Mixed Reality: A model of Mixed Interaction

Céline Coutrix and Laurence Nigay

CLIPS-IMAG Laboratory, University of Grenoble 1, BP 53, 38041 Grenoble Cedex 9, France

33 4 76 51 44 40

{Celine.Coutrix, Laurence.Nigay}@imag.fr

ABSTRACT

Mixed reality systems seek to smoothly link the physical and data processing (digital) environments. Although mixed reality systems are becoming more prevalent, we still do not have a clear understanding of this interaction paradigm. Addressing this problem, this article introduces a new interaction model called Mixed Interaction model. It adopts a unified point of view on mixed reality systems by considering the interaction modalities and forms of multimodality that are involved for defining mixed environments. This article presents the model and its foundations. We then study its unifying and descriptive power by comparing it with existing classification schemes. We finally focus on the generative and evaluative power of the Mixed Interaction model by applying it to design and compare alternative interaction techniques in the context of RAZZLE, a mobile mixed reality game for which the goal of the mobile player is to collect digital jigsaw pieces localized in space.

Categories and Subject Descriptors

D.2.2 [**Software Engineering**]: Design Tools and Techniques - *User interfaces*. H.5.2 [**Information Interfaces And Presentation**] User Interfaces - *Graphical user interfaces, Interaction styles, User-centered design*. I.3.6 [**Computer Graphics**] Methodology and Techniques - *Interaction techniques*.

General Terms

Design, Theory.

Keywords

Augmented Reality-Virtuality, Mixed Reality, Interaction Model, Instrumental Model, Multimodality, Interaction Modality.

1. INTRODUCTION

Mixed reality is an interaction paradigm that seeks to smoothly link the physical and data processing (digital) environments. Although mixed reality systems are becoming more prevalent, we still do not have a clear understanding of this interaction paradigm. Historically, mixed reality systems have been dominated by superimposing visual information on the physical environment. As a proof, consider the taxonomy presented in [8] that defines a Reality-Virtuality continuum in order to classify displays for mixed reality systems. Nevertheless the design and realization of the fusion of the physical and data processing environments (hereafter called physical and digital worlds) may also rely on the use of other interaction modalities than the visual ones. Moreover, the design of mixed reality systems gives rise to new challenges due to the novel roles that physical objects can play in an interactive system; in addition to the design of mixed objects, interacting within such mixed environments composed of physical, mixed and digital objects, involves novel interaction modalities and forms of multimodalities that require new interaction models.

An interaction model [1] aims at providing a framework for guiding designers to create interactive systems. An interaction model can be characterized along three dimensions [1]:

1. descriptive/classification power: the ability to describe a significant range of existing interfaces and to classify them;
2. generative power: the ability to help designers create new designs; and
3. comparative power: the ability to help assess multiple design alternatives.

The article is organized according to these three dimensions. We first present our Mixed Interaction Model and illustrate it with existing mixed reality systems. We then examine its descriptive power by comparing our model with previous classification schemes. We finally study its generative and comparative powers by applying it to design the interaction techniques in RAZZLE, a mobile mixed reality game that we designed and developed.

2. MIXED INTERACTION MODEL

The Mixed Interaction model focuses on the link between the physical and digital worlds and on how the user interacts with the resulting mixed environment. It is based on the notions of physical and digital properties and extends the Instrumental Interaction model [1] by considering the involved mixed objects such as an augmented picture in a museum [14] as well as interaction modalities, such as the manipulation of phicons in the Tangible Geospace [12]. We reuse our definition of a modality [9] as the coupling of a physical device d with an interaction language l: Given that d is a physical device that acquires or delivers information, and l is an interaction language that defines a set of well-formed expressions that convey meaning, a modality m is a pair (d,l). For example, a phicon in the Tangible Geospace [12] is the device d of a modality and the associated language l is the direct manipulation on the table as a reference frame. This definition follows the notion of articulatory and semantic distances of the Theory of Action [10]. We also reuse the different types of composition of modalities defined in [9][22]. For example the manipulation of two phicons in parallel to specify a zoom command corresponds to a case of synergistic use of two modalities (two-handed interaction).

AVI'06, May 23–26, 2006, Venezia, Italy.

The main concept of the Mixed Interaction model is a mixed object. As identified in our ASUR (Adapter, System, User, Real object) design notation for mixed reality systems [4], an object is either a tool used by the user to perform her/his task or the object that is the focus of the task. In other words, a mixed object is either the device *d* of a given interaction modality like a phicon in the Tangible Geospace [12] or the object manipulated by the user by means of interaction modalities, like an augmented picture in a museum [14].

2.1 Mixed Object: Linking Modalities between Physical and Digital Properties

A real object is composed of a set of physical properties and in the same way a digital object is composed of a set of digital properties. A mixed object is then composed of two sets: a set of physical properties linked with a set of digital properties. To describe the link between the two sets of properties we consider the two levels of a modality *(d, l)*. The modalities that define the link between physical and digital properties of an object are called *linking modalities* as opposed to *interaction modalities* used by the user to interact with the mixed environment. Adopting a system point of view, we identify two *linking modalities* for a mixed object as shown in Figure 1:

- An input linking modality (d_o^i, l_o^i) is responsible for
  1. acquiring a subset of *physical properties*, using a device d_o^i (object input device),
  2. interpreting these *acquired physical data* in terms of *digital properties*, using a language l_o^i (object input language).
- An output linking modality is in charge of
  1. generating data based on the set of *digital properties*, using a language l_o^o (object output language),
  2. translating these *generated physical data* into perceivable *physical properties* thanks to a device d_o^o (object output device).

A mixed object may be based on (1) an input linking modality or (2) an output linking modality or (3) input and output linking modalities. In Figure 2, we consider the example of the NaviCam system and we model an augmented picture as a mixed object. A camera captures the physical properties of this object. The image is then translated into the identifier of the recognized picture. The information related to this identified picture is then displayed on the Head-Mounted Display (HMD). The linking modalities of this example are elementary, but input/output linking modalities can also be composed. For characterizing the composition of modalities, we consider the different types of composition based on the CARE (Complementarity, Assignment, Redundancy and Equivalence) framework [9][22]. An example of a composed input linking modality is given in Figure 3: We consider the Mah-Jongg mixed reality game described in [17], in which the player interacts with a Mah-Jongg tile. Since the tile has a location and an orientation from the user's point of view, two input linking modalities (one for location and one for orientation) are combined in order to acquire and interpret data about the position of the tile according to the user's point of view. The resulting digital properties are used for displaying the image of the tile on the HMD. In Figure 3, the composition of the two input linking modalities is represented by a triangle.

To summarize, we can characterize a mixed object by the input and output linking modalities that can be either absent, elementary or composed. Finally we can further characterize a mixed object by reusing characteristics of interaction modalities such as those defined in the theory of modalities [2]: for example an input/output linking modality of a mixed object can be analogue or non-analogue. For instance, the output modality (HMD, l_o^o) of the mixed Mah-Jongg tile, modeled on Figure 3, is analogue (its representation displayed on the HMD being similar to a physical tile). By identifying and characterizing linking modalities, the descriptive power of the Mixed Interaction model is higher than in previous attempts [4], since it goes further than just distinguishing two types of mixed objects, namely tool and object of the task.

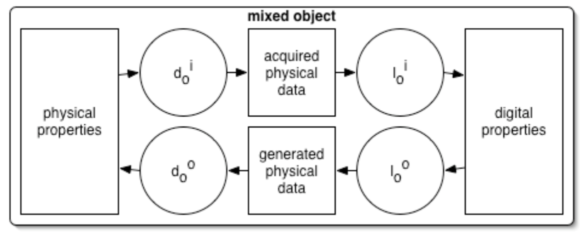


Figure 1. A mixed object.

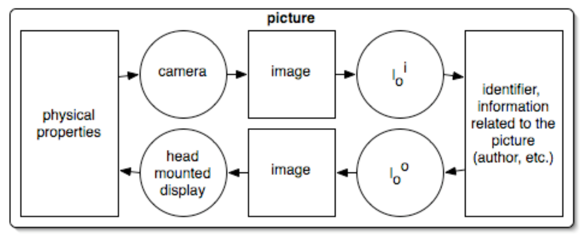


Figure 2. A picture in NaviCam [14].

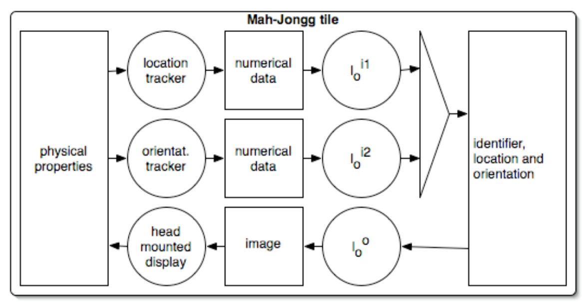


Figure 3. A tile in the Mah-Jongg mixed reality game [17].

2.2 Mixed Interaction

A mixed interaction involves a mixed object. As explained above, a mixed object can either be a tool (i.e., the device of an interaction modality) or be the focus of the task. To model mixed interaction, we extend the Instrumental Interaction model [1] by considering our definition of a mixed object as well as our definition of an interaction modality as the coupling of a device *d* with a language *l*.

In the Instrumental Interaction model, the interaction via a graphical user interface between a user and a domain object is decomposed into two layers as shown in Figure 4: (1) between the user and the instrument, there is the *action* of the user on the instrument, and the *reaction* of the instrument towards the user; (2) between the instrument and the domain object, there is the *command* (or *elementary task*) applied by the instrument onto the domain object, and the *response* of the object to the instrument. Moreover, the domain object can interact directly with the user through the *feedback* it can provide. The instrument or tool is decomposed into a *physical* tool and a *logical* tool. For example in [1], the mouse is a physical tool and a graphical scrollbar is a logical tool. As shown in Figure 4-b, if the physical tool is assigned to a particular elementary task, there is no logical tool. As another example, we consider the paper button used in the DigitalDesk [23]. As shown in Figure 5, "SUM" is written on the paper button and a camera recognizes the written word: it then triggers the computation of the sum of the selected cells. The paper "SUM" button is a dedicated tool, like the physical slider in Figure 4-b. On the contrary, the mouse is a non-dedicated tool and is therefore linked to a logical tool as shown in Figure 4-a.

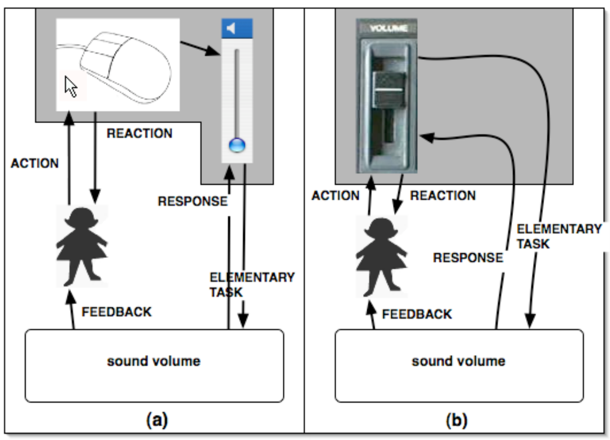


Figure 4. Mixed and logical tools: A non-dedicated (a) vs. a dedicated mixed tool (b).

Figure 5. Dedicated mixed tool in the DigitalDesk [23].

First we extend the Instrumental Interaction model by refining the physical tool as a mixed object called *mixed tool* as well as a domain object as a mixed object called *task object*. Secondly, a tool is the device (*d*) of an interaction modality and a language (*l*) is consequently necessary. For the case containing both physical and logical tools, two languages are required as shown in Figure 6. Indeed a mixed tool is a mixed object, which plays the role of a device of the modality $m_{ti} \rightarrow (mixed\ tool, l_{ti})$. The information conveyed by this modality is related to the digital properties of the logical tool. In turn, another language l_i is required to obtain the elementary tasks from the properties of the logical tool and vice versa to translate the response in terms of digital properties: as a result we obtain a second interaction modality defined as (m_{ti}, l_i). At each level, composition of modalities as defined by the CARE framework [9][22] can be performed.

Figure 6 presents the most general case of a mixed interaction based on an interaction modality, whose physical device is a mixed tool, for manipulating a task object. The user performs an action modifying the physical properties of the mixed tool. The new physical properties are acquired by the tool's input device d_{tool}^i. The acquired physical data are then translated into a set of digital properties of the mixed tool.

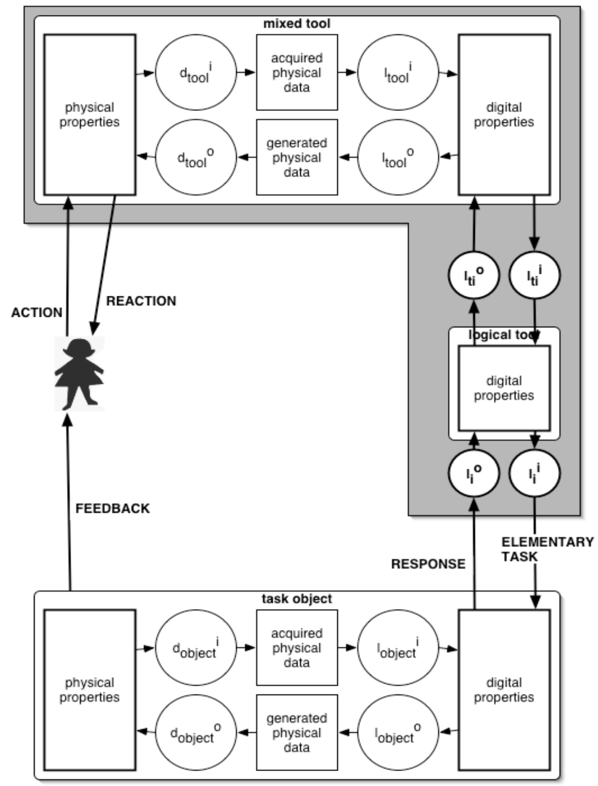


Figure 6. The Mixed Interaction model.

These new digital properties can be perceived by the user through the output linking modality, so that the mixed tool reacts. The digital properties of the mixed tool are then abstracted into the logical tool's digital properties thanks to the input tool's interaction language l_{ti}^i. These digital properties can be perceived by the user thanks to the output tool's interaction language l_{ti}^o and the mixed tool. Finally, based on the input interaction language l_i^i, an elementary task is defined from the digital properties of the logical tool. Moreover an output interaction language l_i^o translates the response from the task object into digital properties, so that the task object can take part in the reaction.

We now illustrate the general case of Figure 6 with two examples. First we consider the example of the DigitalDesk, where the user is pressing the paper button "SUM" of Figure 5: Figure 7 presents the corresponding model of interaction. Secondly, in Figure 8, we model the interaction when the user is manipulating two phicons in the Tangible Geospace [12] for zooming and rotating the map. Two modalities based on mixed tools (i.e., two phicons corresponding to particular buildings) are combined in order to obtain the command, zoom or rotate, that is then applied to the map.

The Mixed Interaction model extends the Instrumental Interaction model by considering the mixed objects and modalities involved in the human-computer interaction. The model underlines two types of modalities, the linking and interaction modalities. We illustrated it by modeling existing mixed reality systems such as the DigitalDesk and the Tangible Geospace. We now examine the unifying and descriptive power of the model by comparing it with existing classification schemes. We will then illustrate its generative and comparative power in Section 4 in the context of the design of a particular mixed reality system, RAZZLE.

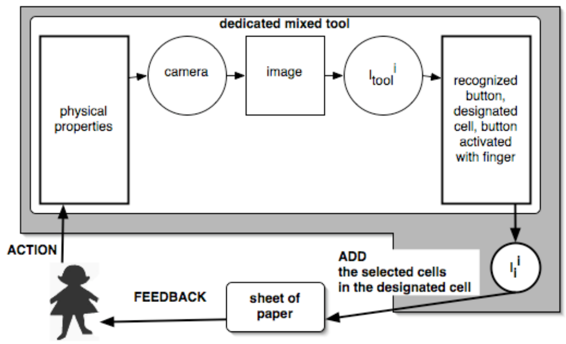


Figure 7. Sum of selected cells using the DigitalDesk.

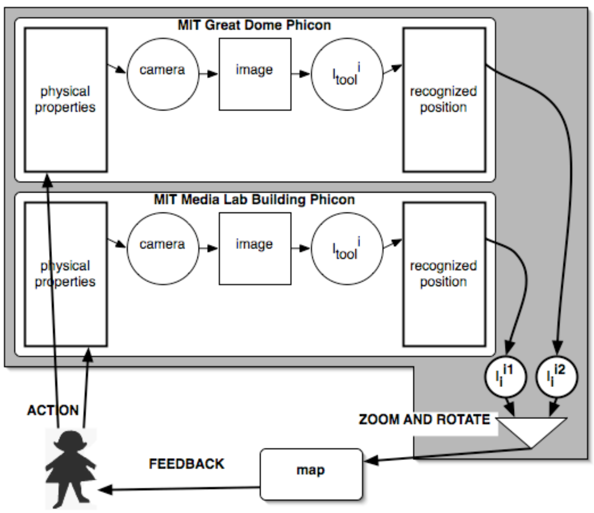


Figure 8. Zooming and rotating the map in the Tangible Geospace.

3. DESCRIPTIVE POWER : COMPARISON WITH RELATED WORK

We have presented the model, and showed that it is well suited for modeling mixed reality systems. In this section we further motivate the model by showing that previous classification schemes of mixed reality systems are accommodated within it, and that the model also reveals fields that were not considered for the evolution of the mixed reality domain. We do so by studying aspects that are related to mixed objects and to mixed interaction.

3.1 Mixed Objects

3.1.1 Physical and digital properties

In [21][16], Tangible User Interfaces (TUI) are described as relationships between tokens, constraints and variables: the TAC model. A token is defined as a graspable physical object; a constraint is a graspable physical object that limits the behavior of the token with which it is associated; a variable is a digital piece of information or a computational function. In our model, a token is described by the physical properties while a variable denotes a digital property. For instance, in the Tangible Geospace, a token corresponds to a phicon (i.e., physical properties) while the variable is the position of the phicon on the table (i.e., digital properties). Moreover the concept of constraint is related to the language of the linking modality l_{tool}^i, by restricting the number of expressions that can be recognized by the language. For example, the table in the Tangible Geospace is a constraint, limiting the manipulation of the phicons to the surface of the table: the position of a phicon (a digital property) will be obtained only if the phicon is on the table. We therefore see how the Mixed Interaction model can accommodate the notions of tokens, constraints and variables as defined in [16].

In [6], a design space of Bricks (i.e., mixed tools) for Graspable User Interfaces is structured along several dimensions. One dimension called "Interaction representation" defines whether an object is physical or digital. We extend this axis by considering three values, digital, physical and mixed. A mixed object is clearly defined in our model by two sets of properties (physical and digital) and linking modalities. Another dimension called "spatially aware", presented in Figure 9, characterizes the presence or absence of spatial information as digital properties of the mixed tool. Nevertheless other digital properties than the spatial ones can define a mixed tool such as the discrete event "open/closed" of the bottle in the ambientROOM [13].

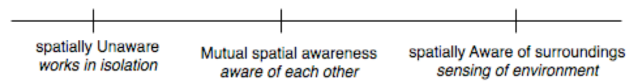


Figure 9. The "spatially aware" dimension in [6].

A last but important aspect concerning the properties of a mixed object is defined by the noun metaphor in [5]: "a <X> in the system is like a <X> in the real world". For example, in the Tangible Geospace [12] (Figure 8), the object that the user manipulates is analogous to MIT's Great Dome, as opposed to a brick in [6] that is a small 1-inch cube. The noun metaphor extends the Mixed Interaction model by further characterizing the physical properties of the mixed object: analogue or non-analogue.

3.1.2 Linking modalities between physical and digital properties

Numerous studies have focused on the link between physical and digital properties.

In the design space of Bricks [6], several dimensions characterize the relationships between physical and digital properties of a mixed object. First the "Input & Output" dimension determines what properties can be sensed and made observable by the user. Based on our definition of an input/output linking modality, we refine the dimension "Input & Output" by considering two levels of abstraction: device and language. For example for the spatial digital property [x, y, z], it can be either the input linking modality *(camera,* l_o^i*)* or *(GPS,* l'^i_o*)*. Another dimension "Bond between Physical & Virtual layers" reveals if the physical and digital layers (physical and digital properties) are tightly coupled or loosely coupled. Such a dimension enriches the Mixed Interaction model by defining a new characteristic of the linking modalities: real-time or batch mode.

In the taxonomy of TUI [5], the Embodiment axis describes how closely the input is tied to the output focus. This axis is to be related to the dimension "Physical & Virtual layers" in the design space of Bricks [6]. They both focus on the spatial continuity. We have previously studied the continuity criterion [4] based on the definition of a modality: perceptual continuity (device level) and cognitive continuity (language continuity). To study continuity within a mixed object, the input and output linking modalities are examined. For instance, for the case of a mouse, spatial continuity is not verified as opposed to the case of an augmented picture in a museum [14] (Figure 2). Continuity is not only spatial but also temporal as we pointed out in [22].

In [7], a "tool" corresponds to a mixed tool while a "container" defines a task object. A container is further described as a generic object that can be reassigned through time. This definition raises the issue of dynamicity of the linking modalities that is currently not covered in our model. Indeed our model describes interaction at a given time. Such an issue is also described in the design space of Bricks by the dimension "Function assignment" along with the three values (permanent, programmable and transient) are identified. In [15] this dimension is refined into three orthogonal axes: Temporality (which can be static or dynamic), interaction mode while defining a mixed object (which can be passive or active from a user's point of view) and interaction mode while modifying a mixed object (which can also be passive or active). For example the mediaBlocks in [18] are mixed objects that are dynamic and the interaction mode for defining or modifying them is active by inserting the mediaBlocks into slots.

About mixed objects, we conclude that our Mixed Interaction model unifies and extends existing frameworks. Moreover by relating our model to previous frameworks, we also identify new characteristics that enrich the model: (1) the noun metaphor [5] for characterizing the physical properties (2) the bond between physical and digital properties (tightly/loosely coupled) [6] as well as the temporality [15] as two additional characteristics of the linking modalities.

3.2 Mixed Interaction

To study mixed interaction in the light of previous frameworks, we first consider mixed modalities (device and language) as well as their combined usages. We then study frameworks that describe the entire interaction process.

3.2.1 Mixed modalities

Both in [1] and in [6], space-multiplexed and time-multiplexed interactions are defined. Space-multiplexed interaction is when several mixed tools, assigned to a task, are available in space at the same instant. Time-multiplexed interaction is when a single mixed tool in the space at a given time can be associated with different logical tools. Inherited from [1], our model underlines this difference by identifying mixed tools and logical tools as illustrated in Figure 4.

Considering the parallel usage of multiple mixed tools at a given time, the design space of Graspable User Interfaces [6] includes two dimensions: "Bricks in use at same time" and "Spatially aware". The first dimension describes the number of bricks that can be used in parallel while the second one presented in Figure 9 identifies one kind of composition, that is the spatial relationship between bricks. This relationship between mixed tools is further refined in [19] by considering three approaches: spatial, relational, and constructive. In spatial approach, the spatial configuration of physical tokens (i.e., mixed tools) is interpreted by the system (often the cartesian position and orientation). Relational approaches map logical relationships between tokens onto computational interpretation. The constructive assembly corresponds to elements connected together mechanically as for the classic LEGO assembly. A system can be spatial, relational, or constructive, or either relational-constructive, etc. In our model, such relationships between mixed tools are studied in the light of composition of modalities at the device or language levels: fusion mechanisms (represented by a triangle in our model) have been extensively studied in the multimodal community. In Figure 10, we present one design space for characterizing the usage of multiple modalities.

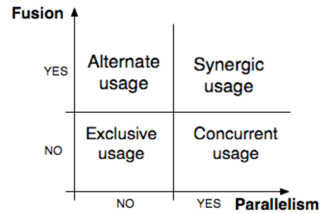


Figure 10. The multimodal system design space [9].

As shown in Figure 11, the fusion can take place at the lower level of abstraction by combining mixed tools. For example in the GraspDraw application [6], rotation is done by manipulating two bricks. Fusion will then be performed at the device level and will define a compound mixed tool. Fusion can also be performed at the language level, such as in Figure 11-b: when the user is manipulating two phicons representing distinct buildings in The Tangible Geospace, the logical properties of the two phicons (i.e., positions on the table) are first interpreted by a language before combining the results (i.e., the new desired positions of the two buildings) in order to obtain the command (zoom, pan, rotate) to be performed on the map.

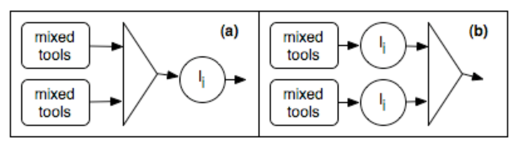


Figure 11: Fusion at the device (a) and language (b) levels.

Finally, we note that the verb metaphor, as defined in [5] "<X>-ing in our system is like <X>-ing in the real world", characterizes the language l_i^i linked to the mixed tool and corresponds to the analogue/non-analogue characteristic of a modality in the theory of modalities [2].

3.2.2 Whole interaction process

A first framework that describes the entire interaction process is the TAC model. As explained in Section 3.1.1, a TAC (Token And Constraints) is the relationship between a token, its variable and one or more constraints. Interaction is described by listing the TACs, each TAC being presented in terms of representation and behavior. Table 1 corresponds to the description of the Tangible Query interface[20], as described in [16]: the user is manipulating sliders in a rack for specifying a query.

Applying our model to the Tangible Query interface, we obtain two dedicated mixed tools (TAC1 et TAC2). They will both be linked to a language l_i^i for interpreting the perceived physical actions on the sliders in terms of query parameters. Results will then be combined to obtain an elementary task, a query (TAC 3). Fusion will be performed at the language level (Figure 11-b). So according to our approach, TAC1 and TAC2 are modeled as two input mixed tools while TAC3 (TAC1 and TAC2) is described as a combined modality for specifying a query. We obtain a similar model as for the Tangible GeoSpace of Figure 8. Within table 1, the distinction between digital properties of a mixed tool and commands or part of commands is not explicit. Moreover the column "observed feedback" in Table 1 does not contain the description of the modalities used to make the feedback perceivable by the user and describes the reaction of the tools as well as the feedback from the system. Table 1 could be extended so that each line describes a pure or combined mixed interaction modality in terms of (Token, Constraint, Digital properties, Commands or Part of commands, Reaction, Feedback). The column "Physical action" could also be moved to the left as in the UAN notation [11]. Finally, it is important to highlight the fact that TAC is dedicated to tangible interaction with digital information. Unlike TAC, the mixed interaction model is dedicated to the design of not only tangible interaction but mixed interaction in general (see for example our RAZZLE system described in the next section).

Another framework for describing mixed interaction is the ASUR notation (Adapter, System, User, Real objects) [4]. For a given task, ASUR describes an interactive system as a set of four kinds of entities, called components:

- Component S: computer System;
- Component U: User of the system;
- Component R: Real object involved in the task (tool (Rtool) or object of the task (Rtask));
- Component A: Adapter (Input Adapter (Ain) or Output Adapter (Aout), bridging the gap between the computer-provided entities (component S) and the real world entities.

Subsequently a relation between two ASUR components describes an exchange between these two components. In our model, the distinction between a mixed tool and a task object is based on the ASUR notation: components Rtool and Rtask. Our model combines the ASUR components R and A for defining a mixed tool or a task object and further characterizes the tool or object by defining the linking modalities. Moreover while ASUR focuses on the bridge between the physical and digital worlds, we model the whole interaction including the interaction modalities (that are parts of the System component in ASUR).

By studying existing frameworks with regard to our model, we have focused on the descriptive power of the model, showing how the model unifies and extends previous frameworks but also how it can be enriched. We now focus on the generative and descriptive power of the model.

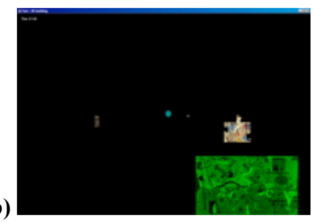
4. GENERATIVE AND COMPARATIVE POWER

To illustrate the generative and comparative power of the model on a concrete example, we consider RAZZLE, a mixed reality system that we designed and developed. Its main features are presented in the next paragraph. The goal of this section is not to show that the mixed interaction model leads to the best solution, but rather, as we stated in the introduction, that the interaction model helps designers to create new designs and it helps them to assess multiple design alternatives.

4.1 RAZZLE

Our study example is the design of RAZZLE. RAZZLE is a mobile augmented game. The goal of the player is to collect the pieces of a digital puzzle. The digital puzzle pieces are scattered all over a modeled playground. The users can access the digital pieces in the physical world thanks to the “augmented field” interaction technique [15]. In a few words, that technique enables users to see digital objects localized in space, if they are well oriented and close enough to the objects. Then, collected digital pieces are added to the puzzle, in order to show the final result. The game ends off when the puzzle is completed. The user wears a see-through Head-Mounted Display (HMD) and is equipped by an orientation sensor (Figure 12-a). We use a wizard of oz technique for simulating location information. Figure 12-b shows a view displayed on the HMD: the user can see the puzzle pieces scattered in space and the puzzle in the foreground.

(a)



(b)

Figure 12. RAZZLE: (a) a player (b) a view displayed on the see-through head-mounted display (black pixels are transparent).

Among the tasks a user can perform with RAZZLE, we only consider the task of collecting a selected puzzle piece. Before focusing on the mixed interaction modalities for collecting a puzzle piece, we first model the task object, that is the puzzle piece. Its model is similar to the one in Figure 3, where we describe the tile in the Mah-Jongg game. Indeed since the puzzle piece has a location and an orientation from the user's point of view, two input linking modalities (one for location and one for orientation) are then combined in order to acquire and interpret data about the position of the piece according to the user's point of view. The resulting digital properties are used for displaying the image of the puzzle piece on the HMD.

4.2 Generative Power

Based on the Mixed Interaction model, we will define several design alternatives for enabling a user to collect a selected puzzle piece in RAZZLE. As shown in Figure 13, the design options

Table 1. A TAC Table for describing the Tangible Query Interface (from [16]).

TAC	REPRESENTATION		BEHAVIOR		
	Token	Constraints	Variable	Physical Action	Observed Feedback
1	Upper slider	Parameter slider, Lower slider	Upper bound variable value in Query	Slide vertically	Updated display
2	Lower slider	Parameter slider, Upper slider	Lower bound variable value in Query	Slide vertically	Updated dislay

consist of describing the mixed tool assuming that a language l_l^i is able to translate the digital properties of the mixed tool into the elementary task <collect the selected puzzle piece>. Thanks to the model, we generated eight different modalities for describing the mixed tool.

A first design option is voice commands, as modeled in Figure 14. The digital properties are therefore the recognized words, while the input linking modality is defined by the pair (microphone, voice recognizer). Another design option is to add an output linking modality to provide a reaction of the mixed tool towards the user. For example, we use speech synthesis: the recognized word is repeated to the user (Figure 14).

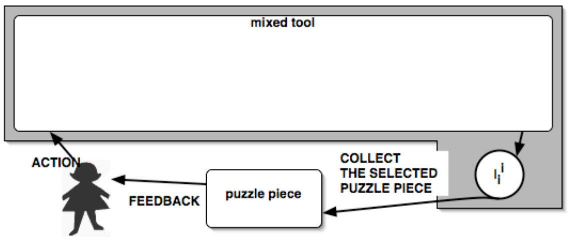


Figure 13. Design options: a frame where to plug in the designed mixed tools.

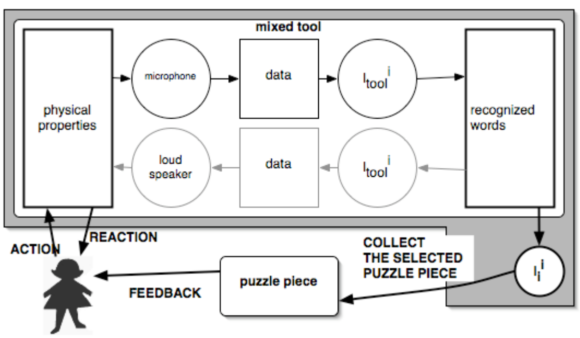


Figure 14 : m1.

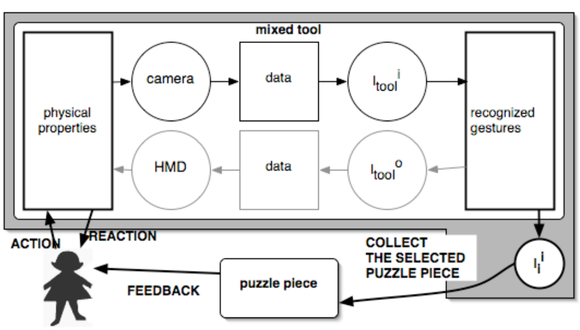


Figure 15. m2.

Another design option is to consider 3D gesture captured by a camera (such as the player in Figure 12-a, who is grabbing a puzzle piece with his hand). The digital properties of the mixed tool are therefore the recognized gestures and the input linking modality is described by the pair *(camera, gesture language)*. Again we can consider an output modality for providing a reaction of the mixed tool towards the user. For instance, we can display on the HMD the name of the recognized gesture (e.g, grabbing, shaking hands, etc.). The corresponding output linking modality is then *(head-mounted display, textual language)* (Figure 15).

A third mixed tool is based on a PDA. As opposed to the two previous design options, the mixed tool is no longer dedicated to a single task and a logical tool is necessary. For example, the RAZZLE player selects with the stylus a graphical button "COLLECT" displayed on the PDA screen. Figure 16 presents the corresponding model of this design solution.

A last design option consists of using a touchpad attached to the wrist, as shown in Figure 17. By simply touching the touchpad, the user collects the puzzle piece. The mixed tool is dedicated to the task of collecting: consequently there is no logical tool.

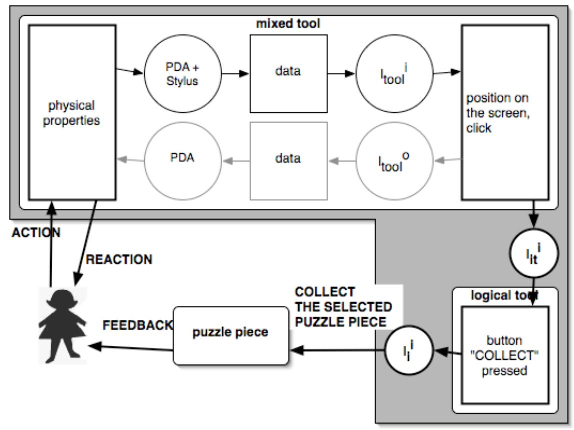


Figure 16. m3.

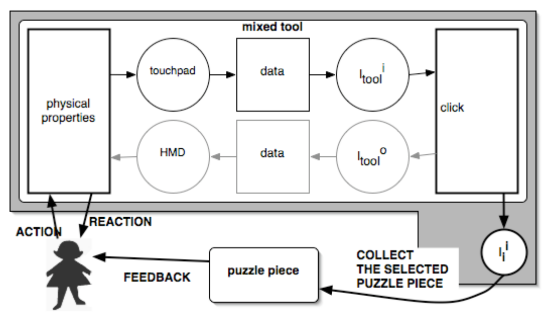


Figure 17. m4.

Based on the model and in particular by focusing on the linking modalities of the mixed tool, several design options can be generated. For example we can also use a cube that will be recognized by a camera: the selected puzzle piece will then be automatically stored in the cube. Having designed several alternatives, we now examine how to compare them.

4.3 Comparative Power

We identified two first criteria that can help assess multiple design alternatives within the model: the continuity and the observability criteria. We examine these criteria in the context of the design of RAZZLE.

We study continuity within a mixed tool as we explained in Section 3.1.2, by considering the input and output linking modalities of the mixed tool. More generally, we can examine all the modalities involved in the interaction for performing the task [4]. For example in the third design option with a PDA, we can conclude that spatial continuity is not verified since the player must always shift between looking at the PDA in order to select the graphical button and looking at the playground. Such design solution can then be eliminated.

For observability, the Mixed Interaction model identifies five different levels: the observability of the state of the task object, the observability of the state of the mixed tool, the observability of the state of the logical tool, the observability of the control of the mixed tool onto the logical tool, and finally, the observability of the control of the logical tool onto the task object. Such refinement of the observability criterion contributes to the evaluative power of the model. For example, while exploring several design options for RAZZLE, the model guided us to consider the output linking modality of the mixed tool in order to provide a reaction of the tool towards the user. In RAZZLE such a reaction is maybe not useful since the player will immediately perceive a feedback from the object, that is the selected puzzle piece disappearing from the playground and being displayed within the puzzle under construction. To increase the evaluative power of the model, empirical results are needed in order to experimentally validate a relevant interaction pattern related to criteria within the model. That is the purpose of the user tests of RAZZLE performed this summer, whose collected data are currently analyzed.

5. CONCLUSION

In this article, we have presented a new interaction model for mixed reality systems. The main contributions of the Mixed Interaction model is (1) to unify several existing approaches on mixed reality systems such as TUI, Augmented Virtuality and Augmented Reality as well as approaches dedicated to more classical GUI and in particular the model of Instrumental Interaction (2) to study mixed reality systems in the light of modality and multimodality.

We intend to further examine the generative power of the model at the design stage, by asking master students to design a particular mixed reality system: one design group applying the Mixed Interaction model while another one without the model. Moreover an interesting research avenue is to study a development tool based on the model. We will further investigate the links between the model and our ICARE tool [3] for developing multimodal interaction. ICARE being based on the definition of a modality as the coupling of a device with a language, the tool should be able to support the development of both linking modalities and interaction modalities.

ACKNOWLEDGMENTS

This work is partly funded by France Telecom R&D, under contract "Mobile AR" and by the SIMILAR European FP6 network of excellence dedicated to multimodality (http://www.similar.cc).

REFERENCES

[1] Beaudoin-Lafon, Designing Interaction, not Interfaces. *AVI'04, 15-22*.

[2] Bernsen, Taxonomy of HCI Systems: State of the Art. *ESPRIT BR GRACE, delivrable 2.1, 1993*.

[3] Bouchet, Nigay, Ganille, ICARE Software Components for Rapidly Developing Multimodal Interfaces. *ICMI'04, 251-258*.

[4] Dubois, Nigay, Troccaz, Consistency in Augmented Reality Systems. *EHCI'01, 117-130*.

[5] Fishkin, A taxonomy for and analysis of tangible interfaces. *Personal Ubiquitous Computing, Vol.8, No.5, 347-358*.

[6] Fitzmaurice, Ishii, Buxton, Bricks: Laying the foundations for Graspable User Interfaces, *CHI'95, 442-449*.

[7] Holmquist, Redström, Ljungstrand, Token-based access to digital information, *HUC'99, 234-245*.

[8] Milgram, Kishino, A Taxonomy of Mixed Reality Visual Displays, *IEICE Transactions on Information Systems, Vol.E77-D, No.12*.

[9] Nigay, Coutaz, The CARE Properties and Their Impact on Software Design. *Intelligence and Multimodality in Multimedia Interfaces: Research and Applications, John Lee, AAAI Press, 1997*.

[10] Norman, Cognitive Engineering. *Book chapter of User Centered System Design, New Perspectives on Human-Computer Interaction, 1986, 31-61*.

[11] Hartson, Siochi, Hix, The UAN: a user-oriented representation for direct manipulation interface designs, *ACM Transactions on Information Systems, Vol.8, No.3, 181-203*.

[12] Ishii, Ullmer, Tangible Bits : Towards Seamless Interfaces between People, Bits and Atoms, *CHI'97, 234-241*.

[13] Ishii, Wisneski, Brave, Dahley, Gorbett, Ullmer, Yarin ambientROOM: Integrating Ambient Media with Architectural Space, *CHI'98, 173-174*.

[14] Rekimoto, Katashi, The World through the Computer: Computer Augmented Interaction with Real World Environments, *UIST'95, 29-36*.

[15] Renevier, Nigay, Bouchet, Pasqualetti, Generic Interaction Techniques for Mobile Collaborative Mixed Systems. *CADUI'2004, 307-320*.

[16] Shaer, Leland, Calvillo, Jacob, The TAC Paradigm: Specifying Tangible User Interfaces, *Personal and Ubiquitous Computing, Vol.8, No.5, 359-369*.

[17] Szalavári, Eckstein, Gervautz. Collaborative Gaming in Augmented Reality. *VRST'98, 195-204*.

[18] Ullmer, Ishii, mediaBlocks: Tangible Interfaces for Online Media. *CHI'99, 31-32*.

[19] Ullmer, Ishii, Emerging Frameworks for Tangible User Interfaces, *Human-Computer Interaction in the New Millenium,, John M. Carroll, ed.; August 2001, 579-601*.

[20] Ullmer, Ishii, Jacob, Tangible Query Interfaces: Physically Constrained Tokens for Manipulating Database Queries. *INTERACT'03*.

[21] Ullmer, Ishii, Jacob, Token+Constraint Systems for Tangible Interaction with Digital Information. *ACM TOCHI, Vol.12, No.1, 81-118*.

[22] Vernier, Nigay, A Framework for the Combination and Characterization of Output Modalities. *DSVIS'00, 32-48*.

[23] Wellner, Interacting with Paper on the DigitalDesk, CACM, *Vol.36, No.7, 87-96*.