A Framework for Inter-referential Awareness in Collaborative Environments

Jeffrey W. Chastine Ying Zhu Jon A. Preston
Department of Computer Science
Georgia State University
Atlanta, GA. USA

***Abstract*—For collaborative environments to be successful, a fundamental requirement is that they provide support for *inter-referential awareness* – or the ability for one participant to refer to a set of objects, and for that reference to be understood by others. Participants in co-located collaboration benefit from the availability of non-verbal communication, including gestures, eye gaze and body movements. However, when geographically separated, they can experience difficulty in communicating - as computer-mediated cues are often compromised. Given the heterogeneity of media and myriad of interaction techniques that exist in groupware, supporting inter-referential awareness can be difficult. We present a unified and systematic way of encapsulating the numerous factors related to this form of awareness through the creation of process-driven ontology. Our framework provides a formal method for describing inter-referential awareness, and serves as approach that interface designers can use to better comprehend the relevant factors involved. This research stems from our previous investigations in inter-referential awareness in collaborative augmented reality environments.**

Keywords- references, framework, ontology

I. INTRODUCTION

Designing interfaces to support awareness in collaborative applications can be difficult; while today's interfaces afford promising methods of communication, they also present interface designers with new challenges by employing a wide variety of devices and interaction techniques. As argued by Sellen, any computer-mediated communication suffers when compared with its face-to-face equivalent [1]. Thus, to support effective communication, designers of distributed collaborative systems are challenged with developing artificial methods to overcome these limitations. For collaboration to be successful, one small, but highly critical type of awareness is *inter-referential awareness* - or the ability for participants to generate accurate references.

Early work by Hall found that over 50% of communication occurs non-verbally [2]. It can be argued that co-located collaborators, therefore, have a natural advantage over their distributed counterparts; sharing physical space provides access to different kinds of communicative information including eye gaze, gesturing and body movements, and thus, deictic terms have meaning. However, when collaborators are geographically separated, to maintain this level of communication, they must rely on computer-mediated techniques. Gutwin et al. recognize the importance of referencing in group spaces, and further define the generic term *workspace awareness* as "*up-to-the-minute knowledge about others' interaction with the task environment*" [3].

The difficulty in generating meaningful references to objects within the environment varies with the communication medium, the application domain and context. Many distributed groupware systems support the use of audio, video, text, 2D and 3D space, and may be either synchronous or asynchronous. Devices for interacting within these interfaces are typically stretched to work across different modalities, such as a mouse interacting with 3D content. Thus, making references to content within these mediums can be complex; objects may be in any number of states, and may be referenced through various criteria. Though we most often focus on the challenges of distributed virtual space, the objects need not be digital - as real-world objects, such as those in mixed reality environments, are part of the natural interaction that exist in shared spaces. These objects acquire additional spatial properties, such as proximity to participants and occlusion by other objects.

The need for inter-referential support is argued throughout the literature, but is highly desirable in systems where asymmetry exists in the level of knowledge between participants (e.g. a teacher/student or expert/technician scenario). Even when participants are assumed to have similar backgrounds, references are still a fundamental requirement – as users must be able to annotate or refer to changes to portions of a design.

Support for inter-referential awareness in collaborative interfaces is crucial, yet it is our observation that no systematic analysis that specifically addresses it has been performed. By creating a framework, we hope to provide methods for better understanding inter-referential awareness, the factors involved, and when referential ambiguity might occur. Our contributions include the creation of a process-driven model for inter-referential and a high-level ontology that encompasses the heterogeneity of media and interaction techniques involved in collaborative systems.

II. BACKGROUND AND RELATED RESEARCH

Dourish and Bellotti's research suggests that referencing is an important activity, and that absence of inter-referential support hinders communication [4]. Participants worked in a multi-windowed shared text editor, and though there was an explicit method of finding the location of other users, no explicit method of generating references was present, forcing one group to develop an alternate technique. Wang and Chee describe deictic referencing as a "*crucial part of the way we communicate in shared spaces*" [5]. They argue that deictic references have no meaning in asynchronous systems, unless participants have knowledge of the context at the time the reference was made. Their application allows participants to create references (annotations) with embedded timestamps.

In their study on interfaces containing combinations of audio and video, Kraut et al. found that "*references to task objects comprised a small but critical proportion of overall messages in each dialogue – objects had to be identified before instructions for working with them could be given*." In audio-only conditions, "*lengthy descriptive sequences were typically required*" and collaborators took more time, suggesting that the visual nature of video allows collaborators to use deictic references in meaningful ways. Finally, they found that participants made use of *feedback* – or that a message is heard or understood - and found that pairs were "*less likely to explicitly acknowledge one another's messages when performing tasks side-by-side*".There are a myriad of other examples; Tartar et al. note that deictic referencing is a ubiquitous task and is viewed as a more efficient means of communication [7]. Newport et al. found that when collaboratively designing 3D objects "*participants used a lot of gesture during the task, especially during acts of reference*", and in the video-mediated version, they used the mouse "*in a deictic manner, to point to parts of a diagram*." [8].

Awareness is a broad term, and several models and tools have been proposed to both conceptualize and support it. Spring et al. developed an awareness tool, CASCADE, attempted to keep the group aware of other participants including their current focus [9]. Rodden generalized and formalized (through the use of graph structures) Benford's spatial model of awareness [10, 11]. Gutwin and Greenberg created a framework of groupware widgets under the umbrella term *workspace awareness*, focusing on the interaction within a shared workspace [3]. Their work shows the power of *radar views* and *telepointers* in groupware, stating that "*pointing to an object is much easier than indicating it through description, but deixis requires that people know what is visible to the other person*" [12].

Groupware has successfully transitioned into the 3D paradigm as well. In their study of virtual 3D spaces, Ott and Dillenbourg discovered that proximity and orientation towards objects help with the interpretation of deictic references [13]. Billinghurst et. al, in their study on 3D CSCW using wearable computing, describe *communication asymmetries* as "*an imbalance in communication introduced by the interface used for communication, the expertise or roles of the people communicating, or the task undertaken*" [14]. They further define a *functionally symmetric* system as one in which the modalities of communication are equal (e.g. both support audio), *socially symmetric* if use of these mediums is equivalent (e.g. video showing the participants' faces), and *information symmetric* if the participants have equivalent knowledge of a domain.

III. The Inter-referential Pipeline

We begin at an abstract level, viewing inter-referential awareness as a sequential process consisting of a *selection*, *representation* and *acknowledgement* phase. In this context, there is an implied set of participants and set of objects.

We describe *selection* as an atomic process in which, through the actions of an individual, an object or collection of objects is chosen. It is possible to decompose selection into a *cognitive cycle* (the mental process of determining the selection), and a *physical cycle* (the resultant act of choosing). Further, *implicit selection does not exist* in our model; after determining which objects are to be selected, a physical cycle *must* eventually occur, though its manifestation may serve an alternate purpose.

We define *representation* as the means through which the attention of others is directed to a set of selected objects. Representation techniques are often visual, such as highlighting or alternate visualizations. Though discussed in more detail below, pointing is a visual representation, just as deictic speech is an auditory representation. The distinction between the selection and representation phase is subtle - especially when addressing non-verbal referencing, such as changing ones pose, gesturing or using deictic speech.

We define *acknowledgement* as the optional act of recognizing a reference and responding; of interest is its formality. In some systems, an acknowledgement may be a gesture, utterance or physical action [6]. In other systems (e.g. distributed collaborative surgery), a guaranteed acknowledgement becomes increasingly important.

IV. The Inter-referential Life Cycle

The pipeline described above serves as the first level of ontology, creating the *inter-referential life cycle*, the boundaries of which are kept shallow. This restriction allows it to more easily merge with existing ontology, and avoids describing domain specific techniques. Figure 1 shows the integration of the concepts of objects and participants as well as the relationships that exist among them. Though more formally defined later, the figure is summarized here for clarity. The process begins with an initiator who has relationships with one or more objects. Through the process of selection, any number of objects are chosen and represented to a set of participants, each of which has relationships with the objects. An acknowledgement may or may not be generated for the initiator of the reference by these participants. Additionally, there exists a context between the initiator and each participant.

Without loss of generality, we make the following assumptions. First, all objects in the system have some method of being uniquely identified, either through a naming scheme, or a set of attributes. Without this property, it becomes difficult (if not impossible) for collaborators to differentiate between objects. Second, time is inherently part of any collaborative environment, and is used here to enforce the order of operations in the pipeline (represented as direction in the graph).

The graph is interconnected by both process and the relationships between the concepts. The environment contains a set of participants Z and an *initiator* $I \in Z$ who intends to generate a reference for one or more receivers $V \subset Z$. Note that V is not defined as $Z - I$, because a reference may not be intended for all recipients, and cannot be defined as $V \subseteq Z$ because it is implied that the initiator already understands the reference.

We define the set of all objects O as a set of artifacts that exist in the shared space. In concordance with the definition of Rodden, objects might be people or other information [10]. These objects contain domain-dependent properties that are separate from their relationship with the participants, such as *ownership*, *virtuality* or *locked* (held by another participant) - though this list is certainly not exhaustive. It is possible, however, to hierarchically organize objects by modality, including visual, auditory and tactile.

The *objects of interest* O_i is a set of 1 or more objects that are intended to be selected by I; thus, $O_i \subseteq O$ and $|O_i| \geq 1$. We further define O_a as the actual object set which contains the resultant set of objects produced by a selection. For clarification, we use standard mathematical definitions to state that two sets are equal if they contain the same elements, regardless of ordering (and thus are equivalent as well).

A set of relationships R_1 exists between I and O. Each relationship contains domain-specific properties such as proximity and whether or not the object is within view. Additionally, relative to the participant, the object may be located within the shared space, in the environment of a remote participant, or within their own environment. A secondary set of relationships R_2 exists between O and V, consisting of similar properties. Thus, $|R_1| = |O|$ and $|R_2| = |O| \times |V|$.

The relationship that exists between the initiator and the other participants is called *context* – or the symmetry that exist between participants [14]. This includes shared history, the amount of common ground participants share (*information symmetry*) as well as the amount of *functional* and *social symmetry*. The role of context is important in generating meaningful references; participants who share similar experience viewing the environment from similar viewpoints will have less difficulty communicating than those where cognitive and communication asymmetries exist [6, 12, 14]. We agree with the claim by Billinghurst et al. that no collaborative application can be perfectly symmetrical, given the varied knowledge and experiences of the participants. To this extent, though it is indeed plausible to devise a metric for context (as a factor of symmetry), we believe this to be beyond the scope of this paper. Finally, we must also define the concept of a *cursor*, which is a location-based identifier within the shared space (e.g. a *telepointer* in Gutwin's work) which can exist in one, two and three dimensions; in co-located collaboration, the cursor may even be the participant's hand.

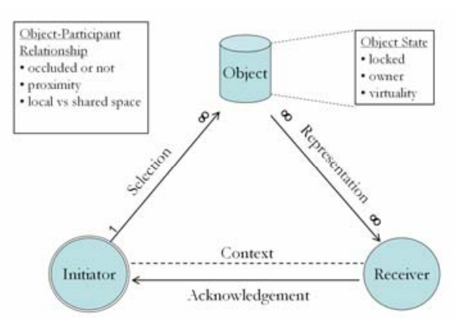

Figure 1. The Inter-referential Life Cycle

A. Selection

We can now formally define *selection* as the specification of a resultant set of target objects using a set of rules for inclusion into (or exclusion from) set O_a; that is, it is a set of functions reducing the universal set of objects $O \rightarrow O_a$. We say that a selection technique is *accurate* if $O \rightarrow O_a = O_i$. The selection technique chosen is highly dependent on context and object type (see Figure 2). For example, a set of rules may specify all objects that contain a specific property (such as *.txt files) or objects whose position fall into a given range (a.k.a - rubberbanding).

Hierarchically, the concept of selection can be further decomposed into *mode* (the medium through which the selection is made) to obtain more specific domains of selection. For example, audio interfaces are becoming more commonplace in mainstream applications. Examples of one-dimensional selection methods include queries, textual selection or 'tabbing' through an interface. Second-dimension techniques include rubber-banding (or bounding box), knife tools, or pixel selection. Third-dimension examples include concepts of gesturing, raycasting, image planes and scaling. The scope of this hierarchy is purposely restricted to abstract concepts, allowing us to leverage off of ontology developed by others. For example, Bowman et al. define a hierarchy for 3D selection techniques [15]. The selection hierarchy is extendable for future methods, including biometric techniques

B. Representation

As defined previously, *representation* is a means through which the attention of others is drawn to objects of interest. Representation techniques are designed to be easily perceived, and are thus inextricably linked to the senses. Given their human-centric nature, representation techniques are open to subjectivity, making representation difficult to define formally. It is the human interpretation of this inference where ambiguity can become a problem. In general, however, we say that any given representation P *infers* a set $O_p \subseteq O$ of 0 or more objects; that is, it is the attempt of the representation to draw the attention of the receiver to O_p. The *intent* behind a representation is that $O_a = O_p$.

Most often, representation techniques are visual - such as changing the appearance of an object through highlighting or by gesturing with a cursor - or auditory, such as deictic speech. However, gesturing and deictic speech should not be confused with *selection*; through these actions, the intent is to draw the attention of others - the definition of representation. The object(s) of reference could also emit audio – the effectiveness of which can be demonstrated in our ability to locate a ringing telephone. Finally, though less common, a representation can be tactile. Examples include force-feedback mechanisms like haptic devices or hand-held game controllers that vibrate.

When designing interfaces to support inter-referential awareness, one must be aware that what can be perceived varies across participants. Visual references may *not* be within view, and auditory references may be too far away to be heard. Therefore, a representation should not only consist of some way of drawing the attention of the receiver(s), but should also *guide* them if the reference is beyond their perception.

It should be made clear that the representation of an object should not be confused with the *feedback* received during the selection; feedback is intended as a confirmation of the selection for the initiator. In many situations, this feedback can be used as a representation for the receivers (e.g. the highlighting of text). However, it is the responsibility of the system to forward this representation in a meaningful way when other participants are in different computing contexts.

In distributed applications, it is evident that for the *system* to generate a representation, it must be aware of the selection. How then is gesturing, movement or deictic speech interpreted in this ontology? They serve as a form of representation, with the argument that one does these actions in order to communicate with others. In the work of Ott, participants were represented by a cone (thus, devoid of orientation information) and used proximity to refer to objects [13]; their change in position was a representation used to make others aware of their selection. Further, if gesture interfaces are used to propagate references to others, the system itself becomes the receiver, and gesturing is still classified as a representation.

Given that collaboration occurs across heterogeneous media, representations can be heterogeneous as well - comprised of a combination of digital representations, non-verbal cues, deictic speech, or movement. It could be argued that redundancy in representations decreases ambiguity.

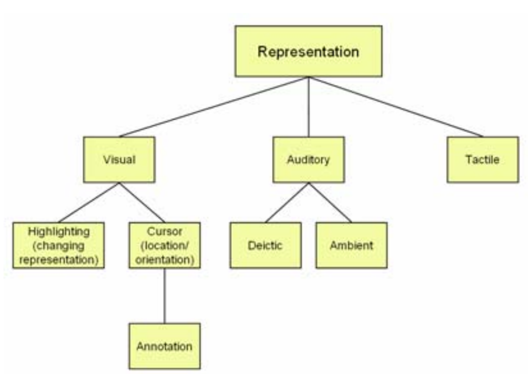

Figure 2. The Representation Hierarchy

An overarching goal for an inter-referential system is to eliminate *referential ambiguity*. We argue that ambiguity occurs from poor selection techniques (during the physical cycle) or weak representations. When poor selection techniques are used, O_a does not equal O_i. Using weak representation techniques, O_p does not equal O_a. Thus, we say that no referential ambiguity exists when $O_i = O_a = O_p$.

C. Acknowledgment

The final phase of the pipeline is *acknowledgement*, and is used as a confirmation for the initiator that a reference is understood by V; depending on the formality of the reference, an acknowledgement is optional. Acknowledgement can take on several forms, including speech or gesture (head-nodding). However, in formal systems where a *guaranteed* reference is required, the transpose of the graph can be taken such that the initiator becomes the receiver (and vice versa) – requiring a complete re-selection of O_p. Though this may seem unnecessary, it becomes increasingly useful in life-dependent scenarios such as collaborative tele-surgery.

V. Definition of an Inter-referntial System

Given the definitions above, it is possible to define a *referential system* $\Re = (I, S, R_1, O, P, R_2, V, C, A)$ where:

- $I \in Z$ is the initiator of the reference
- S represents the selection technique, mapping $O \rightarrow O_a$
- R_1 is the set of relationships that exist between the initiator and O
- O_a is the set of actual objects selected
- P is the representation technique(s) that infer O_p
- R_2 is the set of relationships that exist between O and V
- $V \subset Z$ is the set of receivers of the reference
- C is the context between I and each member of V, and
- A is the optional acknowledgement technique

VI. Domain Study: Inter-Referential Awareness in Collaborative Augmented Reality

This section attempts to validate the core concepts of the ontology against techniques found in collaborative mixed reality interfaces; it is our work in these environments that served as the impetus behind the creation of this framework.

Augmented reality (AR) is the seamless integration of spatially-registered virtual objects into the physical environment. In co-located collaborative augmented reality, multiple participants are provided with their own independent viewpoint, and benefit from the preservation of many non-verbal cues of communication.

Our virtual environment for collaborative molecular modeling (AMMPVis) is governed by a molecular mechanics server, and allows biologists and chemists to manipulate molecular structures while receiving real-time feedback [16]. Biologists were interested in creating an augmented reality version of this application; however, in transitioning from a purely virtual environment to an augmented one, we found the domain to contain additional spatial factors - challenging us to develop new interface techniques for referencing. Co-located participants are now able to gesture both physically and

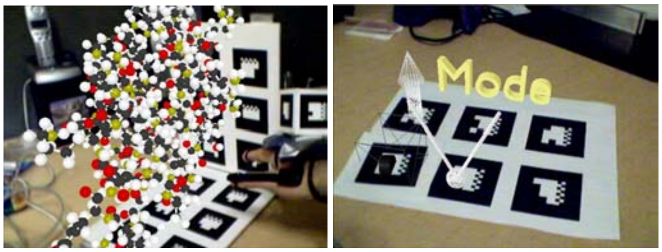

Figure 3. A Collaborative AR Environment for Molecular Modeling

virtually to objects; such interfaces also allow for unusual situations, such as virtual spaces embedded within physical ones (see Figure 4). The relationships between participants and objects have changed as well; objects can be in either the local or remote environment, or be partly or totally occluded by both physical and virtual objects.

The prototype of our interface was built in DART (the Designer's Augmented Reality Toolkit), and provides users with three forms of generating references [17]. Participants manipulate regions of the molecule (such as translating and rotating them) via a data glove. Users are provided with a tangible menu, allowing them to choose from referencing method (see Figure 3).

We have added additional properties to the object and relationship states and have extended the selection hierarchy inter-referential awareness in this domain still adheres to our model; for example, we have included a new instance, the "skew line" technique – a subclass of raycasting – which allows for the referencing of both physical *and* virtual objects. After reviewing much of the literature on collaborative augmented reality, we have found no referential techniques that cannot be instantiated within our model.

VII. Conclusion and Future Work

We have presented a framework to help interface designers better understand inter-referential awareness, which includes a process-driven ontology. Further, we have identified the relevant elements needed in referential systems, and have provided a formal definition to help define and eliminate ambiguous references. In the future, we are interested in developing a metric for both context and referential ambiguity, and are curious about the efficacy of multi-modal representations.

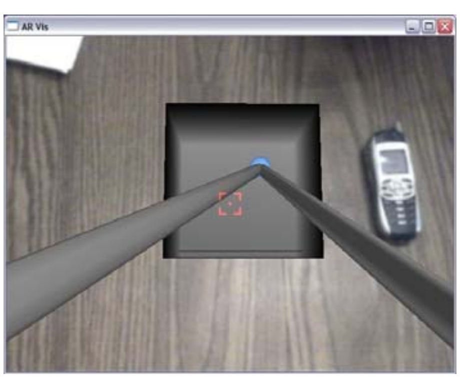

Figure 4. Referencing Virtual Space Beneath a Physical Desktop

References

[1] A. J. Sellen, "Remote Conversations: The Effects of Mediating Talk With Technology," *Human-Computer Interaction*, vol. 10, pp. 401-444, 1995.

[2] E. T. Hall, *The Silent Language*: A Fawcett Premier Book, 1959.

[3] C. Gutwin, S. Greenberg, and M. Roseman, "Workspace Awareness in Real-Time Distributed Groupware: Framework, Widgets, and Evaluation," in *Proceedings of HCI on People and Computers X*, pp. 281-298, 1996.

[4] P. Dourish, and V. Belloti, "Awareness and Coordination in Shared Workspaces," in *Proceedings of the ACM Conference on Computer Supported Collaborative Work (CSCW)*, 1992.

[5] H. Wang, and Y. S. Chee, "Supporting Workspace Awareness in Distance Learning Environments: Issues and Experiences in the Development of a Collaborative Learning System," in *–Proceedings of the Ninth International Conference on Computers in Education*, pp. 1109-1116, 2001.

[6] R. E. Kraut, M. Miller, and J. Siegel, "Collaboration in Performance of Physical Tasks: Effects on Outcomes and Communication," in *Proceedings of the ACM Conference on Computer-Supported Cooperative Work (CSCW)*, 1996.

[7] D. G. Tatar, G. Foster, and D. G. Bobrow, "Design for conversation: Lessons from Cognoter," in *International Journal of Man-Machine Studies*, vol. 34, pp. 185-209, 1991.

[8] A. Newport, A. Anderson, A. Thomson, B. Ion, and N. Dickson, "Using Speech Related Gestures to Aid Referential Communication in Face-to-face and Computer Supported Collaborative Work," in *Proceedings of the First Congress of the International Society for Gesture Studies*, 2002.

[9] M. Spring, R. Andriati, and V. Vathanophas, "Usability of Collaborative Authoring System for Standards Development: Preferences, Problems, and Prognosis," in *Proceedings of the First IEEE Conference on Standardization and Innovation in Information Technology (SIIT)*, 1999.

[10] T. Rodden, "Populating the Application: A Model of Awareness for Cooperative Applications," in *Proceedings of the ACM Conference on Computer-Supported Cooperative Work (CSCW)*,1996.

[11] S. Benford, J. Bowers, L. Fahlén, and C. Greenlaugh, "Managing Mutual Awareness in Collaborative Virtual Environments," in *Proceedings of the ACM Symposium on Virtual Reality Software Technology (VRST)*, 1994.

[12] C. Gutwin and S. Greenberg, "The Effects of Workspace Awareness Support on the Usability of Real-Time Distributed Groupware," *ACM Transactions on Computer-Human Interaction*, vol. 6, pp. 243-281, 1999.

[13] D. Ott and P. Dillenbourg, "Using Proximity and View Awareness to Reduce Referential Ambiguity in a Shared 3D Virtual Environment," in *Proceedings of Conference on Computer Support for Collaborative Learning* (*CSCL)*, 2002.

[14] M. Billinghurst, S. Bee, J. Bowskill, and H. Kato, "Asymmetries in Collaborative Wearable Interfaces," in *Proceedings of the Third International Symposium on Wearable Computers*, pp. 133-140, 1999.

[15] D. Bowman, E. Kruijff, J. LaVoila, and I. Poupyrev, *3D User Interfaces: Theory and Practice*: Addison-Wesley Professional, 2004.

[16] J. W. Chastine, J. C. Brooks, Y. Zhu, G. S. Owen, R. W. Harrison, and I. T. Weber, "AMMP-Vis: A Collaborative Virtual Environment for Molecular Modeling," in *Proceeding of the ACM Symposium on Virtual Reality Software and Technology (VRST)*, 2005.

[17] B. MacIntyre, M. Gandy, J. Bolter, S. Dow, and B. Hannigan, "DART: The Designer's Augmented Reality Toolkit," presented as a demo at the *Second International Symposium on Mixed and Augmented Reality (ISMAR)*, 2003.