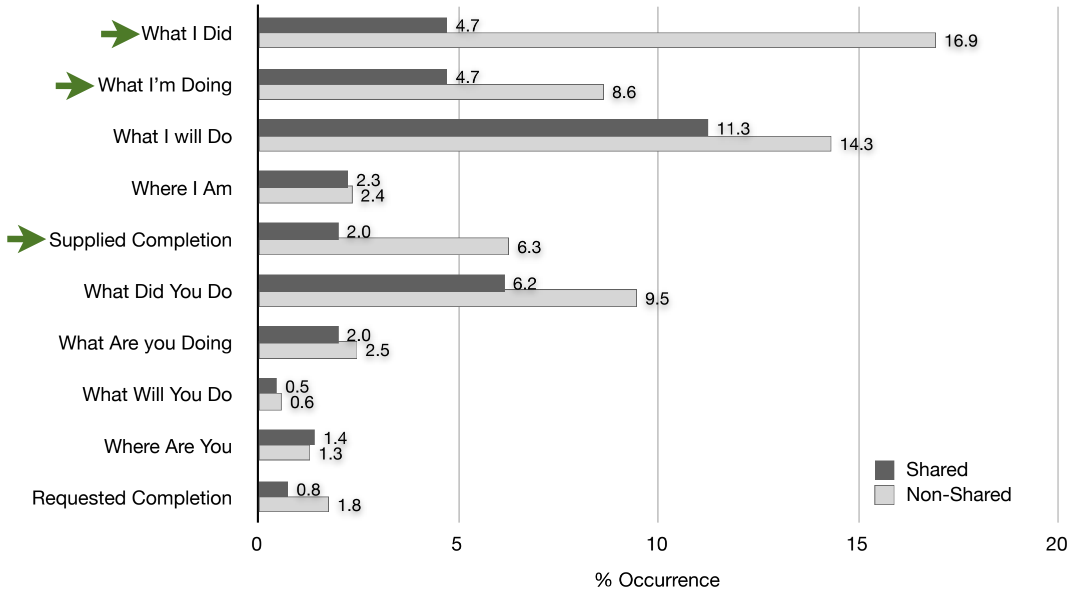


Figure 7.7: Proportions of WA information exchanges in the Shared and Non-shared conditions – arrows highlight statistically significant differences.

7.5.3 Activity Categories Structure and Organisation

Once coded, activity categories were used to capture two aspects of the audio-only collaborative process; 1) the proportions of time spent on each activity, and 2) process organisation in terms of transitions between the activities during the collaborations.

7.5.3.1 Time Distribution

Figure 7.8 summarises the average proportions of time spent on each activity category in each condition. What is immediately noticeable from the figure is that the patterns of time distribution in the two conditions are very similar. In both conditions, almost half of the construction times was dedicated to the *"Other"* activity category, where pairs manipulated diagrams and read through textual descriptions without conversing with one another (48.3% and 48.5% in the Shared and Non-Shared conditions respectively). About 40% of the remaining half was dedicated to discussions about content, both while participants interacted with the diagrams (16.3% in the Shared and 18.2% in the non-Shared condition) and when they did not (24.4% in the Shared and 22.1% in the Non-Shared condition). 6% of the remaining time was divided amongst the remaining six activity categories, with 2% to 3% of the total construction times dedicated to each of the *"Strategy"*, *"Labour"* and *"Execution"* activities.

Comparing the details of the *"Other"* activity category across the two conditions revealed

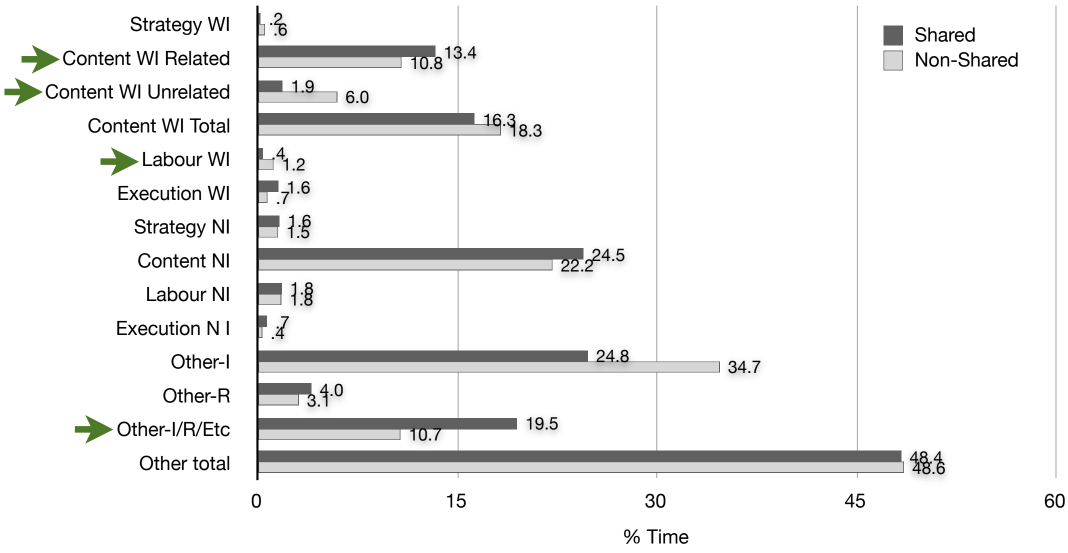


Figure 7.8: Proportions of time spent on each activity category in the Shared and Non-Shared conditions – arrows highlight statistically significant differences.

that participants spent significantly more time on the *"Other-I/R/etc"* activity in the Shared condition than they did in the Non-Shared condition (19.4% versus 10.7%; t=2.248, p<0.05). That is, when audio was delivered through speakers, pairs spent significantly more time where only one participant manipulated the diagram at a time than they did when audio was delivered through headphones. In contrast, pairs spent more time on the *"Other-I"* activity – where participants manipulated diagrams simultaneously – in the Non-Shared condition than they did in the Shared condition (34.7% versus 24.8%), though this difference was not statistically significant (t=-1.29, p=0.21).

Comparing the average time spent on the *"Content-NI"* and *"Content-WI"* activities across the two conditions yielded no statistical significance (16.3% in the Shared versus 18.2% in the Non-Shared conditions for the total *"Content-WI"*; t=-0.561, p=0.58, and 24.4% in Shared versus 22.1% in Non-Shared for *"Content-NI"*; t=0.947, p=0.35). Statistically significant results were found when looking at the details of the *"Content-WI"* activity on its own, however. The average time spent on the *"Content-WI Related"* activity was significantly higher in the Shared condition than it was in the Non-Shared condition (13.3% versus 10.8%; t=3.961, p<0.01), while the average time spent on the *"Content-WI Unrelated"* activity was significantly higher in the Non-Shared condition than it was in the Shared condition (5.9% versus 1.8%; t=-3.99, p<0.01). That is, the proportion of time where the discussed content and the accompanying diagram manipulations were related was significantly higher when audio was delivered through speakers, and the average time where they were unrelated was significantly higher when audio was delivered through headphones.

The differences between the average times spent on the *"Strategy"* and *"Execution"* activities across the two conditions, both while interacting with the diagram and while not, were not statistically significant. The difference between the average times spent on the *"Labour-NI"* activity across the two conditions were also not statistically significant (1.8% in Shared versus 1.7% in Non-Shared; $t=0.8$, $p=0.94$), but pairs spent on average significantly more time discussing labour while interacting with the diagram (*"Labour-WI"*) in the Non-Shared condition than they did in the Shared condition (1.2% versus 0.4%; $t=-2.752$, $p<0.02$).

7.5.3.2 Transitions

The way in which pairs structured and organised their collaborations was also examined in terms of how often they transitioned between the various activity categories when constructing the diagrams. Figure 7.9 shows the proportions of time spent on each activity and the probabilities of first-order transitions between activities in each condition. The figure presents an overview of the structure and organisation of the collaborations and the following addresses the major features captured by this representation.

There were similarities in the way pairs organised their collaborations in the Shared and Non-Shared conditions, but there were also clear differences. According to the transitions captured in Figure 7.9, the *"Strategy-NI"* activity acted as a point of distribution from which pairs moved to other activities. With the exception of movements towards the *"Labour-NI"* activity in the Non-Shared condition, transitions from the *"Strategy-NI"* activity had the same probabilities of occurrence in both conditions. Essentially, regardless of how audio was delivered to the workspace, pairs were more likely to move to discussing content without interacting with the diagrams (*"Content-NI"*) after discussing the construction strategy than they were to move to constructing the discussed content while concurrently discussing it (*"Content-WI"*) or without conversing with one another (*"Other"*). Similar patterns of transitions (towards the *"Content-NI"*, *"Content-WI"* and *"Other"* activities) were also likely to occur after pairs discussed division of labour (*"Labour-NI"*). When in *"Labour-NI"*, pairs were more likely to move towards these three activities than any of the other five activities. This was also the case in both conditions,

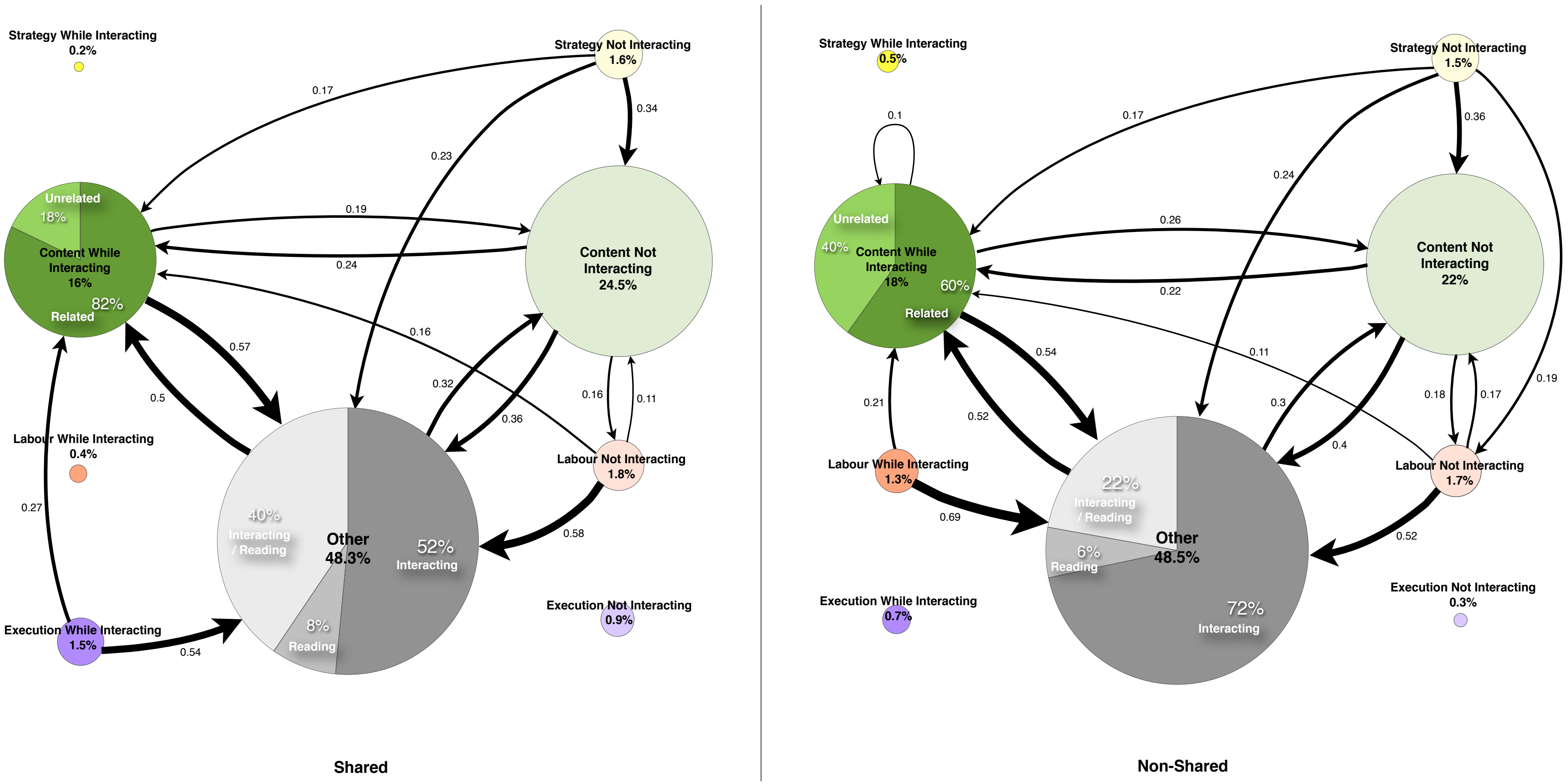


Figure 7.9: Pairs transitions between activity categories in the Shared and Non-Shared conditions. The size of the circles represents time, the arrows represent the probability of transition and the direction of movement. The bigger the circle the longer the pairs spent on the corresponding activity, and the thicker the arrow the more likely they were to move in its direction. The figure only shows transitions with a probability of occurrence higher than 0.1 and those occurring between activities that pairs were actively engaged in for more than 1% of the total construction times. The remaining transitions were omitted in the interest of clarity. The “Content-WI Related” and the “Content-WI Unrelated” categories were merged under one category labelled “Content-WI”. The “Other-I”, “Other-R” and “Other-I/R/etc” categories were merged under the “Other” category to further simplify the illustration.

but unlike the transitions from the *"Strategy-NI"* activity, pairs at this point were more likely to move towards the *"Other"* activity than they were to move towards the *"Content-NI"* and *"Content-WI"* activities. That is, the probability of discussing labour then moving on to executing the assigned tasks without conversing with one another was higher than the probability of discussing labour than moving on to discussing the content to be added unto the diagrams.

As mentioned above, pairs dedicated the largest part of their construction times to the *"Other"* activity. The next largest proportion of construction times was dedicated to discussing diagram content either while interacting with the diagrams (*"Content-WI"*) or when not (*"Content-NI"*). In both conditions, pairs moved fluidly back and forth between these three activities. However, when pairs discussed content while concurrently manipulating the diagrams (*"Content-WI"*), they were more likely to move towards the *"Other"* activity than they were to move towards the *"Content-NI"* activity. The transitions from the *"Other"* activity were also more likely to occur towards the *"Content-WI"* than they were to occur towards the *"Content-NI"* activity in both conditions. That is, pairs were more likely to interrupt their "silent" manipulation of the diagrams with concurrent discussion and manipulations than they were to do so by halting the manipulations all together.

In term of differences, there was a higher probability of transitions from the *"Labour-WI"* activity towards the *"Other"* and the *"Content-WI"* activities than towards any other activities in the Non-shared condition. Pairs, therefore, frequently interrupted their manipulations of the diagrams with discussions about division of labour when audio was delivered through headphones. Conversely, Pairs frequently interrupted their manipulations of the diagrams with discussions about execution (*"Execution-WI"*) when audio was delivered through speakers. Much lower times were dedicated to discussions about execution in the Non-Shared condition, and no transition from this activity was prominent enough to be captured by the representation in Figure 7.9. Also, pairs spent significantly more time where both participants manipulated the diagrams without conversing with one another (*"Other-I"*) when audio was delivered through headphones, and in contrast, they spent significantly more time where only one participant manipulated the diagram at a time when it was delivered through speakers ("Other-I/R/etc"). The ensemble of these differences naturally resulted from different patterns of interactions, which are examined in more detail in the following section.

7.5.4 Patterns of Interaction Styles

Actions performed on the tool were logged and timestamped and this information was used to capture and plot participants activities over the duration of the collaborations. Figure 7.10 shows example extracts from the plotted interaction patterns of pairs 11 and 15. Using this representation, it was possible to track not only which participant interacted with the tool at any particular moment during the collaborations, but also the type of actions that they performed (inspection or editing actions), their duration and the order in which these were executed.

The plots from pairs 11 and 15 show different styles of interaction. The difference is particularly noticeable in the way pairs organised turn taking when interacting with the diagrams in the Shared condition. When audio was delivered through speakers, participants in pair 11 interacted with the system in a sequential manner, where only one participant manipulated the diagram at any given moment. As shown in Figure 7.10, participant P11A issued the first eight diagram editing commands between minutes 2:45 and 6:00 of the collaboration while their partner remained interactively idle. Then, at minutes 6:15, it was participant P11B's turn to take on the editing of the diagram, issuing the next six editing commands (actions labelled 9 to 14 on the figure) between minutes 6:15 and 9:30 while P11A remained interactively idle. This sequential turn taking working style continued throughout the remainder of the extract shown on the figure. On the other hand, pair 15's collaborative interaction in the Shared condition shows a different pattern. Even though participant P15A issued the first three diagram editing commands between minute 2:00 and 2:30, their partner P15B joined in at minute 2:45 of the collaboration and both participants simultaneously executed editing actions 5, 6, 7, 8, 9 and 10. The pair continued working in parallel throughout the extract, both when inspecting the diagram and when editing it. The interactions patterns of pairs 11 and 15's in the Non-Shared condition do not manifest the same evident difference in the way collaborative work was organised. When audio was delivered through headphones, some of the pairs' editing actions occurred sequentially while others were executed in parallel, but almost all editing actions occurred when one participant was inspecting the diagram while their partner was issuing editing commands.

A similar analysis of interaction patterns was applied by the author to the remaining pairs to determine their working styles. The outcome is listed in Table 7.3. The majority of pairs worked either sequentially or in parallel under each condition, though some pairs did not exclusively employ a single style of interaction per condition. For pairs who employed both styles, one style

Figure 7.10: Example plots of interaction patterns for the first 18 minutes of pairs 11 and 15. The time is represented on the horizontal axis with fifteen seconds intervals. The red and green colours differentiate between each participant in a given pair, where lighter green and lighter red colours represent actions related to inspecting the diagram, and darker numbered red and green boxes represent editing actions. The dark red and green boxes that are not numbered are continuations of the editing action that preceded them (i.e. editing actions that took longer than fifteen seconds to perform). The numbers give the order in which editing actions occurred. Black bordered boxes highlight actions that were performed at roughly the same time. For each pair, the top part of the plot depicts patterns from the Shared condition and the bottom part from the Non-Shared condition.

Table 7.3: Pairs working styles in the Shared and Non-Shared conditions. For pairs who used two styles under one condition, the dominant style is mentioned first (before the “/”), the second style (after the “/”) occurred at some points during the collaboration. The (R) stands for Relations; i.e. the elements of the diagram for which the corresponding style was employed.

Pair	Working Style in Shared	Working Style in Non-Shared
Pair 1	Parallel	Parallel
Pair 2	Parallel	Parallel/Sequential Editing (R)
Pair 3	Parallel/Sequential Editing (R)	Parallel/Sequential Editing (R)
Pair 5	Parallel	Parallel
Pair 6	Parallel	Parallel
Pair 7	Parallel	Parallel
Pair 8	Sequential	Parallel
Pair 9	Sequential	Parallel/Sequential Editing
Pair 10	Sequential/Parallel Inspection	Parallel
Pair 11	Sequential	Parallel
Pair 12	Sequential	Parallel/Sequential Editing
Pair 13	Sequential	Sequential Editing/Parallel Inspection
Pair 14	Parallel	Parallel
Pair 15	Parallel	Parallel
Pair 16	Sequential	Sequential Editing/Parallel Inspection

was used more dominantly than the other, where in most cases the less dominant style was used either when editing particular elements of the diagrams or when inspecting rather than editing the diagrams. Particularly, two pairs who chose to work in parallel switched to the sequential style of working when they edited the relations of the diagram (Pairs 2 and 3), and four pairs used a sequential style when editing the diagram and switch to working in parallel when they inspected it (Pairs 9, 10, 12 and 13). Most pairs used a parallel working style as a dominant style in the Non-Shared condition, but worked sequentially and/or in parallel in the Shared condition; thirteen out of fifteen pairs used the parallel style exclusively or as a dominant style in the Non-Shared condition; while seven out of fifteen pairs worked sequentially, and the remaining eight worked in parallel either exclusively or as a dominant style in the Shared condition. Indeed, the proportion of overlapping interaction times – as extracted from the interaction logs – were significantly higher in the Non-Shared condition (42.56% vs 26.22% in the Shared condition;

t=2.841, p=0.013), but this result changed when pairs were grouped on the basis of their dominant working style in the Shared condition. For pairs classified as *Sequential*, the proportion of overlapping interaction was significantly higher in the Non-Shared condition than it was in the Shared condition (43.29% versus 8.03%; t=8.219 at p<0.001). For pairs classified as *Parallel*, the difference of overlapping interaction times between the two conditions was not statistically significant (42.22% Non-Shared versus 42.11% Shared; t=0.2, p=0.98).

The observed difference in working styles and the emergence of two distinct groups based on interaction patterns gave rise to the question of whether and how the results so far reported for all pairs change when each working style group is considered separately. The following sections, then, reports on the results of post-hoc analyses as independently applied to each working style group.

7.5.5 Results for Parallel Pairs

A total of eight pairs were classified as *Parallel* based on their dominant collaborative working style in the Shared condition.

Annotation Task. As shown in Figure 7.11, parallel pairs' mean score on the annotation task was 76.8% (SD=17.3) in the Shared, and 75.3% (SD=15.4) in the Non-Shared condition. The difference was not statistically significant (t=0.248 at p=0.808).

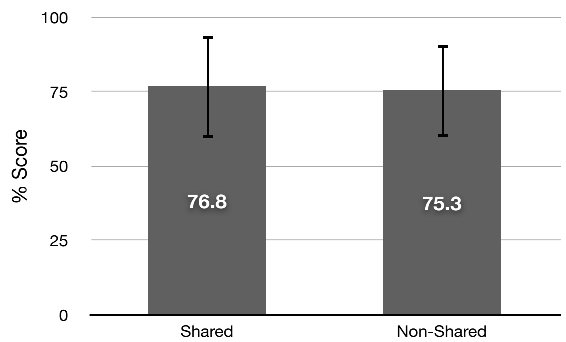


Figure 7.11: Means of parallel pairs' scores on annotation tasks in shared and Non-Shared conditions – error bars show the standard deviations.

Workspace Awareness Information Exchange. Parallel pairs exchanged significantly more WA information in the Non-Shared condition than they did in the Shared condition (W=27.5 for N=16, p<0.02). As shown in Figure 7.12, parallel pairs supplied more WA information to each other than they requested from one another in both conditions, but the difference was statistically

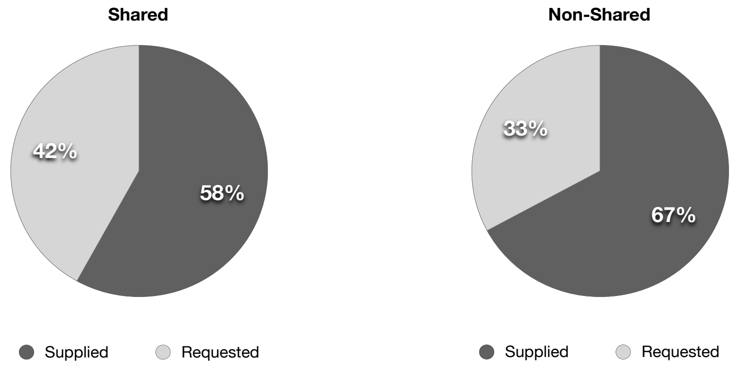


Figure 7.12: Parallel pairs proportions of supplied versus requested types of WA information exchanges in the Shared and Non-shared conditions.

significant in the Non-Shared condition only (58% supplied versus 42% requested in Shared, 67% supplied versus 33% requested in Non-Shared; W=5 for N=15, p<0.01).

A separate comparison of the supplied and requested types of WA information across the two conditions revealed that parallel pairs supplied significantly more WA information in the Non-Shared condition than they did in the Shared condition (W=16.5 for N=16, p<0.01) but differences between the requested types across the two conditions were not statistically significant. Comparing the occurrences of each of the five elements of the supplied types across the two conditions revealed that exchanges of two out of five elements were significantly higher in the Non-Shared condition; *"What I Did"*(W=12.5 for N=14, p<0.01), and *"Supplied Completion Status"* (W=7 for N=10, p<0.01). Figure 7.13 shows these proportions together with the remaining elements of supplied and requested types of WA awareness information.

Activity Categories – Time Distribution. As shown in Figure 7.14, there were no significant differences between the two conditions in terms of the amount of time that parallel pairs dedicated to each activity category. Parallel pairs spent almost 60% of their constructions times on the *"Other"* activity, about 30% on discussing content and 3% to 5% on the remaining activities.

Expectedly, the major portion of the construction times was spent on the *"Other-I"* activity (44.3% Shared and 42.8% Non-Shared) where both participants in a given pair simultaneously manipulated the diagrams without conversing with one another. The difference between these proportions in the two conditions was not statistically significant (t=0.182, p=0.86). Thus, regardless of how audio was delivered to the workspace, up to 44% of parallel collaboration times was dedicate to "silent" parallel manipulations of the diagrams. There was also no statistically

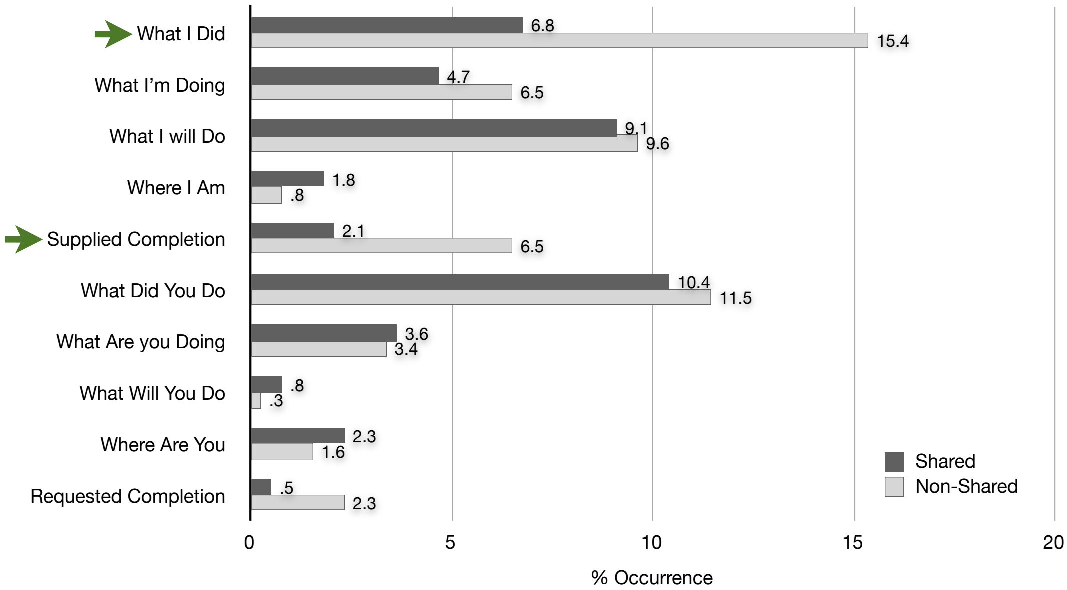


Figure 7.13: Details of parallel pairs proportions of WA information exchanges in the Shared and Non-shared conditions – arrows highlight statistically significant differences.

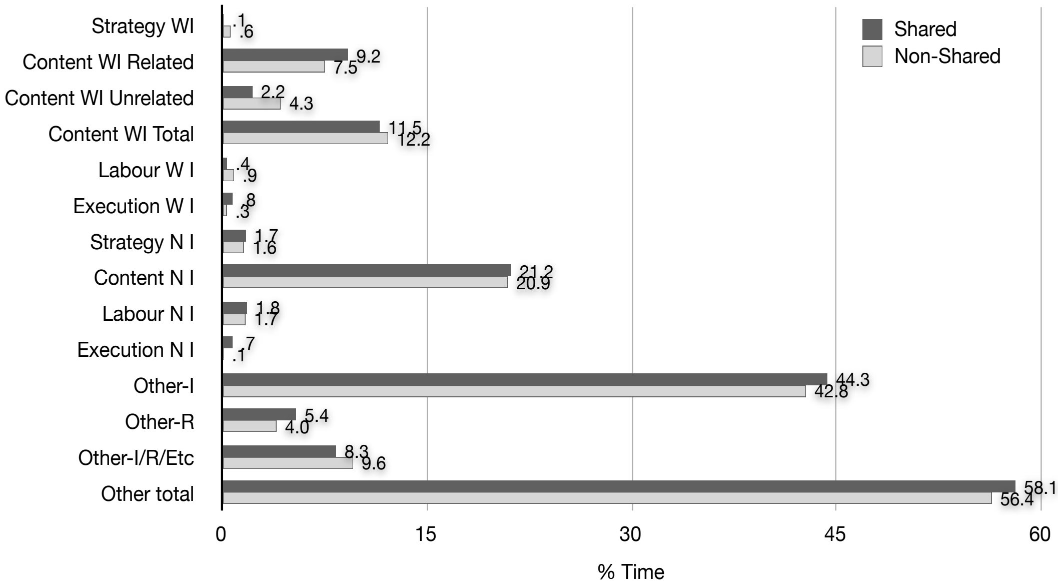


Figure 7.14: Proportions of time spent by parallel pairs on each activity category in the Shared and Non-Shared conditions.

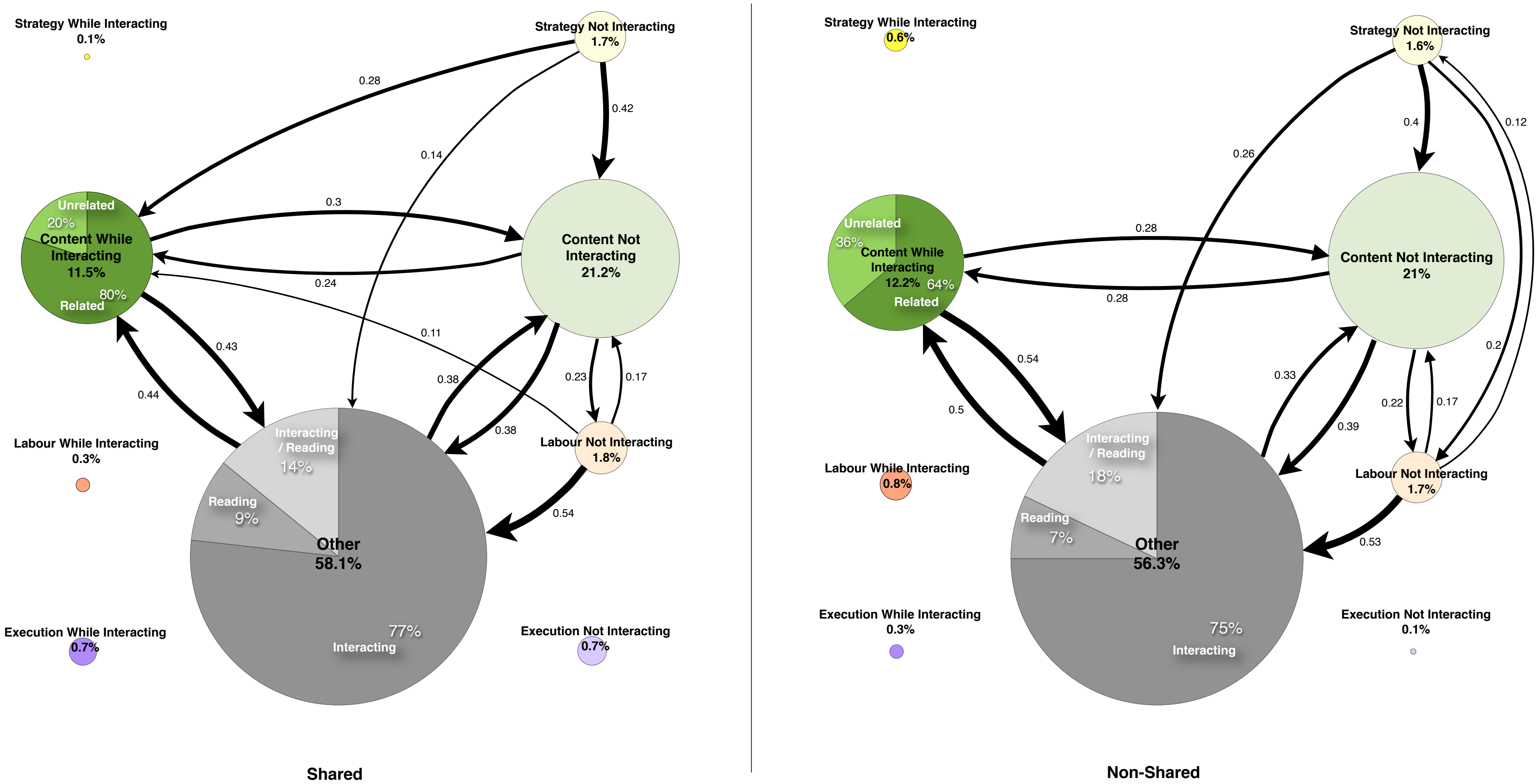


Figure 7.15: Parallel pairs activity categories proportions and transitions in Shared and Non-Shared conditions. The figure only shows transitions with a probability of occurrence higher than 0.1 and those occurring between activities that pairs were actively engaged in for more than 1% of the total construction times. The remaining transitions were omitted in the interest of clarity. The "Content-WI Related" and the "Content-WI Unrelated" categories were merged under one category labelled "Content-WI". The "Other-I", "Other-R" and "Other-I/R/etc" categories were merged under the "Other" category to further simplify the diagram.

significant difference between the amount of time that parallel pairs dedicated to the remaining activity categories in the two conditions. Thus, the way that parallel pairs organised their collaborative work in terms of the time spent on each activity did not change when the means for delivering audio to the workspace changed.

Activity Categories – Transitions. As shown in Figure 7.15, parallel pairs transitions between activities in the Shared and Non-Shared conditions were also similar. Movements between the *"Other"*, *"Content-NI"* and *"Content-WI"* activities were fluid and had similar probabilities of occurrence in both conditions, with the *"Other"* activity more likely to be interrupted by concurrent discussions and manipulations of diagrams (*"Content-WI"*) than by interactively-idle discussions about diagram content (*"Content-NI"*).

Transitions from *"Strategy-NI"* and *"Labour-NI"* activities towards *"Content-WI"* were more likely to occur when audio was delivered through speakers than they were when it was delivered through headphones. Transitions from the *"Strategy-NI"* activity therefore changed from movements towards the *"Content-WI"*, *"Other"* and *"Content-NI"* activities in the Shared condition to movements towards the *"Other"*, *"Content-NI"* and *"Labour-NI"* in the Non-Shared. Thus, parallel pairs were more likely to discuss diagram content after discussing construction strategies when they shared audio through speakers, but unlikely to move in that direction when they shared audio through headphones. Similarly, when discussing labour without interacting with the system, parallel pairs were more likely to move back to discussing strategy of construction when using headphones than they were when using speakers.

7.5.6 Results for Sequential Pairs

Seven out of fifteen pairs used a sequential working style either exclusively or as a dominant style in the Shared condition, and were subsequently classified as *Sequential*.

Annotation Task. As shown in Figure 7.16, sequential pair's mean scores on the annotation task were 58.8% (SD=18.7) in the Shared condition and 70.6% (SD=25.7) in the Non-Shared condition. This difference was statistically significant ($t=-1.798$ at $p<0.01$); pairs working sequentially annotated the diagrams more accurately when they used headphones.

Workspace Awareness Information Exchange. Sequential pairs exchanged significantly more WA information in the Non-Shared condition than they did in the Shared condition ($W=3$ for $N=14$, $p<0.01$). As shown in Figure 7.17, sequential pairs supplied more WA information to

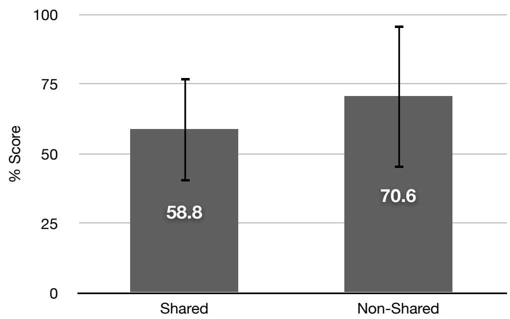


Figure 7.16: Means of sequential pairs scores on diagram annotation tasks in Shared and Non-Shared conditions – error bars show the standard deviations.

each other than they requested from one another in both conditions (84% vs 16% in Shared, and 82% versus 18% in Non-Shared). The proportion of WA information that was supplied and requested in the Non-Shared condition was significantly higher than that in the Shared condition (W=1 for N=13, $p<0.01$ for the supplied; and W=8 for N=12, $p<0.02$ for the requested).

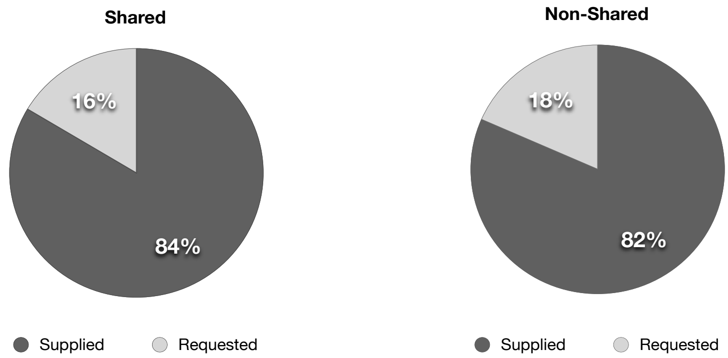


Figure 7.17: Sequential pairs proportions of supplied versus requested types of WA information exchanges in the Shared and Non-shared conditions.

In particular, sequential pairs supplied significantly more WA information of type *"What I Did"* (W=4 for N=14, $p<0.01$), *"What I Am Doing"* (W=17 for N=13, $p<0.05$) and *"Supplied Completion Status"* (W=2 for N=10, $p<0.01$), and requested significantly more WA information of type *"What Did You Do"* (W=2.5 for N=12, $p<0.01$) in the Non-shared condition. Differences in the proportions of the remaining elements were not statistically significant (Figure 7.18).

Activity Categories – Time Distribution. Sequential pairs dedicated almost half of their construction times to the *"Content-NI"* and *"Content-WI"* activities. The next largest share of the

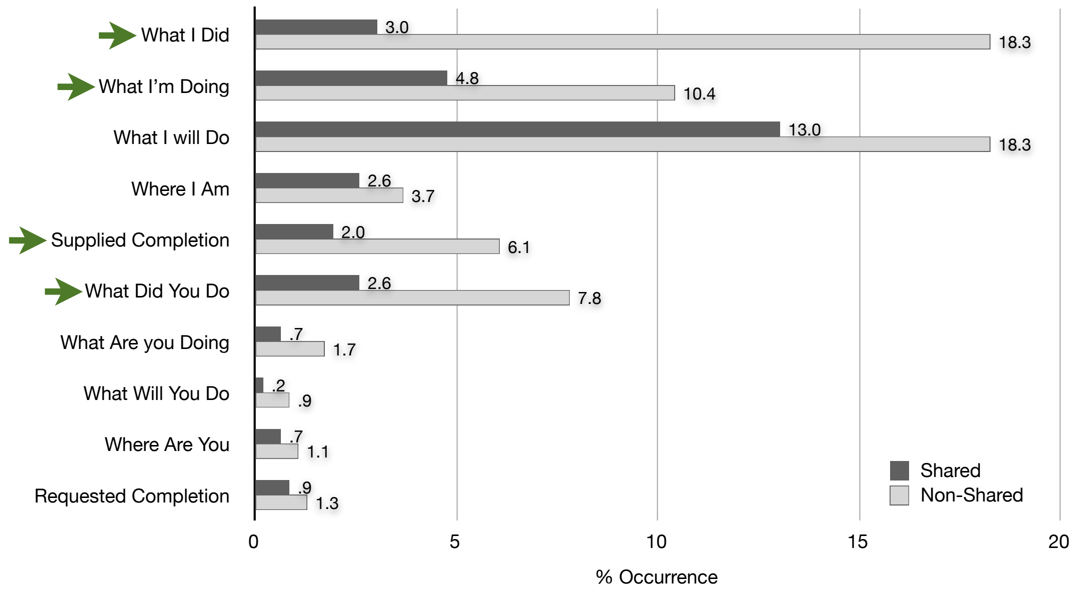


Figure 7.18: Details of sequential pairs proportions of WA information exchanges in the Shared and Non-shared conditions – arrows highlight statistically significant differences.

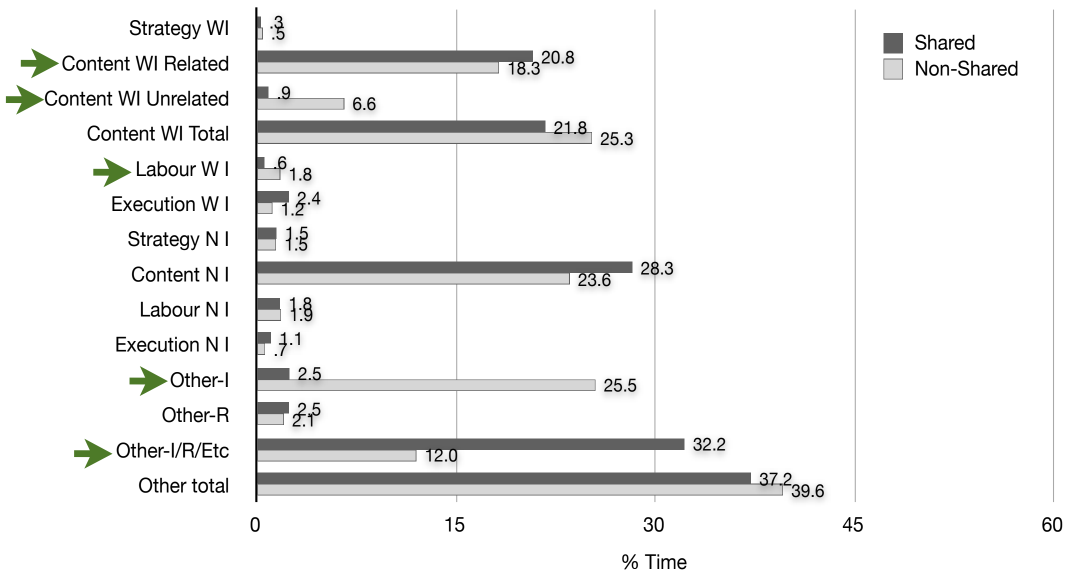


Figure 7.19: Proportions of time spent by sequential pairs on each activity category in the Shared and Non-Shared conditions – arrows highlight statistically significant differences.

construction times was dedicated to the "Other" activity with just under 40% (37.2% in Shared, and 39.6% in Non-Shared). About 7% of the remaining time was divided up between the remaining activities. As shown in Figure 7.19, there were a number of significant differences between the ways in which sequential pairs divided their construction times in the Shared and Non-Shared conditions. First, they spent significantly more time discussing labour while concurrently interacting with the diagram (*'Labour-WI"*) in the Non-shared condition (t=-2.313, $p<0.06$). Second, the average proportion of times where concurrent discussions about diagram content and diagram manipulations were related (*"Content-WI Related"*) was significantly higher in the Shared condition (t=6.639 at $p<0.001$). In contrast, the average proportion of times where they were unrelated (*"Content-WI Unrelated"*) was significantly higher in the Non-Shared condition (t=-6.142 at $p<0.001$). Third, the average time where both participants in a sequential pair interacted with the tool without conversing with each other (*"Other-I"*) was significantly higher in the Non-Shared condition (25.5% versus 2.5%; t=-6.016 at $p<0.001$), and, as expected, they spent significantly more time where only one participant interacted with the tool at a time (*"Other-I/R/etc"*) in the Shared condition (32.2% versus 12%; t=6.218 at $p<0.001$).

Activity Categories – Transitions. When discussing content without interacting with the system, sequential pairs were as likely to move to the *"Other"* activity as they were to move towards the *"Content-WI"*. However, when discussing content and concurrently manipulating the diagrams, they were more likely to move towards the *"Other"* activity than they were to move towards the *"Content-NI"* activity. The transitions from the *"Other"* activity were also more likely to occur towards the *"Content-WI"* than they were to occur towards the *"Content-NI"* activity. That is to say that sequential pairs were more likely to interrupt silent manipulations of the diagrams with concurrent discussions and manipulations of diagram content than they were to do so while halting the manipulations all together, and this was the case regardless of whether audio was delivered through headphones or speakers.

After discussing strategy in the Shared condition, sequential pairs were most likely to move towards the *"Other"* activity, then towards the *"Content-NI"* activity and, with a relatively less likelihood probability, towards the *"Content-WI"* activities. This was different in the Non-Shared condition where movements from the *"Strategy-NI"* activity were more varied, with transitions towards the *"Labour-NI"* activity equally likely to occur as the transitions towards the *"Content-WI"* and *"Other"* activities. Transitions from and towards the *"Labour-NI"* activity were also

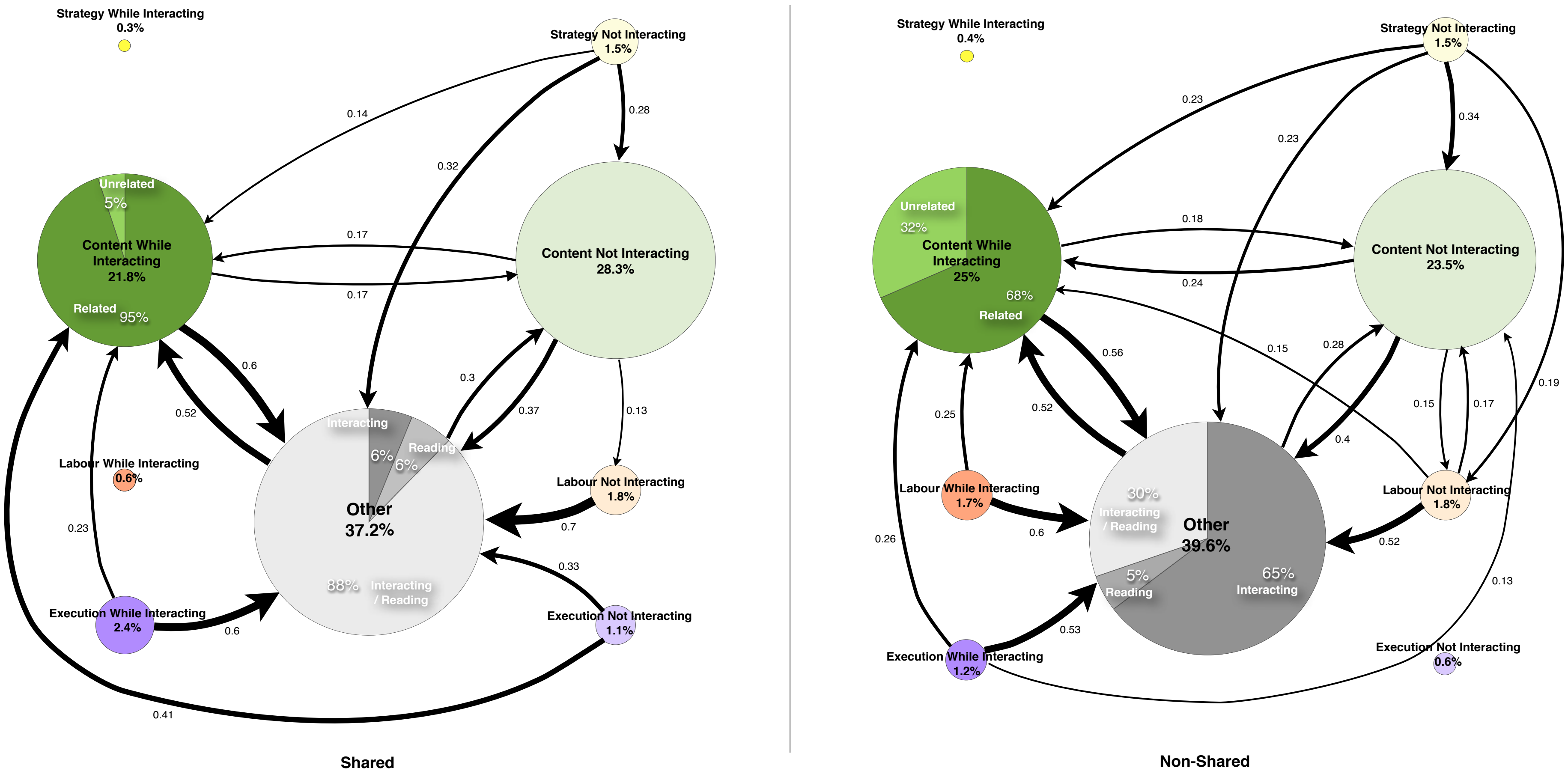


Figure 7.20: Sequential pairs activity categories proportions and transitions in Shared and Non-Shared conditions. The figure only shows transitions with a probability of occurrence higher than 0.1 and those occurring between activities that pairs were actively engaged in for more than 1% of the total construction times. The remaining transitions were omitted in the interest of clarity. The "Content-WI Related" and the "Content-WI Unrelated" categories were merged under one category labelled "Content-WI". The "Other-I", "Other-R" and "Other-I/R/etc" categories were merged under the "Other" category to further simplify the diagram.

different in the two conditions. In the Shared condition, discussions about labour typically followed interactively idle discussions about diagram content (*"Content-NI"*) and were also most likely to lead to the *"Other"* activity than any other activity. This was not the case in the Non-Shared condition, where discussions about labour were likely to occur after discussions about both strategy (*"Strategy-NI")* and diagram content (*"Content-NI"*).

Another noticeable difference between the two conditions is related to the *"Labour-WI"*, *"Execution-NI"* and *"Execution-WI"* activities. Sequential dedicated about 3.5% of their time to the *"Execution-NI"* and *"Execution-WI"* activities, and movements from such activities were likely to occur towards two main activities; the *"Content-WI"* and the *"Other"* activities. Thus, when audio was delivered through speakers, sequential pairs frequently interrupted their manipulations of the diagrams with discussions about how such manipulations are executed. Similar movements occurred in the Non-Shared condition, but the direction of movements from the *"Execution-WI"* activity in this case included all of the *"Other"*, with the highest probability of occurrence, then the *"Content-WI"* and *"Content-NI"* activities. Also, the majority of discussions about execution that occurred in the Non-Shared condition occurred while the pairs concurrently manipulated the diagrams, which was not the case in the Shared condition, where discussion about execution occurred both when pairs actively manipulated the diagram and when they remained interactively idle. Finally, discussions about *"Labour-WI"* were more prominent in the Non-Shared condition, frequently interrupting the *"Other"* activity, which was not the case the Shared condition.

7.5.7 Parallel versus Sequential pairs

Annotation Task. As shown in Figure 7.21, parallel pairs' mean score on the annotation task was significantly higher than that of sequential pairs in the Shared condition (76.8% versus 58.9%; $t=-2.721$ $p<0.02$) but not in the Non-Shared condition (75.4% versus 70.6%; $t=-0.619$; $p=0.54$). Thus, parallel pairs were more accurate at highlighting self and partner's contributions.

Workspace Awareness Information Exchange. A Mann-Whitney test revealed that there was no significant difference in the overall amount of WA information exchanged between parallel and between sequential pairs in the Shared condition. However, comparing the supplied and requested types of exchange separately in this condition revealed that parallel pairs requested significantly more WA information from one another than did the sequential pairs ($U=58.5$, $p<0.05$),

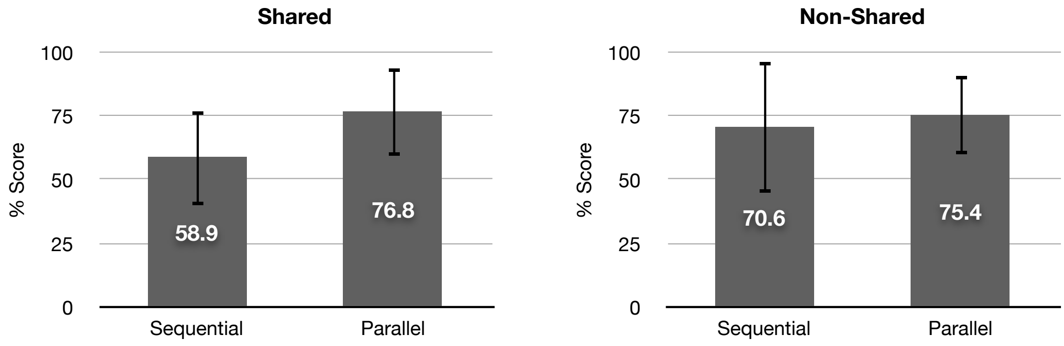


Figure 7.21: Means of parallel versus sequential pairs scores on diagram annotation tasks in Shared and Non-Shared conditions error bars show the standard deviations.

particularly of type *"What Did You Do"* (U=56.5 p<0.02).

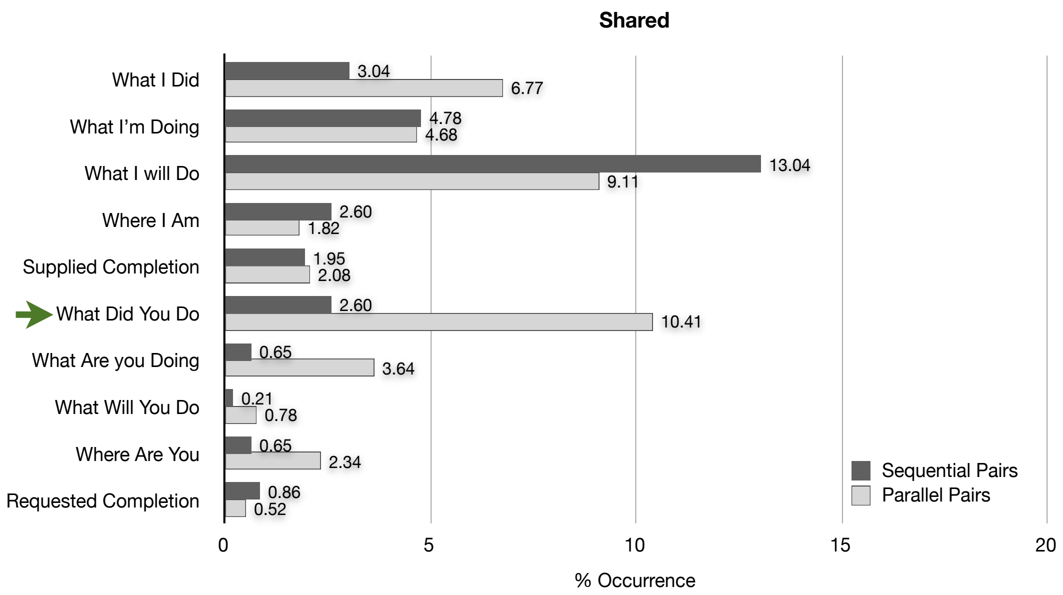


Figure 7.22: Parallel versus sequential pairs proportions of workspace awareness information exchanges in the Shared condition – arrows highlight statistically significant differences

In the Non-Shared condition, sequential pairs exchanged significantly more WA information than the parallel pairs (U=50, p<0.02). Comparing the supplied and requested types separately in this condition revealed that this difference was significant for the supplied type (U=38, p<0.02) but not the for requested type (U=102.5). In particular, sequential pairs supplied significantly more WA information of the type *"What I Will Do"* (U=40.5, p<0.02). No significant difference was found for the remaining WA elements (Figures 7.22 and 7.23).

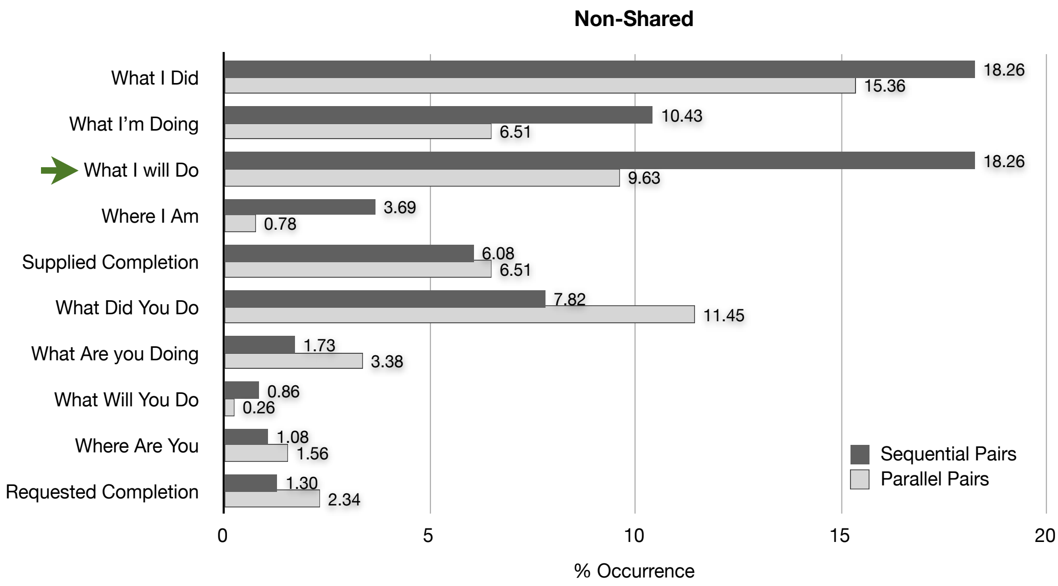


Figure 7.23: Parallel vs sequential pairs proportions of workspace awareness information exchanges in the Non-Shared condition – arrows highlight statistically significant differences

Activity Categories – Time Distribution. As shown in Figures 7.24 and 7.25, the main differences in time distributions between parallel and sequential pairs were in relation to the *"Content-WI"*, *"Content-NI"* and *"Other"* activities. In the Shared condition, and as expected, parallel pairs spent significantly more time on the *"Other-I"* (44.3% versus 2.5%; t=-6.48 at $p<0.0001$) and *"Other-R"* (5.4% versus 2.5%; t=-2.212 at $p<0.05$) activities, while sequential pairs spent significantly more time on the *"Other-I/R/etc"* activity (32.3% versus 8.3%; t=5.712 at $p<0.0001$). Additionally, sequential pairs spent significantly more time on the *"Content-WI"* (21.8% versus 11.5%; t=2.655 at $p<0.02$) and *"Content-NI"* (28.3% versus 21.2%; t=2.337 at $p<0.05$) activities than parallel pairs did. In the Non-Shared condition, sequential pairs spent significantly more time on the *"Content-WI"* activity than the parallel pairs did (25.3% versus 12.2%; t=3.036 at $p<0.01$). Differences in the proportions of time spent on the remaining activities were not statistically significant in either conditions.

Activity Categories – Transitions. As shown in Figure 7.26, there were differences in the way parallel and sequential pairs organised their collaborative work in terms of transitions amongst activities. In the Shared condition, sequential pairs' movements from and towards the *"Other"* activity were more varied. In particular, the *"Other"* activity was frequently interrupted with discussions about execution (2.4% on the *"Execution-WI"* and 1.2% on the *"Execution-NI"* activities). This was in contrast to parallel pairs' who dedicated only 0.7% of their overall construction times to each of the *"Execution-NI"* and *"Execution-WI"* activities.

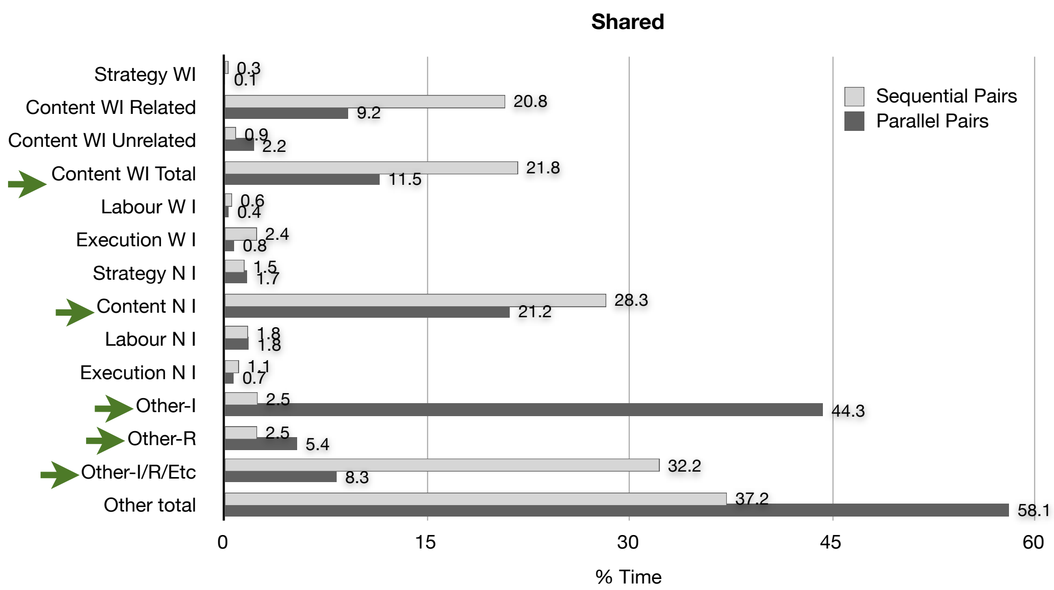


Figure 7.24: Parallel versus sequential pairs proportions of time spent on each activity category in the Shared condition – arrows highlight statistically significant differences.

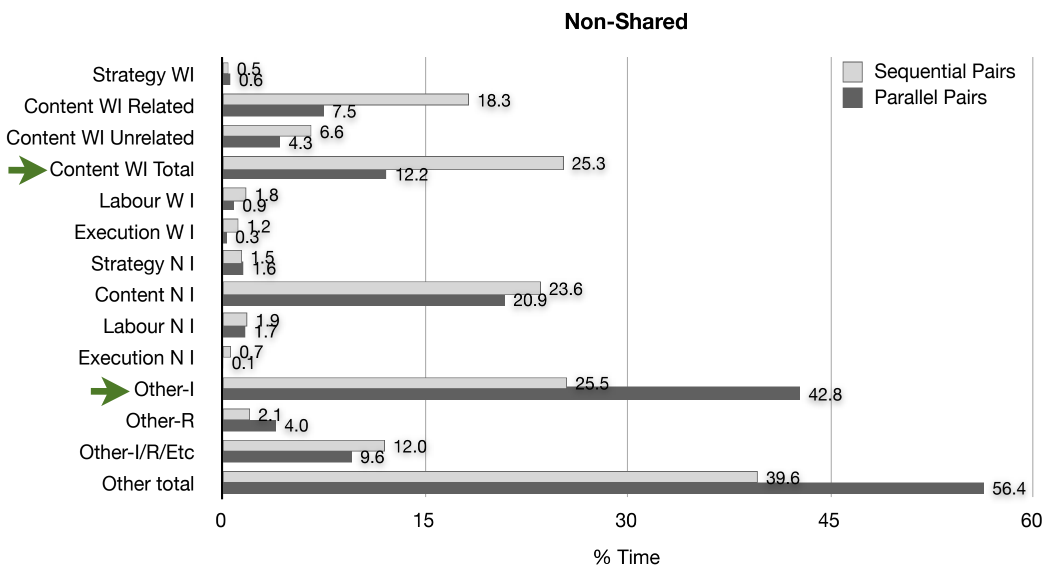


Figure 7.25: Parallel versus sequential pairs proportions of time spent on each activity category in the Non-Shared condition – arrows highlight statistically significant differences.

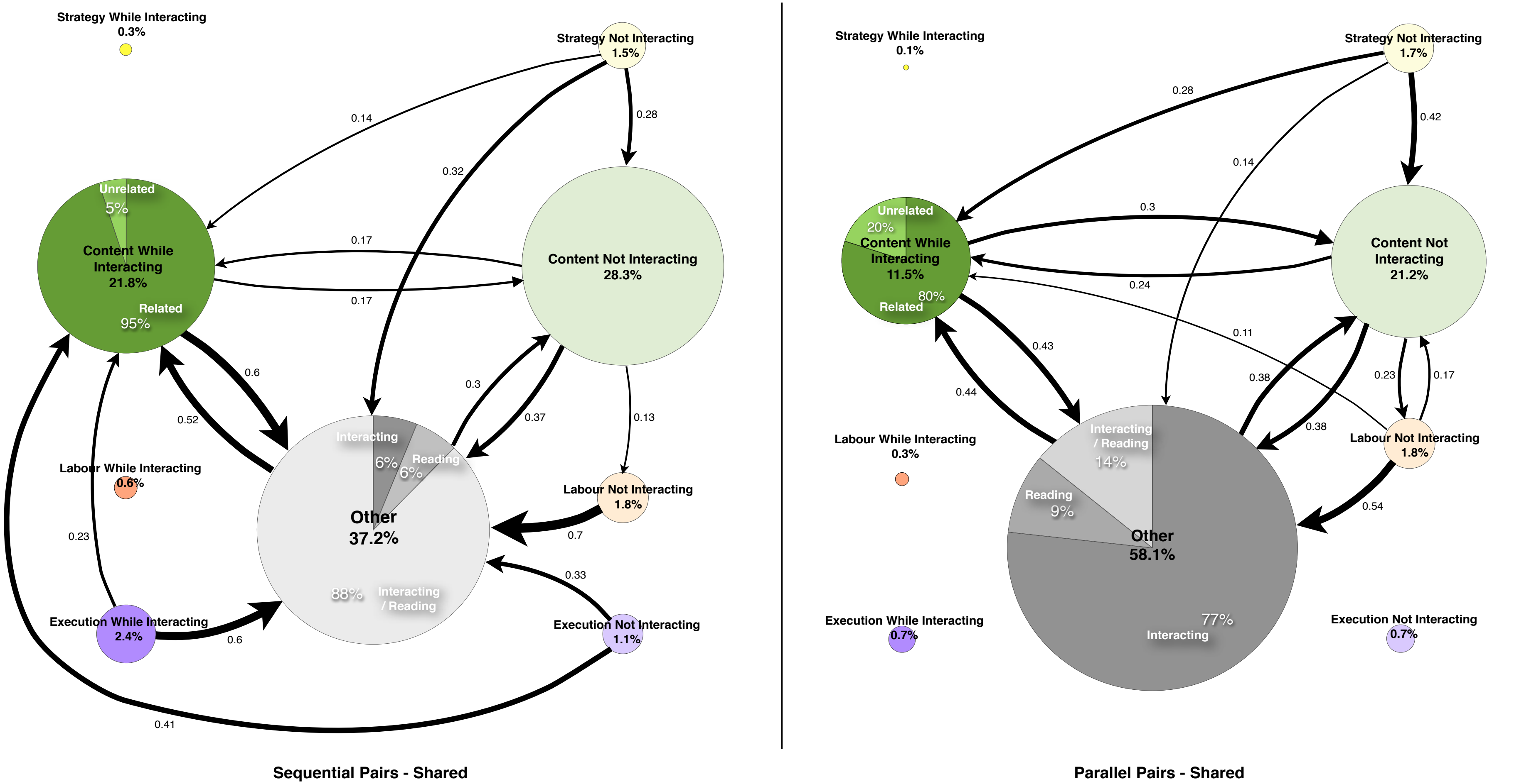


Figure 7.26: Parallel versus sequential pairs activity categories proportions and transitions in Shared condition. The figure only shows transitions with a probability of occurrence higher than 0.1 and those occurring between activities that pairs were actively engaged in for more than 1% of the total construction times. The remaining transitions were omitted in the interest of clarity. The Content-WI Related and the Content-WI Unrelated categories were merged under one category labelled Content-WI. The Other-I, Other-R and Other-I/R/etc categories were merged under the Other category to further simplify the diagram.

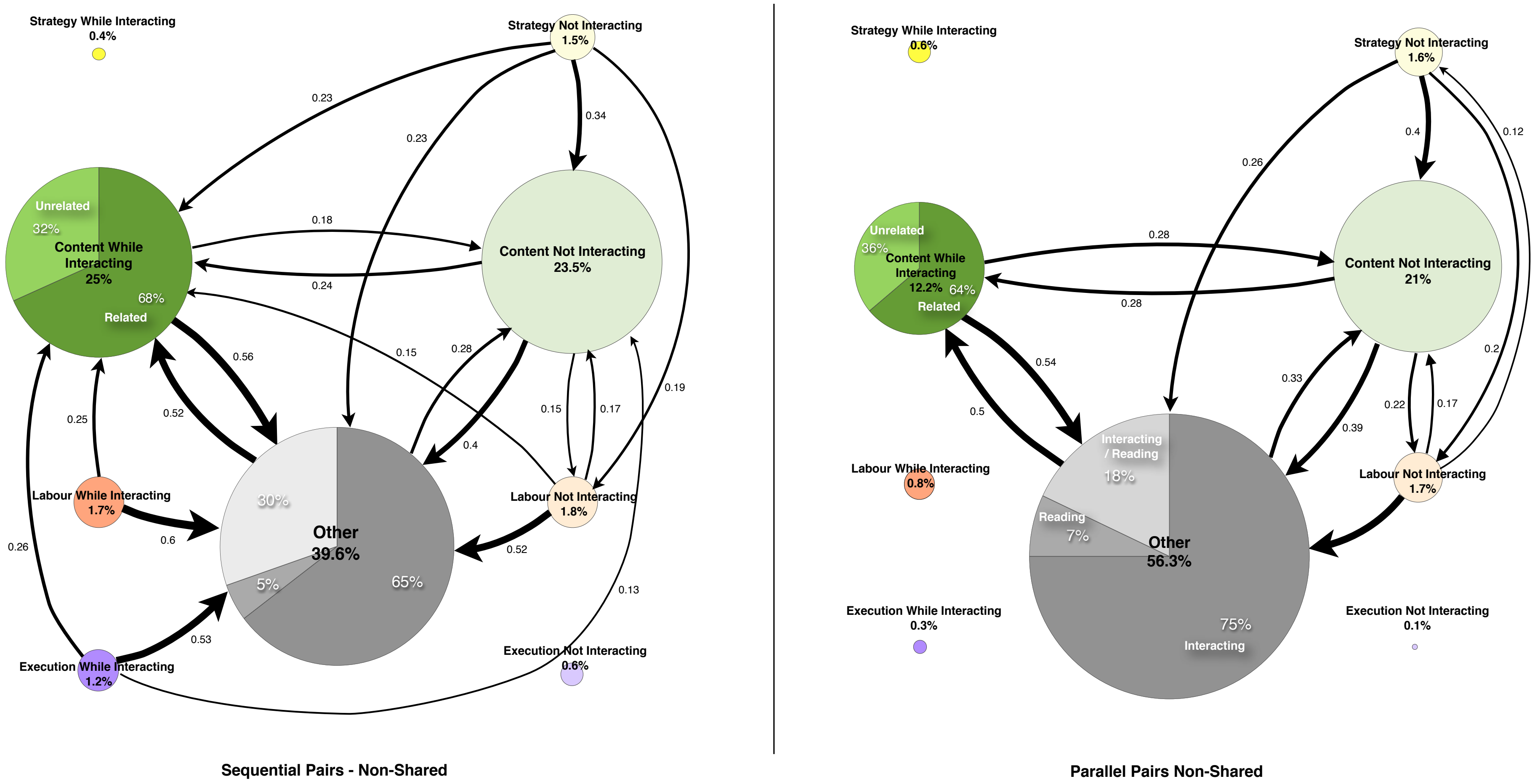


Figure 7.27: Parallel versus sequential pairs activity categories proportions and transitions in Non-Shared condition. The figure only shows transitions with a probability of occurrence higher than 0.1 and those occurring between activities that pairs were actively engaged in for more than 1% of the total construction times. The remaining transitions were omitted in the interest of clarity. The Content-WI Related and the Content-WI Unrelated categories were merged under one category labelled Content-WI. The Other-I, Other-R and Other-I/R/etc categories were merged under the Other category to further simplify the diagram.

Furthermore, Sequential pairs dedicated half of their construction times to the *"Content-NI"* and *"Content-WI"* activities, while parallel pairs dedicated just over 30% of their time to these activities. Movements from and towards the *"Content-WI"* activity were also more varied for the sequential pairs than they were for the parallel pairs. In particular, sequential pairs were most likely to move back and forth between the *"Other"* and the *"Content-WI"* activity, but also to interrupt such activities with discussions about execution. For parallel pairs, the main movements from the *"Content-WI"* activity were towards the *"Other"* and the *"Content-NI"* activities; that is, discussing content then constructing it without conversing with one another. Also, when sequential pairs discussed construction strategies, they were more likely to move towards the *"Other"* and the *"Content-NI"* activities than towards the *"Content-WI"* activity. In contrast, parallel pairs were less likely to move towards the *"Other"* activity after discussing construction strategies, and more likely to move towards the *"Content-NI"* and *"Content-WI"* activities.

Both parallel and sequential pairs discussed labour when not interacting with the diagrams (*"Labour-NI"*) much more than they did when concurrently interacting with it (*"Labour-WI"*). However, their movements from and towards this activity were different; sequential pairs were likely to discuss division of labour after having discussed diagram content in the *"Content-NI"* activity, then, from there, they were most likely to move on to the *"Other"* activity than towards any other activity. On the other hand, parallel pairs' discussions about division of labour were likely to be followed by either one of the *"Other"*, *"Content-NI"* or *"Content-WI"* activities. In the Non-Shared condition, sequential pairs' transitions towards the *"Content-WI"* activity were much more varied than was the case for pairs working in parallel. In particular, sequential pairs interrupted the *"Content-WI"* activity with discussion about labour, both when interacting with the diagram and when interactively idle, and with discussions about construction strategies and execution. In contrast, parallel pairs' most likely movements from and towards the *"Content-WI"* activity included only two out of the eight activity categories; the *"Other"* and the *"Content-NI"* activities. Further, discussions about execution (*"Execution-WI"*) were much more prominent in sequential pairs' collaborations, where discussions about execution were likely to interrupt all of the *"Other"*, *"Content-WI"* and *"Content-NI"* activities. Similarly, the *"Labour-WI"* activity was much more prominent for the sequential pairs than it was for the parallel pairs.

7.6 Summary and Conclusion

This chapter presented Study 3, which aimed to examine audio-only collaborative interaction with diagrams; i.e. addressing Research Question 3 of this thesis. The results showed that pairs of sighted participants were able to use a hierarchy-based audio-only tool to collaboratively construct and edit shared diagrams in an audio-only workspace. The results also showed that varying the means for delivering audio to such a workspace had an impact on how participants structured and organised their work as well as on how they exchanged and used workspace awareness information. Additionally, an examination of participants interaction patterns distinguished between two different working styles; a *Parallel* style, where participants worked concurrently; and a *Sequential* style, where they took turns in contributing to the shared task. Pairs were retrospectively grouped on the basis of their dominant working style in the Shared condition, and the observed impact of the means for delivering audio on the collaborations was found to be different when data from each working style group was considered independently.

Awareness of Contributions. An annotation task measured participants' awareness of self and partner's contributions to the shared task and the results showed no significant difference between participants' scores when all pairs were considered. No significant difference was found between parallel pairs' scores but sequential pairs scored significantly higher on this task when audio was delivered through headphones.

Workspace Awareness. A coding scheme was used to capture instances in the collaborations where participants explicitly exchanged information pertaining to workspace awareness. This provided a means for establishing which elements of workspace awareness information were used during the collaborations and for quantifying such information. The results confirmed that delivering audio through headphones to the audio-only workspace increased participants exchange of workspace awareness information. Participants supplied significantly more information to each other than they requested from one another, but examining the details of such exchange revealed that this significance was in supplying WA information of types *What I Did* (past actions), *What I am Doing* (current actions) and *Completion Status* (activity level). Considering each working style group independently revealed differences in the details of these exchanges:

- Parallel pairs supplied as much WA information to each other as they requested from one another when audio was delivered through speakers. Sequential pairs supplied significantly more information to each other than they requested from one another in both conditions.

- When audio was delivered through headphones, parallel pairs supplied significantly more WA information of type *What I Did* (past actions), whereas sequential pairs supplied significantly more WA information of type *What I Am Doing* (current actions) and requested significantly more WA information of type *What Did You Do* (partner's past actions).

- When compared against each other, sequential pairs were found to supply significantly more WA information of type *What I Will Do* (intentions) than parallel pairs did when audio was delivered through headphones.

Activity Categories. A coding scheme was used to categorise the various activities that participants engaged in the course of their collaborations. The categories provided a means for capturing the structure and organisation of the audio-only collaborations in terms of the amount of time allocated to each activity and probabilities of transition between one activity and another. Similar time distributions were observed in the two conditions. Almost half the construction times was dedicated to silent interactions; where participants manipulated the diagrams or read through diagram descriptions without conversing with one another. Participants spent the second largest share of their time discussing diagram content, either while concurrently manipulating it or when interactively idle. The remaining time was dedicated to discussing construction strategies, division of labour and technicalities of executing system commands.

There were differences in the details of these distributions, however. When audio was delivered through speakers, pairs spent significantly more time where only one participant manipulated the diagrams at a time. Concurrent manipulations and discussions about diagram content were also significantly more related in this condition. Additionally, and although not statistically significant, system related discussions were more prominent when speakers were used. In contrast, when audio was delivered through headphones, pairs spent significantly more time where both participants simultaneously manipulated the diagrams, and concurrent manipulations and discussions about diagram content were significantly more unrelated. These differences were amplified when each working style group was considered independently. Parallel pairs showed no significant difference in the time allocated to each activity regardless of whether audio was delivered through headphones or speakers, while sequential pairs discussions about diagram content were significantly more related to their manipulations of the diagrams when they used speakers and significantly unrelated when they used headphones. When compared against each other, sequential pairs were found to discuss diagram content and system execution significantly more often than the parallel pairs did when audio was delivered through speakers.

The next chapter presents a thorough analysis and discussion of the these results highlighting their implications for the design of collaborative audio-only interaction with relational diagrams.

Chapter 8

Analysing Audio-only Collaboration

8.1 Introduction

The previous chapter presented a study that examined collaborative interaction with diagrams in an audio-only workspace. The results showed that pairs of sighted participants were able to use sounds as the only means to communicate, access and edit shared diagrams and that varying the means for delivering audio in such a workspace impacts collaborative organisation and exchange and use of workspace awareness information. Pairs employed parallel and sequential working styles to collaborate through sounds and the observed impact was found to be more prominent when data from each working style group was considered independently.

This chapter presents a detailed discussion of the findings reported in Chapter 7 and compiles a set of design lessons learnt from them, which are presented in pop-out boxes throughout the chapter and will be summarised later in Chapter 9. Aspects of the collaborative process are discussed in Section 8.2, addressing the approach followed in categorising participants' activities, their strategies for tackling and organising shared tasks. Sections 8.3, 8.4 and 8.5 then examine aspects of awareness in audio-only collaboration. Particularly, the sections deal with the impact of working styles on; collaborators' ability to keep an awareness of self and partner's contributions to shared tasks; exchange of workspace awareness information during the collaborations; and the way in which such information was extracted from the audio-only workspace and used in the collaborations. Section 8.6 and 8.7 examine the use of audio as a shared representation and the impact of working style in audio-only collaboration. Section 8.8 concludes the chapter.

8.2 Collaborative Process

The collaborative process of constructing diagrams in an audio-only workspace was captured using a set of categories that coded participants activities during the course of their collaborations. The following sections examines the general characteristics of their collaborative process.

8.2.1 Activity Categories

Collaborating pairs examined textual descriptions and used an audio-only tool to construct ER diagrams that captured the described information. To do this, pairs discussed four main topics; the potential *Content* of the diagrams, construction *Strategies*, division of *Labour* to agree on individual task assignments, and the technicalities of using the provided tool to *Execute* editing actions. There were additional aspects of the collaborations that involved no discussion, including instances where both participants in a pair silently interacted with the diagrams (*Other-I*), remained interactively idle (*Other-R*) (e.g. reading the descriptions), or where only one participant interacted with the diagrams (*Other-I/R/etc*). The pairs were explicitly instructed to consult the provided textual descriptions and construct diagrams using the provided tool; it was therefore natural that they engaged in the activity categories mentioned above in order to get the task done.

The tool did not include any means for controlling the collaborative process. Participants were thus free to choose how to organise and structure their work. Because of the lab-based nature of the task, participants did not know which diagrams they were going to construct, and were aware that they were not going to reuse them once their participation was over. This aspect of the task is likely to have influenced the nature of the activities that the participants chose to engage in. For example, participants did not need to consider the wider context of the diagrams, such as the suitability, feasibility or even quality of their designs. Indeed, the outcome of their collaborations was not assessed in terms of such criteria since the aim of Study 3 was to examine the collaboration process rather than its outcome. The above are issues that would need careful consideration and planning if the context of the task was different. For instance, if it was part of an academic assessment or a real company commissioned project.

8.2.2 Construction Strategies

Collaborating pairs typically initiated the construction process by reading through the textual descriptions, either silently or out loud to one another, then examining the initial diagram already loaded on the tool to establish which portions of the descriptions are already in place, then devised strategies for completing the diagrams and executed them. Their strategies varied, and an examination of the video footage and transcripts revealed that those were of two main types:

8.2.2.1 Description-based Strategies

Some pairs used the order in which diagram elements were presented on the textual descriptions to organise the construction process. This strategy was used by both parallel and sequential pairs and consisted of going through the provided descriptions and adding items onto the diagrams as soon as these are encountered on the text and agreed on by the participants. When working in parallel, it was typical for each participant in a pair to construct the items available on their own description before prompting their partner to check what have been done independently. This often meant that extra editing was needed to reconcile the complementary information contained in each description. When working sequentially, participants typically went through the order of the text together, jointly extracting relevant information about a particular item from both descriptions before adding it onto the diagram. Reconciliation of the information was thus achieved before the editing of a given diagram element began.

8.2.2.2 Diagram-based Strategies

Other pairs used the elements of the diagram to be constructed to organise the construction process rather than the order in which these appear on the textual descriptions. In this strategy, participants would choose to focus on a particular element, and scan through the descriptions a number of times extracting all relevant information related such element each time around. For example, participants might go through their descriptions to retrieve all information related to a particular entity, such as the names of its attributes, add these onto the diagram, then move on to the next entity, and so on. This strategy of construction was also executed by both parallel and sequential pairs.

Not all pairs orchestrated their work using one of these strategies exclusively, however. Some pairs used a mixture of the two strategies to execute different parts of the diagram. For instance, using a description-based strategy to complete the entities part of the diagram and a diagram-based strategy to complete its relations. In all cases, discussions about the choice of construction strategies typically occurred at moments during the collaborations where an episode of construction was completed [1].

[1] A construction episode is one where a pair completes the execution of an agreed set of actions and moves on

8.2.3 Time Distributions and Transitions

Pairs dedicated almost half of the construction times to reading through the textual descriptions and manipulating the diagrams without conversing with one another. The second largest share of the construction times was dedicated to discussing diagram content (40%), and the remaining times (10%) were dedicated to discussions about construction strategies, division of labour, and the technicalities of executing editing actions. Allocating times in such a manner is also reflective of the lab-based nature of the task. Since they were instructed to do so, it is natural that participants concentrated much of their time on using the tool and discussing diagram content during their collaborations, allocating relatively little time to the remaining activities.

As well as capturing the time distributions, a similar approach to that of Olson et al. (1992)'s analysis of collaborative design meeting practices was used to capture the frequency of transitions between collaborative activities. Such information provided insights into the flow of activities during the collaborations, and served as a further resource to inform the comparisons of the collaborative process in the two experimental conditions. Although informative, observations based on the transition diagrams should be taken with caution as a number of constraints were imposed to produce such representations. Particularly, the imposed constraints meant that micro level transitions between activities and the chronological structure of the collaborations were not captured, all of which might give rise to further insights into how pairs organised their work.

8.2.3.1 The impact of the means for delivering audio

Whether audio was "publicly" available through speakers or "privately" received through headphones made a difference to the way participants spent their times during their collaborations. In particular, pairs spent more time where only one participants manipulated the diagrams – and hence produced sounds – when using speakers. The opposite was true, pairs spent more time where they simultaneously manipulated the diagrams when they used headphones. Related to this is the fact that concurrent manipulations of the diagrams and accompanying discussions about diagrams content often matched in substance when speakers were used, significantly more so than when headphones were used. This was also true for discussions about execution, which were significantly more frequent when speakers where used.

In general, discussions about execution were of two types; *Reminders* and *Guidance*. Reminders occurred when a participant forgot which system command corresponded to performing

to a new set. It is similar to the concept of *closure* (Dix et al., 1998) that a single user experiences during interaction after completing a given task and moving on to the next.

a particular action on the diagram. Their partner in such a case reminded them of which key combination to use (for example that pressing on the key *P* while holding the *command* key sets an attribute as a "primary key"). The partner would do this either when prompted or when they managed to detect that a problem is encountered with the execution of a given action[2]. On the other hand, guidance involved supplying the details of the step by step execution of a series of commands or procedure. The partner in this case followed and supplied information relevant to each step until the desired actions were completed[3].

The majority of discussions about execution that occurred when audio was delivered through headphones were of the first type, whereas those that occurred when audio was delivered through speakers included both types. Here too, the means for delivering audio made a difference. Guiding a partner through the process of executing a procedure requires an ability to keep track of their progress at every single step of such a process. While such information could be readily available in an environment where audio feedback is publicly displayed through speakers, it is harder to obtain when the audio is presented through headphones. In the latter case, a participant would need to keep supplying their partner with information about various aspects of their interaction (location, outcome of actions, system feedback, etc.) so that, in turn, the partner can provide appropriate guidance that corresponds to the progress of execution[4].

Design Lesson 12 – Detect guidance of partners and supply awareness information accordingly. An audio collaborative system should be able to either detect guidance modes of interaction or allow users to manually switch to such a mode. If using headphones, the guidance mode should explicitly convey and/or allow users to explicitly track awareness information of types *"What Did You Do"*, *"What Are You Doing"* and actions' *"Completion Status"*.

Thus, the means for delivering audio changed the way participants went about their collaborations in terms of both when to manipulate the diagrams and what to discuss during such manipulations. A parallel drawing to this observation is Olson et al. (1993)'s work on analysing how shared representations support small group design meetings. In their work, the addition of a visual shared tool to co-located group collaborations had an effect on how the groups went about organising and structuring their work. Olson et al. (1993) report on how shared access to the tool

[2] A detailed example of this interaction is describe later in the chapter in Table 8.9 on page 207.

[3] A detailed example of this interaction is describe later in the chapter in Table 8.12 on page 210.

[4] Examples showing contrasts in the use of audio to exchange awareness information in each condition are described in more detail in Section 8.5.2.2 on page 209 of this chapter.

changed the focus of the groups' discussions, how they allocated times to different activities, as well as made the groups more homogeneous in their work. Study 3 also provided collaborating pairs with equal shared access to an audio-only tool in both conditions. However, changes to pairs discussions and transitions between activities occurred when both the means for delivering audio to the workspace and the pairs choice of working style changed. This raises the question of what it really means to *share* access to a workspace in an audio-only setting. The next section examines the impact of working styles on the collaborations.

8.2.3.2 The impact of working style

The impact of the means for delivering audio on the time distributions was further apparent when pairs were retrospectively separated into *Sequential* and *Parallel* working style groups and data from each group was considered independently.

The observation that collaborating pairs use different working styles to complete joint tasks is inline with previous work on audio-enhanced collaboration. For example, Morris et al. (2004) describe a change from a "divide-and-conquer" strategy when co-located users manipulated visual artefacts on a tabletop and received audio feedback through headphones, to a serial strategy when the feedback was delivered through shared speakers. The parallel and sequential working styles reported in Study 3 were not exclusive to a particular condition, however. What is evident from the results reported in Chapter 7 is that the choice of working style had an impact on how pairs went about their work in both conditions. McGookin and Brewster (2007b) report a similar result, where pairs of participants used either a divide-and-conquer or a turn-taking strategy to complete audio-haptic graph-building tasks, and where the choice of strategy influenced participants perceived advantage of having a publicly shared audio representation of their collaboration.

There are a number of reasons that could have influenced participants' choice to work in parallel or in sequence in Study 3. Having two sets of speakers emitting audio at the same time in the same place can be distracting or confusing[5]. Some participants might have been more susceptible to such distraction or confusion and subconsciously chose a working style that would eliminate such inconveniences. It could also be the case that participants consciously changed their working style so as not to cause disturbance during a construction episode, or simply preferred to work jointly rather than independently. Regardless of whether participants' choice of working style was a conscious or a subconscious decision, what is of interest – besides the fact

[5] Note that volume levels of the audio output in both conditions were adjusted to match participants' preferences, with the exception that, in the Shared condition, both partners in a pair could clearly hear each other's audio.

that a change of style did occur – is that it had an evident impact on the way pairs organised their collaborations. Pairs who worked sequentially when using speakers spent significantly more time discussing diagram content than pairs who worked in parallel in the same condition. When such discussions occurred while one or both participants simultaneously interacted with the diagram, sequential pairs' interactions matched the discussed content 95% of the times. This suggests that using speakers does not of itself encourage matching interactions and discussions about diagram content unless pairs "use" the publicly available audio in a manner that increases this likelihood. Further, sequential pairs spent significantly more time discussing execution than parallel pairs did; a difference that can also be explained by the choice of working style. When only one participant uses the tool at a time, there is only one audio output source in the environment that both participants can attend to. This leaves room to discuss not only content that is relevant to what is auditorally displayed, but also to exchange tips on how to execute the appropriate editing actions that manipulate such content. Participants in a sequential pair were thus more likely to have joint attention over the various aspects of the interactions surrounding the construction of a diagram.

Since each participant in a parallel pair executed independent editing actions, it was unlikely that discussions about the detailed step by step execution of such actions occur. This is also likely to be because there was less elements that could potentially trigger such discussions; the audio feedback of individual actions was only available to the participant executing the action, either through headphones or because the focus of each participant was directed away from their partner's speakers and towards their own. Participants in parallel pairs were thus less likely to have joint attention over their interactions.

Design Lesson 13 – Provide collaborators with both a private and a public workspace area and the ability to switch between them. Collaborators should be able to work privately, such that they control whether their partners can hear their output, but also be able to expose their audio to their partner, for example, to support guidance interaction. If the collaboration involves more than two people, then collaborators should be provided with a means to select a workspace to switch to from available users, or invite specific users to access their workspace.

8.3 Awareness of Contributions

Scores on the annotation task measured participants ability to differentiate their contributions to the constructed diagrams from their partner's. Hypothesis H6 predicted that scores on such a task will be higher when audio is delivered through speakers because, under such condition, the auditory output of individual editing actions are publicly displayed, and hence could be attended to by both partners. However, contrary to what was expected, there was no significant difference between participants' scores on the annotation task in the two conditions. Scores from the two working style groups yielded different results, however. While there was still no significant difference between the two conditions for parallel pairs, sequential pairs scored significantly higher when audio was delivered through headphones. According to this result, what had an effect on participants ability to accurately differentiate their contributions from their partner's seems to be the combination of working style and the means for delivering audio.

Parallel pairs. The proportions of times that parallel pairs dedicated to discussing division of labour was 2.1% in the Shared condition and 2.5% in the Non-Shared condition. The difference between these proportions was not statistically significant and transitions from and towards the discussions about division of labour were also equally varied across the two conditions[6], which is likely to have contributed to the non-significant difference in their scores on the annotation task across the two conditions.

Hancock et al. (2005) use a similar method for assessing awareness of individual and group contributions to shared tasks where audio is delivered to three tabletop collaborators through localised speakers (adjacent to each user, i.e. similar to the setup in the Shared condition) or a single speaker outputting audio emitted from all collaborators. Their results contrast those reported here, however, showing that users who focus on their own task (equivalent to a parallel working style) are less likely to be aware of the overall group performance. This contrast can be related to the number of collaborators in a given group (three versus two in Study 3), but also to the type of the post-test task used in Hancock et al. (2005)'s study and that used in Study 3. Participants in Study 3 were asked to identify contributions on the final product of the collaborations (i.e the diagram), whereas those in Hancock et al. (2005)'s study were asked to recall events that occurred during the course of the collaborations and which did not form part of the final product (e.g. how many error were made by each group member during the collaborations). Additionally, while Study 3 examined the effect of using speakers and headphones on awareness of contributions, Hancock et al. (2005) did not consider the use of headphones at all, postulating that their use *"may not allow for improved group awareness"*(p.46), a claim that is also contested

[6] Refer back to Figure 7.15 on page 168 for an illustration of parallel pairs transitions between activities.

by the results obtained from Study 3.

Sequential pairs. Not only did sequential pairs score significantly less on the annotation task when they used speakers than they did when they used headphones, but they also scored significantly less than pairs who used a parallel working style in the Shared condition. A close examination of sequential pairs collaborative organisation can explain these differences. Sequential pairs dedicated more time to discussing labour in the Non-Shared condition (3.5%) than they did in the Shared condition (2.4%), and the difference was statistically significant for the *"Labour-WI"* activity, i.e. dividing labour while one or both participants concurrently interacted with the diagram. Thus, the significant difference in the times allocated to discussing labour could have contributed to the significant increase in awareness of individual contributions in the Non-Shared condition. Furthermore, Sequential pairs movements from and towards the *"Labour-NI"* activity were more varied in the Non-Shared condition[7]. Thus, sequential pairs frequently interrupted their manipulations of a diagram, reading of the descriptions and discussions about diagram content with discussions about division of labour when they used headphones. The frequent interruptions of the main collaborative activities is likely to have contributed to emphasising the division of labour at various points during the collaboration, and might have therefore led to sequential pairs increased awareness of contributions in the Non-Shared condition.

Parallel versus sequential pairs. But why is it that pairs working in parallel scored significantly higher than pairs working sequentially under the same condition? A closer examination of how each working style group organised their collaborations revealed that parallel pairs typically agreed on task assignments then concurrently executed them. This meant that, within each construction episode, each participant in a parallel pair could focus on an individual editing action independently from the activity that their partner was engaged in. On the other hand, after agreeing on a task assignment, both participants in a sequential pair would jointly focus on the step by step execution of that particular task. Thus, by the end of a construction process, both participants in a sequential pair would have focused on the detailed execution of most, if not all, editing actions that occurred during the collaborations. In contrast, each participant in a parallel pair would have focused on only a subset of the total editing actions that occurred during the collaboration – those actions assigned to them individually. The level of joint attention between participants on editing actions during the construction process was therefore higher when pairs

[7] Refer back to Figure 7.20 on page 173 for an illustration of sequential pairs transitions between activities.

used a sequential working style, and the emphasis during such joint attention seemed to be on the actions themselves rather than who was performing them.

A further characteristic that could have contributed to this outcome is related to the organisation of sequential and parallel collaborations. Parallel pairs' transitions from the *"Labour-NI"* activity were more varied than sequential pairs' transitions from that same activity[8]. This means that parallel pairs' frequently interrupted the main collaborative activities with discussions about labour, appearing at points during discussions about diagram content and during "silent" manipulations of the diagram and consultations of the textual descriptions. For sequential pairs, movements from the *"Labour-NI"* activity were less varied, mainly proceeding discussions about diagram content when interactively idle and preceding "silent" diagram manipulations and consultations of the textual descriptions. This difference in transition variability is an indication that parallel pairs discussed labour at more frequent points during the collaborations than sequential pairs did, which could have enforced their knowledge of self and partner's contributions throughout the collaborations and consequently led to significantly higher scores on the annotation task.

Design Lesson 14 – Provide a log of collaborators' contributions to the shared task. Although the ability to keep track of self and partner's contributions to the task was not a factor that could impact performance on or the outcome of the studied task, it is clear that both the choice of working style and the means for delivering audio to the workspace had an impact on such ability. In situations where such ability is crucial to the task at hand, the collaborative tool should be designed to convey awareness information about which action took place and by which user, i.e. of types *"What Did You Do"/ "What I Did"* or allow users to requested such information, for example, by browsing a timed log of actions.

8.4 Workspace Awareness Information Exchange

Hypothesis H7, which predicted that participants would exchange more WA information in the Non-Shared condition was supported, both when considering data from all pairs and from each working style group independently. When audio was delivered through headphones, information about partner's actions, intentions and locations was not explicitly available in the workspace environment, requiring participants to supply and/or request such information from their partners.

[8] Refer back to Figure 7.26 on page 178 for an illustration of parallel and sequential pairs transitions between activities in the Shared condition.

8.4.1 The Need for Awareness Information

In both conditions, participants supplied significantly more WA information to each other than they requested from one another. First, these proportions show that WA information exchanges did occur during the collaborations. That is, participants needed to know about their partner's actions, intentions and locations when they constructed the diagrams, and evidently felt the need to tell their partner about their own actions, intentions and locations. In most cases (70% to 75%) this need was fulfilled when participants supplied information to their partners without being prompted to do so. When this was not the case (the remaining 25% to 30%), participants expressed explicit requests to their partners to provide them with the information they felt they needed.

The proportions of supplied information in the Non-Shared condition was significantly higher than in the Shared condition. This means that, in general, participants in Study 3 felt a greater need to let their partners know about their own actions, intentions and locations on the diagram when they used headphones. On the other hand, there was no significant difference in the proportions of requested information across the two condition, which means that the need to know about a partner's actions, intentions and locations was similar when using speakers and when using headphones. But not all types of supplied information were equally needed. The proportions of only three out of the five elements of the supplied types of WA information (*"What I Did"*, *"What I Am Doing"* and *"Supplied Completion Status"*) were significantly higher when audio was delivered through headphones. Differences between the proportions of the remaining elements were not statistically significant. It could therefore be argued that the availability of the audio in the workspace environment through speakers afforded the deliverance of these three types of information. Not surprisingly, sharing audio through speakers provided participants with the ability to obtain information about their partner's past and current actions, and about the progress of those actions without them having to explicitly ask for it.

Design Lesson 15 – Types of workspace awareness information when using headphones. If headphones are used during collaboration, the collaborative system should be design to convey workspace awareness information of types *"What I Did"/"What Did You Do"*, *"What I Am Doing"/"What Are You Doing"* and actions *"Completion Status"*. Alternatively, the system should be designed to allow users to manually request such information.

8.4.2 Impact of Working Styles

The use of the affordances described above varied across the two working style groups. In the Shared condition, the differences in the proportions of WA information supplied and requested by parallel pairs were not statistically significant. Thus, parallel pairs' need to supply WA information and to request it was equivalent when audio was delivered through speakers (58% supplied and 48% requested). On the other hand, pairs working sequentially supplied more information than they requested in the Shared condition (84% supplied and 16% requested). Here too, the choice of working style had an effect on the rates of WA information exchange. Parallel pairs worked concurrently and as a result were unable to keep track of each other's actions, intentions and locations, therefore requesting far more WA information than sequential pairs did when using speakers. Sequential pairs had joint attention over the execution of editing actions throughout most of the construction process, and so did not feel a great need to request WA information from one another.

Table 8.1: Extract from pair 16's collaboration in the Shared condition – Speak-aloud utterances: supplying awareness information in the form of description of actions; <user action>; *(non-speech sound)*; *"system speech output"*.

Time	P16A	P16A's Audio	P16B	P16B's Audio
27:49	<down>	*(down)"journey"*		
27:51	Journey <down>	*(down)"driver"*		
27:52	Driver			
27:53	**i'm gonna select Driver**			
27:55	control S			
27:56			control shift S	
27:57	control shift S <select>			
27:58		*(success)"driver selected"*		
27:59	**Driver selected now**			
28:00	**i wanna select Stage**			
28:01	<down>	*(down)"garage"*	yeah	
28:02	<down>	*(down)"bus"*		
28:03	<down>	*(down)"stage"*		
28:04	Stage control..			
28:05	..shift S <select>	*(success)"stage selected"*		
28:06	ok now go to			
28:07	to Relations		<down>	*(down)"relations"*

An examination of the transcripts showed that, in some cases, the proportions of WA information that was supplied by sequential pairs in the Shared condition were in fact speak-aloud utterances. Table 8.1 shows an extract from a sequential pair's collaboration in the Shared condition illustrating this behaviour. In this extract, pair 16 are creating a relation between two entities "Journey" and "Stage" having divided their labour so that participant P16A locates and selects the entities to be related, and participants P16B creates the relation, names it and edits its cardinalities. P16A moves around the hierarchy describing what they are doing and what they intend to do at every step. Meanwhile P16B remains interactively idle, follows the selection process, acknowledges actions (at minute 28:01), and provides his partner with information about the correct system commands to use (at minute 27:56). Both participants are thus jointly engaged in the process of editing this relation, and using "outlouds" (Heath and Luff, 1992; Gutwin and Greenberg, 2002) to intentionally communicate workspace awareness information to one another.

Table 8.2: Extract from pair 1's collaboration – Supplying awareness information in the form of updates; [non interactive user action]; *(editing)* is a continuous ambient sound reflecting the mode of interaction.

Time	P1A	P1A's Audio	P1B	P1B's Audio
35:04	<down>	*(down)"journal"*	<down>	*(down)"article"*
35:05			Article	
35:06	<down>	*(down)"number"*		
35:08	<set primary>	*(success)"attribute*		
35:09		*number set as primary"*	<set cardinality>	*(editing)"choose*
35:10		*"attributes open"*		*(editing)cardinality"*
35:11	**i've done the err**		<press one>	*(editing)"one selected"*
35:12	err			
35:13	[reading]		<press M>	*(editing)"many selected"*
35:16			<set>	*"cardinality many*
35:17	**i've finished the err**		relations?	*added"*
35:18	**the entities**			
35:21			<down>	*(down)"article"*
35:22			<down>	*(down)"gene'*
35:23	<close>	*(close)"attributes"*	**ok i'm finishing**	
35:24			**the last relation**	

Consider, in contrast, the interaction extract in Table 8.2. Here, pair 1 are finalising their diagram; participant P1A setting an attribute of one of the entities as a primary key, and participant P1B adding cardinalities to a relation. The participants in this pair are working in parallel, and updating each other about the editing actions that they have completed. Each participant is engaged with their own task, and unless an update is supplied, the participants remain unaware of each other's progress. In fact, the pair in this example are so disengaged and unaware of one another's progress that when participant P1A updates his partner at minute 35:11 with an unfinished utterance: *"i've finished the err"*, P1B completes the utterance with: *"relations?"*, which is far from a correct assumption about what his partner has been up to.

Supplying awareness information while both participants are jointly engaged with one task is different from supplying it when each of them is engaged with an independent task. In the former case, exemplified in Table 8.1, the information is supplied in a form of descriptions of what is currently occurring or what is about to occur in the immediate future (i.e. information of types *"What I Am Doing"* and *"What I Will Do"*). Note that these descriptions are supplied while expecting the partner to be listening to one's audio output. In contrast, information in the latter case is mostly supplied in a form of updates (i.e. information of type *"What I Did"*). Furthermore, information about participants' actions and locations in the former case is duplicated by their verbal descriptions; i.e. the participant repeats what the system displays in audio. Thus, even if not verbally described, such information could potentially be retrieved from the workspace if a participant listens to a partner's audio. When working in parallel on the other hand, each participant is engaged with their own task, and the information about actions and locations disappears from the environment as soon as auditorally displayed. Unless supplied by verbal descriptions, such information is not available and is thus likely to be requested if the need for it arises.

Indeed, parallel pairs requested significantly more WA information of type *"What Did You Do"* than sequential pairs did in the Shared condition. Parallel pairs therefore felt a greater need to find out about each other's past actions, which is again a direct consequence of the working style. A participant in a parallel pair worked concurrently with, and did not follow their partner's progress, yet such information is needed to coordinate future editing actions. The need to know about what a partner has done in such cases increased, which led to a significant increase in the requests of the *"What Did You Do"* type. Participants working sequentially did not have as much a need to request WA information from their partner, but this was not because partners supplied the needed information, unprompted, at higher proportions than partners in parallel pairs did. There was, in fact, no significant difference in the overall proportions of supplied information between parallel and sequential pairs in the Shared condition. It is therefore more likely that sequential pairs' joint attention with the execution of editing actions has led to decreasing the need to request WA information. Sequential pairs were thus engaged in what could be described as close and focussed collaboration (Gaver, 1991; Dourish and Bellotti, 1992), characterised by intense levels of awareness and manifested in Study 3 by a significant decrease in the exchange of WA information. In contrast, parallel pairs collaborations were, to use Dourish and Bellotti (1992)'s term, loose, characterised by the focus on individual rather than joint work and manifested by the significant increase in the exchange of WA information.

> ***Design Lesson 16 – Adapt both the type and amount of workspace awareness information to match collaborators' working style.*** When collaborative interactions occur in parallel, the system should provide users with updates about each others actions, i.e. emphasising information of types *"What I Did"/"What Did You Do"* and actions *"Completion Status"*. The collaborative system should also be designed to detect sequential interaction and reduce the amount of workspace awareness information it conveys to users, since this information can be redundant in such a case. Alternatively, the system should provide users with a means to control the amount of awareness information that is conveyed to match their needs.

8.5 Workspace Awareness Through Auditory Display

The measurements used to account for the types and rates of workspace awareness information exchange in the two conditions highlighted the affordances of the Shared and Non-Shared audio-only settings. Significantly higher rates of exchanging a given type of awareness information in the Non-Shared condition could be considered an indication that delivering audio through speakers made such information accessibly available. The following examines two further points; 1) from which sounding elements in the audio-only workspace did the participants extract such WA information, and 2) how did they use it in their collaborations.

As previously described, the WA information that was captured from the video transcripts was related to participants' actions, intentions and locations. Actions and intentions captured supplied and requested information about past, current and future editing and inspection activities, whereas locations captured supplied and requested information about a participant's "physical" position on the diagram or the hierarchy.

8.5.1 Extracting Workspace Awareness Information from Sounds

When audio was delivered through speakers, information about partners' actions and locations could potentially be extracted from the environment if a participant attentively listened to their partner's audio output. Since such output was not available in the Non-Shared condition, the alternative there would be for a partner to verbally describe the details of their activity. When pairs worked sequentially in the Shared condition, there was only one set of speakers outputting audio at any given moment, and so extracting information about partner's actions could be done by paying attention to and interpreting the information displayed through this output. When pairs worked in parallel, participants needed to somehow pay attention to both their audio output source and their partner's, or switch their attention between the two if they were to be able to extract detailed information about their partner's actions and locations.

Truax (2001) describes three types of listening attention to account for various listening experiences. An active level of listening, *listening-in-search*, involves a conscious search in the environment for significant cues. This type of listening provides a means to extract specific pieces of information without the need to consider other auditory information that might also be present in the environment but is irrelevant to the task at hand. An example of this is the so called "cocktail party effect" (Handel, 1989; Arons, 1992), which describes the ability to focus on one sound and ignore others in noisy environments. Attention in such instances can be controlled at will, voluntarily directed to a particular sound or set of sounds to extract meaning from it. In other instances, attention to sounds can be involuntary. Truax (2001) uses the term *listening-in-readiness* to refer to an intermediate level of listening where *"the attention is in readiness to receive significant information, but where the focus of one's attention is probably directed elsewhere"*(p.22). A sound in such a case can fade in the listener's background attention but grabs it if something about it is deemed significant (including if it disappears from the environment). In the third type, *background listening*, the occurrence of a sound bears no special or immediate significant to the listener, and might as a result remain in the listener's background attention. In such instances a listener may still be aware of a sound and *"if asked whether [they] have heard it, [they] could probably respond affirmatively, as long as the event were not too distant in the past"* (Truax, 2001, p.24). The following will refer to these different types of listening to examine how information about workspace awareness was extracted from the sounds during the audio-only collaborations in Study 3.

8.5.1.1 Extracting information about actions

When a participant interacted with the collaborative tool, they were either inspecting or editing the diagrams. The tool displayed a mixture of speech and non-speech sounds to assist interaction in the two modes. Two auditory display techniques were used to distinguish between the inspection and editing modes. The first was the timbre of an ambient sound that was continuously displayed while a user was in a given mode of interaction; the second is the gender of the speech display that alternates between a male voice for the inspection mode and a female voice for the editing mode. To know about current actions, a participant would listen out for either one or both of these audio clues and to the detailed of the spoken information. Depending on how much attention a participant is paying to the audio output, or on how much an audio output grabs the attention of a participant, different information is extracted from such output.

Table 8.3: Extract from pair 6's collaboration – Audio output grabbing attention – listening in readiness.

Time	P6A	P6A's Audio	P6B	P6B's Audio
04:47			<new att.>	*(editing)"new attribute*
04:48			<typing>	*(editing) enter name"*
04:49	<open>	*(open)*	<typing>	
04:53	<down>	*(down)"attributes"*		*(success)"attribute*
04:55	<open>	*(open)"two"*		*capacapacity added*
04:56	<down>	*(down)"name"*		*"to entity bus"*
04:58	<down>	***(down) "capacapacity"***		*"bus"*
04:59	[laughs]		[laughs]	

Table 8.3 shows an extract from pair 6's collaboration, and illustrates how a participant's attention is grabbed by an auditory event that is displayed through their partner's audio output. In this extract, pair 6 are working in parallel in the Shared condition; participant P6B is adding an attribute to an entity, while participant P6A is, incidentally, browsing the part of the diagram currently being edited by P6B. At minute 4:47, P6B issues the command to add the attribute "Capacity" to the entity "Bus" but misspells the name of the attribute and types "Capacapacity" instead. His partner, P6A, then encounters the newly edited attributes and the misspelled attribute name is displayed in speech at minute 4:58. The unusual pronunciation of the name as spoken by the speech synthesiser and displayed from P6A's output grabs P6B's attention and both participants burst with laughter. In the case of P6A, the unusual pronunciation was displayed from

his output source as a result of his inspection action, it is therefore natural that he was attentively listening to it. On the other hand, P6B's exhibits a listening-in-readiness behaviour; his attention was grabbed by the "funny" speech output even though he was engaged in a separate task and his audio output was concurrently displaying spoken information about his current position on the hierarchy (the entity "Bus", at minute 4:58).

Table 8.4: Extract from pair 6's collaboration – Extracting detailed information about partner's actions from their audio output source – attentive listening, listening-in-search.

Time	**P6A**	**P6A's Audio**	**P6B**	**P6B's Audio**
05:10			<down>	*"capacapacity"*
05:12			<delete>	
05:13				*"capacapacity removed"*
05:18			<new attr.>	*(editing)"new attribute*
05:19				*(editing) enter name"*
05:20			[typing]	
05:24				*(success)"attribute capacity*
05:26	nice one you			*added to entity bus"*
05:27	deleted Capacapacity			

In the next extract, shown in Table 8.4, participant P6B rectifies the spelling mistake by removing the attribute "Capacapacity" and replacing it with a correctly spelled one. Notice how this time around, P6A partner halts his interaction altogether and remains interactively idle while P6B completes the editing action. Here, P6A switches to a listening-in-search mode and, as soon as the action is completed, he acknowledges it by saying: *"Nice one, you deleted Capacapacity"*. The presence of the audio output in the Shared condition thus communicates partner's activity information at two levels. At an initial level, detecting that audio output is coming out of a partner's speakers enables a user to establish that their partner is actively interacting with the diagram. At the second level, the details of the audio output allows them to extract detailed information about what activity their partner is engaged in. Each level of information communication requires a different level of attention, and thus a different type of listening.

The first level of information can also be used to establish that a partner is not active when no audio output is detected from their output source. The extract in Table 8.5 illustrates an example where the lack of audio is used in this manner. In this example, pair 8 are working sequentially in the Shared condition and are trying to collectively interpret the sound of a relation's cardinality [9]. Participant P8A inspects the hierarchy to display the sound in question while participant P8B listens. Both participants attempt to interpret the sound, initially misinterpreting it for a "One" sound, before P8B realises the mistake and explains to her partner why they made such mistake at minute 33:25: *"wait it's once only the Gene [spoken sound] is vibrating [giving the illusion of multiple taps] not the one [sound]"*. P6B then suggests that the pair collectively listen to the sound in question again, emphasising: *"listen properly"*. At this point, both participants are listening-in-search, expecting the partner to inspect the part of the hierarchy that contains the information and display the sound in question.

Table 8.5: Extract from pair 8's collaboration – Determining inactivity from the lack of sound.

Time	P8A	P8A's Audio	P8B	P8B's Audio
33:14	there is a sound			
33:15	for once or was it			
33:16	more than once?			
33:16	<down>	*(one)"therapy"*		
33:17			it's more than one	
33:18	<down>	*(one)"gene"*		
33:19			no?	
33:20	<up>	*(one)"therapy"*	[leans forward]	
33:21	<down>	*(one)"gene"*		
33:22	yeah			
33:25			wait it's once only	
33:26			the Gene is vibrating	
33:28			not the one	
33:30			listen properly	
...				
33:34			the Gene part	
33:35	yeah play it			

Both participants in pair 8 wait quietly for the next four seconds while, presumably, waiting for the partner to display the sound. P6B then, assuming that her partner did not understand which sound she meant for them to display, provides him with a clarification: *"the Gene part"*, to which the partner, who was also expecting P6B to play the sound, responds: *"yeah, play it"*.

[9] Recall that the cardinality "One" is displayed by the sound of a single tap, while the cardinality "Many" is displayed by the sound of multiple taps.

The part of interest here is the four seconds between minute 33:30 and 33:34, during which both participants expected the other to act. Because this was an audio-only working environment – participants could not see each other – the only source of information for determining the level of a partner's activity is how much sound they make. There was clearly a misunderstanding with relation to who should act next in this sequence, and the lack of audio output communicated a lack of activity, subsequently triggering participants to rectify the mis-coordination. Using the lack of audio output as a clue to determine the level of partner's activity is not possible in the Non-Shared condition, where a user had access to their audio output through headphones.

Table 8.6: Extract from pair 8's collaboration – Determining partner's location on the diagram based on context ambient sounds.

Time	**P8A**	**P8A's Audio**	**P8B**	**P8B's Audio**
16:21		***(editing)****"relating patient*		
16:22		*and tag"*		
16:24		*"enter new relation name"*		
16:26			ok enter new relation	
16:27			name	
16:29			Associate? i can't think	
16:30			of any other name	
16:32	ok then [typing]			
16:33			Associate?	
16:34	yeah fine			
16:35		*(success)"relation*		
16:36		*associate added"*		
16:37	err	*"two relations"*		
16:38		***(inspecting relations)***		
16:40			we are in Relations now	
16:41			go to Associate	
16:42	<down>	*(down)"associate"*		

8.5.1.2 Extracting Information about Locations

Two auditory display techniques were used to communicate information about a user's position on the hierarchy. The first is the the non-speech continuous ambient sounds that distinguish between interaction modes and hierarchical perspectives; as described by most participants in Study 3, exploring the "Entities" perspective was accompanied by a "bubbly" ambient sound, while exploring the "Relations" perspective was accompanied by a "hissing" ambient sound. A third continuous ambient sound dominated the display when in the editing mode. The second is the mixture of speech and non-speech sounds that display the content of the nodes on the hierarchy. Detecting and recognising the timbre of the ambient sounds could potentially help a user infer which branch on the hierarchy their partner is on, whereas listening to the details of the displayed speech and non-speech sounds enables them to establish details about their exact position.

Table 8.6 is an example where a participant uses the audio displayed form their partner's output to infer their position. In this extract, pair 8 are working sequentially in the Shared condition to create a relation. Participant P8A is manipulating the diagram and his partner P8B is following his interaction while interactively idle[10]. Participant P8B issues the command to create the relation at minute 16:21 and a system prompt for inputting a name to the new relation is displayed in speech. The partner, P8B repeats the displayed speech output and suggests a potential name to input for the new relation.

Participant P8B was listening-in-search, therefore able to correctly establish knowledge about the progress of her participant's action, correctly inferring what next step was required to complete the editing action, and assisting her partner with a potential input information. The mixture of the spoken message and the ambient editing sound communicated position information in this case, but it is not clear exactly which of the two sets of sounds assisted P8B in detecting her partner's position – though repeating the system speech output could be considered an indication that she was attentively listening to what was spoken. When a relation is successfully created, the collaborative tool automatically switches the hierarchy perspective to display the "Relations" view, subsequently triggering a switch in the continuous ambient sound to display the "hissing" sound. Participant P8B picks on this switch at minute 16:40 in the extract and makes the correct inference that: *"we are in Relations now"*. Having been attentively listening and following P8A's interaction, participant P8B was able to correctly interpret the audio output to determine the position of her partner without being explicitly supplied with such information. The inference in this case can be said to be a result of two activities; the first is extracting correct information from the audio displayed through the partner's audio output; the second is correctly matching

[10] Recall that the process of creating a relation using the audio-only collaborative tool involves three steps; first, the entities to be related must be selected from the list of existing entities, then a relation is created and given a name, and finally, the cardinalities are added to the relation. The extract on Table 8.6 shows the second step of this process.

such information to one's own mental model of the structure of the interaction (in terms of the steps required to execute a given action) and the structure of the hierarchy.

Table 8.7: Extract from pair 13's collaboration – Determining partner's location on the diagram based on their audio output and one's own mental model of the hierarchy.

Time	P13A	P13A's Audio	P13B	P13B's Audio
04:43			<down>	*(down)"design"*
04:44	so a Tag can be used			
04:45	by many Doctors			
04:46	for many Therapies			
04:47			<open>	*(open)"three"*
04:48			<down>	***(down)"therapy"***
04:50	Therapy		erm	
04:51	ok stay on Therapy		yeah	

Table 8.7 is another illustration of using audio output together with one's own mental model of the hierarchy to determine partner's position. In this extract, pair 13 are working sequentially in the Shared condition, and, having successfully created a relation, the pair are in the process of discussing the suitable cardinalities that should be added onto it. Participant P13B manipulates the diagram while participant P13A is interactively idle. While P13A is describing the relationship between the related entities, her partner browses the relation in question reaching one of its entities "Therapy" at minute 4:48. P13A hears the displayed spoken name of this entity and immediately asks her partner to remain in that position: *"Therapy, ok stay on Therapy"*. P13A in this case has correctly inferred the position of her partner and, by matching such information to her own understanding of the hierarchical structure, inferred that this position is the right position to be at in order to execute the next step in the current editing action; i.e. suggesting that her partner stays on that position while they reason through the content to establish the appropriate cardinality to add to "Therapy".

8.5.1.3 Extracting Information about Intentions

The collaborative tool did not provide the user with a means to communicate information about their future actions, or intentions, through an auditory display. Information about intentions was therefore only explicitly present in the actors mind, and it was natural that it could only be made available when explicitly requested from and/or verbally supplied by the actor. It could be argued, however, that establishing knowledge about planned actions through a “What I Will Do” supplied WA information might help a participant follow their partner and extract information about future actions based on knowledge about current ones. Indeed, competent actors in a given collaborative task often use their expertise to anticipate colleagues actions (Heath and Luff, 1992; Schmidt and Simone, 2000). To do this, knowledge about activity progress could be established by retrieving information about current actions through similar means as those described above. For example, if a participant hears their partner selecting entities (either through the audio display or their partner’s speak-aloud) they can infer that the partner is intending to create a relation.

The above examples clearly show that the principal means for extracting WA information is listening – at various attention levels – to the sounds made and triggered by partners. The described extracts are thus examples of what is referred to as *Artefact Feedthroughs* (Dix et al., 1998; Gutwin and Greenberg, 2002), where the manipulation of an artefact communicates information about what actions incured changes to its state, and *Outlouds* (Heath and Luff, 1992), where verbal communication is intentionally and explicitly exchanged. Gutwin and Greenberg (2002) highlighted such means as one of the mechanisms for gathering workspace awareness information[11], and in the case of audio-only collaboration, sharing audio in the workspace, by attending to it, becomes the principle source for such information. Here too, the recommendations outlined in *Design Lessons 13* and *16* above should apply. That is, audio-only collaborative systems should be design to provide users with both private and public workspaces and the ability to switch between them. Additionally, the collaborative system should be designed to detect when users are sharing the audio in the same workspace and adapt the amount of workspace awareness information it conveys accordingly.

8.5.2 Using the Extracted Information in Audio-only Collaboration

In general, participants used the WA information extracted from the audio output of their partners interactions in the Shared condition in a variety of ways including; 1) to detect errors, 2) to provide guidance to their partners, and 3) to coordinate collaborative actions.

8.5.2.1 Detecting Errors

Invalid inspection commands were displayed with a non-speech error sound, while invalid editing commands were displayed with the same error sound accompanied with a spoken message de-

[11] Refer back to Section 6.4.2 on page 134 for a full description of the mechanisms for gathering workspace awareness information in a collaborative environment.

scribing the cause of the error. Participants made a variety of errors when using the collaborative tool and the auditory display of such errors enabled their partners to detect them. Such awareness typically led to providing assistance and explanations, suggesting solutions, or overtaking the executing of the task that is posing difficulty to a partner.

Table 8.8: Extract from pair 8's collaboration – Detecting an interaction error through the error sound.

Time	**P8A**	**P8A's Audio**	**P8B**	**P8B's Audio**
17:21	<down>	*(down)"Associate"*		
17:22	<open>	*(open)"two"*		
17:23	<open>	***(error)***		
17:24	<open>	***(error)***		
17:25			what's the problem?	

In the extract on Table 8.8 for instance, participant P8B was able to identify that her partner is encountering a problem after hearing the error sound displayed twice in a row, this prompted her to enquire: *"What's the problem?"*. Her partner had encountered an *interaction error* having issued an invalid inspection command. Similarly, participant P16A in the extract in Table 8.9 deletes an entity when he intended to create a new one. In this case the error was not a result of issuing an invalid command, but of issuing an unintended command. His partner heard the outcome of the editing action and realised that the action was not the desired one and explains: *"you pressed control E, should've pressed command E"*. This is an *execution error*, and was detected through interpreting the spoken output of an editing action.

Table 8.9: Extract from pair 16's collaboration – Detecting an execution error through speech output.

Time	**P16A**	**P16A's Audio**	**P16B**	**P16B's Audio**
3:49	so now i add an entity			
3:52	<delete>	***(success)"entity driver***		
3:53		***removed"***		
3:54	ah		you pressed ctrl E	
3:55			should've pressed cmd E	

A third kind of errors detected by participants in Study 3 was a *procedural error*, in which a participant mixes the order of the steps required to execute a given editing action. The extract in Table 8.10 shows an example of this. Participant P8A intended to create a relation between two entities, but to do that she first needed to locate and select the entities to be related on the Entities branch of the hierarchy. Instead, P8A switches to the Relations perspective. Her partner, participant P8B, detects this move and immediately rectifies it by instructing her with the right procedure: *"go to entities first"*. P8B was able to detect such an error while listening to his partner's audio output which displayed a non-speech "switch" sound accompanied by a spoken description of the destination branch.

Table 8.10: Extract from pair 9's collaboration – Detecting a procedural error through the speech and non-speech audio output.

Time	**P9A**	**P9A's Audio**	**P9B**	**P9B's Audio**
58:31			one route	
58:32			many stages	
58:34	ok			
58:36			go for it	
58:37	<switch>	***(switch)"relations"***		
58:38			go to entities first	

Table 8.11: Extract from pair 11's collaboration – Detecting a content error through the speech and non-speech audio output.

Time	**P11A**	**P11A's Audio**	**P11B**	**P11B's Audio**
33:49			<new entity> so bus	*(editing)"new entity*
33:50			has a make	*(editing) enter name"*
33:51			[typing]	
33:52				***(success)"entity make***
33:53				***added"***
33:54	wait wouldn't that			***"three entities"***
33:55	be entity Bus and			
33:56	then attribute Make?			

A final kind of errors that participants detected from the auditory display of their partners' interactions is a *content error*. In such a case, the partner would have issued the correct command in the correct procedural order, but fails to input the correct content on the diagram. The extract in Table 8.11 exemplifies the occurrence of this type of error. Participant P11B of pair 11 had came across the description of an entity and initiated an editing action to add it on the diagram.

P11B proceeds to create an entity with the name *"Make"*. His partner detects the completion of this editing action and responds by stating the correct content that should have been added onto the diagram: *"wait wouldn't that be entity Bus and then attribute Make?"*. Having listened attentively to their partner's actions, P11A was able to swiftly pick up on the content error and to suggest rectifications.

In the above examples, participants were able to extract relevant information about partners' progress and use such information to provide them with appropriate assistance without being requested to do so. This was highlighted by Gutwin and Greenberg (2002) as one of the activities in which workspace awareness plays an important role. Participants in Study 3 were thus able to provide timely assistance while relying on audio as the only means for doing so.

8.5.2.2 Guiding Partners

Participants also used the WA information extracted from the auditory display to provide their partners with detailed guidance. This included guiding their movements around the hierarchy, as well as through the steps required for executing editing actions. The extract in Table 8.12 illustrates an example where a participant guided their partner's movements on the hierarchy. Participant P8B in pair 8 was assigned the task of setting the attribute "Number" of entity "Article" as a primary key and is guided by P8A to complete this task. At every step of the interaction, P8A listens attentively to his partner's auditory output to obtain information about her current location, and instructs her movements accordingly (*"go inside", "come down"*). When he realises that she had reached the appropriate position on the hierarchy, he instructs her to issue the corresponding editing command for setting the primary key (*"cmd P"*).

Notice how the participant executing the editing action provided her partner with little to no explicit verbal clues about her actions or locations. This is is an indication that she was expecting him to be attentively listening to her interaction, and so did not feel the need to supply him with any WA information. The participant doing the guidance, in turn, did not request any information form his partner; he was indeed able to collect the needed information directly from the environment, and only needed to match the obtained information to his own mental model of the interaction and of the structure of the hierarchy in order to provide corresponding, accurate, instructions that would complete the task.

Guiding partners did not occur exclusively in the Shared condition nor only between pairs who worked sequentially. There were instances where guidance occurred between pairs working

Table 8.12: Extract from pair 8's collaboration – Guiding partner's movements on the hierarchy in the Shared condition.

Time	P8A	P8A's Audio	P8B	P8B's Audio
34:55				*"article open"*
34:56	go inside		ok	
34:57			<down>	*(down)"attributes"*
34:59	come down			
35:00			<open>	*(open)"one"*
35:01			<down>	*(down)"number"*
35:02	yeah control err			
35:03	cmd P			
35:04			hmm <set primary>	*(success)"attribute number*
35:05				*set as primary"*

in parallel in the Shared condition, and between pairs working in the Non-Shared condition. But these had noticeable differences. Parallel pairs working in the Shared condition typically switched to the sequential working style during guidance interaction, and, expectedly, guidance in the Non-Shred condition involved extensive exchange of awareness information (both supplied and requested), particularly of the types *"Where I Am/ Where Are You"* and *"What I Have Done/ What Have you Done"*.

Table 8.13 shows an extract where a participant guides their partner in the Non-Shared condition. In this example, participant P11B is confused as to how to create a relation between two entities, and his partner P11A guides him through this process. As soon as P11B expresses confusion, his partner, P11A, halts the parallel interaction and focuses on the verbal information he receives from P11B. After each instruction, P11B supplies detailed information about his locations and current action and his partner uses the supplied information to time his instructions accordingly. Notice how there is a slight offset between the supplied awareness information and the guiding instructions. Participant P11A, for instance, had to repeat the same instruction twice (at minutes 21:21*"and do cmd S"* and again at 21:25 *"do cmd S which selects it"*) to match the WA information that was supplied by his partner (*"yeah i'm"* at minute 21:19 and again *"ok i'm at therapy"* at minute 21:24). Notice also how the participant providing guidance explicitly requests information in order to confirm the current status of his partner's progress at minute 21:28 of the interaction: *"does it say it's selected"*, and only provides the next instruction once his partner provides the sought confirmation: *"therapy selected, yes"*. There is clearly more utterances dedicated to coordination in this example than there was in the extract on Table 8.12 above. Granted that the task in this example is slightly more complex, participants supplied and requested more WA information at each step of the guidance than did the pair in Table 8.12.

The speak-aloud utterances in Table 8.13 were crucial for keeping the partner focused on the activity, for seeking confirmation about execution steps and assistance when encountering difficulties; activities that were completed through the system audio output alone in in Table 8.12. Here, information that was provided in the form of feedthroughs when using speakers was replaced by explicit requests and outlouds (Gutwin and Greenberg, 2002) when using headphones.

Table 8.13: Extract from pair 11's collaboration – Guiding partner in the Non-Shared condition.

Time	**P11A**	**P11A's Audio**	**P11B**	**P11B's Audio**
21:16			let's start again	
21:17	go into Entities		<switch>	*(switch)"entities"*
21:18			**hmm**	
21:19	Therapy		**yeah i'm**	
21:21	and do..			
21:22	..cmd..		<down>	*(down)"therapy"*
21:23	..S			
21:24			**ok i'm in therapy**	
21:25	do cmd S			
21:26	which selects it			
21:27			**cmd and S** <select>	*(success)"therapy selected"*
21:28	**does it say**			
21:29	**it's selected?**		it says	
21:30			**therapy selected yes**	
21:31	then go to Gene			
...				
21:37			<down>	*(end of list)"gene"*
21:38			**now i'm in gene**	
21:39	then select it		**and i select it**	
21:40	and do cmd R		<select>	*(success)"gene selected"*
21:41			<new rel.>	*(editing)"relating gene*
21:42	and that's			*(editing) and therapy"*
21:43	your relation			*(editing)"enter name"*

> ***Design Lesson 17 – Provide support for guidance interaction when using headphones***. In addition to the recommendation outlined in *Design Lesson 12* above, when audio is delivered through headphones, collaborative systems should provide workspace awareness information of types *"Where Are You"* and *"What Have You Done"* if guidance mode is detected or triggered by the users.

8.5.2.3 Coordinating Collaborative Actions

Extracting information about partners' actions and locations from the auditory display allowed pairs to coordinate their collaborations in a variety of ways; information about the content to be edited could be exchanged swiftly, when needed, and unprompted; construction episodes could be organised fluidly; and interdependent editing actions could be coordinated efficiently.

Coordinating Content Exchange. Participants were able to detect when their partner reached a stage where content information needed to be input onto the diagram by listening to their partner's interactions. Table 8.14 shows an extract from pair 8's collaboration where one participant was editing diagram content based on the information described on their partner's text. The partner, unprompted, supplies the content information at the exact moment when it is needed.

Table 8.14: Extract from pair 8's collaboration – Providing diagram content during an editing action.

Time	P8A	P8A's Audio	P8B	P8B's Audio
05:54	i add one more			
05:55	attribute right?			
05:56			add one more attribute	
05:57	<new attr.>	*(editing)"new attribute"*		
05:58		*(editing)"enter name"*		
05:59			Tissue Type	
06:00	[typing]			

The availability of the audio output in the environment allowed for two aspects of the collaboration in this particular case to occur fluidly. First, supplying relevant information at the moment of the interaction where it is relevant, not before and not long after, was a direct consequence of the ability to accurately judge the progress of a partner's activity. Second, the audio allowed for the two participants to have joint attention over the progress of a task, which would otherwise have been exclusively accessible to one participant. WA information about actions and locations in this case are not explicitly requested nor supplied, but extracted and used efficiently in the collaboration. This is a manifestation of an ability to move between focussed individual work and joint work (Gaver, 1991; Dourish and Bellotti, 1992), an activity referred to as efficient management of coupling (Gutwin and Greenberg, 2002).

Coordinating Construction Episodes. Discussions about new content and sometimes labour typically marked the transition from one construction episode to the next. Effectively extracting information about partners' actions and locations facilitated such transitions.

An example of this is illustrated in Tables 8.15 and 8.16. In the first example, participant P9B detects that the process of adding attributes to the entity "Driver" and setting it as "Primary" have been completed at minute 46:47, and that it was time to move on to the next piece of content; the "Journey" entity. Similarly, participant P10A in the second example detects the point that marks the end of the construction episode of the relation "Use" at minute 29:32, and initiates the discussion about the next diagram content to be addressed. In both these examples the pair moved fluidly between construction episodes without explicitly requesting or supplying information about the progress or completion; they were able to detect the completion of the editing actions through a feedthrough that otherwise would have been supplied or requested as *"Completion Status"* updates.

Table 8.15: Extract from pair 9's collaboration – Organising episodes of construction.

Time	**P9A**	**P9A's Audio**	**P9B**	**P9B's Audio**
46:42	<down>	*(down)"employee ID"*		
46:45	<set primary>	*(success)"employee ID*		
46:46		*set as primary"*		
46:47			ok cool err	
46:48			so that's Driver	
46:49	Journeys		should we move go to the	

Coordinating Interdependent Actions. The ability to fluidly move between editing different parts of the diagrams was also observed when the initiation of a new editing task depended on the completion of the current one and where each task was assigned to a different participant. In such cases, a signalling of the completion of the first task is needed to coordinate the initiation of the next one. This could of course be achieved by supplying or requesting information of type *"Completion Status"*. There were instances in the Shared condition were this was not necessary and information about completion status was extracted from the environment instead.

An example of this is illustrate on Table 8.17. Participant P10A in this case was assigned the task of relating entities, while his partner was assigned the task of locating the relation on the "Relations" perspective once created and editing its cardinalities. The extract shows that P10B remained interactively idle, attentively listening to his partner's interaction, suggesting content where appropriate, then fluidly moving on to his task when the completion of his partner's task is detected. P10B in this case was able to extract information about completion status at minute 33:15 of the collaboration and to use such information to coordinate his next move.

Table 8.16: Extract from pair 10's collaboration – Organising episodes of construction.

Time	P10A	P10A's Audio	P10B	P10B's Audio
29:25			<confirm>	
29:26				*(success)"cardinality many*
29:27				*added to Tag" "use open"*
29:32	so now there is			
29:33	also err Doctor			
29:34	related to Therapy		ok	

Table 8.17: Extract from pair 10's collaboration – Coordinate interdependent actions.

Time	P10A	P10A's Audio	P10B	P10B's Audio
33:08		*(editing)"enter new*		
33:09		*(editing)"relation name"*	i think	
33:10			it's Analysis	
33:11	Analysis?			
33:12	[typing]		yeah	
33:14		***(success)"relation***		
33:15		***analysis added"***		
33:16			ok	
33:17			<switch>	*(switch)"relations"*

Design Lesson 18 – Provide users with a means to display awareness information of their choice to their partners' workspace "views". In order to support coordination, particularly when collaborators work in parallel or use headphones, users should be provided with a means for choosing which awareness information to display as well as when such information is displayed to their partners.

8.6 Sound as a Shared Representation

Collaborating pairs in Study 3 had equivalent access to shared diagram representations. That is, both the content of a diagram and the way this was hierarchically structured were available to both participants in a pair. However, each participant had independent control over which part of the diagrams they wished to be displayed in audio and delivered to their individual headphones or speakers. Thus, the setting of the study imposed a shared structure of diagrams' content, but did not impose a single shared auditory "view" of such content.

Participants who worked sequentially during all or parts of their collaborations, or those who switched their attention to their partner's audio output when speakers were used, were essentially sharing more than diagrams content and structure. The previous sections examined how the ability to hear the audio output of partners' interactions was used to establish knowledge about partners' actions, locations and to some extent predict their intentions. This section presents further examples of how participants exploited the ability to interactively control the audio output to use it as a shared representation. Using audio as a shared representation, therefore, refers to instances where only one output displayed audio in the workspace and pairs made collective use of the displayed information.

8.6.1 Exploring Diagrams Content

Participants used the audio displayed from a single output to jointly explore and discover diagram content. The decision as to which part of the diagram to explore in such cases came from either participant in a pair and not necessary from the one manipulating the diagram. The extract on Table 8.18 shows an example where one participant explicitly requests their partner to explore a particular part of the diagram while they jointly make use of the produced output. In this extract, P8B of pair 8 asks her partner to list the current entities of the diagram, and as her partner goes through the list of entities, they both listen and exchange confirmations about the discovered content. At minute 40:15 of the extract, P8B effectively picks up on the *(end of list)* non-speech sound to infer that all entities have been visited.

8.6.2 Pointing to/ Highlighting Items

There were times where participants wanted to draw their partners' attention to a particular item or bulk of items on the diagram. But since auditory events are transient, and the hierarchical

Table 8.18: Extract from pair 8's collaboration – Sharing the representation to check content.

Time	P8A	P8A's Audio	P8B	P8B's Audio
40:01			what are the five	
40:02			entities can you list	
40:03			them out please	
40:04	<down>	*(down)"gene"*		
40:05	Gene		hmm	
40:06	<down>	*(down)"therapy"*		
40:07	Therapy		hmm	
40:08	<down>	*(down)"patient"*		
40:09			hmm	
40:10	ok Patient <down>	*(down)"tag"*		
40:11			Tag	
40:12	Tag		yeah	
40:13	<down>	*(down)"article"*		
40:14			article yeah	
40:15	<down>	***(end of list)**"gene"*	and that's it	

structuring of diagram content did not encode any spatial information, participants needed an alternative to what otherwise would be pointing to objects. They did this by visiting the item they wished to highlight at moments where they assumed their partner was attentively listening to their interaction. The audio output of the item of interest in this case would be interactively triggered, and if necessary re-triggered, while the participant uses a verbal clue to attract their partners' attentions to the displayed item.

Tables 8.19 and 8.20 show two examples where items on the diagram are "pointed" to or "highlighted" using this method. When participant P10A enquired about whether the entity "Therapy" have already had two attributes added to it in extract 8.19, his partner locates the items in question on the hierarchy and displays them to show that there were indeed two such items. *"Yeah"* – he replied while triggering his audio output to display: *"attributes, two"* – *"see [then browses to the first attribute] Method [and browses to the second] and Type"*. Similarly, when participant P12A was prompted by his partner to add the "Many" cardinality to one of a relation's ends as shown in extract 8.20, he asserted that such content already existed. But when challenged by his partner, he explicitly draws their attention: *"look there is a tapping sound here"* then displays the item in question on the hierarchy: *"(many) journey"*. The combination

Table 8.19: Extract from pair 10's collaboration – "Pointing" to items on the diagram.

Time	P10A	P10A's Audio	P10B	P10B's Audio
09:08	you have two			
09:09	attributes for Therapy?			
09:10			**yeah**	
09:18			<up>	*(up)"attributes"*
09:21			<open>	*(open)"two"*
09:22			**see**	
09:23			<down>	*(down)"method"*
09:24			method <down>	*(down)"type"*
09:25			and type	
09:26	ok			

of the words *"see"* and *"here"* in these examples coupled with the immediate triggering of an audio output was a means for highlighting content through a shared auditory representation of such content. Participants in such instances needed to make sure that their partners' attention is directed towards the output source in a listening-in-search mode.

Table 8.20: Extract from pair 12's collaboration – "Pointing" to diagram elements.

Time	P12A	P12A's Audio	P12B	P12B's Audio
35:22			put many	
35:23	it's there already			
35:23			it's not	
35:25	look there is			
35:26	a tapping sound			
35:27	**here**			
35:28	<down>	*(many)"journey"*		
35:29			ok	
35:29			go to Driver	
35:32	<down>	*(one)"Driver*		
35:33			ok fine	

Participants in Study 3 used this method of pointing in two prominent ways; to seek clarification about the meaning of a displayed sound; and to make joint decisions about how content should be edited.

Table 8.21: Extract from pair 6's collaboration – Sharing audio to seek clarification

Time	P6A	P6A's Audio	P6B	P6B's Audio
35:07			stop for a second	
35:08	yeah?			
35:09			<request>	***(request)"o o"***
35:12			o o ?	
35:14			what is o o ?	
35:15	what is that in?			
35:22			<close>	*(close)"relations"*
35:24			<open>	*(open)"five relations"*
35:26			five relations yeah?	
35:27	yeah			
...				

Seeking Clarification. The extract on Table 8.21 shows an example where a participant encountered an ambiguous item on the diagram; a relation with the label "o o". Pair 6 in this extract are working in parallel in the Shared condition, but participant P6B asks his partner to halt the parallel interaction and to listen to his auditory display, thus explicitly directing his partner's attention to the appropriate audio output source. He points him to the ambiguous content by replaying the sound and asking for a clarification: *"o o?"*, *"what is o o?"*. His partner then asks for further details about this "strange" item's location on the hierarchy, having effectively switched his attention to P6B's output. From that point onwards – too long to include in this extract – the pair collectively establish that the ambiguous object was a relation with a misspelled label, and assign tasks to rectify this mistake. Thus, a parallel working pair engaged in a sequential interaction to inspect, discuss and manipulate their diagram, essentially using one auditory ouptut as a shared representation to support these processes.

Making Joint Decisions. A further use of this pointing method was to highlight items on the diagram where joint decisions needed to be made about how such content should be edited. Table 8.22 shows an instance where a participant consults with his partner to decide on which attribute to use as a primary key for a particular entity. P12B goes through a list of existing attributes then displays his choice of attribute twice in a row, while suggesting: *"we can make this a primary key"*. Again, In this example, the participant "points" to the item of interest by first requesting their partner's attention to be directed towards the appropriate audio output

Table 8.22: Extract from pair 12's collaboration – Making decisions about diagram content.

Time	P12A	P12A's Audio	P12B	P12B's Audio
10:27			<down>	*(down)"attributes"*
10:28			<open>	*(open)"two"*
10:30			<down>	*(down)"name"*
10:31			<down>	*(down)"address"*
10:32			ok	
10:34			<up>	*(up)"name"*
10:37			we can make **this**	
10:38			a primary key	
10:39	yes			

source, then replaying the audio they wished to point to or highlight. The use of the pronoun *"this"* in this case to refer back to the item in question is an example of using deictic reference to simplify communication (Gutwin and Greenberg, 2002). This indicates that the participant had assumed that their partner was indeed listening to their audio output, and also that they had mutual knowledge about the referent. The partner's affirmation indicates that such mutual understanding was indeed present; both participants conceived of the auditory displayed output as a shared object that could be referred to.

Design Lesson 19 – Provide a means for pointing to items on the shared workspace. Audio collaborative systems should include a mechanism for users to highlight items on the workspace and draw partners' attention to them. This could be achieved by, for example, giving users control over the audio output that is displayed to their partners as outlined in *Design Lesson 13*. A mechanisms for handling conflicts of audio output should be included with such an implementation.

8.6.3 Structuring Tasks

The way in which pairs organised the structure of a given editing task was sometimes influenced by the information that was displayed from a shared audio output source. As pairs devised a strategy for a construction episode, they used the shared audio, either implicitly or explicitly, to determine how to address the editing required to complete such episode. An example clarifies this further.

Having added all the entities to the diagram, pair 13 in the extract on Table 8.23 are reading

Table 8.23: Extract from pair 13's collaboration – Structure of the task determined by the order of auditory presentation.

Time	P13A	P13A's Audio	P13B	P13B's Audio
49:53	[typing] <confirm>	*(success)*"attributes name		
49:54		*added to gene"*		
49:56	ok			
49:57	<close>	*(close)"gene"*		
49:58	that's it for gene		yeah	
50:00	<down>	***(down)"therapy"***		
50:01			Therapy erm [reading]	
50:08			Therapy targets one Gene	
50:10			which sounds like	
50:11	yeah		a relationship	

through the textual descriptions to extract and add the attributes of each entity. P13A manipulates the diagram while P13B listens to the displayed audio. When P13A finishes adding the attributes to the entity "Gene" she announces the end of this construction episode: *"that's it for Gene"* and hits the browse command to display the next entity on the list; *"Therapy"*. Her partner hears the displayed output and uses it to implicitly decide that the displayed entity should be the focus of the next episode of construction. The pair could have used the order in which content is presented on the provided textual descriptions, or indeed any other random order, but instead, the shared audio output triggered by P13A was in this instance used as a basis for organising the structure of this editing task. A similar example is shown in the extract of pair 10's collaboration on Table 8.24. Here too, the order in which content is addressed is based on what is auditorally displayed at the moment of initiating a new construction episode. Participant P10A hits the browse command to discover the next item in the list, and uses the displayed information to enquire about content from his partner. Again, the way in which the representation presented information dictated the order in which this part of the construction was addressed.

In general, some pairs conceived of the diagram construction process as a whole to be a shared task, and tended to do so by sharing a single audio output source during all or most of the collaborations. A further characteristic that could be considered an indicator of such conception is the use of language that employs plural pronouns when discussing actions that have been completed or need to be executed in the future (e.g. *"we can make this a primary key"* from

Table 8.24: Extract from pair 10's collaboration – Structure of the task determined by the order of auditory presentation.

Time	P10A	P10A's Audio	P10B	P10B's Audio
07:22	we now need to add			
07:23	attributes to the entities			
07:25			err yes	
07:28	ok <open>	*(open)"six entities*		
07:31	<down>	***(down)"gene"***		
07:32	gene ok..			
07:33	..what does gene have?			
07:35			an ID i think	

extract 8.22; *"should we move go to"* from extract 8.15), and to refer to locations on the diagram (e.g. *"we are in Relations now"* from extract 8.6). This is in contrast to other pairs who conceived of the construction process as a series of independent individual tasks, typically using singular pronouns to refer to completed actions or movements on the hierarchy (e.g. *"ok, I'm finishing the last relation"* from extract 8.2).

8.6.4 Alternative Audio-only Interaction Strategies

Pairs who worked in parallel in the Shared condition, or in the Non-Shared condition, were sharing the same diagram content and structure but not its audio presentation. Yet participants in such instances still needed to engage with their partners during their collaborations for similar reasons to those mentioned above; to explore content, highlight items of interest, make joint decision or structure editing tasks. A number of alternative interaction strategies were deployed by participants under such conditions to overcome the "inability" of using the audio as a shared representation.

Follow. The first of such alternative is referred to as the "Follow" strategy. Participants typically used this strategy to keep track of each other's actions in the Non-Shared condition. Essentially, a participant employing this strategy uses their existing knowledge of where on the hierarchy their partner will act next – established through a division of labour or through explicitly supplied or requested WA information – to *"follow"* their interaction. Following is done by simultaneously exploring the same part of the hierarchy that a partner is currently editing, repeatedly triggering the audio display of such part until a change in content is detected. Doing this enables a participant to acquire information about their partner's progress on a particular editing action – i.e. a means for waiting for a feedthrough that indicates that an action has taken place.

Table 8.25: Extract from pair 13's collaboration – Following partner's interaction in the Non-Shared condition.

Time	**P13A**	**P13A's Audio**	**P13B**	**P13B's Audio**
30:52			then Many	
30:53			for Stage	
30:55	yeah			
30:56	<up>	*(one)"driver"*		
30:57	<down>	*"stage"*	<new card.>	*(editing)"choose cardinality*
30:58	<down>	*(end of list)(one)"driver"*		*(editing)*
30:59	<down>	*"stage"*		*(editing)*
31:00	<down>	*(end of list)(one)"driver"*	<many>	*(editing)"many selected"*
31:01	<down>	*"stage"*	<many>	*(editing)"many selected"*
31:02	<down>	*(end of list)(one)"driver"*		
31:03	<down>	*"stage"*		
31:04	<down>	*(end of list)(one)"driver"*	ok <confirm>	*(success)"cardinality many*
31:05	<down>	***(many)**"stage"*		*added to stage"*
31:06	yeah			*"assign open"*
31:07	<close>	*(close)"assign"*		
31:08	**great**			

The example listed in Table 8.25 illustrates this behaviour. Pair 13 are working in the Non-Shared condition, having assigned the task of adding cardinalities to participant P13B. Instead of awaiting updates from her partner, participant P13A explores the part of the hierarchy currently being edited by P13B, by looping through the list of the relation's ends, until she hears the ***(many)**"stage"* at minute 31:05. This confirms that the cardinality "Many" has been added to the relation end "Stage" because the shared content is updated with a new cardinality. Thus, this strategy replaces the ability to follow a partner's actions by listening to their interaction. McGookin and Brewster (2007b) developed an explicit haptic means to allow a user to follow their partner in a non-visual collaborative workspace that is equivalent to the follow strategy observed in Study 3. Participants in their study found such mechanism particularly useful to locate their partners on the workspace. A similar audio-only mechanism could have been used here where, for instance, a user is able to control which audio output they receive through their headphones effectively switching between theirs or their partner's.

Table 8.26: Extract from pair 11's collaboration – Synchronising interaction in the Non-Shared condition.

Time	**P11A**	**P11A's Audio**	**P11B**	**P11B's Audio**
19:11	what does it say?			
19:12			right go back	
19:13			and go into entities	
19:14			<switch>	*(switch)"entities"*
19:15	<switch>	*(switch)"entities"*	erm	
19:18	then to Therapy?		<down>	*(down)"gene"*
19:19	<down>	*(down)"gene"*	yeah	
19:20	<down>	*(down)"therapy"*	<down>	*(down)"therapy"*
19:21			ok go inside	
19:22	<open>	*(open)*	<open>	*(open)*
19:23	into Therapy			
19:24	<down>	*(down)"related"*	into attributes <down>	*(down)"related"*
19:25	it says related		<down>	*(down)"attributes"*
19:26	<down>	*(down)"attributes*	to attributes	
19:27	<open>	*(open)"one"*	then go right <open>	*(open)"one"*
19:28	it says it has one attribute			

Synchrony. While pairs working sequentially in the Shared condition used a single audio output to collaboratively explore diagram content, those working in parallel and/or in the Non-Shared condition used the strategy of *"Synchrony"*. This is essentially done by synchronising one's interaction with the partner's to simultaneously explore the same part of the hierarchy that it is currently being discussed or edited.

Table 8.26 shows an example where pair 11 synchronised their interaction to match the discussed content in the Non-Shared condition by simultaneously exploring the same content. Typically, such synchronised interaction allows for the shared content to be accessed at the same time, thus allowing participant to clarify misunderstandings about ambiguous content, highlight problems or make joint editing decisions; i.e. replacing the "pointing" method used by participants who shared a single audio output source. Since no feedback is available about where each participant is at during such synchronised interaction, it is not surprising that participants typically increase their exchange of WA information to efficiently execute this strategy.

Design Lesson 20 – Provide a means for following partners' interactions in a workspace. Audio collaborative systems should provide users with the ability to allow their partners to control the audio output that they receive, particularly when using headphones, i.e. handing the control of their display to their partners. Just as *Design Lesson 18* recommended allowing users to displays their audio output to their partners, this should allow users to monitor their partners activity.

8.7 Working Styles and Interaction in Audio-only Collaboration

As described in section 7.5.4, participants in Study 3 employed two different working styles during their collaborations, referred to as *parallel* and *sequential*. A parallel working style was characterised by overlapping interactions and concurrent manipulations of a diagram, whereas a sequential style was characterised by successive turn taking in the manipulation of diagrams. Participants used these styles in both the Shared and Non-shared conditions, either exclusively or in combination. This observation is in line with previous studies which referred to the emergences of similar collaborative working styles in auditorally augmented workspaces (Morris et al., 2004; McGookin and Brewster, 2007b). Note that while the choice of working style might have been explicitly agreed upon by participants, the way labour was divided could also have implicitly influenced such choice. This is particularly the case when interdependent tasks are assigned to each participant, automatically requiring the participant with the dependent task to wait until the task they depend on is completed by their partner, hence resulting in a sequential pattern of interaction. Indeed, whether the participant with the dependent task chooses to remain interactively idle while their partner completes their task, or to concurrently execute other non-dependent actions (such as inspecting the diagram) overrides the implicit influence of the division of labour on the patterns of interaction.

The two working styles that were extracted from the interaction logs were used to retrospectively group the pairs into Sequential and Parallel groups, based on which style they used dominantly in the Shared condition. Other groupings were also possible. For instance, pairs could have been grouped on the basis of whether they combined parallel and sequential working styles in either conditions or used one style exclusively. Groupings could also have been based on the strategy of construction employed by the participants; those who used a description-based versus those who used diagram-based strategies. Grouping pairs on the basis of their dominant working style in the Shared condition was chosen to allow for examining the role of audio in supporting co-located audio-only collaboration. That is, grouping pairs in a way that highlights instances where the use of the audio, when present in the workspace, was distinctively different.

The previous sections have shown how the choice of working style impacts on pairs' collaborative organisation and exchange and use of workspace awareness information. The following sections use data captured from observations and the post-test interviews to examine a number of issues that were associated with each working style – irrespective of which condition it was employed in – as well as with working in an audio-only workspace in general.

8.7.1 Issues With Working in Parallel

When participants in a pair worked in parallel, they focused on and concurrently executed different tasks, updating each other about the progress of their actions during or when execution is completed. Coordination was thus necessary to make sure that information from individual diagram descriptions is collated effectively. If coordination is poor, the outcome of concurrent actions are likely to be erroneous.

Coordination was thus one of the main issues when working in parallel. Participants in Study 3 sometimes created duplicate or conflicting content due to poor coordination. An example of this which frequently occurred in Study 3 is conflict during the process of adding relations to a diagram. This is because such a process typically involved the simultaneous handling of the same diagram content. As described in section 7.3, the collaborative tool supported two means for selecting entities; public and private selections. Creating a relation using publicly selected entities is different from creating it using privately selected entities, but some participants were not careful with the use of such selection mechanisms when working in parallel. This resulted in the construction of inappropriate content; e.g. involuted, ternary or even quaternary instead of binary relations. If not carefully rechecked, such inappropriate unintended content often went unnoticed by the participants.

There were also instances where a subset of diagram content provided on the textual descriptions was never addressed by pairs whose working style was dominantly parallel. Such an issue was often the result of conflicting interactions and/or inattention to partners' utterances. Conflicting interactions occurred when a participant attempted to edit an item on the diagram that was currently being edited by their partner. As described earlier, the tool in such a case "protects" the element in question to prevent conflicting edits, and issues a warning message to the participant attempting to access the item currently in use. Blocked participants were then faced with two choices; to return to the element in question when their partner "releases" it; or to request their partner to execute the intended editing action instead of them while they still have a lock on the item. If a participant forgot to do either of these steps, planned edits were likely to be missed or misconstrued. Inattentions to partner's utterances occurred when a participant was addressed while they were engaged in executing an independent editing action (i.e. attentively listening to the audio emitted from their headphones or speakers). There were instances where the participant in such a case ignored or did not notice that their partner was addressing them. Consequently, planned edits were also likely to be missed or misconstrued.

In general, participants who used a parallel working style in Study 3 found using speakers to be distracting. This is a similar observation to that made in (McGookin and Brewster, 2007b). When both partners' speakers simultaneously output audio, and each participants attempts to focus on one output at a time, the second source becomes a source of noise. The audio thus takes an interfering role rather than a helpful one. This was also equivalent to instances where a participant was addressed by their partner while they were manipulating diagrams using headphones. The partner's voice in such instance interferes with the sound received through the headphones and, similarly, becomes a source of disturbance. The recommendations outlined in *Design Lesson 13*, *18* and *19* – i.e. to provide the users with the ability to switch back and forth between private and shared "views" in the auditory workspace and to display their audio and monitor their partners' – could address the issues associated with the parallel working style.

8.7.2 Issues With Working Sequentially

As described previously, participants who worked sequentially took turns manipulating the diagrams and often had joint focus on the execution of editing and inspection actions. Such joint attention sometimes led to biased interpretation of audio output. There were instances in sequential collaborations where a participant influenced their partner's opinion on what a sound means (see for example the extract on Table 8.5). When such interpretation was incorrect, the bias could lead to inappropriate or incorrect interactions with the diagrams.

While allowing for a variety of shared activity and joint attention, the presence of audio through speakers in the workspace sometimes hindered participants ability to freely interact with the diagrams. Some participants described how they found it inappropriate or difficult to interact with the diagram at the same time as their partners, and preferred to remain interactively idle. They did this either to avoid triggering sounds that might interfere with their partner's interactions or because they themselves could not concentrate when more than one output source was displaying audio. This means that the choice of a sequential working style was not always a choice of preference, but one of convenience. That is, there were instances where participants did not want to sit there and wait while their partners interacted with a diagram but the audio in the environment "blocked" them from concurrently manipulating the diagram. As one participant puts it: *"[using speakers] was more boring, because there is nothing to do, and i can't play around when [my partner] is doing her thing"*. This and similar remarks can perhaps provide a further explanation as to why the parallel working style was used by the majority of pairs in the Non-Shared condition.

8.7.3 Issues in Audio-only Collaboration

Participants in Study 3 collaborated in an audio-only setting; they could not see each other but could hear each other's voices in both conditions, and each other's audio output in the Shared condition. There were two issues directly related to working in this setting; referencing problems; and conflicting sounds.

8.7.3.1 Referencing Problems

Participants had access to a shared auditory representation of diagram content. However, there were other representations in the environment that contained information about the same content, albeit in a different form and format; namely the textual descriptions. This made the participants prone to clash when making references to such information.

The extract on Table 8.27 shows an instance where a participant refers to content on the textual description while their partner refers to the same content but as accessed through the audio-only tool. Working in the Non-Shared condition, P2A announced that the entity "Gene" did not have any attributes, referring to its current state on the constructed diagram. His partner P2B, however, turns to the textual description of the diagram to check and asserts that the entity does have attributes. When negated, P2B further asserts: *"I have it"*. Each participant in this example is referring to a different representation of the information in question, and both are being vague about which of these representations they are referring to. The collaborative setting in this case did not help; if P2A could see that his partner had his head down reading the description when he made the reference, the problem could have been avoided. Indeed participants were often explicit when referring to content, or when requesting information from their partner to

Table 8.27: Extract from pair 2's collaboration – Referencing problem across representations in the Non-Shared condition.

Time	P2A	P2A's Audio	P2B	P2B's Audio
03:57	Gene doesn't have			
03:58	any attributes			
03:59			wait	
04:01	<down>	*(down)"therapy"*	[reading]	
04:02	<down>	*(loop)"gene"*	[reading]	
04:03	<open>	*(open)*	[reading]	
04:04	<down>	*(down)"related'*	[reading]	
04:05	<down>	*(loop)"related"*	**it has a**	
04:06			**name**	
04:07	**no i can't**			
04:08			i have it	

distinguish between the textual and the auditory representation; *"do you mean on the system or the text*, replied P2B on another instance in the collaboration when his partner vaguely described the existence of two entities on the shared diagram.

There was, however, another type of referencing problems that occurred even when both participants were aware of the source of information, specifically, when they both referred to the auditory representation. Table 8.28 shows an example of such a case. Here, participant P8A explores the attributes of an entity on the diagram, while his partner P8B listens to his interaction. P8B did not understand the spoken output of the attribute name that was displayed second in the list of three attributes and asks: *"what was that?"*, to which P8A responds by stating the attribute name that was displayed last:*"MSP"*. The two participants referred to two different auditorally displayed items; P8B to an item heard during their partner's interaction, and P8A to the item heard last. This referencing problem is somewhat a counter effect of the method adopted by the participants in Study 3 to point to auditory object on the diagram (described on page 215). To point to items, a participant would draw their partner's attention with a verbal cue while interactively displaying the item of interest; i.e. highlighting the last auditorally displayed item. The temporary confusion exemplified in this extract was essentially a result of the timing of the request, which occurred after the word *"MSP"* was spoken, (incorrectly) rendering it as the potential referent object.

Table 8.28: Extract from pair 8's collaboration – Referencing problem within one representation in the Shared condition.

Time	P8A	P8A's Audio	P8B	P8B's Audio
40:53	<down>	*(down)"name"*		
40:54	<down>	*(down)" tissue type*		
40:55	<down>	*(primary)"MSP"*		
40:57			what was that?	
40:58	MSP			
40:59			the tissue one	
41:00			i didn't get it	

Thus, there were overall two types of referencing problems likely to occur in the audio-only collaboration; the first is when referring to items within one representation (audio output, or text) and the second when referring to item across representations (the audio and the text). The recommendations outlined in *Design Lesson 19* for providing a means to point or highlight items on the workspace could address the issues associated with these kinds of referencing problems.

8.7.3.2 Conflicting Sounds

Collaborating in a workspace where sound is the only means of communication and representation increases the likelihood of sounds produced from the various audio producing sources to conflict[12]. There were three types of such conflicts; 1) audio from one source conflicting with that of another; 2) audio from one or both sources conflicting with participants' speech; and 3) participants speech conflicting with the audio displayed from one source. Naturally, the first instance only occurred in the Shared condition, whereas the latter two occurred in both the Shared and Non-Shared conditions.

When speakers were used, audio from a total of four sources could potentially be produced at the same time; two sets of speakers and the two participants' speech. This number was reduced to three when headphones were used; i.e. excluding the partner's audio output source. The choice of working style also changes the number of potential audio producing sources; working sequentially reduces it to three in the Shared condition and to two in the Non-Shared condition, whereas working in parallel keeps it at four in the Shared condition and three in the Non-Shared condition. It is therefore not surprising that the working style influenced the likelihood of the occurrence of

[12] Conflict here refers to instances where two or more sounds are simultaneously displayed from more than one audio producing source.

conflict during the collaborations, with participants working in parallel being particularly prone to encountering conflicting sounds.

The occurrence of conflicting sounds that were observed in Study 3 did not lead to misinterpretations of the information displayed as much as leading to participants' irritation an discontent. Participants typically managed to effectively reduce the number of audio outputting sources to one in order to be able to discern what was displayed or discussed. They did this by asking their partner to halt their interaction to handle the first kind of conflict; by halting their own interaction to handle the second kind; and by asking their partner to hold their comment until they completed the task at hand to handle the third. Conflicting sounds were not always effectively handled, however. Particularly, there were instances where participants did not pay attention to their partner addressing them, both when using speakers and headphones. As described above, this typically led to issues with the constructed content (duplicate; inappropriate; or missing content). What is important to note is that using speakers was not the sole nor the most likely source of conflict. In fact, as expressed by some participants, using headphones was uncomfortable precisely because they could not anticipate when their partner would speak to them and hence cause conflicting sounds. The recommendations outlined in *Design Lesson 13, 18* and *20* above could thus be used to address the issues associated with these kinds of conflicts.

8.8 Summary and Conclusion

This chapter described a detailed analysis and discussion of the results obtained from Study 3. Pairs of sighted participants used an audio-only tool to construct diagrams in two audio-only settings; where audio was delivered through speakers or headphones. The analysis examined how the means for delivering audio in each setting changed the way participants went about their collaborations in terms of when to manipulate the diagram and what to discuss during such manipulations.

The examination showed that the mere presence or absence of the audio from the workspace did not necessarily elicit such a change unless collaborators chose to work using a style that exploited the characteristics of the shared audio-only workspace. Participants working sequentially were more likely to have joint attention over the various aspects that surround the construction of diagrams because they both relied on a single audio output source for interaction. When pairs worked in parallel, they focused on independent actions and thus on separate audio output sources, which decreased the number of potential items that could have triggered joint attention or discussions during the collaborations. The level of joint attention dedicated to the shared task blurred the distinctions between one's own contributions to the shared task and one's partner's. Pairs working sequentially were often biased to treat their partners actions as their own, whereas pairs working in parallel were often not. The analysis postulated that the combined effect of the means of delivering audio on the collaborations (particularly on discussions about division of labour) and the choice of working style influenced such bias; where the emphasis during higher levels of joint attention over the execution of tasks tended to be on the actions themselves rather than on who is executing them.

It was described in Chapter 6 that the naturalness and efficiency of a collaboration is influenced by how much collaborators are aware of the events that occur in a shared workspace and of each other's interactions with it. Addressing the awareness problem in groupware design is an important and difficult problem. Gutwin and Greenberg (2002) suggested that part of the solution to such a problem is to provide users with more information about their collaborators, but designers must carefully determine what information is most important in a collaboration. The second part of the presented analysis addressed this latter point, and provided an understanding of how workspace awareness information was communicated through sound in an audio-only workspace, and how it was extracted and used to support collaborative interaction. An audio-only collaborative workspace where sound is present through speakers made information about partners' activity and progress readily available, but the use of this information varied depending on the working style. Particularly, the *loose* character of parallel pairs' collaborations often meant that participants felt a greater need to find out about each other's past actions and frequently supplied each other with information in the form of updates about what has happened. On the other hand, sequential pairs' collaborations were *focused*, and information was often supplied in the form of descriptions about what was currently happening or what was about to happen in the immediate future.

When audio was shared through speakers, different types of listening were employed to extract information about partner's interaction. Once extracted, awareness information was used to detect and address a variety of individual and collaborative errors, to guide partners around the workspace and through the execution of editing actions, and to coordinate collaborative actions. Audio was also used as a shared representation around which collaborative action was organised and executed. By interactively controlling the audio display of various parts of the diagrams, participants in Study 3 were able to draw each other's attention to areas of interest, to collaboratively explore diagram content, and to structure their tasks in synchrony with the shared representation. When audio was not shared in this way, participants often used alternative strategies, such as following each other around the workspace seeking auditory feedthroughs, or synchronising their interaction while exchanging verbal instructions. Working in parallel amplified the challenge of coordination in an audio-only workspace. This was often overcome by increasing the exchange of workspace awareness information and deploying alternative interaction strategies. In some instances, the challenge persisted, however, and resulted in missing or inappropriate content being added onto the diagrams. Working sequentially exploited audio as a shared resource to overcome coordination issues, but sharing audio was also associated with negative experiences. For some participants, their perception of the system forced them to sit and wait while their partners interacted so as not to disturb them or increase the potential for conflicting sounds, and while providing workspace awareness, the information displayed through speakers was not always desired, and became a source of noise when no use was made of it.

The analysis and discussion presented in this chapter was thus also used to compile a set of lessons for designing support for audio-only collaborative interaction. This concludes Part II of this thesis. Chapter 9 presents a general discussion and conclusions of the research presented in this thesis, as well as the implications of the overall results reported in Parts I and II.

Chapter 9

General Discussion and Conclusions

This chapter concludes the thesis by summarising the major findings and contributions of the presented work and discussing their implications for auditory interface design in HCI and CSCW. Outstanding issues are also discussed, as well as suggestions for future work.

9.1 Overview of the Thesis

Sound complements other sensory information to enrich our everyday interactions with the world, but its inclusion in our interactions with and through computers, and digital technology in general, remains limited. The potential of the auditory modality to overcome some of the limitations of GUIs in contexts where vision is not the optimal channel for interaction is demonstrated by efforts in the areas of Auditory Display and multimodal interaction. The practicality of audio as a sole medium of collaborative interaction remains largely unknown, however, and the aim of this thesis is to reinforce auditory design knowledge by expanding awareness of the practicality of audio as a medium of representation and interaction. Its central premise is that, by itself, audio can be an effective means for supporting collaborative interaction with diagrammatically represented information.

Chapter 2 presented a brief historical account of audio output in computers and listed the benefits and issues surrounding the use of sound in HCI. An overview of the field of Auditory Display was also presented, together with descriptions of existing research-based techniques that have been found to successfully convey information through sounds at the user interface. Additionally, Chapter 2 reviewed current approaches to supporting non-visual interaction with visual displays. This review focussed on the representational models used by such approaches and highlighted the lack of research on the use of hierarchical models to support non-visual collaborative interaction with visual displays.

In order to practically use audio in collaborative interaction with diagrams, collaborators must be able both to communicate with one another through sounds and to contribute to a shared auditory workspace where they can access and manipulate diagrams. Therefore, during the course of conducting the presented work, focus was directed first towards developing such a workspace, that is, to exploring the transformation of diagrams into an audio accessible form, and to supporting individual interaction with the resulting form. As such, Part I of this thesis developed and evaluated a model for supporting audio-only interaction with diagrams. Chapter 3 reviewed the nature of diagrams as a form of representation and the way in which they are used to encode information. The review was used to inform the design of a multiple perspective hierarchy-based model that translates a given relational diagram from a graphical to a form that could be accessed in audio. Chapter 4 then presented two user studies that evaluated the feasibility of the developed model in supporting the inspecting and constructing of diagrams in audio. The results of these studies were thoroughly analysed and discussed in Chapter 5, which also complied a set of lessons for designing support for audio-only interaction with relational diagrams using hierarchical models of representation.

Focus then turned towards exploring how the developed model could function as a shared auditory workspace for supporting collaborative access and manipulation of diagrams. This inquiry was addressed in Part II of the thesis. Chapter 6 reviewed existing research on the use of audio in co-located and remote computer-supported collaboration, highlighting the link between the auditory modality and the concept of awareness in collaboration. This link motivated the focus of the audio-only collaborative study presented in Chapter 7, which explored the characteristics of collaborative interaction with diagrams where pairs of individuals use audio as the only means to communicate, access and edit shared diagrams. Chapter 8 analysed and discussed the obtained results, exploring the characteristics of collaborative interaction in terms of diagram construction processes, awareness of contributions to shared tasks, and exchange of workspace awareness information during collaboration. Chapter 8 also compiled a set of lessons from the presented analysis to design support for audio-only collaborative interaction with and through diagrams.

9.2 Major Results and Contributions

This thesis has presented the first investigation exploring the use of hierarchy-based audio-only displays as the sole means for supporting collaborative interaction with diagrammatically represented information. That is, collaborations where both the means for communication and for accessing and manipulating a shared diagram is auditory[1], and where the hierarchical model of representation is used as a metaphor for supporting interaction. The main question that the thesis sought to answer was:

> **"Can audio be a practical medium to support individual and collaborative interaction with diagrammatically represented information?"**

The answer to this question is reflected in the extent to which the research questions outlined in Section 1.2 were tackled by the theoretical developments and the empirical evaluations described in the subsequent chapters of the thesis. The following sections reflects on each research question while explicitly highlighting the contributions of the thesis.

9.2.1 Research Question 1

"RQ1 : How can a given diagram be translated from a graphical form to an alternative form that can be accessed in audio?"

9.2.1.1 Summary of motivation and literature review

Existing approaches to supporting non-visual computer-based interaction with visual displays, such as GUIs, tables, graphs and diagrams, use either one or a combination of two models of representation; a *Spatial* model, typically implemented using a combination of auditory and haptic displays to preserve the spatial properties of the original visual display; and a *Hierarchical* model, typically implemented using the auditory modality to capture and organise the semantics of the original visual display in terms of groupings and parent-child relationships. The work presented in this thesis focused on using a hierarchical model of representation as the basis for exploring audio-only interaction with relational diagrams. This choice was based on the following rationale:

[1] Recall that according to the definitions of Auditory Display (specified in Section 2.3 on page 16), and audio-only collaborative workspace (specified in Section 7.1 on page 143) the work presented in this thesis is concerned with auditory feedback that is emitted by a system to its user(s) and excludes that which is produced from user(s) to the system (e.g through voice control).

- The focus of this thesis is on exploring the potential of the auditory modality in interaction, therefore choosing a representational model that has been found to be effectively conveyed solely in sounds (e.g Brewster, 1994; Mynatt, 1995; Leplatre, 2002).

- Theoretical accounts of human perception suggest that perceptual representation is hierarchically organised. For example, visual form is analysed at various embedded levels of structural organisation (Palmer, 1977) and auditory streams are used to group auditory elements into meaningful wholes (Bregman, 1994).

- Hierarchical organisation supports the notion that obtaining an overview should precede any exploratory interaction with a given dataset, expressed by Shneiderman (1996)'s Visual Information Seeking Mantra and the equivalent Auditory Information-Seeking Principle (Zhao et al., 2004). A hierarchy can thus be used to present different levels of details at different levels of hierarchical depth.

- As discussed in Section 3.4.2, a hierarchical organisation can capture the hierarchical relations found to exist between the scale types represented by relational diagrams.

- Existing work on supporting non-visual collaborative interaction with diagrams have mainly focused on using spatial models of representation. Consequently, the practicality of hierarchical models for supporting such interaction is largely unknown.

9.2.1.2 Original work and Contributions

The thesis proposed the use of *multiple perspective hierarchies* as an approach to translating diagrams from a graphic to an audio accessible form. This approach is based on understanding when and why diagrams are better than other forms of representation, and on grasping how, exactly, information is encoded in diagrams. It is motivated by two main questions; first, which information should be captured about a diagram when translating it into an alternative modality? and second, how should the captured information be hierarchically structured? While previous research has suggested similar approaches to translating visual displays into hierarchical structures (Mynatt, 1995; Bennett, 1999; Brown et al., 2004), the proposed approach is unique in its focus on examining diagrams in terms of their properties as *external representations*.

Chapter 3 addressed the two questions above by examining the *functional* properties of diagrams to establish when and why they are a good form of representation, and their *structural* properties to establish how they encode information. In terms of functional properties, the reviews presented in Sections 3.2.2 and 3.2.3 revealed that diagrams facilitate searching for and recognising information using locational indexing. Additionally, diagrams help inference processes by making some conclusions more apparent than others, particularly when they are interactively accessed and manipulated. In terms of structural properties, Section 3.3 used Zhang (1996)'s taxonomy for Relational Information Displays to establish that relational diagrams encode information by combining two sets of features:

- *Modality-independent* features, which are reflected in the represented scale types without being specific to any particular medium of presentation, and
- *Modality-dependent* features, which are reflected in the dimensional representations that implement the scale types using graphical structural tokens.

The proposed approach applied this analysis to capture the former set of features when translating a relational diagram from the graphical to an alternative form. Capturing modality-independent features preserves the essence of the represented information, which can then be implemented using alternative structural tokens that are appropriate to the desired modality of translation. Furthermore, the proposed approach structures the captured features such that they can be interactively explored and easily searched. To achieve this, scale type indexing is used instead of locational indexing to hierarchically organise the captured information. Depending on their scale types, related items are grouped together in the resulting hierarchy, and together provide access to the same relational information from more than one perspective, which makes it easy to anticipate where information would be located on the hierarchy. Interactive exploration gives the user control over which information is presented to them at any given moment of the interaction, and scale type indexing creates a semi-fixed hierarchical structure, which has been found to facilitate orientation.

9.2.2 Research Question 2

"RQ2: How can the activities of inspecting and constructing diagrams be supported through the resulting auditory translation?"

9.2.2.1 Summary of motivation and literature review

The work on RQ1 developed a model to capture and structure the information represented in a given relational diagram so that it could be accessed in audio. The work on RQ2 investigated whether presenting the developed model in audio provides a practical means for inspecting and constructing such information. Inspecting diagrams in audio was considered to be the ability to search through and correctly interpret auditory presentation of their content, while constructing diagrams in audio was considered to be the ability to both inspect and alter their content. Focus was placed on the activities of inspecting and constructing diagrams because these are essential for successful participation in collaborative interaction with shared representations (Cherubini et al., 2007). Evaluating the effectiveness of the developed model for supporting such activities was thus necessary before the question of collaborative use could be addressed.

Previous work has investigated the use of audio-based hierarchical models to inspect a variety of visual displays, such as GUIs (Mynatt, 1995), phone-based menus (Brewster, 1997; Leplatre and Brewster, 2000) and graph-based diagrams (Bennett, 1999; Brown et al., 2004), but no previous work has investigated the use of hierarchical models to support constructing and editing diagrams in audio.

9.2.2.2 Original work and Contributions

Chapter 4 presented two user studies, recapped here, that evaluated the effectiveness of the developed model in supporting audio-only interaction with diagrams. Study 1 focused on inspecting relational diagrams in audio, and Study 2 focused on their construction. The obtained results were thoroughly analysed and discussed in Chapter 5.

Study 1 – Audio-only diagram inspection:

Section 4.2.2 described two presentation modes that were developed as part of the work on RQ2 to display the hierarchy in audio. The two modes differed in the amount of speech used to convey the hierarchically structured diagram content. A *high-verbosity mode* used only speech output to display diagram content, while a *low-verbosity mode* used a mixture of speech, earcons and parameterised auditory icons. This thesis contributed a detailed study that compared the effectiveness of the hierarchy-based model for supporting the inspection of relational diagrams when presented using the two auditory presentation modes. Sighted users completed tasks that required them to inspect the content of UML Class diagrams by navigating through the hierarchy, locating information of interest and correctly interpreting the displayed audio.

The multiple perspective hierarchy. Section 5.2.1 examined the usability of the hierarchy-based model in supporting inspection tasks. The scale type indexing and grouping of diagram elements allowed users to anticipate where specific diagram content would be located on the hierarchy.

For example, when asked to locate information about a particular relation on a diagram, users switched to the *relations' perspective* of the hierarchy and searched through the list of items on the corresponding branch. This ability to interactively explore the hierarchy allowed users to control the flow of information to match their needs and helped them orient themselves within the hierarchy. The semi-fixed structure of the hierarchy also aided the task of orientation by keeping navigation possibilities at a manageable level regardless of the complexity of the diagram being inspected.

The audio output. Section 5.2.2 examined users' reactions to the sounds used to display the hierarchy. Not all sounds were equally useful for supporting diagram inspection. In particular, sounds that alerted users to the occurrence of unexpected events, such as reaching the end of a list or making interaction errors, were found to be most informative and were actively used for orientation within the hierarchy. On the other hand, sounds that communicated feedback about events that users expected to occur, such as expanding or collapsing a branch on the hierarchy, were mostly appreciated for the added aesthetics.

The non-speech sounds used to display diagram content in the low-verbosity presentation mode were well received, and were particularly intuitive when communicating relations' types directions and attributes. Continuous ambient sounds, which were designed to communicate contextual information, were found to be more effective in conveying transitional information. That is, users were most aware of such sounds and found them to be most useful when they switched between perspectives rather than when they were within a particular perspective. In general, the level of verbosity of the audio presentation mode did not impact the time it took participants to complete inspection tasks. Instead, this was found to depend on users' strategies for going about completing such tasks.

Learning. The thesis contributed a method for analysing how participants develop expertise in using the hierarchy-based model to inspect diagrams. The analysis used the concept of *interaction traps* (Blandford et al., 2003) to examine and compare the efficiency of executing diagram inspection tasks over a number of scenarios involving the same set of inspection tasks but different and increasingly complex diagrams. The efficiency with which users executed an inspection task was determined by analysing whether the user deviated from following the optimal hierarchical path for locating and retrieving diagram content from the hierarchy. A user's performance on an inspection task was classified as either *efficient* if they followed the optimal path; *inefficient* if they deviated from it; or *less efficient* if they followed it but made interaction errors along the way. Using this classification, Section 5.2.3 examined users' expertise development and learning patterns over three inspection tasks and across four different scenarios. The method captured the rate of users' efficiency in completing inspection tasks, which increased from one scenario to the next. The rate of increase was equivalent across the two audio presentation modes. This method also served as a means for identifying causes for inefficient and less efficient interactions, which were often the result of misinterpreted auditory cues that conveyed hierarchical position and navigation possibilities.

Study 2 – Audio-only diagram construction:

Section 4.3.2 described two interaction strategies that were developed as part of the work on RQ2 to support the construction of diagrams through the hierarchy-based model. This is an original development in the use of hierarchical models to support the editing of graphically represented information. Prior to this, similar research only focused on supporting inspection activities. The two strategies differed in the ways in which they combine or isolate inspecting and editing modes of interaction. A *Guided* interaction strategy isolates the two modes and assists the user in executing editing actions by guiding them through the steps necessary to complete them. On the other hand, a *Non-Guided* strategy combines inspection and editing modes of interaction, ceding control to the user to act directly on the hierarchy when executing editing actions. This thesis compared the effectiveness and usability of the two strategies in supporting the ability to construct and edit relational diagrams. Sighted users used the hierarchy-based model to construct Entity-Relationship diagrams from text-based descriptions, and their performances and preferences were thoroughly analysed.

Usability issues. Section 5.3.1 addressed aspects of users' performances on diagram construction tasks in terms of completion time and diagram comprehension. There was a difference between the amount of time it took to construct diagrams when using the two interaction strategies, with users spending significantly longer times with the Guided strategy. This was found to be a consequence of the way in which editing actions are specified in each strategy. Additionally, isolating inspection and editing into two distinct modes in the Guided strategy made the execution of editing actions more flexible as users did not need to be at a specific location on the hierarchy in order to edit it. This, however, made it difficult for users to integrate newly added and existing diagram content. The Non-guided strategy situated participants within the content of the diagram being edited, affording integration of new and existing content, but at the cost of restricting movement within the hierarchy. These differences, however, did not affect users' comprehension of the constructed diagrams.

Users' preferences were used to categorise editing actions into two types; *Global* and *Local* actions. Section 5.3.2 discussed the usability of the Guided and Non-guided interaction strategies for supporting these two types of editing actions. Global editing actions affected higher levels of the hierarchy and no particular preference was noted towards either strategy when executing such actions. However, the Non-guided strategy was preferred for executing Local editing actions, which affected deeply nested nodes in the hierarchy. This was because the Non-guided strategy matched the systematic way in which users planned to execute editing actions that targeted those parts of the hierarchy. The Guided strategy was found to be cumbersome for executing Local editing actions because it involved longer editing steps, which made it difficult for users to keep track of where they were at in an editing process, often getting lost, disoriented and frustrated.

Interaction modes. The Guided strategy isolated inspection and editing actions into two distinct modes of interaction and switching between the two modes made users prone to mode errors. A detailed analysis of such errors was presented in Section 5.3.3. Mode errors occurred when participants got distracted from the construction task, mainly by shifting their attention between manipulating the diagram and reading through the text-based descriptions, and were due to the inadequacy of the audio cues to convey mode information. Potential improvements to the sound design were suggested, for instance, by developing more explicit speech messages, or adapting ambient contextual sounds to convey transitions between modes rather than just between perspectives.

9.2.3 Design Lessons from Study 1 and 2

The following design lessons were compiled throughout the analysis and discussion of the results obtained from Studies 1 and 2 as described in Chapter 5, page numbers are included for reference to the context from which each lesson was derived:

Design Lesson 1 – Use grouping by type to structure information. When possible, enforce a semi-fixed structure of the hierarchy by grouping items together by their categorical types, with the parent node labelled to reflect its content. Also, shortcut commands should be provided to allow users to jump to the main parent types from anywhere on the hierarchy. (See Page 88).

Design Lesson 2 – Allow for active control of the display. The user should be allowed to control which information is auditorally displayed by actively browsing the hierarchy. Content information should thus only be displayed when the hierarchy is interrogated by the user. (See Page 89).

Design Lesson 3 – Hierarchical path selection. To further support orientation within the hierarchy, the selection of a hierarchical path should only be changed in response to a user's action or when accompanied by a detailed message that explicitly highlights the occurrence of a change and the new position on the hierarchy. (See Page 89).

Design Lesson 4 – Convey three types of information. Three types of information should be communicated when a hierarchy-based model is presented in audio; *content*, *navigational*, and *contextual* to capture the richness of information that is contained in a hierarchical structure. (See Page 92).

Design Lesson 5 – Emphasise the occurrence of unexpected events. The occurrence of unexpected events should be explicitly highlighted; for example users in Study 1 found the sounds used to highlight reaching the end of a list and the occurrence of an illegal move particularly useful for orientation within the hierarchy. Sounds communicating feedback about expected events (e.g. moving between nodes, expanding and collapsing branches) should not be excluded unless they interfere with other sounds in the interface. (See Page 93).

Design Lesson 6 – Use non-speech sounds to display iconic content. Using timbre to display relation types and varying the order of the short and long sounds to display a relation's direction was more intuitive than using speech to display iconic content. Where possible, describing iconic content of relational diagrams in speech should be avoided and replaced with equivalent non-speech sounds. (See Page 95).

Design Lesson 7 – Avoid displaying context information using spoken descriptions. Contextual information should be conveyed through less intrusive means than spoken descriptions, particularly when lengthy messages need to be displayed. In the case of our evaluations, lengthy context messages may push users to avoid requesting such information all together. (See Page 96).

Design Lesson 8 – Use ambient sounds to convey hierarchical perspectives. Users in Studies 1 and 2 were more aware of the continuous ambient sounds at points of the interaction where they switched perspectives. Mapping the timbre of such sounds to match the hierarchical branch was thus an effective means to convey context, but the ambient sounds were more effective at communicating transitional information and should therefore be gradually faded out to a minimum amplitude when user movements are limited within a single branch. (See Page 96).

Design Lesson 9 – Constrain navigational possibilities to reflect context. Movements on the hierarchy should be constrained depending on where the user is located on the hierarchy. In particular, movements between cousin nodes should be disabled such that a user loops to the first child of a list rather than move to the next cousin node when reaching the end of such a list. Movement to parent nodes should also be constrained; to avoid confusion, the interface controller used to move to parent nodes should be different to the controller used to move to sibling nodes. For example, if a 4-way navigation controller is used to navigate the hierarchy such as a joystick or the keyboard cursor keys, then the functions of the keys for moving within the hierarchy should be mapped to match the layout of the hierarchy. (See Page 98).

Design Lesson 10 – Combine aspects from the Guided and Non-guided strategies to support editing diagrams through a hierarchy. This could be achieved by tracking users editing actions and position on the hierarchy, the sequence of guiding steps could then be triggered if the action and location bare no relevance to one another; the non-guided process should proceed otherwise. Additionally, allow users to specify multiple edits to the same local area of the hierarchy when using the Guided strategy. (See Page 106).

Design Lesson 11 – Manage transitions between interaction modes explicitly. The Guided strategy breaks the interaction into two independent modes; an editing mode and an inspection mode. Care must therefore be taking to prevent the user from falling into mode confusion. Explicit auditory cues should be designed to convey mode information; for instance, using a distinctive continuous ambient sound that conveys mode status, or other cues that are contingent to each mode, such that a user hears an accompanying sound with every keystroke to correspond to the current mode of interaction. (See Page 111).

9.2.4 Research Question 3

"RQ3: What are the characteristics of collaborative interaction with diagrammatic representations when audio is the only medium of interaction?"

9.2.4.1 Summary of motivation and literature review

The work on RQ2 developed and evaluated auditory means for supporting individual interaction with diagrammatically represented information. This was a necessary step to complete before arriving at a stage where an informed study of collaborative audio-only interaction with diagrams could be conducted. Having gained insight into how diagrams can be translated, accessed and manipulated in sound, the work on RQ3 focused on investigating how pairs of users collaborate to construct diagrams in an audio-only workspace.

Audio has previously been used to augment computer-supported collaborative environments with both open channels for verbal communication and auditory representations of events that occur in such environments (e.g Gaver et al., 1992; Olson et al., 1995; Mynatt et al., 1998). Previous research has reported that the addition of sound for these purposes benefits collaborations by increasing collaborators' awareness of their interactions with each other and with shared artefacts (e.g. Hindus et al., 1996; Gaver et al., 1991; Cohen, 1993). The concept of *awareness* is often highlighted as a crucial element of successful collaboration (Dourish and Bellotti, 1992; Gutwin and Greenberg, 1996) and additional evidence points to the potential impact of the means for delivering audio to shared workspaces on the support for awareness in collaboration (e.g. Morris et al., 2004; McGookin and Brewster, 2007b). The above evidence is based on studies of collaboration where audio is used as an additional modality to otherwise multimodal shared spaces. It therefore remains unclear whether and how such benefits and impact extend to situations where collaboration takes place in audio-only settings, or indeed whether, by itself, audio is a practical means for supporting collaborative interaction.

9.2.4.2 Original Work and Contributions

The thesis contributed an original study that examined collaborative interaction with diagrammatically represented information where audio is used as the sole means for mediating communication and representing shared artefacts. Chapter 7 described how the hierarchy-based model was extended to support simultaneous access and manipulation of diagrams to serve as a shared audio-only workspace. Pairs of users used the model to construct Entity-Relationship diagrams in a lab setting where they could only communicate with one another verbally. The audio output of the hierarchy-based model was delivered to the workspace in one of two ways; through speakers in a *Shared* setting, such that collaborators could hear each others' interactions with the model; and through headphones in a *Non-shared* setting, such that each user could only hear their individual interaction with the model.

Overall, the results of the study showed that pairs of sighted users are able to collaboratively construct relational diagrams using sounds as the sole means for communicating with one another and for accessing and manipulating a shared diagram. The results also showed that varying the means for delivering audio to an audio-only workspace affected how users structured and organised their work and how they exchanged workspace awareness information to maintain their collaborations. The thesis has made two specific contributions in this regard. The first is characterising the audio-only collaborative process by capturing which activities collaborators engaged in, for how long, and how often they switched activities. The second is a detailed description of the workspace awareness information exchanged during the collaborations. To this end, the thesis used Gutwin and Greenberg (2002)'s workspace awareness framework to establish; 1) exactly which information is supplied or requested by collaborators from one another and which information is extracted from the available audio output; 2) how such information is used in collaboration; and 3) how this changes when varying the means for delivering audio to the audio-only workspace.

Additionally, Section 7.5.4 examined collaborators' patterns of interaction with the shared workspace and distinguished between two working styles that were used for collaboration; a *Parallel* style, in which pairs worked concurrently; and a *Sequential* style, in which they took turns contributing to the shared workspace. When retrospectively examined in Sections 7.5.5, 7.5.6 and 7.5.7, the impact of the means for delivering audio on the collaborative process and the exchange of workspace awareness information was found to be different for each working style group. The obtained results were thoroughly analysed and discussed in Chapter 8.

Structure and organisation. Section 8.2 analysed the audio-only collaborative process. In general, the means for delivering audio changed the way in which collaborators went about their work in terms of when they manipulated shared diagrams and what they discussed during such manipulations. This was further amplified by the choice of working style; when collaborators worked sequentially, they were more likely to have joint attention over the interactions that surrounded the construction process because they relied on a single audio output source. However, when they worked in parallel – regardless of whether they used speakers or headphones – they focused on independent actions and thus on separate audio output sources, which decreased the number of potential items that might have triggered joint attention and discussion.

Workspace awareness exchange. Section 8.5 examined workspace awareness information exchange and how collaborators used such information to support their collaborations. When information about one's partner's activity and progress was readily available in a Shared audio-only setting, the use of such information depended on the working style employed by the collaborators. When working in parallel, collaborations were *loose* and consequently collaborators felt a greater need to know about each other's past actions and frequently supplied each other with such information in the form of explicit *updates* about which actions occurred in the workspace. On the other hand, collaborations were *focused* when the collaborators worked sequentially, in which case they often supplied each other with information in the form of *descriptions* of what was currently happening in the workspace or what was about to happen in the immediate future.

Further, when collaborators worked sequentially in the Shared setting, they extracted as much awareness information as possible from the speakers. Once extracted, they used this information to detect individual and collaborative errors, to guided their partner around the hierarchy or through the execution of an editing action, and to coordinate collaborative actions. Furthermore, when collaborators worked sequentially, they used the shared audio output as a shared resource around which collaborative action was organised and executed. By interactively controlling the audio display of various parts of the hierarchy, collaborators were able to draw each other's attention to areas of interest, to collaboratively explore diagram content, and to structure their tasks in synchrony with the shared representation. When audio was not shared in this way – either because pairs worked in parallel or used headphones – collaborators often used alternative strategies to make up for the lack of awareness. Examples of alternative strategies included following each other around the workspace seeking auditory feedthroughs, or synchronising their interactions while exchanging verbal instructions.

9.2.5 Design Lessons from Study 3

The following design lessons were compiled throughout the analysis and discussion of the results obtained from Study 3 as described in Chapter 8, page numbers are included for reference to the context from which each lesson was derived:

Design Lesson 12 – Detect guidance of partners and supply awareness information accordingly. An audio collaborative system should be able to either detect guidance modes of interaction or allow users to manually switch to such a mode. If using headphones, the guidance mode should explicitly convey and/or allow users to explicitly track awareness information of types *"What Did You Do"*, *"What Are You Doing"* and actions' *"Completion Status"*. (See Page 188).

Design Lesson 13 – Provide collaborators with both a private and a public workspace area and the ability to switch between them. Collaborators should be able to work privately, such that they control whether their partners can hear their output, but also be able to expose their audio to their partner, for example, to support guidance interaction. If the collaboration involves more than two people, then collaborators should be provided with a means to select a workspace to switch to from available users, or invite specific users to access their workspace. (See Page 190).

Design Lesson 14 – Provide a log of collaborators' contributions to the shared task. Although the ability to keep track of self and partner's contributions to the task was not a factor that could impact performance on or the outcome of the studied task, it is clear that both the choice of working style and the means for delivering audio to the workspace had an impact on such ability. In situations where such ability is crucial to the task at hand, the collaborative tool should be designed to convey awareness information about which action took place and by which user, i.e. of types *"What Did You Do"/ "What I Did"* or allow users to requested such information, for example, by browsing a timed log of actions. (See Page 193).

Design Lesson 15 – Types of workspace awareness information when using headphones. If headphones are used during collaboration, the collaborative system should be design to convey workspace awareness information of types *"What I Did"/"What Did You Do"*, *"What I Am Doing"/"What Are You Doing"* and actions *"Completion Status"*. Alternatively, the system should be designed to allow users to manually request such information. (See Page 194).

Design Lesson 16 – Adapt both the type and amount of workspace awareness information to match collaborators' working style. When collaborative interactions occur in parallel, the system should provide users with updates about each others actions, i.e. emphasising information of types *"What I Did"/"What Did You Do"* and actions *"Completion Status"*. The collaborative system should also be designed to detect sequential interaction and reduce the amount of workspace awareness information it conveys to users, since this information can be redundant in such a case. Alternatively, the system should provide users with a means to control the amount of awareness information that is conveyed to match their needs. (See Page 198).

Design Lesson 17 – Provide support for guidance interaction when using headphones. In addition to the recommendation outlined in *Design Lesson 12* above, when audio is delivered through headphones, collaborative systems should provide workspace awareness information of types *"Where Are You"* and *"What Have You Done"* if guidance mode is detected or triggered by the users. (See Page 212).

Design Lesson 18 – Provide users with a means to display awareness information of their

choice to their partners' workspace "views". In order to support coordination, particularly when collaborators work in parallel or use headphones, users should be provided with a means for choosing which awareness information to display as well as when such information is displayed to their partners. (See Page 214).

Design Lesson 19 – Provide a means for pointing to items on the shared workspace. Audio collaborative systems should include a mechanism for users to highlight items on the workspace and draw partners' attention to them. This could be achieved by, for example, giving users control over the audio output that is displayed to their partners as outlined in *Design Lesson 13*. A mechanisms for handling conflicts of audio output should be included with such an implementation. (See Page 219).

Design Lesson 20 – Provide a means for following partners' interactions in a workspace. Audio collaborative systems should provide users with the ability to allow their partners to control the audio output that they receive, particularly when using headphones, i.e. handing the control of their display to their partners. Just as *Design Lesson 18* recommended allowing users to displays their audio output to their partners, this should allow users to monitor their partners activity. (See Page 224).

9.3 Outstanding Issues and Future Work

As is the case with any research endeavour, the presented work has generated more questions than it provided answers. Despite the above mentioned contributions, there are limitations associated with the research conducted to address the questions that the thesis set out to investigate, which also translate into future avenues for further research to expand on the reported findings.

The hierarchy-based model. The proposed approach of using multiple perspective hierarchies to capture and structure information was found to be effective for supporting audio-only interaction relational diagrams. Providing access to the same information from more than one perspective has previously been reported to be beneficial in non-visual interaction (Bennett, 2002; McGookin and Brewster, 2006b). However, it is not clear what advantages are exactly offered by the multiple perspective aspect of the developed model when compared to other approaches that use hierarchies to support non-visual interaction with graphically represented information. Future work should explore how the hierarchy-based model compares to other hierarchies, particularly those which employ factors other than scale types to organise and present information from one rather than multiple perspectives.

Additionally, the hierarchy-based model translated relational diagrams that encoded nominal data. The audio display used to present the hierarchy was thus limited to conveying this particular type of data. The question remains as to how to use the model to capture and display diagrams that encode other types of data. It is easy to image how the hierarchical structure could be overlaid with auditory display techniques that convey data dimensions associated with a specific item within a given category of items. For instance, encountered items on a particular branch of the hierarchy could be accompanied by sonifications that convey quantitative dimensions associated with them. Applying and evaluating the hierarchy-based model with diagrams that represent data on a variety of scale types is thus another important venue for future work.

The audio display. The audio display of the model could be improved in a number of ways. The first study, which compared high-verbosity and low-verbosity auditory displays, found no significant differences between task completion times, comprehension and learning patterns. The study also reported some issues related to navigation and orientation within the hierarchy. Two questions remain unaddressed. First, do the reported results hold when speech output is used to convey not only content but also navigational and contextual information in the high-verbosity display, or when other data types are presented using non-speech sounds? Second, how can other auditory display techniques improve the interaction with the hierarchy? For example, earcons have been shown to be an effective means for supporting navigation of hierarchical structures (Brewster, 1998), and spearcons have been found to increase the speed and accuracy of locating items within menus (Walker et al., 2006). Future research should explore how such techniques could be incorporated within the hierarchy-based model to improve performance.

Evaluations. The user studies focused on evaluating performance of exploring and retrieving information from the hierarchy. It is not clear how the hierarchy-based model supports other types of tasks, particularly those requiring spatial information. It could be predicted that the model would perform poorly when supporting such tasks since information about spatiality is discarded during the developed translation approach. Another interesting avenue for future research is thus exploring how the spatial information of a diagram can be encoded within a hierarchical structure. Existing research has addressed this question in two main ways; augmenting the nodes on a hierarchy with coordinates information (Bennett, 2002), and using earcons to annotate or convey an overview of the content and structure of a diagram (Brown, 2008). An alternative method could display items encountered on the hierarchy using 3D audio to reflect their spatial location on the original diagram. 3D sound has already been explored to display menu-based interfaces (Savidis et al., 1996; Brewster et al., 2003; Frauenberger and Noisternig, 2003), and could therefore be a potentially effective means for conveying the spatial properties of a diagram when captured through a hierarchical structure.

Another outstanding issue associated with the evaluations is the analysis of learning patterns that was conducted as part of the work on RQ2 in Part I of the thesis. The analysis used a classification that examined users' interaction logs to identify the hierarchical paths followed by each user when retrieving information from the hierarchy. These logs were then classified in terms of efficiency by assessing whether paths deviated from optimal trajectories. However, the author conducted this classification alone, and future work should develop such a method into a scheme that could be used by multiple investigators to classify users' interactions.

Lab-based versus the real-world. While the evaluation of the hierarchy-based model has demonstrated its practicality for supporting audio-only interaction with relational diagrams, a number of scalability issues remain unaddressed. The evaluations were conducted in laboratory conditions, which does not demonstrate the practicality of the developed model in real-world contexts of interaction where complex diagrams are often encountered. The evaluations focused on small to medium size relational diagrams (up to seven nodes and seven relations). As described in Section 5.2.1, the developed model uses a semi-fixed hierarchical structure that expands in breadth but not depth. It could therefore be predicted that the model could scale up to diagrams representing a large number of categorical scale types, where each type would be represented with a distinct branch on the hierarchy. In such a case, a means for supporting more efficient navigation within the resulting complex structure would be necessary, for instance by introducing bookmarking features and user defined shortcuts. Future work should therefore investigate performance on more complex diagrams.

Collaborating through sounds. The reported findings of audio-only collaboration are limited to instances where pairs of users collaborate to construct diagrams. It remains unclear if such findings would extend to collaborations involving more than two members. Certain aspects of the observed findings are likely to reoccur in such scenarios, such as collaborators choice of working styles and its influence on the collaborative process and dynamics of interaction. For instance, collaboration between a large group of users could break up into smaller subgroups who focus on subtasks. As highlighted by the design lessons outlined above, audio collaborative systems should be designed to automatically detect or allow users to manually switch to working styles modes between subgroups with the larger user group and convey appropriate workspace awareness information to match the dynamics of each subgroup. However, such postulations should be empirically explored, and an important avenue for future research is therefore to explore the practicality of the developed model for supporting collaborations between larger groups of users.

Existing research on supporting non-visual and multimodal collaboration focuses on employing spatial models to support interaction. This thesis demonstrated that hierarchical models are also a practical means for supporting audio-only collaboration. The question remains as to how the two models compare in terms of supporting collaborative interaction. What are the advantages and disadvantages associated with each model? How does this change if hierarchical models are augmented with 3D auditory displays or other means to convey spatiality? How can they be combined for optimal non-visual interaction? Additionally, the assumption that a spatial model is better suited to support collaboration involving visual and non-visual displays can now be informatively evaluated. This is a particularly interesting venue for future research since real world collaborative scenarios are likely to involve collaborators who have differing access to modalities due to situational or perceptual impairments, or because they use devices with differing display capabilities to connect to shared spaces.

Awareness through sounds. Part of the work on RQ3 addressed the question of which awareness information is exchanged, extracted and used in audio-only collaboration. The findings are limited to only two audio-only settings in which audio is delivered to the workspace through headphones or speakers. Future work should explore other audio-only settings. Example of alternative configurations include connecting headphones so that users can hear their partner's interactions; using bone conduction headphones; using speakers arranged so that collaborators have explicit control over their own and their partner's output; and limiting audio output so that only one output can be heard at any given moment, much like sharing a pen to edit a shared whiteboard, to name but a few possibilities. Establishing an understanding of which awareness information is exchanged and used in collaboration, as contributed by this thesis, is only the first step in exploring how awareness in audio-only interaction, and collaboration in general, can be supported by auditory means. An important way to further the findings reported in this thesis is thus to develop and evaluate auditory display techniques that convey awareness information in audio-only or multimodal and cross-modal environments.

9.4 General Implications and Conclusion

The following concludes this thesis by reflecting on some of its major results and contributions and their general implications to the fields of Auditory Display, HCI and CSCW.

9.4.1 Using Hierarchies for Audio-only Interaction

The work presented in this thesis showed that information represented in some relational diagrams is not "inherently" visual and could be captured and translated into the auditory modality. Instead, these relational diagrams are "conveniently" visual as they exploit the highly sophisticated visual sense to encode and convey information. Translating a diagram into an alternative modality should therefore be driven by both understanding the underlying nature of the information it represents and considering the strengths and weaknesses of the target modality.

One of the main issues associated with using the auditory modality in HCI is the transient character of sound. A diagrammatic representation is persistent in space, therefore serving as an external extension to memory and other cognitive processing mechanisms. Indeed, as reviewed in this thesis, many of the advantages associated with using diagrams as a form of representation stem from the persistent character of graphical display. The work presented in this thesis showed that *interactivity* can play an important role in – at least partly – compensating for the transient nature of sound. The ability to play and replay an audio output allows the information it conveys to be revisited over and over again, thus playing a similar role to that of gazing at the same part of an interface more than once. As observed in the individual and collaborative studies reported in this thesis, this level of control allows users to interact with sounds in ways that are important to the task at hand, such as drawing a partner's attention by "pointing to" objects of interest.

But while the ability to control the replay of audio in a user interface is important, it might not be enough if the auditory message is too long. For example, in the context of the reported research, playing and replaying a spoken description of a complex diagram might be inconvenient to the user due to the linear nature of presentation. Linearity is also one of the main disadvantages associated with using screen-reader technology to access modern visual displays (Stockman and Metatla, 2008). A hierarchical organisation of information can overcome the disadvantages associated with the linearity of its auditory presentation. The use of scale type indexing instead of location indexing in the reported work helped to change the nature of the task of searching and locating diagrammatically represented information when these are accessed in audio. Instead of searching in space, users search through lists, grouped together to capture the relational information in the original representation.

The strength of hierarchically structuring information is thus in the organisation it imposes on the captured information which, when interactively accessed, can serve two functions that are important in any human-computer interface; 1) ceding control to the user to choose which parts of the information to interact with at any given moment, and 2) aiding orientation within the displayed information. The question of how information should be hierarchically organised is then another crucial aspect of using hierarchies to support auditory interaction with information. Finer control over which audio is displayed empowers the user who, instead of passively receiving information, can actively explore and search through an information space, much like gazing around a graphical display to examine different parts of the information it represents.

9.4.2 Collaborating Through Sounds

The work presented in this thesis showed that audio, by itself, can be used as a shared resource to support collaborative interaction with diagrams. The naturalness and efficiency of a collaboration is influenced by the degree to which collaborators are aware of the events that occur in a shared workspace including each other's interactions with it. Addressing the awareness problem in groupware design is an important and difficult problem and part of the solution is to provide collaborators with more information about what is going on in the shared space (Gutwin and Greenberg, 2002). But designers must carefully determine which information should be conveyed. This is particularly crucial when audio is used to support collaboration because sound can be annoying when it does not convey useful or relevant information (Buxton, 1989). Indeed, annoyance and interference are two issues that often stand in the way of exploiting the auditory modality in HCI (Frauenberger et al., 2007), and "relevance" is a determining factor between a good and a bad representation (Zhang, 1991; Norman, 1993).

While previous research has highlighted the important role that audio plays in conveying awareness in collaborative interaction, the work presented in this thesis went a step further to establish exactly which information is relevant in an audio-only collaboration in order to support adequate levels of awareness. The findings reported in this thesis are thus useful when designing explicit support for awareness in an audio-only collaborative environment or in a shared environment augmented by the auditory modality.

Overall, the presented work questioned and sought answers to the meaning of *sharing* an auditory representation as the sole resource in collaborative interaction. One of the important conclusions of the reported work is the insight that the mere physical presence of audio in a shared space does not necessary imply that it is being "shared" by those present in such a space. This might seem to be an obvious point to state, but it has crucial implications for the relevance of the awareness information identified to be important in an audio-only collaborative space. Essentially, while the reported work identified relevant awareness information, it also showed that such relevance is not static, but dynamically changes according to how collaborators choose to work with sounds. The choice of working style affects how collaborators attend to the sounds present in a collaborative space, which in turn influences how they structure and organise their interactions, and hence which information is rendered relevant to their collaboration and which information is not. Designers must therefore cater for these dynamics when designing auditory support for awareness in collaboration.

Appendix A

Study 1 Materials – Inspecting Diagrams In Audio

Introduction

This appendix contains the materials used in Study 1, which examined audio-only inspection of UML CLass diagrams. The appendix includes the pre-test questionnaire, the diagram inspection tasks, and the collected raw data that is reported as part of Chapter 4.

A.1 Information Sheet

Hearing Diagrams Study - Information Sheet

INFORMATION SHEET

YOU WILL BE GIVEN A COPY OF THIS INFORMATION SHEET

Title of Study: Investigating Strategies for Inspecting Diagrams in Audio-Only Environments

We would like to invite you to participate in this original research project. You should only participate if you want to; choosing not to take part will not disadvantage you in any way. Your decision will not affect your access to treatment or services. Before you decide whether you want to take part, it is important for you to understand why the research is being done and what your participation will involve. Please take time to read the following information carefully and discuss it with others if you wish. Ask us if there is anything that is not clear or if you would like more information. If you do decide to take part, please let us know beforehand if you have been involved in any other study during the last year.

Description:

This study is an evaluation of an audio-only interface for inspecting UML Class diagrams. We have developed a number of ways to create, navigate, and edit a ER diagram using only the computer keyboard and computer speakers, and we would like to test how efficient these techniques are for allowing interaction with diagrams.

If you choose to participate in this study you will be expected to attend a session which last for an hour and a half to two hours at the Department of Computer Science, Queen Mary University. You will be given a short training to familiarise yourself with the application that you will be using. Once you feel comfortable using the application, you will be asked to complete a number of inspection tasks on four diagrams, one at a time. You will also be asked to fill a short questionnaire at the beginning of the session to tell us about yourself and you will be interviewed at the end of the session to tell us about your experience using the application.

We would like you to remember at all time that we are evaluating the application that you will be using rather than your skills in using it.

As part of the experiment, we will log how you use the system and video record you as you use it. All records will be anonymised.

You will receive a payment of £15 for your participation in this study.

Anyone can apply to take part and there are no risks associated with this study. Although, it is unlikely that you will be asked any personal questions, any information that you disclose will be strictly confidential and any information collected will be handled in accordance with the provisions of the Data Protection Act 1998.

‘It is up to you to decide whether or not to take part. If you do decide to take part you will be given this information sheet to keep and be asked to sign a consent form. If you decide to take part you are still free to withdraw at any time and without giving a reason.

In the event of you suffering any adverse effects as a consequence of your participation in this study, you will be compensated through Queen Mary University of London’s ‘No Fault Compensation Scheme’.

Figure A.1: The information sheet provided to participants at the beginning of the sessions in Study 1.

A.2 Pre-test Questionnaire

Hearing Diagrams Study - Pre-test Questionnaire

PRE-TEST QUESTIONNAIRE

Please fill in this short questionnaire before the start of the session. Please answer all questions and hand back to the observer when you are finished.

1. What is your name?

____

2. Please rate your expertise in UML modeling. *(Tick as appropriate)*

None [ ] Beginner [ ] Intermediate [ ] Expert [ ]

3. Please rate your expertise with using screen-readers. *(Tick as appropriate)*

None [ ] Beginner [ ] Intermediate [ ] Expert [ ]

4. Do you play/ed a musical instrument, read music scores? *(Tick as appropriate)*

None [ ] Beginner [ ] Intermediate [ ] Expert [ ]

Please give more details if other than none: ____

End of Pre-test Questionnaire. Thank You.

Figure A.2: The pre-test questionnaire that collected basic information about participants in Study 1.

A.3 Tasks

Hearing Diagrams Study - Tasks

TASKS

Scenario 1: Animals

Task 1:
Find the connections of the class SHEEP
What is the type of each connection and to which other class it links?________________

Task 2:
Find the classes connected through EATS ________________
What is the direction of EATS? From ________________ To ________________

Task 3:
How many classes are there in the diagram?________________
How many relations are there in the diagram?________________

Scenario 2: Geometry

Task 1:
Find the connections of the class POINT
What is the type of each connection and to which other class it links?________________

Task 2:
Find the classes connected through CONTAINS ________________
What is the direction of CONTAINS? From ________________ To ________________

Task 3:
How many classes are there in the diagram?________________
How many relations are there in the diagram?________________

Scenario 3: Bookshop

Task 1:
Find the connections of the class COPY
What is the type of each connection and to which other class it links?________________

Task 2:
Find the classes connected through REGISTER ________________
What is the direction of REGISTER? From ________________ To ________________

Task 3:
How many classes are there in the diagram?________________
How many relations are there in the diagram?________________

Scenario 4: Taxi Company

Task 1:
Find the connections of the class VEHICLE
What is the type of each connection and to which other class it links?________________

Task 2:
Find the classes connected through DELIVER ________________
What is the direction of DELIVER? From ________________ To ________________

Task 3:
How many classes are there in the diagram?________________
How many relations are there in the diagram?________________

Figure A.3: The diagram Inspection tasks used in Study 1.

A.4 Raw Data

Study 1 - Task Completion Times (seconds)

Task 1		Task 2		Task 3	
	High-Verbosity		High-Verbosity		High-Verbosity
P1	45.5	P1	29.75	P1	29
P2	45.5	P2	29.75	P2	29
P3	62	P3	31	P3	37.75
P4	62.25	P4	33	P4	26.25
P5	107	P5	87.5	P5	47
P6	81	P6	37.5	P6	30.75
P7	67.5	P7	78	P7	37.25
P8	72	P8	24.25	P8	37.5
P9	68	P9	45.75	P9	28.25
P10	52.5	P10	32.5	P10	22.75
	Low-Verbosity		Low-Verbosity		Low-Verbosity
P11	38.5	P11	13.25	P11	23
P12	88.75	P12	50.25	P12	27.75
P13	56	P13	42.33	P13	15.67
P14	47.67	P14	53.67	P14	43.33
P15	65	P15	43	P15	23
P16	70.75	P16	51.75	P16	21
P17	64.25	P17	69.5	P17	17
P18	50.26	P18	19.25	P18	18.75
P19	81.67	P19	39	P19	25
P20	91	P20	48	P20	16

Study 1 - Scores on Diagram Reading Tasks (%)

Task 1		Task 2		Task 3	
	High-Verbosity		High-Verbosity		High-Verbosity
P1	87.5	P1	100	P1	100
P2	100	P2	100	P2	100
P3	95.75	P3	100	P3	100
P4	100	P4	100	P4	100
P5	78.25	P5	100	P5	87.5
P6	100	P6	100	P6	100
P7	100	P7	87.5	P7	100
P8	100	P8	100	P8	100
P9	81.25	P9	100	P9	100
P10	100	P10	100	P10	100
	Low-Verbosity		Low-Verbosity		Low-Verbosity
P11	100	P11	100	P11	100
P12	100	P12	100	P12	100
P13	94.33	P13	100	P13	100
P14	100	P14	100	P14	100
P15	99	P15	96	P15	99
P16	100	P16	100	P16	100
P17	100	P17	87.5	P17	100
P18	100	P18	100	P18	93.75
P19	100	P19	83.33	P19	100
P20	100	P20	100	P20	100

Study 1 - % Interaction Efficiencies Per Scenario Averaged Across Tasks

High-Verbosity				
	Scenario 1	Scenario 2	Scenario 3	Scenario 4
Inefficient	26.66	6.66	0	3.33
Less Efficient	40	36.66	23.33	26.66
Efficient	33.33	56.66	76.66	70

Low-Verbosity				
	Scenario 1	Scenario 2	Scenario 3	Scenario 4
Inefficient	20	20	13.33	3.33
Less Efficient	43.33	10	36.33	20
Efficient	36.66	70	50	76.66

Figure A.4: Raw data from Study 1 showing task completion times and scores on the diagram inspection tasks per each participant and the proportions of interaction efficiencies in each scenario.

	Study 1 - Classification of Interaction Efficiencies Per Participant for each Task and each Scenario											
CONDITION 1	High-Verbosity		(i)	Inefficient		(L)	Less Efficient		(E)	Efficient		
Scenario 1	P1	P2	P3	P4	P5	P6	P7	P8	P9	P10	Totals	30
Task 1	(L)	(E)	(L)	(E)	(i)	(i)	(i)	(E)	(i)	(i)	Inefficient	8
Task 2	(E)	(E)	(L)	(E)	(i)	(L)	(L)	(E)	(i)	(E)	Less Efficient	12
Task 3	(L)	(L)	(L)	(L)	(i)	(L)	(L)	(L)	(E)	(E)	Efficient	10
Scenario 2												
Task 1	(E)	(i)	(L)	(E)	(L)	(E)	(E)	(L)	(L)	(E)	Inefficient	2
Task 2	(L)	(E)	(E)	(E)	(E)	(E)	(L)	(E)	(i)	(E)	Less Efficient	11
Task 3	(L)	(L)	(E)	(L)	(E)	(E)	(L)	(L)	(E)	(E)	Efficient	17
Scenario 3												
Task 1	(E)	(E)	(E)	(L)	(L)	(L)	(E)	(L)	(E)	(E)	Inefficient	0
Task 2	(E)	(E)	(E)	(E)	(E)	(E)	(E)	(E)	(L)	(E)	Less Efficient	7
Task 3	(L)	(E)	(E)	(E)	(E)	(E)	(L)	(E)	(E)	(E)	Efficient	23
Scenario 4												
Task 1	(E)	(E)	(E)	(E)	(L)	(E)	(L)	(E)	(L)	(L)	Inefficient	1
Task 2	(E)	(E)	(E)	(E)	(i)	(E)	(E)	(E)	(E)	(E)	Less Efficient	8
Task 3	(E)	(E)	(E)	(L)	(L)	(E)	(L)	(L)	(E)	(E)	Efficient	21
CONDITION 2	Low-Verbosity											
Scenario 1	P11	P12	P13	P14	P15	P16	P17	P18	P19	P20	Total	
Task 1	(E)	(i)	(L)	(E)	(L)	(i)	(L)	(L)	(L)	(L)	Inefficient	6
Task 2	(E)	(L)	(E)	(i)	(i)	(E)	(L)	(E)	(E)	(E)	Less Efficient	13
Task 3	(i)	(L)	(E)	(E)	(i)	(E)	(L)	(L)	(L)	(L)	Efficient	11
Scenario 2												
Task 1	(E)	(i)	(L)	(E)	(L)	(E)	(E)	(E)	(E)	(E)	Inefficient	6
Task 2	(E)	(E)	(E)	(E)	(i)	(L)	(E)	(E)	(E)	(E)	Less Efficient	3
Task 3	(E)	(E)	(E)	(i)	(i)	(E)	(i)	(i)	(E)	(E)	Efficient	21
Scenario 3												
Task 1	(L)	(E)	(i)	(E)	(L)	(L)	(L)	(L)	(E)	(E)	Inefficient	4
Task 2	(E)	(L)	(L)	(L)	(L)	(E)	(E)	(E)	(E)	(i)	Less Efficient	11
Task 3	(E)	(L)	(E)	(L)	(i)	(E)	(i)	(E)	(E)	(E)	Efficient	15
Scenario 4												
Task 1	(E)	(E)	(L)	(E)	(L)	(E)	(E)	(E)	(E)	(E)	Inefficient	1
Task 2	(E)	(L)	(E)	(E)	(L)	(E)	(E)	(L)	(E)	(E)	Less Efficient	6
Task 3	(E)	(E)	(E)	(L)	(i)	(E)	(E)	(E)	(E)	(E)	Efficient	23

Figure A.5: Raw data from Study 1 showing the interaction efficiencies of each participant on each diagram inspection task.

Appendix B

Study 2 Materials – Constructing Diagrams In Audio

Introduction

This appendix contains the materials used in Study 2, which examined audio-only construction and editing of ER diagrams. The appendix includes the information sheet, the pre-test questionnaire, the diagram textual descriptions, and the collected raw data that is reported as part of Chapter 4.

B.1 Information Sheet

Editing and Constructing Diagrams in Audio - Information Sheet

INFORMATION SHEET

REC Protocol Number QMREC2007/65

YOU WILL BE GIVEN A COPY OF THIS INFORMATION SHEET

Title of Study: Investigating Strategies for Editing and Constructing Diagrams in Audio-Only Environments

We would like to invite you to participate in this original research project. You should only participate if you want to; choosing not to take part will not disadvantage you in any way. Your decision will not affect your access to treatment or services. Before you decide whether you want to take part, it is important for you to understand why the research is being done and what your participation will involve. Please take time to read the following information carefully and discuss it with others if you wish. Ask us if there is anything that is not clear or if you would like more information. If you do decide to take part, please let us know beforehand if you have been involved in any other study during the last year.

Description:

This study is an evaluation of an audio-only interface for constructing and editing entity-relationship (ER) diagrams. We have developed a number of ways to create, navigate, and edit a ER diagram using only the computer keyboard and computer speakers, and we would like to test how efficient these techniques are for allowing interaction with diagrams.

If you choose to participate in this study you will be expected to attend a session which last for an hour and a half to two hours at the Department of Computer Science, Queen Mary University. You will be given a short training to familiarise yourself with the application that you will be using. Once you feel comfortable using the application, you will be asked to create up to two diagrams, one at a time. You will also be asked to fill a short questionnaire at the beginning of the session to tell us about yourself and you will be interviewed at the end of the session to tell us about your experience using the application.

We would like you to remember at all time that we are evaluating the application that you will be using rather than your skills in using it.

As part of the experiment, we will log how you use the system and video record you as you use it. All records will be anonymised and only your hands will show of the videos.

You will receive a payment of £15 for your participation in this study.

Anyone can apply to take part and there are no risks associated with this study. Although, it is unlikely that you will be asked any personal questions, any information that you disclose will be strictly confidential and any information collected will be handled in accordance with the provisions of the Data Protection Act 1998.

'It is up to you to decide whether or not to take part. If you do decide to take part you will be given this information sheet to keep and be asked to sign a consent form. If you decide to take part you are still free to withdraw at any time and without giving a reason.

In the event of you suffering any adverse effects as a consequence of your participation in this study, you will be compensated through Queen Mary University of London's 'No Fault Compensation Scheme'.

Figure B.1: The information sheet provided to participants at the beginning of the sessions in Study 2.

B.2 Pre-test Questionnaire

Editing and Constructing Diagrams in Audio - Pre-test Questionnaire

PRE-TEST QUESTIONNAIRE

Please fill in this short questionnaire before the start of the session. Please answer all questions and hand back to the observer when you are finished.

1. What is your name?

____

2. Please rate your expertise in Entity-Relationship modeling. *(Tick as appropriate)*

None [ ] Beginner [ ] Intermediate [ ] Expert [ ]

3. Please rate your expertise with using screen-readers. *(Tick as appropriate)*

None [ ] Beginner [ ] Intermediate [ ] Expert [ ]

4. Do you play/ed a musical instrument, read music scores? *(Tick as appropriate)*

None [ ] Beginner [ ] Intermediate [ ] Expert [ ]

Please give more details if other than none: ______________________________________

End of Pre-test Questionnaire. Thank You.

Figure B.2: The pre-test questionnaire that collected basic information about participants in Study 2.

B.3 Diagrams Textual Descriptions

Editing and Constructing Diagrams in Audio - Scenarios

SCENARIOS

Retail Store
A worker has a unique ID number, a name and an address. many workers participate in projects. Each project is identified by a unique ID and a name. In general, a worker can participate in any number of projects.

A worker belongs to a particular department. Each department has a name a department ID and an address. an order is usually made by a worker. Each order contains a number of products, and has a delivery date and an ID by which it is identified. Each product is identified by a code.

A supplier is responsible for delivering a specific order. the same supplier can deliver many orders. A supplier is identified by an ID and a name.

Bioinformatics
A patient has a name, a tissue type, and a unique MSP number. Each patient is associated with tags. A tag has a unique number and a unique sequence. In general, the same tag can be associated with any number of patients.

A tag maybe be mapped to a gene. Each gene has a unique gene name and type. In general, multiple tags maybe mapped to the same gene. A doctor then uses a number of tags to design a number of therapies. Each therapy targets one gene, and has a method and a unique type by which it is identified.

An article may analyse multiple genes and a gene may be analysed by multiple articles. An article is identified by a unique number and a journal name.

Bus Company
A bus has a make, is identified by a number, and can carry different numbers of passengers. Each bus is allocated to a particular route. A route is identified by a route number. Several busses can be allocated to the same route.

A route includes a number of stages. Each stage is identified by a number. In general, routes include more than one stage. A driver is assigned to each stage to drive a number of journeys through some or all of the towns on a given route. Each driver has an employee number, name and address

Some towns have a garage where busses are kept. Many buses can be kept at the same garage, which is identified by a name and an address.

Figure B.3: The textual descriptions provided to participants in Study 2, which they used to complete the construction of the ER diagrams.

Blank Modelling Sheet

Editing and Constructing Diagrams in Audio - Consent Form

BLANK MODELING SHEET

Entities

Name	Attributes	Primary? (yes/no)

Relations

Name	Entities	Cardinalities

Figure B.4: The blank modelling sheet used in the control condition of Study 2 (note that the actual version had more rows and that participants were provided with extra blank modelling sheets if they required it.

B.4 Tasks

Editing and Constructing Diagrams in Audio - Tasks

TASKS

A - About the diagram

1. How many entities were there in the diagram? *Total Entities* = __________
2. How many relations were there in the diagram? *Total Relations* = __________
3. Please list the entities of the diagram: ____
____
4. Please list the relations of the diagram: ____
____

B - Some Details

1. Which entity had the biggest number of relations? *Entity* ________________
2. Which entity had the biggest number of attributes? *Entity* ________________
3. Name two entities which had at least one primary key: ___________________________
____

Which of their attributes were primary? ____
____

C - More Details

1. Are entities <...................> and <.....................> related? (YES) (NO)
2. Entities <...................> and <...................> are related by relation ___________?
3. Entity <...................> is related to entities_____________________________________
____
4. Entity <...................> is related to which entities, via which relations?
____
____
5. What is the name of the relation which relate entities <.................> and <.................>?
____
6. Relation <...................> relates entities ____

With cardinalities ____

Figure B.5: Study 2 post-construction tasks. The blank <...........> were filled by the experimenter with information that corresponded to the content of the constructed diagrams.

B.5 Raw Data

	Study 2 - Diagram Construction Times (minutes)	
	Guided	Non-Guided
P1	19	10
P2	32	17
P3	17	16
P4	18	14
P5	19	13
P6	15	19
P7	15	11
P8	26	14
P9	26	17
P10	25	33
P11	29	21
P12	26	22
P13	19	16
P14	28	26
P15	30	13
P16	22	18
P17	25	22
P18	26	16
P19	26	25
P20	17	13
P21	34	21
P22	18	23
P23	31	33
P24	19	28

Study 2 - Scores on Post-Construction Tasks (%)			
	Control	Guided	Non-guided
P1	95	83	84
P2	72.5	68	56
P3	78	70	78
P4	85	84	81
P5	77	77	77
P6	80	86	91
P7	51	71	60
P8	74	73	80
P9	97	88	73
P10	82	87	83
P11	64	65	68
P12	75	75	82
P13	82	90	86
P14	79	61	62
P15	74	67	63
P16	91	80	70
P17	95	86	76
P18	74	78.5	93
P19	72.5	79	91
P20	95	64.5	83
P21	90	76	97
P22	90	82	86
P23	78	66	79
P24	77	74	60

	Study 2 - Preference Group (Guided) Scores on Post-Construction Tasks (%)		
	Control	Guided	Non-guided
P7	51	71	60
P9	97	88	73
P12	75	75	82
P14	79	61	62
P19	72.5	79	91
P24	77	74	60

	Study 2 - Preference Group (Non- Guided) Scores on Post-Construction Tasks (%)		
	Control	Guided	Non-guided
P1	95	83	84
P3	78	70	78
P4	85	84	81
P5	77	77	77
P6	80	86	91
P8	74	73	80
P10	82	87	83
P11	64	65	68
P13	82	90	86
P15	74	67	63
P16	91	80	70
P17	95	86	76
P18	74	78.5	93
P20	95	64	83
P21	90	76	97
P23	78	66	79

Figure B.6: Raw data from Study 2 showing diagram construction times and scores on post-construction tasks for each participants, in addition to preference groups.

Appendix C

Study 3 Materials – Collaborating Through Sounds

Introduction

This appendix contains the materials used in Study 3 which examined audio-only collaborative construction and editing of ER diagrams. The appendix includes the information sheet, the pre-test questionnaire, the construction scenarios (i.e the complementary diagram descriptions given to each participant in a collaborating pair) and the collected raw data that is reported as part of Chapter 7.

C.1 Information Sheet

Collaborative Diagram Construction in Audio - Information Sheet

INFORMATION SHEET

REC Protocol Number QMREC2008/34

YOU WILL BE GIVEN A COPY OF THIS INFORMATION SHEET

Title of Study: Collaborative diagram construction through an audio-only interface

We would like to invite you to participate in this original research project. You should only participate if you want to; choosing not to take part will not disadvantage you in any way. Your decision will not affect your access to treatment or services. Before you decide whether you want to take part, it is important for you to understand why the research is being done and what your participation will involve. Please take time to read the following information carefully and discuss it with others if you wish. Ask us if there is anything that is not clear or if you would like more information. If you do decide to take part, please let us know beforehand if you have been involved in any other study during the last year.

Description:

This study is an evaluation of an audio-only interface for collaborative construction and editing of entity-relationship (ER) diagrams. We have developed a number of ways to create, navigate, and edit a ER diagram using only the computer keyboard and computer speakers, and we would like to test how efficient these techniques are.

If you choose to participate in this study you will be expected to attend a session which last for up to two hours or so at the Department of Computer Science, Queen Mary University. You will be given a short training to familiarise yourself with the interface that you will be using. Once you feel comfortable using it, you will be asked to create up to two diagrams together with your partner, one diagram at a time. You will also be asked to fill a number of questionnaires throughout the session to tell us about yourself and your experience using the interface.

We would like you to remember at all time that we are evaluating the application that you will be using rather than your skills in using it.

As part of the experiment, we will log how you use the system and video record you as you use it. All records will be anonymised.

You will receive a payment of £15 for your participation in this study.

Anyone can apply to take part and there are no risks associated with this study. Although, it is unlikely that you will be asked any personal questions, any information that you disclose will be strictly confidential and any information collected will be handled in accordance with the provisions of the Data Protection Act 1998.

‘It is up to you to decide whether or not to take part. If you do decide to take part you will be given this information sheet to keep and be asked to sign a consent form. If you decide to take part you are still free to withdraw at any time and without giving a reason.

In the event of you suffering any adverse effects as a consequence of your participation in this study, you will be compensated through Queen Mary University of London’s ‘No Fault Compensation Scheme’.

Figure C.1: The information sheet provided to participants at the beginning of the sessions in Study 3.

C.2 Pre-test Questionnaire

Collaborative Diagram Construction in Audio - Pre-test Questionnaire

PRE-TEST QUESTIONNAIRE

Please fill in this short questionnaire before the start of the session. Please answer all questions and hand back to the observer when you are finished.

1. What is your name?

____

2. Please rate your expertise in Entity-Relationship modeling. *(Tick as appropriate)*

None [ ] Beginner [ ] Intermediate [ ] Expert [ ]

3. Please rate your expertise with using screen-readers. *(Tick as appropriate)*

None [ ] Beginner [ ] Intermediate [ ] Expert [ ]

4. Do you play/ed a musical instrument, read music scores? *(Tick as appropriate)*

None [ ] Beginner [ ] Intermediate [ ] Expert [ ]

Please give more details if other than none: ____

End of Pre-test Questionnaire. Thank You.

Figure C.2: The pre-test questionnaire that collected basic information about participants in Study 3.

C.3 Collaborative Scenarios

Collaborative Diagram Construction in Audio - Scenarios

SCENARIOS

Bioinformatics

(Part 1)

A patient has a name, a tissue type, and a unique MSP number. Each patient is associated with tags. A tag maybe be mapped to a gene. A doctor uses a number of tags to design a number of therapies. Each therapy targets one gene. A therapy has a method and a unique type by which it is identified. An article may analyse multiple genes and a gene may be analysed by multiple articles.

(Part 2)

A tag has a unique number and a unique sequence. In general, the same tag can be associated with any number of patients. Each gene has a unique gene name. In general, multiple tags maybe mapped to the same gene. A doctor uses a number of tags to design a number of therapies. Each therapy targets one gene. An article is identified by a unique number and a journal name.

Bus Company

(Part 1)

A bus has a make, a capacity, and is identified by a number. Each bus is allocated to a particular route. A route includes a number of stages. A driver is assigned to a stage to drive a number of journeys. Each journey covers a number of stages. Some towns have a garage where busses are kept, many buses can be kept at the same garage.

(Part 2)

A route is identified by a route number. In general, several busses can be allocated to the same route. Each stage is identified by a number. In general, routes include more than one stage. A driver is assigned to a stage to drive a number of journeys. Each driver has an employee number, and address. A garage is identified by a name and an address.

Figure C.3: The complementary diagram descriptions provided to collaborating pairs in Study 3, which they used to complete the construction of the ER diagrams. Each participant was given one part of each description.

C.4 Raw Data

All pairs scores on the annotation task

		Shared	Non-Shared
Pair 1	P1A	90	100
	P1B	63	60
Pair 2	P2A	59	50
	P2B	97	92
Pair 3	P3A	74	75
	P3B	90	85
Pair 5	P5A	67	90
	P5B	51	97
Pair 6	P6A	83	63
	P6B	59	67
Pair 7	P7A	83	60
	P7B	75	64
Pair 8	P8A	42	88
	P8B	33	13
Pair 9	P9A	58	77
	P9B	76	100
Pair 10	P10A	76	67
	P10B	82	75
Pair 11	P11A	45	24
	P11B	53	87
Pair 12	P12A	62	71
	P12B	57	89
Pair 13	P13A	40	66
	P13B	100	87
Pair 14	P14A	100	60
	P14B	90	77
Pair 15	P15A	48	90
	P15B	100	76
Pair 16	P16A	56	50
	P16B	44	95

Parallel pairs scores on the annotation task

		Shared	Non-Shared
Pair 1	P1A	90	100
	P1B	63	60
Pair 2	P2A	59	50
	P2B	97	92
Pair 3	P3A	74	75
	P3B	90	85
Pair 5	P5A	67	90
	P5B	51	97
Pair 6	P6A	83	63
	P6B	59	67
Pair 7	P7A	83	60
	P7B	75	64
Pair 14	P14A	100	60
	P14B	90	77
Pair 15	P15A	48	90
	P15B	100	76

Sequential pairs scores on the annotation task

		Shared	Non-Shared
Pair 8	P8A	42	88
	P8B	33	13
Pair 9	P9A	58	77
	P9B	76	100
Pair 10	P10A	76	67
	P10B	82	75
Pair 11	P11A	45	24
	P11B	53	87
Pair 12	P12A	62	71
	P12B	57	89
Pair 13	P13A	40	66
	P13B	100	87
Pair 16	P16A	56	50
	P16B	44	95

Figure C.4: Participants scores on diagrams annotation tasks.

SHARED		Supplied					Requested				
	Participants	What I Did	What I'm Doing	What I Will Do	Where I Am	Completion	What Did You Do	What Are You Doing	What will You Do	Where Are You	Completion
Pair1 (Pilot2)	P1A	3	6	3	0	0	2	2	0	0	0
	P1B	0	1	1	0	0	10	0	0	2	0
Pair2	P2A	0	0	4	0	1	2	2	0	2	0
	P2B	4	1	2	0	0	5	3	2	2	1
Pair3	P3A	1	0	1	3	1	0	0	0	0	0
	P3B	4	5	3	3	0	1	1	0	2	1
Pair5	P5A	1	0	0	0	0	3	0	0	0	0
	P5B	2	0	7	0	3	1	4	0	0	0
Pair6	P6A	0	0	2	0	2	1	0	1	0	0
	P6B	0	0	2	0	0	1	0	0	0	0
Pair7	P7A	3	0	0	0	0	1	0	0	0	0
	P7B	0	1	0	0	1	7	0	0	0	0
Pair8	P8A	0	0	3	0	0	2	1	0	0	3
	P8B	1	0	0	1	0	0	0	0	0	0
Pair9	P9A	0	4	3	1	3	0	0	0	0	1
	P9B	0	0	1	0	0	0	0	0	0	0
Pair10	P10A	5	1	1	0	1	1	0	0	0	0
	P10B	1	5	4	0	1	0	0	0	0	0
Pair11	P11A	0	1	8	0	1	3	0	0	0	0
	P11B	0	0	6	2	1	0	0	0	0	0
Pair12	P12A	1	0	5	5	0	3	1	0	3	0
	P12B	1	1	15	3	2	3	1	0	0	0
Pair13	P13A	0	1	2	0	0	0	0	0	0	0
	P13B	0	1	2	0	0	0	0	0	0	0
Pair14	P14A	1	4	2	1	0	3	0	0	1	0
	P14B	4	0	4	0	0	3	1	0	0	0
Pair15	P15A	2	0	3	0	0	0	0	0	0	0
	P15B	1	0	1	0	0	0	1	0	0	0
Pair16	P16A	5	7	6	0	0	0	0	1	0	0
	P16B	0	1	4	0	0	0	0	0	0	0

Figure C.5: The amount workspace awareness information exchanged by each participant in the Shared condition.

NON-SHARED		Supplied					Requested				
	Participants	What I Did	What I’m Doing	What I Will Do	Where I Am	Completion	What Did You Do	What Are You Doing	What will You Do	Where Are You	Completion
Pair1 (Pilot2)	R1	2	2	1	1	1	1	0	0	0	0
	L1	4	2	3	0	0	3	3	0	2	0
Pair2	R2	0	0	5	0	4	0	2	0	0	1
	L2	3	1	3	0	2	3	0	0	0	5
Pair3	R3	5	0	1	0	3	2	0	0	0	1
	L3	9	3	4	2	3	4	5	0	2	0
Pair5	R5	4	3	0	0	3	0	0	1	0	0
	L5	7	3	2	0	3	4	0	0	1	1
Pair6	R6	3	2	0	0	1	2	0	0	0	0
	L6	5	0	1	0	2	9	0	0	0	0
Pair7	R7	1	4	3	0	0	8	0	0	0	0
	L7	4	0	3	0	0	5	2	0	1	0
Pair8	R8	1	3	2	0	2	5	1	1	0	0
	L8	4	1	5	0	3	2	0	0	0	3
Pair9	R9	5	6	7	1	3	7	1	1	5	0
	L9	13	2	9	0	2	4	0	1	0	1
Pair10	R10	2	0	1	0	3	2	0	0	0	0
	L10	8	1	5	0	2	0	2	0	0	0
Pair11	R11	4	1	5	2	1	5	1	1	0	1
	L11	14	7	6	7	5	2	0	0	0	1
Pair12	R12	8	3	9	0	2	5	1	0	0	0
	L12	7	2	11	1	1	2	0	0	0	0
Pair13	R13	3	10	4	3	3	0	0	0	0	0
	L13	4	4	0	1	0	0	0	0	0	0
Pair14	R14	2	1	1	0	1	1	1	0	0	0
	L14	4	2	2	0	1	2	0	0	0	0
Pair15	R15	1	2	4	0	1	0	0	0	0	0
	L15	5	0	4	0	0	0	0	0	0	1
Pair16	R16	9	6	15	0	0	1	1	0	0	0
	L16	2	2	5	2	1	1	1	0	0	0

Figure C.6: The amount of workspace awareness information exchanged by each participant in the Non-Shared condition.

SHARED	While Interacting				Not Interacting				Other			
	Strategy	Content	Labour	Execution	Strategy	Content	Labour	Execution	Total	Interacting	Reading	Interacting/Read
Pair1	0	14.88	0.81	1.06	0	26.98	1.11	0.48	48.24	42.19	3.26	2.78
Pair2	0.66	16.76	0.22	0	1.09	26.36	2.31	0.55	45.57	31.6	4.4	9.57
Pair3	0.29	21.91	0.54	4.53	7.29	25.54	4.6	0.17	27.71	7.81	6.47	13.42
Pair5	0	10.69	0.18	0	5.04	12.21	0.97	0	65.51	53.78	11.73	0
Pair6	0	8.36	0.98	0	0	8.49	0.47	4.19	79.75	74.14	3.78	1.82
Pair7	0	8.98	0.16	0	0.24	20.32	0.5	0	67.08	60.42	2.73	3.92
Pair8	0	36.54	0.55	2.69	0.81	31.8	0.6	0.35	22.36	0.83	2.05	19.46
Pair9	0.68	16.33	1.23	0.25	2.67	27.61	2.37	0	45.24	0	6.87	38.36
Pair10	0	10.36	1.29	1.27	2.12	34.5	1.37	2.04	39.46	11.24	2.04	26.17
Pair11	0	23.98	0.22	4.47	1.87	21.17	2.67	0	36.63	1.88	0	34.75
Pair12	0.18	17.74	0.71	2.72	2.04	27.01	1.52	3.56	41.52	0.6	0.99	39.93
Pair13	1.49	17.76	0.26	0.42	1.14	28.74	1.53	0.23	45.81	0.51	3.38	41.91
Pair14	0	6.69	0	0.64	0	23.94	3.07	0.53	63.13	43.62	6.26	13.24
Pair15	0	3.93	0	0	0.24	25.45	1.66	0	67.94	41.13	4.78	22.02
Pair16	0	29.67	0	5.31	0	27.26	2.42	1.57	29.52	2.5	1.95	25.07

NON-SHARED	While Interacting				Not Interacting				Other			
	Strategy	Content	Labour	Execution	Strategy	Content	Labour	Execution	Total	Interacting	Reading	Interacting/Read
Pair1	3	15.71	1.6	0	1.09	12.38	1.26	0	50.74	37.36	4.25	9.12
Pair2	1.96	7.07	2.03	0	3.3	24.74	0.72	0.77	55.26	41.49	7.27	6.48
Pair3	0	28.29	0	1.52	1.71	28.78	2.84	0	30.43	19.59	9.47	1.36
Pair5	0	8.8	0.16	0	2.1	19.41	3.77	0	59.83	35.29	0.97	23.55
Pair6	0	13.55	0.18	1.22	3.25	13.34	0.12	0	77.45	68.92	0	8.53
Pair7	0	6.83	1.81	0	0	13.73	0.47	0	64.36	50.46	2.32	11.57
Pair8	1.13	17.6	0.31	0.22	3.99	29.57	2.84	0.39	38.82	33.16	0	5.66
Pair9	0.65	35.39	1.54	0	0.65	19.94	0	0	38.92	21.8	2.38	14.73
Pair10	0	15.86	1.89	0.32	1.44	30.08	1.96	0	44.79	34.67	0.65	9.47
Pair11	0	26.02	2.37	2.66	0.63	26.71	2.4	0	34.88	27.45	1.02	6.4
Pair12	1.39	18.74	2.78	2.57	2.34	11.52	2.78	2.61	51.8	32.13	2.08	17.58
Pair13	0.14	27.29	3.57	2.65	1.25	24.24	0.12	0.92	36.26	22.01	2.89	11.35
Pair14	0	13.72	0	0	0	27.18	2.11	0	54.31	48	3.65	2.65
Pair15	0	3.32	1.23	0	1.29	27.92	2.41	0	58.61	41.03	3.97	13.6
Pair16	0	35.93	0	0	0	22.98	2.85	0.67	31.94	7.44	5.46	19.02

Figure C.7: The proportions of time spent on each activity category by each collaborating pair in the Shared and Non-Shared conditions.

SHARED	C-NI	C-WI	E-NI	E-WI	L-NI	L-WI	S-NI	S-WI	Other
C-NI	0.0542	0.2419	0.0108	0.0325	0.1661	0.0469	0.0758	0.0108	0.3610
C-WI	0.1990	0.0765	0.0102	0.0510	0.0281	0.0383	0.0204	0.0026	0.5740
E-NI	0.0625	0.3750	0.0625	0.1250	0.0000	0.0000	0.0000	0.0000	0.3750
E-WI	0.0411	0.2740	0.0137	0.0548	0.0685	0.0274	0.0000	0.0000	0.5205
L-NI	0.1125	0.1625	0.0375	0.0625	0.0125	0.0125	0.0000	0.0125	0.5875
L-WI	0.0000	0.2000	0.0000	0.0400	0.0000	0.0000	0.0200	0.0000	0.7400
S-NI	0.3478	0.1739	0.0000	0.0435	0.0652	0.0652	0.0435	0.0217	0.2391
S-WI	0.0000	0.3333	0.0000	0.0000	0.0000	0.0000	0.2222	0.0000	0.4444
Other	0.3284	0.5085	0.0148	0.0508	0.0339	0.0318	0.0233	0.0085	0.0000

Figure C.8: All pairs probability transition matrix for movements between the activity categories in the Shared condition; C = Content, E = Execution, L = Labour, S = Strategy, NI = Not Interaction, WI = While Interaction. Shaded cells shows the probabilities addressed in the reported analysis.

NON-SHARED	C-NI	C-WI	E-NI	E-WI	L-NI	L-WI	S-NI	S-WI	Other
C-NI	0.0356	0.2622	0.0178	0.0000	0.1867	0.0400	0.0489	0.0089	0.4000
C-WI	0.2261	0.1019	0.0064	0.0127	0.0446	0.0382	0.0159	0.0064	0.5478
E-NI	0.2222	0.1111	0.1111	0.1111	0.0000	0.0000	0.0000	0.0000	0.4444
E-WI	0.1250	0.2500	0.0000	0.0000	0.0000	0.0000	0.0000	0.0625	0.5625
L-NI	0.1744	0.1163	0.0116	0.0000	0.0349	0.0349	0.0698	0.0349	0.5233
L-WI	0.0476	0.2143	0.0000	0.0000	0.0000	0.0238	0.0000	0.0238	0.6905
S-NI	0.3659	0.1707	0.0000	0.0000	0.1951	0.0244	0.0000	0.0000	0.2439
S-WI	0.0769	0.0000	0.0000	0.0000	0.1538	0.0000	0.1538	0.0000	0.6154
Other	0.3067	0.5200	0.0027	0.0293	0.0480	0.0373	0.0400	0.0160	0.0000

Figure C.9: All pairs probability transition matrix for movements between the activity categories in the Non-Shared condition; C = Content, E = Execution, L = Labour, S = Strategy, NI = Not Interaction, WI = While Interaction.Shaded cells shows the probabilities addressed in the reported analysis.

SHARED	C-NI	C-WI	E-NI	E-WI	L-NI	L-WI	S-NI	S-WI	Other
C-NI	0.0400	0.2400	0.0000	0.0200	0.2300	0.0300	0.0600	0.0000	0.3800
C-WI	0.3056	0.0926	0.0185	0.0278	0.0185	0.0556	0.0463	0.0000	0.4352
E-NI	0.0000	0.1429	0.0000	0.1429	0.0000	0.0000	0.0000	0.0000	0.7143
E-WI	0.0909	0.5455	0.0909	0.0909	0.0000	0.0000	0.0000	0.0000	0.1818
L-NI	0.1714	0.1143	0.0286	0.0857	0.0286	0.0000	0.0000	0.0286	0.5429
L-WI	0.0000	0.0625	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.9375
S-NI	0.4286	0.2857	0.0000	0.0000	0.0714	0.0714	0.0000	0.0000	0.1429
S-WI	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.3333	0.0000	0.6667
Other	0.3835	0.4436	0.0226	0.0075	0.0677	0.0376	0.0150	0.0226	0.0000

Figure C.10: Parallel pairs probability transition matrix for movements between the activity categories in the Shared condition; C = Content, E = Execution, L = Labour, S = Strategy, NI = Not Interaction, WI = While Interaction.Shaded cells shows the probabilities addressed in the reported analysis.

NON-SHARED	C-NI	C-WI	E-NI	E-WI	L-NI	L-WI	S-NI	S-WI	Other
C-NI	0.0396	0.2871	0.0099	0.0000	0.2277	0.0198	0.0198	0.0000	0.3960
C-WI	0.2800	0.1040	0.0000	0.0000	0.0240	0.0320	0.0160	0.0000	0.5440
E-NI	0.0000	1.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
E-WI	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	1.0000
L-NI	0.1707	0.0732	0.0000	0.0000	0.0488	0.0244	0.1220	0.0244	0.5366
L-WI	0.0000	0.1429	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.8571
S-NI	0.4000	0.0667	0.0000	0.0000	0.2000	0.0667	0.0000	0.0000	0.2667
S-WI	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.2500	0.0000	0.7500
Other	0.3377	0.5033	0.0000	0.0066	0.0596	0.0397	0.0265	0.0265	0.0000

Figure C.11: Parallel pairs probability transition matrix for movements between the activity categories in the Non-Shared condition; C = Content, E = Execution, L = Labour, S = Strategy, NI = Not Interaction, WI = While Interaction.Shaded cells shows the probabilities addressed in the reported analysis.

SHARED	C-NI	C-WI	E-NI	E-WI	L-NI	L-WI	S-NI	S-WI	Other
C-NI	0.0753	0.1712	0.0205	0.0342	0.1370	0.0685	0.0959	0.0205	0.3767
C-WI	0.1745	0.0723	0.0085	0.0681	0.0255	0.0340	0.0085	0.0043	0.6043
E-NI	0.0833	0.4167	0.0833	0.0833	0.0000	0.0000	0.0000	0.0000	0.3333
E-WI	0.0392	0.2353	0.0000	0.0196	0.0588	0.0392	0.0000	0.0000	0.6078
L-NI	0.0882	0.1176	0.0588	0.0000	0.0000	0.0294	0.0000	0.0000	0.7059
L-WI	0.0000	0.2188	0.0000	0.0625	0.0000	0.0000	0.0313	0.0000	0.6875
S-NI	0.2857	0.1429	0.0000	0.0714	0.0000	0.0714	0.0714	0.0357	0.3214
S-WI	0.0000	0.5000	0.0000	0.0000	0.0000	0.0000	0.1667	0.0000	0.3333
Other	0.3095	0.5272	0.0136	0.0782	0.0136	0.0306	0.0238	0.0034	0.0000

Figure C.12: Sequential pairs probability transition matrix for movements between the activity categories in the Shared condition; C = Content, E = Execution, L = Labour, S = Strategy, NI = Not Interaction, WI = While Interaction.Shaded cells shows the probabilities addressed in the reported analysis.

NON-SHARED	C-NI	C-WI	E-NI	E-WI	L-NI	L-WI	S-NI	S-WI	Other
C-NI	0.0320	0.2480	0.0240	0.0000	0.1520	0.0560	0.0720	0.0160	0.4000
C-WI	0.1823	0.0990	0.0104	0.0208	0.0573	0.0417	0.0156	0.0104	0.5625
E-NI	0.2500	0.0000	0.1250	0.1250	0.0000	0.0000	0.0000	0.0000	0.5000
E-WI	0.1333	0.2667	0.0000	0.0000	0.0000	0.0000	0.0000	0.0667	0.5333
L-NI	0.1739	0.1522	0.0217	0.0000	0.0217	0.0435	0.0217	0.0435	0.5217
L-WI	0.0714	0.2500	0.0000	0.0000	0.0000	0.0357	0.0000	0.0357	0.6071
S-NI	0.3462	0.2308	0.0000	0.0000	0.1923	0.0000	0.0000	0.0000	0.2308
S-WI	0.1000	0.0000	0.0000	0.0000	0.2000	0.0000	0.2000	0.0000	0.5000
Other	0.2870	0.5247	0.0045	0.0448	0.0404	0.0404	0.0493	0.0090	0.0000

Figure C.13: Sequential pairs probability transition matrix for movements between the activity categories in the Non-Shared condition; C = Content, E = Execution, L = Labour, S = Strategy, NI = Not Interaction, WI = While Interaction.Shaded cells shows the probabilities addressed in the reported analysis.

Bibliography

M. S. Ackerman, B. Starr, D. Hindus, and S. D. Mainwaring. Hanging on the 'wire: a field study of an audio-only media space. *ACM Trans. Comput.-Hum. Interact.*, 4(1):39–66, 1997. ISSN 1073-0516. 118, 119

M. J. Adams, Y. J. Tenney, and R. W. Pew. Situation awareness and the cognitive management of complex systems. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 37:85–104(20), 1995. 128

M. Adcock and S. Barrass. Cultivating design patterns for auditory displays. In S. Barrass and P. Vickers, editors, *Proceedings of the 10th International Conference on Auditory Display (ICAD2004)*, Sydney, Australia, 2004. International Community for Auditory Display (ICAD). 21

J. L. Alty and D. I. Rigas. Communicating graphical information to blind users using music: the role of context. In *CHI '98: Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 574–581, 1998. ISBN 0-201-30987-4. 7

B. Arons. A review of the cocktail party effect. *Journal of the American Voice I/O Society*, 12: 35–50, 1992. 199

C. Asakawa, H. Takagi, S. Ino, and T. Ifukube. Maximum listening speeds for the blind. In *Proceedings of the 9th International Conference on Auditory Display (ICAD2003)*, pages 276–279. Boston University Publications Production Department, 2003. 102

F. Avanzini and P. Crosato. Haptic-auditory rendering and perception of contact stiffness. In *Haptic and Audio Interaction Design*, volume 4129/2006, pages 24–35, 2006. ISBN 978-3-540-37595-1. 24

R. M. Baecker. *Readings in GroupWare and Computer-Supported Cooperative Work: Assisting Human-Human Collaboration*. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 1994. ISBN 1558602410. 116

G. Baier, T. Hermann, S. Sahle, and U. Stephani. Sonified epileptic rhythms. In T. Stockman, L. V. Nickerson, C. Frauenberger, A. D. N. Edwards, and D. Brock, editors, *Proceedings of the 12th International Conference on Auditory Display (ICAD2006)*, pages 148–151, London, UK, 2006. Department of Computer Science, Queen Mary, University of London, UK. 19

K. Barnicle. Usability testing with screen reading technology in a windows environment. In *CUU '00: Proceedings on the 2000 conference on Universal Usability*, pages 102–109, 2000. ISBN 1-58113-314-6. 13

S. Barrass. *Auditory Information Design*. PhD thesis, The Australian National University, 1998. 3, 21, 22

M. I. Bauer and P. N. Johnson-Laird. How diagrams can improve reasoning. *Psychological Science*, 4(6):372–378, 1993. 38

M. Beaudouin-Lafon and A. Karsenty. Transparency and awareness in a real-time groupware system. In *UIST '92: Proceedings of the 5th annual ACM symposium on User interface software and technology*, pages 171–180, 1992. ISBN 0-89791-549-6. 132

S. Benford, J. Bowers, L. E. Fahlen, J. Mariani, and T. Rodden. Supporting Cooperative Work in Virtual Environments. *The Computer Journal*, 37(8):653–668, 1994. 131

D. Bennett. *Presenting Diagrams in Sound for Blind People*. PhD thesis, University of York, 1999. 7, 58, 236, 238

D. Bennett. Effects of navigation and position on task when presenting diagrams to blind people using sound. *Diagrammatic Representation and Inference*, pages 307–319, 2002. 27, 47, 58, 65, 87, 91, 248, 249

S. Bennett, S. McRobb, and R. Farmer. *Object-oriented Systems Analysis and Design Using UML*. McGraw Hill Higher Education, 2005. ISBN 0077110005. 53

V. Best, A. Van Schaik, and S. Carlile. Two-point discrimination in auditory displays. In E. Brazil and B. Shinn-Cunningham, editors, *Proceedings of the 9th International Conference on Auditory Display (ICAD2003)*, pages 17–20. Boston University Publications Production Department, 2003. 15

A. F. Blackwell. *Metaphor in Diagrams*. PhD thesis, University of Cambridge, 1998. 4, 36

T. Blaine and T. Perkis. The jam-o-drum interactive music system: a study in interaction design. In *DIS '00: Proceedings of the 3rd conference on Designing interactive systems*, pages 165–173, 2000. ISBN 1-58113-219-0. 123

A. Blandford, H. Stelmaszewska, and N. Bryan-Kinns. Use of multiple digital libraries: a case study. In *JCDL '01: Proceedings of the 1st ACM/IEEE-CS joint conference on Digital libraries*, pages 179–188, 2001. ISBN 1-58113-345-6. 66

A. Blandford, H. Thimbleby, and N. Bryan-Kinns. Understanding interaction traps. In H. J. Johnson, editor, *Proceedings of HCI 2003: Designing for Society, Volume 2*, pages 57–60. Research Press International, 2003. 66, 239

M. M. Blattner, D. A. Sumikawa, and R. M. Greenberg. Earcons and icons: their structure and common design principles. *Hum.-Comput. Interact.*, 4(1):11–44, 1989. ISSN 0737-0024. 17

P. Blenkhorn and G. D. Evans. Using speech and touch to enable blind people to access schematic diagrams. *Journal of Network and Computer Applications*, 21(1):17–29, 1998. ISSN 1084-8045. 23, 88

S. A. Bly. Presenting information in sound. In *CHI '82: Proceedings of the 1982 conference on Human factors in computing systems*, pages 371–375, 1982. 14, 16

S. A. Bly, S. R. Harrison, and S. Irwin. Media spaces: bringing people together in a video, audio, and computing environment. *Commun. ACM*, 36(1):28–46, 1993. ISSN 0001-0782. 116, 119, 129, 138

A. S. Bregman. *Auditory scene analysis: The perceptual organization of sound*. MIT Press, 1994. 26, 236

S. Brewster, J. Lumsden, M. Bell, M. Hall, and S. Tasker. Multimodal 'eyes-free' interaction techniques for wearable devices. In *CHI '03: Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 473–480, 2003. ISBN 1-58113-630-7. 250

S. A. Brewster. *Providing a Structured Method for Integrating Non-Speech Audio into Human-Computer Interfaces*. PhD thesis, University of York, 1994. 14, 20, 236

S. A. Brewster. Navigating telephone-based interfaces with earcons. In *HCI 97: Proceedings of*

HCI on People and Computers XII, pages 39–56, London, UK, 1997. Springer-Verlag. ISBN 3-540-76172-1. 25, 58, 238

S. A. Brewster. Using nonspeech sounds to provide navigation cues. *ACM Trans. Comput.-Hum. Interact.*, 5(3):224–259, 1998. ISSN 1073-0516. 58, 98, 249

S. A. Brewster. Nonspeech auditory output. In *The human-computer interaction handbook: fundamentals, evolving technologies and emerging applications*, pages 220–239, Mahwah, NJ, USA, 2002. Lawrence Erlbaum Associates, Inc. ISBN 0-8058-3838-4. 12, 14, 16, 17

S. A. Brewster, P. C. Wright, and A. D. N. Edwards. A detailed investigation into the effectiveness of earcons. In G. Kramer, editor, *Auditory Display: Sonification, Audification, and Auditory Interfaces*, pages 471–498, Reading, MA, 1994. Addison-Wesley. 18

S. A. Brewster, P. C. Wright, and A. D. N. Edwards. Experimentally derived guidelines for the creation of earcons. In *Adjunct Proceedings of HCI95*, pages 155–159, 1995. 20

S. A. Brewster, V.-P. Räty, and A. Kortekangas. Earcons as a method of providing navigational cues in a menu hierarchy. In *HCI '96: Proceedings of HCI on People and Computers XI*, pages 169–183, London, UK, 1996. Springer-Verlag. ISBN 3-540-76069-5. 17

A. Brown. *Non-visual interaction with graphs*. PhD thesis, University of Manchester, 2008. 7, 27, 249

A. Brown, S. Pettifer, and R. Stevens. Evaluation of a non-visual molecule browser. In *Assets '04: Proceedings of the 6th international ACM SIGACCESS conference on Computers and accessibility*, pages 40–47, 2004. ISBN 1-58113-911-X. 28, 58, 87, 89, 236, 238

A. Brown, R. Stevens, and S. Pettifer. Audio representation of graphs: a quick look. In T. Stockman, L. V. Nickerson, C. Frauenberger, A. D. N. Edwards, and D. Brock, editors, *Proceedings of the 12th International Conference on Auditory Display (ICAD2006)*, pages 83–90, London, UK, 2006. Department of Computer Science, Queen Mary, University of London, UK. 91

L. M. Brown and S. A. Brewster. Drawing by ear: Interpreting sonified line graphs. In E. Brazil and B. Shinn-Cunningham, editors, *Proceedings of the 9th International Conference on Auditory Display (ICAD2003)*, pages 152–156. Boston University Publications Production Department, 2003. 20, 22

N. Bryan-Kinns and F. Hamilton. Identifying mutual engagement. *Behaviour and Information Technology*, pages 1–25, 2009. First published on: 03 December 2009 (iFirst). 131

N. Bryan-Kinns, P. G. T. Healey, D. Papworth, and A. Vaduuva. Cues to mutual knowledge. In L. Bannon, I. Wagner, C. Gutwin, R. Harper, and K. Schmidt, editors, *Proceedings of the 10th European Conference on Computer-Supported Cooperative Work*, 2007. ISBN 978-1-84800-030-8. 150

D. A. Burgess. Techniques for low cost spatial audio. In *UIST '92: Proceedings of the 5th annual ACM symposium on User interface software and technology*, pages 53–59, 1992. ISBN 0-89791-549-6. 19

W. Buxton. Introduction to this special issue on nonspeech audio. *Human-Computer Interaction*, 4(1):1–9, 1989. 15, 253

W. Buxton. Less is more (more or less). In P. Denning, editor, *The Invisible Future: The seamless integration of technology in everyday life*, pages 145–179, New York, 2001. McGraw Hill. 2, 3

J. M. Carroll, D. C. Neale, P. L. Isenhour, M. B. Rosson, and D. S. McCrickard. Notification and awareness: synchronizing task-oriented collaborative activity. *International Journal of Human-Computer Studies*, 58(5):605–632, 5 2003. 130, 131

P. H. Carstensen and K. Schmidt. Computer supported cooperative work: New challenges to systems design. In K. Itoh, editor, *Handbook of Human Factors/Ergonomics*. Asakura Publishing, 2003. 116

A. Chang and C. O'Sullivan. Audio-haptic feedback in mobile phones. In *CHI '05: CHI '05 extended abstracts on Human factors in computing systems*, pages 1264–1267, 2005. ISBN 1-59593-002-7. 3

A. Chapanis. Interactive human communication: Some lessons learned from laboratory experiments. Technical report, 2002. 116

P.-H. Cheng, R. Lowe, and M. Scaife. Cognitive science approaches to understanding diagrammatic representations. *Artificial Intelligence Review*, 15(1):79–94, 2001. 36

M. Cherubini, G. Venolia, R. DeLine, and A. J. Ko. Let's go to the whiteboard: how and why software developers use drawings. In *CHI '07: Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 557–566, 2007. ISBN 978-1-59593-593-9. 4, 6, 115, 238

J. Cohen. "kirk here:": using genre sounds to monitor background activity. In *CHI '93: INTERACT '93 and CHI '93 conference companion on Human factors in computing systems*, pages 63–64, 1993. ISBN 0-89791-574-7. 120, 141, 244

J. Cohen. Out to lunch: Further adventures monitoring background activity. pages 15–20, Santa Fe, NM, U.S., 1994. International Community for Auditory Display, International Community for Auditory Display. 141

T. M. Connolly and C. Begg. *Database Systems: A Practical Approach to Design, Implementation, and Management*. Addison-Wesley Longman Publishing Co., Inc., Boston, MA, USA, 2001. ISBN 0201708574. 71, 80

R. Cox and P. Brna. Supporting the use of external representation in problem solving: the need for flexible learning environments. *J. Artif. Intell. Educ.*, 6(2-3):239–302, 1995. ISSN 1043-1020. 34

A. Crossan and S. A. Brewster. Multimodal trajectory playback for teaching shape information and trajectories to visually impaired computer users. *ACM Trans. Access. Comput.*, 1(2):1–34, 2008. ISSN 1936-7228. 25, 124

T. Dingler, J. Lindsay, and B. N. Walker. Learnabiltiy of sound cues for environmental features: Auditory icons, earcons, spearcons, and speech. In *Proceedings of the 14th International Conference on Auditory Display*, Paris, France, 2008. 18

A. Dix. Human-computer interaction: A stable discipline, a nascent science, and the growth of the long tail. *Interacting with Computers*, 22(1):13 – 27, 2010. ISSN 0953-5438. Special Issue: Festschrift for John Long. 2, 3

A. J. Dix, J. E. Finlay, G. D. Abowd, and R. Beale. *Human-Computer Interaction*. Prentice Hall, 1998. ISBN 0-13-239864-8. 187, 206

P. Dourish and V. Bellotti. Awareness and coordination in shared workspaces. In *CSCW '92: Proceedings of the 1992 ACM conference on Computer-supported cooperative work*, pages 107–114, 1992. ISBN 0-89791-542-9. 127, 129, 138, 198, 213, 244

P. Dourish, A. Adler, V. Bellotti, and A. Henderson. Your place or mine? learning from long-term use of audio-video communication. *Computer Supported Cooperative Work (CSCW)*, (1):33–62, 03 1996. 117, 118, 119, 123, 129

M. R. Endsley. Toward a theory of situation awareness in dynamic systems. *Human Factors*, 37 (1):32–64, 1995. 128

T. Erickson, D. N. Smith, W. A. Kellogg, M. Laff, J. T. Richards, and E. Bradner. Socially translucent systems: social proxies, persistent conversation, and the design of "babble". In *CHI '99: Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 72–79, 1999. ISBN 0-201-48559-1. 131

S. Fertig, E. Freeman, and D. Gelernter. Lifestreams: an alternative to the desktop metaphor. In *CHI '96: Conference companion on Human factors in computing systems*, pages 410–411, 1996. ISBN 0-89791-832-0. 3

J. Fish and S. Scrivener. Amplifying the mind's eye: Sketching and visual cognition. *Leonardo*, 23(1):117–126, 1990. ISSN 0024094X. 37

R. S. Fish, R. E. Kraut, and B. L. Chalfonte. The videowindow system in informal communication. In *CSCW '90: Proceedings of the 1990 ACM conference on Computer-supported cooperative work*, pages 1–11, 1990. ISBN 0-89791-402-3. 116, 117, 119

C. Frauenberger. *Auditory Display Design: An Investigation of a Design Pattern Approach*. PhD thesis, Queen Mary University of London, 2008. 21, 22

C. Frauenberger and M. Noisternig. 3d audio interfaces for the blind. In E. Brazil and B. Shinn-Cunningham, editors, *Proceedings of the 9th International Conference on Auditory Display (ICAD2003)*, pages 280–283. Boston University Publications Production Department, 2003. 250

C. Frauenberger and T. Stockman. Auditory display design–an investigation of a design pattern approach. *International Journal of Human-Computer Studies*, 67(11):907 – 922, 2009. ISSN 1071-5819. Special issue on Sonic Interaction Design - SI: Sonic Interaction Design. 3, 21

C. Frauenberger, T. Stockman, and M.-L. Bourguet. A survey on common practice in designing audio in the user interface. In *BCS-HCI '07: Proceedings of the 21st British CHI Group Annual Conference on HCI 2007*, pages 187–194, Swinton, UK, 2007. British Computer Society. ISBN 978-1-902505-94-7. 3, 20, 253

J. P. Fritz and K. E. Barner. Design of a haptic data visualization system for people with visual impairments. *Rehabilitation Engineering, IEEE Transactions on*, 7(3):372–384, Sep 1999. 24

W. Gaver, T. Moran, A. MacLean, L. Lövstrand, P. Dourish, K. Carter, and W. Buxton. Realizing a video environment: Europarc's rave system. In *CHI '92: Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 27–35, 1992. ISBN 0-89791-513-5. 116, 117, 244

W. W. Gaver. Auditory icons: using sound in computer interfaces. *Hum.-Comput. Interact.*, 2 (2):167–177, 1986. ISSN 0737-0024. 120

W. W. Gaver. The sonicfinder: an interface that uses auditory icons. *Hum.-Comput. Interact.*, 4 (1):67–94, 1989. ISSN 0737-0024. 17

W. W. Gaver. Sound support for collaboration. In *ECSCW'91: Proceedings of the second conference on European Conference on Computer-Supported Cooperative Work*, pages 293–308, Norwell, MA, USA, 1991. Kluwer Academic Publishers. ISBN 0-7923-1439-5. 119, 120, 129, 138, 141, 143, 198, 213

W. W. Gaver. The affordances of media spaces for collaboration. In *CSCW '92: Proceedings of the 1992 ACM conference on Computer-supported cooperative work*, pages 17–24, 1992a. ISBN 0-89791-542-9. 117, 118, 119, 120, 129

W. W. Gaver. *Using and creating auditory icons*. SFI studies in the sciences of complexity. Addison Wesley Longman, 1992b. 17

W. W. Gaver. Auditory interfaces. In M. G. Helander, T. K. Landauer, and P. V. Prabhu, editors, *Handbook of Human-Computer Interaction*, pages 1003–1041. Elsevier, 1997. 17

W. W. Gaver, R. B. Smith, and T. O'Shea. Effective sounds in complex systems: the arkola simulation. In *CHI '91: Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 85–90, 1991. ISBN 0-89791-383-3. 119, 120, 244

D. Gentner and J. Nielsen. The anti-mac interface. *Commun. ACM*, 39(8):70–82, 1996. ISSN 0001-0782. 3

C. A. Gurr. Effective diagrammatic communication: Syntactic, semantic and pragmatic issues. *Journal of Visual Languages & Computing*, 10(4):317–342, 1999. 37

C. A. Gurr, J. Lee, and K. Stenning. Theories of diagrammatic reasoning: Distinguishing component problems. *Minds and Machines*, 8(4), 12 1998. 37, 38, 50

C. Gutwin and S. Greenberg. Workspace awareness for groupware. In *CHI '96: Conference companion on Human factors in computing systems*, pages 208–209, 1996. ISBN 0-89791-832-0. 129, 130, 134, 142, 244

C. Gutwin and S. Greenberg. A descriptive framework of workspace awareness for real-time groupware. *Computer Supported Cooperative Work (CSCW)*, 11(3):411–446, 09 2002. xiii, 129, 134, 135, 136, 137, 138, 139, 140, 142, 152, 153, 154, 196, 206, 209, 211, 213, 219, 231, 245, 253

M. S. Hancock, C. Shen, C. Forlines, and K. Ryall. Exploring non-speech auditory feedback at an interactive multi-user tabletop. In *GI '05: Proceedings of Graphics Interface 2005*, pages 41–50, School of Computer Science, University of Waterloo, Waterloo, Ontario, Canada, 2005. Canadian Human-Computer Communications Society. ISBN 1-56881-265-5. 122, 191

S. Handel. *Listening: An Introduction to the Perception of Auditory Events*. MIT Press, 1989. 199

F. Harary, R. Z. Norman, and D. Cartwright. *Structural Models: An Introduction to the Theory of Directed Graphs*. John Wiley and Sons Inc, 1965. ISBN 978-0471351306. 39

L. Harrar and T. Stockman. Designing auditory graph overviews: An examination of discrete vs. continuous sound and the influence of presentation speed. In G. P. Scavone, editor, *Proceedings of the 13th International Conference on Auditory Display (ICAD2007)*, pages 299–305. Schulich School of Music, McGill University, 2007. 19, 22

C. Heath and P. Luff. Collaboration and controlcrisis management and multimedia technology in london underground line control rooms. *Computer Supported Cooperative Work (CSCW)*, (1):69–94, 03 1992. 122, 130, 132, 133, 138, 142, 196, 206

T. Hermann. *Sonification for Exploratory Data Analysis*. PhD thesis, The Technical Faculty of the University of Bielefeld, 2002. 14

T. Hermann. Taxonomy and definitions for sonification and auditory display. In *Proceedings of the 14th International Conference on Auditory Display*, Paris, France, 2008. 16, 19

D. Hindus, M. S. Ackerman, S. Mainwaring, and B. Starr. Thunderwire: a field study of an audio-only media space. In *CSCW '96: Proceedings of the 1996 ACM conference on Computer supported cooperative work*, pages 238–247, 1996. ISBN 0-89791-765-0. 118, 119, 141, 143, 244

D. Hindus, S. D. Mainwaring, N. Leduc, A. E. Hagström, and O. Bayley. Casablanca: designing social communication devices for the home. In *CHI '01: Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 325–332, 2001. ISBN 1-58113-327-8. 117

E. Hoggan, A. Crossan, S. A. Brewster, and T. Kaaresoja. Audio or tactile feedback: which modality when? In *CHI '09: Proceedings of the 27th international conference on Human factors in computing systems*, pages 2253–2256, 2009. ISBN 978-1-60558-246-7. 3

S. Holland, D. R. Morse, and H. Gedenryd. Audiogps: Spatial audio navigation with a minimal attention interface. *Personal Ubiquitous Comput.*, 6(4):253–259, 2002. ISSN 1617-4909. 13

L. E. Holmquist. *Breaking the screen barrier*. PhD thesis, Göteborg University, 2000. 2

M. Horstmann, M. Lorenz, A. Watkowski, G. Ioannidis, O. Herzog, A. A1 King, D. G. Evans, C. Hagen, C. Schlieder, A. M. Burn, N. King, H. Petrie, S. Dijkstra, and D. Crombie. Automated interpretation and accessible presentation of technical diagrams for blind people. In *New Review of Hypermedia and Multimedia*, volume 10, pages 141–163. Taylor & Francis, 2004. ISBN 1361-4568. 26

E. Hutchins. Metaphors for interface design. In M. M. Taylor, F. Neel, and D. G. Bouwhuis, editors, *The Structure of Multimodal Dialogue*, pages 11–28. North-Holland, Amsterdam, 1989. 73, 75, 76

E. Hutchins. *Cognition in the wild*. Cambridge, MA, US: The MIT Press, 1995. 34, 128

P. Irani, P. Shajahan, and C. Kemke. Voicemarks: restructuring hierarchical voice menus for improving navigation. *International Journal of Speech Technology*, 9(3):75–94, 12 2006. 93

P. Janata and E. Childs. Marketbuzz: Sonification of real-time financial dataa. In S. Barrass and P. Vickers, editors, *Proceedings of the 10th International Conference on Auditory Display (ICAD2004)*, Sydney, Australia, 2004. International Community for Auditory Display (ICAD). 19

J. Johnson and G. Engelbeck. Modes survey results. *SIGCHI Bull.*, 20(4):38–50, 1989. ISSN 0736-6906. 107

H. G. Kaper, E. Wiebel, and S. Tipei. Data sonification and sound visualization. *Computing in Science and Engineering*, 1:48–58, 1999. ISSN 1521-9615. 15

A. R. Kennel. Audiograf: a diagram-reader for the blind. In *Assets '96: Proceedings of the second annual ACM conference on Assistive technologies*, pages 51–56, 1996. ISBN 0-89791-776-6. 88

J. Kildal and S. Brewster. Exploratory strategies and procedures to obtain non-visual overviews using tablevis. *International Journal on Disability and Human Developmen*, 5(3):285–294, 2006. 23, 24, 91

M. Kirsch-Pinheiro, J. V. de Lima, and M. R. S. Borges. A framework for awareness support in groupware systems. *Computers in Industry*, 52(1):47–57, 2003. ISSN 0166-3615. 134

P. Kortum. *HCI Beyond the GUI: Design for Haptic, Speech, Olfactory, and Other Nontraditional Interfaces*. Morgan Kaufmann, 2008. ISBN 0123740177. 3

G. Kramer. *Auditory Display: Sonification, Audification and Auditory Interfaces*. Addison-Wesley Publishing Cmpany, Reading, MA, USA, 1994a. ISBN 0201626039. 3, 14, 16, 22

G. Kramer. An introduction to auditory display. In G. Kramer, editor, *Auditory Display: Sonification, Audification and Auditory Interfaces*, pages 1–77. Addison-Wesley, 1994b. 14

G. Kramer, B. Walker, T. Bonebright, P. Cook, J. Flowers, N. Miner, and J. Neuhoff. Sonification report: Status of the field and research agenda. Technical report, 1997. available: `http://www.icad.org/websiteV2.0/References/nsf.html` Last checked July 2010,. 4, 18

J. H. Larkin. Display-based problem solving. *Complex information processing: The impact of Herbert A. Simon*, pages 319–341, 1989. ISSN 0805801790. 34

J. H. Larkin and H. A. Simon. Why a diagram is (sometimes) worth ten thousand words. *Cognitive Science*, (1):65–100, 1987. 37, 38, 50

T. Le. The brain: revolutionary interface for next-generation digital media. In *FDG '09: Proceedings of the 4th International Conference on Foundations of Digital Games*, pages xviii–xviii, 2009. ISBN 978-1-60558-437-9. 3

G. Leplatre. *The design and evaluation of non-speech sounds to support navigation in restricted display devices*. PhD thesis, The University of Glasgow, 2002. 25, 236

G. Leplatre and S. Brewster. Designing non-speech sounds to support navigation in mobile phone menus. In P. R. Cook, editor, *Proceedings of the 6th International Conference on Auditory Display (ICAD2000)*. International Community for Auditory Display, 2000. 25, 238

D. Lunney and R. C. Morrison. High technology laboratory aids for visually handicapped chemistry students. *Journal of Chemical Education*, 58(3), 1981. 16

S. Mann. Smart clothing: the shift to wearable computing. *Commun. ACM*, 39(8):23–24, 1996. ISSN 0001-0782. 13

D. L. Mansur, M. M. Blattner, and K. I. Joy. Sound graphs: A numerical data analysis method for the blind. *Journal of Medical Systems*, 9(3):163–174, 1985. 16, 19, 22

A. B. Markman and E. Dietrich. Extending the classical view of representation. *Trends in Cognitive Sciences*, 4(12):470–475, 2000. 32

L. M. Mauney and B. N. Walker. Individual differences and the field of auditory display: Past research, a present study, and an agenda for the future. In G. P. Scavone, editor, *Proceedings of the 13th International Conference on Auditory Display (ICAD2007)*, pages 386–390. Schulich School of Music, McGill University, 2007. 15

D. McGookin. *Understanding and Improving the Identification of Concurrently Presented Earcons*. PhD thesis, University of Glasgow, 2004. 3, 16, 19

D. McGookin and S. A. Brewster. Advantages and issues with concurrent audio presentation as part of an auditory display. pages 44–50, London, UK, 2006a. Department of Computer Science, Queen Mary, University of London, UK. 20

D. McGookin and S. A. Brewster. Soundbar: exploiting multiple views in multimodal graph browsing. In *NordiCHI '06: Proceedings of the 4th Nordic conference on Human-computer interaction*, pages 145–154, 2006b. ISBN 1-59593-325-5. 47, 248

D. McGookin and S. A. Brewster. Graph builder: Constructing non-visual visualizations. In *People and Computers XX Engage*, pages 263–278, 2007a. 8, 24, 126

D. McGookin and S. A. Brewster. An initial investigation into non-visual computer supported collaboration. In *CHI '07: CHI '07 extended abstracts on Human factors in computing systems*, pages 2573–2578, 2007b. ISBN 978-1-59593-642-4. 8, 126, 127, 141, 143, 189, 222, 224, 226, 244

O. Metatla, N. Bryan-Kinns, and T. Stockman. Using hierarchies to support non-visual access to relational diagrams. In *BCS-HCI '07: Proceedings of the 21st British CHI Group Annual Conference on HCI 2007*, pages 215–225, Swinton, UK, UK, 2007. British Computer Society. ISBN 978-1-902505-94-7. 22

O. Metatla, N. Bryan-Kinns, and T. Stockman. Constructing relational diagrams in audio: the multiple perspective hierarchical approach. In *Assets '08: Proceedings of the 10th international ACM SIGACCESS conference on Computers and accessibility*, pages 97–104, 2008. ISBN 978-1-59593-976-0. 22

G. Miller. The magical number seven, plus or minus two: Some limits on our capacity for processing information. *The Psychological Review*, 63:81–97, 1956. 91

E. N. Mitsopoulos. *A Principled Approach to the Design of Auditory Interaction in the Non-Visual User Interface*. PhD thesis, The University of York, 2000. 3, 21, 22

A. Monk. Mode errors: a user-centered analysis and some preventative measures using keying-contingent sound. *Int. J. Man-Mach. Stud.*, 24(4):313–327, 1986. ISSN 0020-7373. 110

B. C. J. Moore. *An Introduction to the psychology of hearing*. Academic Press, 2003. ISBN 0125056281. 15

M. R. Morris, D. Morris, and T. Winograd. Individual audio channels with single display groupware: effects on communication and task strategy. In *CSCW '04: Proceedings of the 2004 ACM conference on Computer supported cooperative work*, pages 242–251, 2004. ISBN 1-58113-810-5. 123, 141, 143, 189, 224, 244

C. Müeller-Tomfelde and S. Steiner. Audio-enhanced collaboration at an interactive electronic whiteboard. In J. Hiipakka, N. Zacharov, and T. Takala, editors, *Proceedings of the 7th International Conference on Auditory Display (ICAD2001)*, pages 267–271. Laboratory of Acoustics and Audio Signal Processing and the Telecommunications Software and Multimedia Laboratory, Helsinki University of Technology, 2001. 122, 141

E. Murphy. *Designing Auditory Cues for a Multimodal Web Interface: A Semiotic Approach*. PhD thesis, Queens University Belfast, 2007. 3, 20

B. A. Myers. A brief history of human-computer interaction technology. *interactions*, 5(2): 44–54, 1998. ISSN 1072-5520. 1

E. D. Mynatt. Designing with auditory icons: how well do we identify auditory cues? In *CHI '94: Conference companion on Human factors in computing systems*, pages 269–270, 1994a. ISBN 0-89791-651-4. 20

E. D. Mynatt. Designing with auditory icons. In *Proceedings of the 2nd International Conference on Auditory Display (ICAD1994)*, page 109119, 1994b. 20

E. D. Mynatt. *Transforming Graphical Interfaces into Auditory Interfaces*. PhD thesis, Georgia Institute of Technology, 1995. 7, 20, 22, 27, 58, 236, 238

E. D. Mynatt and K. W. Edwards. The mercator environment: A nonvisual interface to x windows and unix workstations. Technical Report GVU Tech Report GIT-GVU-92-05, 1992. 26

E. D. Mynatt and G. Weber. Nonvisual presentation of graphical user interfaces: contrasting two approaches. In *CHI '94: Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 166–172, 1994. ISBN 0-89791-650-6. 23, 24, 58, 87

E. D. Mynatt, M. Back, R. Want, M. Baer, and J. B. Ellis. Designing audio aura. In *CHI '98: Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 566–573, 1998. ISBN 0-201-30987-4. 19, 120, 121, 141, 244

M. A. Nees and B. N. Walker. Relative intensity of auditory context for auditory graph design. In T. Stockman, L. V. Nickerson, C. Frauenberger, A. D. N. Edwards, and D. Brock, editors, *Proceedings of the 12th International Conference on Auditory Display (ICAD2006)*, pages 95–98. Department of Computer Science, Queen Mary, University of London, UK, 2006. 19

M. A. Nees and B. N. Walker. Auditory interfaces and sonification. In C. Stephanidis, editor, *The Universal Access Handbook*, pages 507–521. New York: CRC Press Taylor and Francis, 2009. 3

K. V. Nesbitt and S. Barrass. Evaluation of a multimodal sonification and visualisation of depth of market stock data. In R. Nakatsu and H. Kawahara, editors, *Proceedings of the 8th International Conference on Auditory Display (ICAD2002)*. Advanced Telecommunications Research Institute (ATR), Kyoto, Japan, 2002. 19

C. Neustaedter and S. Greenberg. The design of a context-aware home media space for balancing privacy and awareness. *UbiComp 2003: Ubiquitous Computing*, pages 297–314, 2003. 117

C. M. Neuwirth, J. H. Morris, S. H. Regli, R. Chandhok, and G. C. Wenger. Envisioning communication: task-tailorable representations of communication in asynchronous work. In *CSCW '98: Proceedings of the 1998 ACM conference on Computer supported cooperative work*, pages 265–274, 1998. ISBN 1-58113-009-0. 130

D. A. Norman. Categorization of action slips. *Psychological Review*, 88(1):1–15, 1981. 66, 106

D. A. Norman. *The Psychology of Everyday Things*. Basic Books, 1988. ISBN 0465067107. 33

D. A. Norman. Cognitive artifacts. In *Designing interaction: psychology at the human-computer interface*, pages 17–38, New York, NY, USA, 1991. Cambridge University Press. ISBN 0-521-40056-2. 32, 34

D. A. Norman. *Things That Make Us Smart: Defending Human Attributes in the Age of the Machine*. Addison Wesley Publishing Company, 1993. ISBN 0201626950. 32, 33, 34, 35, 38, 253

W. Nöth. Representation in semiotics and in computer science. *Semiotica*, 115(3-4):203–214, 1997. 32

G. M. Olson, J. S. Olson, M. R. Carter, and M. Storrøsten. Small group design meetings: an analysis of collaboration. *Hum.-Comput. Interact.*, 7(4):347–374, 1992. ISSN 0737-0024. 150, 187

J. S. Olson, G. M. Olson, M. Storrøsten, and M. Carter. 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, 1993. ISSN 1046-8188. 145, 188

J. S. Olson, G. M. Olson, and D. K. Meader. What mix of video and audio is useful for small groups doing remote real-time design work? In *CHI '95: Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 362–368, 1995. ISBN 0-201-84705-1. 117, 119, 244

S. Oviatt, R. Coulston, and R. Lunsford. When do we interact multimodally?: cognitive load and multimodal communication patterns. In *ICMI '04: Proceedings of the 6th international conference on Multimodal interfaces*, pages 129–136, 2004. ISBN 1-58113-995-0. 13

D. S. Pagani and W. E. Mackay. Bring media spaces into the real world. In *ECSCW'93: Proceedings of the third conference on European Conference on Computer-Supported Cooperative Work*, pages 341–356, Norwell, MA, USA, 1993. Kluwer Academic Publishers. ISBN 0-7923-2447-1. 118

D. K. Palladino and B. N. Walker. Efficiency of spearcon-enhanced navigation of one dimensional electronic menus. In *Proceedings of the 14th International Conference on Auditory Display*, Paris, France, 2008. 18

S. E. Palmer. Hierarchical structure in perceptual representation. *Cognitive Psychology*, 9(4): 441 – 474, 1977. ISSN 0010-0285. 26, 28, 236

S. E. Palmer. Fundamental aspects of cognitive representation. *Cognition and Categorization*, pages 259–303, 1978. 32

P. Parente. *Clique: Perceptually Based, Task Oriented Auditory Display for GUI Applications*. PhD thesis, University of North Carolina, 2008. 95, 111

S. Pauletto and A. Hunt. The sonification of emg data. In T. Stockman, L. V. Nickerson, C. Frauenberger, A. D. N. Edwards, and D. Brock, editors, *Proceedings of the 12th Inter-*

national Conference on Auditory Display (ICAD2006), pages 152–157, London, UK, 2006. Department of Computer Science, Queen Mary, University of London, UK. 19

H. Petrie, S. Morley, and G. Weber. Tactile-based direct manipulation in guis for blind users. In *CHI '95: Conference companion on Human factors in computing systems*, pages 428–429, 1995. ISBN 0-89791-755-3. 24

H. Petrie, V. Johnson, S. Furner, and T. Strothott. Design lifecycles and wearable computers for users with disabilities. In *Proceedings of the first international workshop of Human Computer Interaction with mobile devices*, Glasgow, UK, 1998. 17

H. Petrie, C. Schlieder, P. Blenkhorn, G. Evans, A. King, A.-M. OâNeill, G. Ioannidis, B. Gallagher, D. Crombie, R. Mager, and M. Alafaci. Tedub: A system for presenting and exploring technical drawings for blind people. *Computers Helping People with Special Needs*, pages 47–67, 2002. 89

H. Petrie, C. Harrison, and S. Dev. Describing images on the web: a survey of current practice and prospects for the future. In *Proceedings of 3rd International Conference on Universal Access in Human-Computer Interaction*, 2005. 44

B. Plimmer, A. Crossan, S. A. Brewster, and R. Blagojevic. Multimodal collaborative handwriting training for visually-impaired people. In *CHI '08: Proceeding of the twenty-sixth annual SIGCHI conference on Human factors in computing systems*, pages 393–402, 2008. ISBN 978-1-60558-011-1. 124

I. Pollack and L. Ficks. The information of elementary multidimensional auditory displays. *The Journal of the Acoustical Society of America*, 26(1):155–158, 1954. 16

I. Poupyrev and S. Maruyama. Tactile interfaces for small touch screens. In *UIST '03: Proceedings of the 16th annual ACM symposium on User interface software and technology*, pages 217–220, 2003. ISBN 1-58113-636-6. 3

H. C. Purchase, L. Colpoys, M. McGill, D. Carrington, and C. Britton. Uml class diagram syntax: an empirical study of comprehension. In *APVis '01: Proceedings of the 2001 Asia-Pacific symposium on Information visualisation*, pages 113–120. Australian Computer Society, Inc., 2001. ISBN 0-909925-87-9. 90

R. Ramloll and S. Brewster. An environment for studying the impact of spatialising sonified graphs on data comprehension. *Information Visualisation, International Conference on*, 0: 167, 2002a. ISSN 1093-9547. 127

R. Ramloll and S. Brewster. A generic approach for augmenting tactile diagrams with spatial non-speech sounds. In *CHI '02: CHI '02 extended abstracts on Human factors in computing systems*, pages 770–771, 2002b. ISBN 1-58113-454-1. 23

K. Rassmus-Gröhn, C. Magnusson, and H. E. Eftring. Iterative design of an audio-haptic drawing application. In *CHI '07: CHI '07 extended abstracts on Human factors in computing systems*, pages 2627–2632, 2007. ISBN 978-1-59593-642-4. 8, 24

J. Rekimoto. Time-machine computing: a time-centric approach for the information environment. In *UIST '99: Proceedings of the 12th annual ACM symposium on User interface software and technology*, pages 45–54, 1999. ISBN 1-58113-075-9. 3

H. Rheingold. *Virtual Reality: The Revolutionary Technology of Computer-Generated Artificial Worlds*. Simon & Schuster, 1992. ISBN 0671778978. 3

M. Rittenbruch. Atmosphere: A framework for contextual awareness. *International Journal of Human-Computer Interaction*, 14(2):159–180, 2002. 134

A. Roginska, E. Childs, and M. K. Johnson. Monitoring real-time data: a sonification approach. In T. Stockman, L. V. Nickerson, C. Frauenberger, A. D. N. Edwards, and D. Brock, editors, *Proceedings of the 12th International Conference on Auditory Display (ICAD2006)*, pages 176–181, London, UK, 2006. Department of Computer Science, Queen Mary, University of London, UK. 19

R. W. Root. Design of a multi-media vehicle for social browsing. In *CSCW '88: Proceedings of the 1988 ACM conference on Computer-supported cooperative work*, pages 25–38, 1988. ISBN 0-89791-282-9. 117

D. E. Rumelhart and D. A. Norman. Representation of knowledge. In A. M. Aitkenhead and J. M. Slack, editors, *Issues in Cognitive Modeling*, pages 15–62. Lawrence Erlbaum Associates, 1985. 32

A. Savidis, C. Stephanidis, A. Korte, K. Crispien, and K. Fellbaum. A generic direct-manipulation 3d-auditory environment for hierarchical navigation in non-visual interaction. In *Assets '96: Proceedings of the second annual ACM conference on Assistive technologies*, pages 117–123, 1996. ISBN 0-89791-776-6. 250

N. Sawhney and A. Murphy. Espace 2: an experimental hyperaudio environment. In *CHI '96: Conference companion on Human factors in computing systems*, pages 105–106, 1996. ISBN 0-89791-832-0. 19

M. Scaife and Y. Rogers. External cognition: How do graphical representations work. *International Journal of Human-Computer Studies*, 45(2):185–213, 1996. 32, 34, 39, 50, 66

C. Scaletti. Sound synthesis algorithms for auditory data representations. In G. Kramer, editor, *Auditory Display: Sonification, Audification and Auditory Interfaces*, pages 223–252. Addison-Wesley, 1994. 18

K. Schmidt. The problem with 'awareness': Introductory remarks on 'awareness in cscw'. *Comput. Supported Coop. Work*, 11(3):285–298, 2002. ISSN 0925-9724. 129, 130, 133

K. Schmidt and C. Simone. Mind the gap! towards a unified view of cscw. In *In The Fourth International Conference on the Design of Cooperative Systems COOP*, pages 23–26. The IOS Press, 2000. 133, 136, 206

A. J. Sellen. Remote conversations: the effects of mediating talk with technology. *Hum.-Comput. Interact.*, 10(4):401–444, 1995. ISSN 0737-0024. 117, 118

A. J. Sellen, B. Buxton, and J. Arnott. Using spatial cues to improve videoconferencing. In *CHI '92: Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 651–652, 1992a. ISBN 0-89791-513-5. 117

A. J. Sellen, G. P. Kurtenbach, and W. A. S. Buxton. The prevention of mode errors through sensory feedback. *Hum.-Comput. Interact.*, 7(2):141–164, 1992b. ISSN 0737-0024. 110, 111

P. Shajahan and P. Irani. Representing hierarchies using multiple synthetic voices. In *IV '04: Proceedings of the Information Visualisation, Eighth International Conference*, pages 885–891, 2004. ISBN 0-7695-2177-0. 58, 93, 98

A. Shimojima. Operational constraints in diagrammatic reasoning. In *Logical reasoning with diagrams*, pages 27–48, New York, NY, USA, 1996. Oxford University Press, Inc. ISBN 0-19-510427-7. 38

B. Shneiderman. Direct manipulation: A step beyond programming languages (abstract only). In *CHI '81: Proceedings of the joint conference on Easier and more productive use of computer systems. (Part - II)*, page 143, 1981. ISBN 0-89791-064-8. 1

B. Shneiderman. The eyes have it: A task by data type taxonomy for information visualizations. In *VL '96: Proceedings of the 1996 IEEE Symposium on Visual Languages*, page 336, Washington, DC, USA, 1996. IEEE Computer Society. ISBN 0-8186-7508-X. 26, 236

H. A. Simon. Cognitive science: The newest science of the artificial. *Cognitive Science*, 4(1):33 – 46, 1980. ISSN 0364-0213. 34, 38, 50

A. C. Smith, J. S. Cook, J. M. Francioni, A. Hossain, M. Anwar, and M. F. Rahman. Nonvisual tool for navigating hierarchical structures. *SIGACCESS Access. Comput.*, (77-78):133–139, 2004. ISSN 1558-2337. 56

D. A. Smith and B. N. Walker. Tick-marks, axes, and labels: The effects of adding context to auditory graphs. In R. Nakatsu and H. Kawahara, editors, *Proceedings of the 8th International Conference on Auditory Display (ICAD2002)*, Kyoto, Japan, 2002. Advanced Telecommunications Research Institute (ATR), Kyoto, Japan. 22

S. D. Speeth. Seismometer sounds. *The Journal of the Acoustical Society of America*, 33(7): 909–916, 1961. 16

M. F. Spiegel and L. Streeter. Applying speech synthesis to user interfaces. In M. G. Helander, T. K. Landauer, and P. V. Prabhu, editors, *Handbook of Human-Computer Interaction*, pages 1061–1084. Elsevier, 1997. 17

G. Stahl, T. Koschmann, and D. Suthers. *Computer-supported collaborative learning: An Historical Perspective*, pages 409–426. Cambridge University Press, Cambridge, UK, 2006. 4

M. Stefik, D. G. Bobrow, G. Foster, S. Lanning, and D. Tatar. Wysiwis revised: early experiences with multiuser interfaces. *ACM Trans. Inf. Syst.*, 5:147–167, April 1987. ISSN 1046-8188. 127

K. Stenning. Distinctions with differences: Comparing criteria for distinguishing diagrammatic from sentential systems. *Theory and Application of Diagrams*, pages 43–50, 2000. 37

R. D. Stevens, A. D. N. Edwards, and P. A. Harling. Access to mathematics for visually disabled students through multimodal interaction. *Hum.-Comput. Interact.*, 12(1):47–92, 1997. ISSN 0737-0024. 25, 89

S. S. Stevens. On the theory of scales of measurement. *Science*, 103(2684), 1946. 41

J. Stewart, B. B. Bederson, and A. Druin. Single display groupware: a model for co-present collaboration. In *CHI '99: Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 286–293, 1999. ISBN 0-201-48559-1. 121

T. Stockman. The design and evaluation of auditory access to spreadsheets. In S. Barrass and P. Vickers, editors, *Proceedings of the 10th International Conference on Auditory Display (ICAD2004)*. International Community for Auditory Display (ICAD), 2004. 2, 12, 19, 22

T. Stockman. Listening to people, objects and interactions. In *Proceedings of ISon 2010, 3rd Interactive Sonification Workshop*, pages 3–8, 2010. 11, 15

T. Stockman and O. Metatla. The influence of screen-readers on web cognition. In *Proceedings of ADDW 2008, Accessible Design in the Digital World Conference*, 2008. 13, 252

R. Stults. Media space. Technical report, Xerox Palo Alto Research Centre, 1986. 116

D. D. Suthers and C. D. Hundhausen. An experimental study of the effects of representational guidance on collaborative learning processes. *Journal of the Learning Sciences*, 12(2):183–218, 2003. 4

J. C. Tang. Why do users like video? studies of multimedia-supported collaboration. Technical report, Mountain View, CA, USA, 1992. 117, 118

E. Tanhua-Piiroinen, V. Pasto, R. Raisamo, and E.-L. Sallnäs. Supporting collaboration between visually impaired and sighted children in a multimodal learning environment. *Haptic and Audio Interaction Design*, pages 11–20, 2008. 141

H. Thimbleby. Character level ambiguity: consequences for user interface design. *International Journal of Man-Machine Studies*, 16(2):211 – 225, 1982. ISSN 0020-7373. 107

B. Truax. *Acoustic Communication*. Ablex Publishing, 2nd edition, 2001. ISBN 1-56750-536-8. 199

R. Upson. Sonifications as mathematics teaching tools. In J. Hiipakka, N. Zacharov, and T. Takala, editors, *Proceedings of the 7th International Conference on Auditory Display (ICAD2001)*, pages 217–221, Espoo, Finland, 2001. Laboratory of Acoustics and Audio Signal Processing and the Telecommunications Software and Multimedia Laboratory, Helsinki University of Technology. 19

A. H. Vera and H. A. Simon. Situated action: A symbolic interpretation. *Cognitive Science*, 17 (1):7 – 48, 1993. ISSN 0364-0213. 34

B. N. Walker, A. Nance, and J. Lindsay. Spearcons: speech-based earcons improve navigation performance in auditory menus. In T. Stockman, L. V. Nickerson, C. Frauenberger, A. D. N. Edwards, and D. Brock, editors, *Proceedings of the 12th International Conference on Auditory Display*, pages 63–68. Department of Computer Science, Queen Mary, University of London, UK, 2006. 18, 102, 249

S. A. Wall and S. A. Brewster. Tac-tiles: multimodal pie charts for visually impaired users. In *NordiCHI '06: Proceedings of the 4th Nordic conference on Human-computer interaction*, pages 9–18, 2006. ISBN 1-59593-325-5. 23, 24

G. Weber. Adapting direct manipulation for blind users. In *CHI '93: INTERACT '93 and CHI '93 conference companion on Human factors in computing systems*, pages 21–22, 1993. ISBN 0-89791-574-7. 23

E. M. Wenzel. Localization in virtual acoustic displays. *Presence: Teleoperators and Virtual Environments.*, 1(1):80–107, 1992. ISSN 1054-7460. 11

F. Winberg. Supporting cross-modal collaboration: Adding a social dimension to accessibility. *Haptic and Audio Interaction Design*, pages 102–110, 2006. 25, 124, 125, 126

F. Winberg and J. Bowers. Assembling the senses: towards the design of cooperative interfaces for visually impaired users. In *CSCW '04: Proceedings of the 2004 ACM conference on Computer supported cooperative work*, pages 332–341, 2004. ISBN 1-58113-810-5. 8, 25, 124, 125, 141

F. Winberg and S. O. Hellström. Qualitative aspects of auditory direct manipulation. a case study of the towers of hanoi. In J. Hiipakka, N. Zacharov, and T. Takala, editors, *Proceedings of the 7th International Conference on Auditory Display (ICAD2001)*. Laboratory of Acoustics and Audio Signal Processing and the Telecommunications Software and Multimedia Laboratory, Helsinki University of Technology, 2001. 124

W. Yu, K. Kangas, and S. A. Brewster. Web-based haptic applications for blind people to create virtual graphs. *Haptic Interfaces for Virtual Environment and Teleoperator Systems, 2003. HAPTICS 2003. Proceedings. 11th Symposium on*, pages 318–325, March 2003. 17, 24

J. Zhang. The interaction of internal and external representations in a problem solving task. In *Proceedings of the Thirteenth Annual Conference of Cognitive Science Society*. Hillsdale: NJ: Erlbaum, 1991. 32, 253

J. Zhang. A representational analysis of relational information displays. *International Journal of Human-Computer Studies*, 4:59–74(16), 1996. 40, 41, 42, 237

J. Zhang and D. A. Norman. Representations in distributed cognitive tasks. *Cognitive Science*, 18(1):87–122, 1994. ISSN 0364-0213. 34, 36, 40

J. Zhang and D. A. Norman. A representational analysis of numeration systems. *Cognition*, 57 (3):271 – 295, 1995. ISSN 0010-0277. 40

H. Zhao, C. Plaisant, B. Schneiderman, and R. Duraiswami. Sonification of geo-referenced data for auditory information seeking: Design principle and pilot study. International Community for Auditory Display (ICAD), 2004. 26, 91, 236