Modeling Multimodal Interaction for Performance Evaluation

Emile Verdurand[1,2], Gilles Coppin[1], Franck Poirier[2], and Olivier Grisvard[1,3]

[1] Institut Télécom; Télécom Bretagne; UMR CNRS 3192 Lab-STICC
Technopôle Brest-Iroise CS 83818,
29238 Brest, France
{emile.verdurand,gilles.coppin,
olivier.grisvard}@telecom-bretagne.eu
[2] UBS; VALORIA; Campus de Tohannic, Bât. Yves Coppens, BP 573
56017 Vannes, France
franck.poirier@univ-ubs.fr
[3] THALES Aerospace; Radar & Warfare Systems

Abstract. When designing multimodal systems, the designer faces the problem of the choice of modalities to optimize the system usability. Based on modeling at a high level of abstraction, we propose an evaluation of this choice during the design phase, using a multi-criteria principle. The evaluation focuses on several points of view simultaneously, weighted according to the environment and the nature of the task. It relies on measures estimating the adequacies between the elements involved in the interaction. These measures arise from a fine decomposition of the interaction modalities.

Keywords: modeling, evaluation, modality, context adequacy, interaction language, multimodal interaction.

1 Introduction

Usability is one of the fundamental issues in designing human-system interaction. This issue becomes critical when the systems' complexity increases, for instance in the defense area (e.g. airborne maritime surveillance) or in the medical domain (e.g. assisted surgery). For the time being, users must often carry out tasks under heavy constraints, possibly in critical environmental conditions, while remaining able to handle secondary tasks, such as interacting within a team. In this context, multimodality can allow for more natural and efficient interactions. However, in order to design multimodal human-system interaction on a rational and motivated basis, it is crucial to focus on predicting the usefulness and usability of the modalities, or of their combination, in the context of use. Otherwise, multimodality may not be used [15].

Most of the currently used evaluation methods of multimodal applications are based on empirical analysis that test interaction performances on prototypes or final versions of the system. The major drawback of empirical approaches is a potentially high development cost, notably because of their lack of predictive analysis capabilities [16].

Of course, many analytical methods are proposed in the literature, but one must notice that most of them only offer a restricted choice amongst a limited set of relevant

J.A. Jacko (Ed.): Human-Computer Interaction, Part II, HCII 2009, LNCS 5611, pp. 103–112, 2009.

modalities. Furthermore, they do not allow for predictive evaluation of the real conditions of use of multimodal interaction.

The purpose of this paper is to propose such an approach for predictive evaluation of multimodal interaction, based on some existing models that we propose to extend and federate in a global one. First, we present a brief state of art of the modeling approaches dedicated to evaluation. Then we present our model and we introduce some measures devoted to assessment of the adequacy of modalities to the context of use.

2 Interaction Modeling and Evaluation

Human-machine interaction modeling has given rise to a rich and diverse literature open to many various objectives. Before focusing on models and current needs specifically related to evaluation, we present a brief survey of some existing models, to better situate our work.

2.1 Interaction Models

Interaction models describe different mechanisms implemented to set a dialogue between a user and a system. One can classify these models according to the point of view they favor: some are "task centered" (with cognitive or behavioral modeling), others are "interaction objects centered", and others again are "input / output devices centered" or "interaction situation centered". We shall not elaborate on the first three categories that focus on limited parts of the interaction and thus do not allow for a global analysis such as we aim to achieve[1].

Some models, presenting a global vision, come from different interaction paradigms. We can cite, in relation to the hybrid interaction paradigm, the ASUR notation [8]. It helps describing the various facets of interaction of a user with a mixed system (interaction paradigm for augmented reality / augmented virtuality) and facilitates the exploration of solutions. In the context of interaction with mobile and wearable devices, we can also mention the constraints model presented by Bürgy & Garrett [5] which is based on the one hand on constraints arising from the situation and the activity, and on the other hand on constraints arising from the tasks, the environment, the application, the user and the device.

Finally, in relation to the multimodal interaction paradigm we can cite the model presented by Bouchet [4] for multimodal interaction. It proposes to model the input interaction modalities and their possible compositions, in a global framework for multimodal interaction including task, user and environment modeling. The major limitation of these models is that they are essentially oriented towards the choice of modalities (or modality combinations) in the design phase without evaluative justifications. Some models try to tackle interaction in a predictive way but, as we will see in the next section, in a limited manner.

2.2 Interaction Models Dedicated to Evaluation

Among "restricted" interaction models dedicated to evaluation, we can mention models of the "Goals Operators Methods Selection" family [10]. The goal that the user

[1] See [1] for more information.

wants to achieve is split into sub-goals that can be achieved with adequate methods consisting of an operators series (corresponding to actions). Evaluation is conducted on a limited number of measures (task execution time, procedure learning time for NGOMS, error number). When confronted to the complexity of the phenomena involved in the current systems, these models become increasingly complex as shown in EPIC [12] for example. One of the main critics of these methods relates to the time needed to learn and use them.

Besides, even when models try to consider the global interaction situation, they do not seem to be adequate with an evaluative approach. Thus the approach in [5] is criticized on the coarse granularity of the constraint definition and on the fact that only some specific interaction devices are supported. The approach presented in [13] is not intended to assess a specific interaction task between the system and the users, but mainly the way the task can be carried out along with the interaction device in a wearable computing interaction paradigm.

Actually, it seems there is no work on models that would consider the overall situation of interaction dedicated to a predictive evaluation within the framework of a multimodal interaction paradigm. As such, existing models may suffer from one or several of the following defaults:

- The lack of global modeling of the interaction situation: the users and the environment are frequently missing;
- The lack of relevant human factor consideration: the task nature is not taken into account;
- An insufficient description of interaction modalities, that does not allow for their representation in an uniform way and the implementation of an appropriate evaluation;
- An insufficient number of evaluation criteria used to address the complexity of the interaction situation;
- A superficial interaction evaluation (e.g. completion time): it would be profitable, for a predictive evaluation (or *a priori* evaluation), to address the causes of poor performance, such as the quality of the relations between the elements interacting within the global system.

3 Multimodal Interaction Model

Our model's main objective is to assess the interaction quality allowing the designer to refine the interaction solutions envisioned or make choices among them[2]. We present in the following paragraphs the foundations of our model, and then describe more in details the entities that compose it.

3.1 Overview and Foundations

We use the term *interaction* to designate a mediated communication activity between a user and a system, characterized as a sequence of information and/or action

[2] Our approach registers in the context of both MDE and HCI areas [17] with promising results illustrating the relevance of such approaches. Models developed in model driven design methods can be completed and used in a predictive evaluation phase.

exchanges. Interaction is modeled through the description of the user, the application (interactive task and domain concept), the environment and of course the interaction modalities. This approach federates existing specialized approaches such as of Bouchet [4], Nigay [14] and Beaudouin-Lafon [2], that we complete to provide the grounds for evaluation.

We illustrate in figure 1 how an interaction can be decomposed. The user performs a task (e.g. *file selection*) acting on one or several domain objects (e.g. selection indicator, an attribute of the file : *"is selected"*) amending the internal state of the system. The user manipulates an interaction modality (e.g. *a trackball associated with a pointe*r), or a combination of modalities, and interacts with the device using an interaction language (e.g. *movement language in the plan*). Then, the new system internal state is expressed through an output modality or a combination of output modalities and perceived and interpreted by the user[3].

The entities of the model are expressed along three different abstraction levels, usually identified in the literature, namely: semantic level, articulatory level and physical level (*cf.* figure 1).

3.2 Entities of the Model

We will focus mainly on the two main sub-parts of the modeling which are more directly related to our predictive evaluation goal: the interaction modality on the one hand and the environment on the other hand. The application and the user, modeled along with both his/her physical and cognitive capabilities, will not be described here.

Interaction Modality. Classically, an interaction modality is defined as a couple (d, l), with d an interaction device and l an interaction language [14]. Another definition is given in the context of instrumental interaction [2] as "a mediator or a transducer in

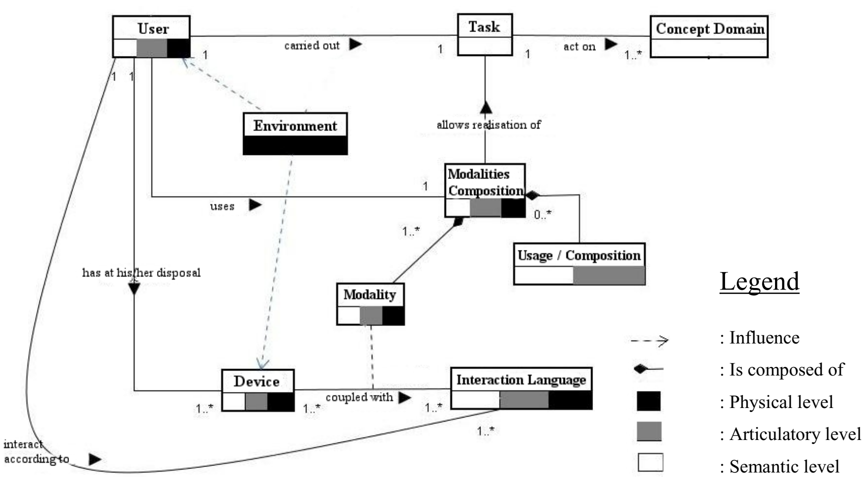

Fig. 1. Global and schematic presentation of the model

[3] For simplicity reasons, output interaction is not represented on figure 1.

both directions between the user and the domain objects (…) the instrument is composed of a physical part, the input device and a logical part, the instrument representation at software level and on the screen".

Our definition aims at associating these approaches and enabling the representation of a broad homogeneous set of possible modalities, allowing for the evaluation of the adequacy of modalities to the context of use. The main composing elements of our model are (see Figure 2):

- The *physical device*: a real world object allowing for information transmission to the computing system (e.g. a microphone);
- The *logical device:* a tool that, combined with information provided by the physical device, allows for the manipulation of objects or domain concepts. For input, it is usually an entry processing algorithm (e.g. a voice recognition engine for the vocal modality);

- The *interaction language* (or representational system) is a "structured system of signs that provides a function of communication between the human and the computer system" [14] (e.g. constrained natural command language for surveillance systems).

We decompose the interaction language into three levels of abstraction:

- *High level* or *semantic interaction language*: it consists of meaning units conveyed by the physical level and expressed in a structured manner by the logical level. It subsumes the two other levels (physical and logical) and because it is inseparable from them, it is expressed through them;
- *Intermediate level* or *logical interaction language*: this language expresses the structure of the information expression conveyed by the modality, and it is constituted by well-constructed interaction concepts (see definition hereafter).

  We define the interaction concepts (see Figure 2) as abstract, generic elements. They can be of event type (with an action dimension), object type or state type, allowing to designate or to represent domain concepts which are precise object or event instances.
- *Low level* or *physical interaction language*: this language covers all the physical actions that convey sense and that can be applied to the physical interaction object or the user input mode (see Bellik's definition [3]). Thus, for a mouse selection, the physical interaction language will be the manual movements applied in the plan on the mouse, and the pressure and click carried out. In the case of the vocal modality, the voice or the sound waves indirectly manipulate the microphone physical device.

Representations of interactions concepts (see Figure 2) are instances of the physical interaction language. The properties characterizing these representations are physical (properties of the sensors and the effectors of interaction modalities). They can be spatial (positions, distances between reference objects, movement, objects size, etc.), or intensity nature (such as volume, color, strength in haptic, etc.). These properties are independent of the information conveyed during the task, but characterize the conveyor.

We will see that the breakdown of the interaction language following these abstraction levels, when associated to the environment description, allows for the representation in one framework, assessable, of various interactions.

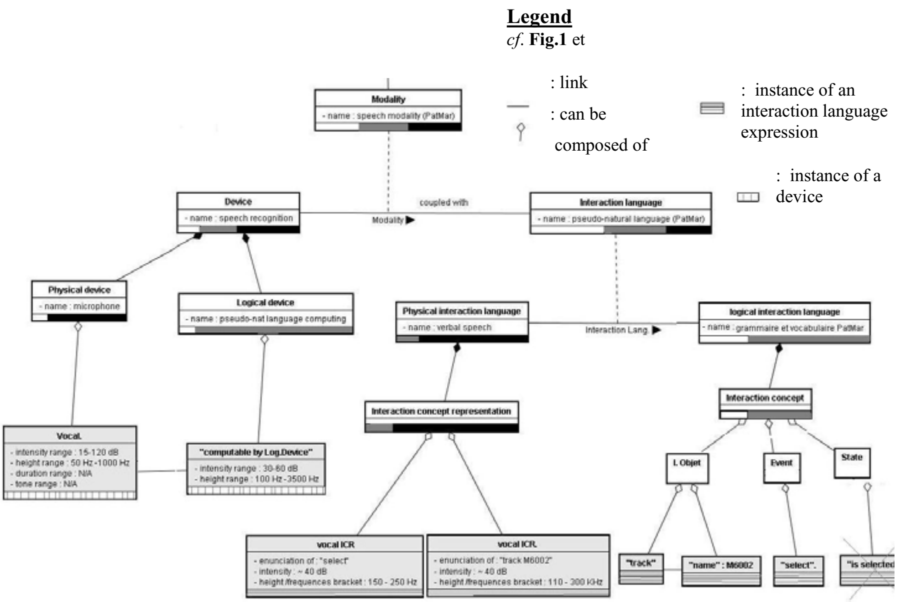

Fig. 2. Modeling of an interaction modality

Environment. We consider the concept of *environment* as close to the broad interaction context definition used by Bellik [3]. Environment is any information relating to a being, an object, a place or a state modification deemed relevant to influence the relations between the user and the system's interaction modalities. We do not define and model the environment in an exhaustive manner, but through all the elements having a significant influence on the user sensitive abilities and the physical input devices.

From the previous descriptions of the modeling of interaction modalities and the environment, it is now possible to set up some measures which allow to compare the level of adequacy of a modality to the task to be performed, among others, and thus to predicatively assess the different design choices[4]. We describe these elements in the next section.

4 Evaluation Method

Traditional approaches focus mainly on the evaluation of results, with the limits we mentioned previously (see 2.3). We propose to tackle the evaluation question through another angle; we focus on the cause of potentially poor interaction performance: the nature of the relations between the elements involved in the interaction. We propose adequacy measures based on properties characterizing the elements of the model.

[4] Our evaluation method is limited, until now, to the comparison of modalities two-by-two through some measures. Modality combination effect evaluation is carried out with more traditional ergonomic rules. We think that it is possible to define adequacy measures based on a large number of these rules.

4.1 A Global Approach

Our approach is in the same line as the existing works of Karsenty [11] and Beaudouin-Lafon [2]. The first distinguishes four classes of modality choice determinants (the task-modality link, a subjective performance evaluation function, the initial preferences, and the user experience). The second presents an approach providing a mean for interaction techniques comparison, introducing the concepts of instrument properties that are at the origin of our own adequacy notion. We propose to generalize this comparison mean, and consequently to lay down the problem differently in assessing adequacy of the overall configuration to the task. Our goal is thus to identify discriminating criteria of modality characteristics allowing to carry out a task in a optimal manner, in function of the user capacities and the environment. This assessment is done through a set of compatibility measures working at different levels of abstraction (semantic, articulatory and physical).

4.2 Adequacy Measures

Characteristics allowing to set up measures. Our adequacy measures rely on some characteristics attached to the model's entities or to elementary components constituting these entities. An interaction device is thus decomposed into a physical device and a logical device, and these two components are – besides others – characterized by their respective degrees of freedom. One can either define the level of effort required for a mental task through three possible values. Thus, the relative effort to an mental interaction task could be estimated by the designer to *medium* for selecting an item in three imbricate sub-menus but to *low* for elocution with a vocal modality. Many of these characteristics or properties are present in the literature [4], [6] and [7], and are used in design frameworks.

Abstraction Levels. We propose to classify the identified adequacy measures according to the three levels of abstraction mentioned previously. We present briefly these three sets of measures.

Top Level Measures – Semantics. These measures[5] allow for indicating adequacy between the expression power of entities implied in the interaction and the semantic level required by the task and the interaction language. They focus notably on the metaphorical intensity highlighted by Fishkin [9], the purpose nature or expressive capacities of the language.

Intermediate Level Measures – Articulatory. Adequacy measures[6] at this level are used to characterize the quality of relationships between entities in different dimensions. These dimensions, briefly mentioned above, group together properties that characterize modeling elements and are used as resources during the interaction.

[5] M^1_S = f(metaphorical intensity, user expertise), M^2_S = f(metaphorical homogeneity), M^3_S = f(grounding, anaphoric references, feedback), M^4_S = f(goal nature, modality).

[6] M^1_A = f(spatial relationship, temporal relationship), M^2_A = f(logical device dimensions, physical device dimensions), M^3_A = f(operator expertise level, language expertise level), M^4_A = f(solicited cognitive capacities, required of interaction language), M^5_A = f(mental efforts ratio: craft mental task - interaction).

One measure incorporates the *integration degree* presented in [2] initially designed for graphical modality. It is defined as the ratio between the number of degrees of freedom (also called dimensions) of the physical device (PD) and the number of dimensions of the logical device (LD) useable with the interaction language.

Another measure concerns the "expertise" homogeneity. The expertise level of the user should be similar to the required expertise level of the interaction language. To relate it to the Cognitive Load Theory [18], we can consider that this measure expresses the adequacy of the user mental schemas with the interaction language used.

Low Level Measures – Physic. Adequacy measures[7] at this level focus on the relationship between the user physical abilities and the physical device precision capabilities, for input as well as output. Through these measures we are also interested in the precision required from the device by the interaction language.

Measures Weighting. Generally speaking, an interaction modality is not good or bad, it can be assessed according to a goal to reach. Therefore, the adequacy measures, alone, do not allow for a meaningful evaluation and should be considered along with the task to achieve. We propose to describe every task with its characteristics "speed", "reliability", "effort" and "required satisfaction" mentioned by Karsenty [11]. All the adequacy measures are representative, in whole or partly, of one or more of the above characteristics. Thus, we can weight every measure with the importance of these characteristics for the task achievement. This weighting allows to meaningfully aggregate measures in a single evaluation indicator.

4.3 Validation

We have started to validate our approach with the results of interviews obtained during a classic empirical evaluation campaign on a prototype of a multimodal maritime surveillance system. The evaluation aimed at showing the contributions of multimodality when compared to a functionally equivalent mono-modal system. We have compared the theoretical results of our predictive evaluation method to the results of the interviews. We have modeled, for a designation task, three different input interaction modalities (vocal, trackball, trackball and vocal combination). We have applied the adequacy measures described in this paper, considering a neutral environment and two different types of tasks (and consequently two different weighting values for measurement): one task with a high reliability and low required effort, the other with a significant required execution speed. The trends observed are compatible with the forecast which can be made out of our modeling.

5 Conclusion

We propose an interaction model of multimodal systems, laid down on several levels of abstraction, describing the entities involved in multimodal interaction and

[7] M^1_P = f(user physical capacities, device physical capacities), M^2_P = f(characteristics « interaction concepts – logical / physical device », user modes characteristics), M^3_P = f(implementation quality of logical device, confidence factor of physical data, effort), M^4_P = f(channel size, quantity of information to transmit, amplitude).

characterizing them by their properties. This detailed description allows us to carry out a predictive evaluation of multimodal interaction performance and to identify its main characteristics and mechanisms.

Taking into account the available resources gives us the possibility to study both inter and intra-device relations and the overall transmission of information. This evaluation is simultaneously performed along several points of view, weighted according to the environment and the task nature. Among the three criteria – descriptive, evaluative and generative – allowing for the evaluation of the quality of the interaction modeling, we have chosen to emphasize the evaluative, contrary to traditional methods which favor the design phase and consequently the generative one.

We plan to conduct experiments focused on a more formal validation of the approach in order to highlight its contributions. The objective, in the end, is to offer an assisting decision tool guiding the designer in his/her preliminary choice of a multimodal configuration during the design of the system.

References

1. Appert, C.: Modélisation, Évaluation et Génération de Techniques d'Interaction. PhD Thesis. Paris: Université Paris Sud (2007)
2. Beaudouin-Lafon, M.: Instrumental interaction: an interaction model for designing post-WIMP user interfaces. In: Proceedings of the SIGCHI conference on Human factors in computing systems. ACM Press, The Hague (2000)
3. Bellik, Y.: Présentation Multimodale de l'Information. HDR. In: LIMSI-CNRS. Paris: Université d'Orsay Paris-Sud (2006)
4. Bouchet, J.: Ingénierie de l'interaction multimodale en entrée. Approche à composants ICARE. PhD Thesis. Grenoble Université Joseph-Fourier - Grenoble I (2006)
5. Bürgy, C., Garrett, J.H.J.: Situation-aware Interface Design: An Interaction Constraints Model for Finding the Right Interaction for Mobile and Wearable Computer Systems. NIST Special Publication (2003)
6. Coutrix, C., Nigay, L.: Interagir avec un objet mixte: Propriétés physique et numérique. In: IHM 2007, Paris (2007)
7. Dubois, E., Gray, P.: A Design-Oriented Information-Flow Refinement of the ASUR Interaction Model. In: Engineering Interactive Systems (EHCI-HCSE-DSVIS 2007), IFIP, Salamanca, Spain (2007)
8. Dubois, E., et al.: Un modèle préliminaire du domaine des systèmes mixtes. In: IHM 2004 Namur, Belgium (2004)
9. Fishkin, K.P.: A taxonomy for and analysis of tangible interfaces. In: Personal Ubiquitous Computing, vol. 8, pp. 347–358. Springer, Heidelberg (2004)
10. John, B.E., Kieras, D.E.: Using GOMS for User Interface Design and Evaluation: Which Technique? ACM Transactions on Computer-Human Interaction 3, 287–319 (1996)
11. Karsenty, L.: Les déterminants du choix d'une modalité d'interaction avec une interface multimodale. In: ERGO-IA 2006 (2006)
12. Kieras, D.E., Wood, S.D., Meyer, D.E.: Predictive engineering models based on the EPIC architecture for a multimodal high-performance human-computer interaction task. ACM Press, New York (1997)
13. Klug, T., Mühlhäuser, M.: Modeling Human Interaction Resources to Support the Design of Wearable Multimodal Systems. In: ICMI 2007. ACM, Nagoya (2007)

14. Nigay, L.: Modalité d'interaction et multimodalité. Université Joseph Fourier, Grenoble (2001)
15. Oviatt, S.: Ten myths of multimodal interaction 42(11), pp. 74–81 (1999)
16. Oviatt, S., Coulston, R., Lunsford, R.: When do we interact multimodally? cognitive load and multimodal communication patterns. In: Proceedings of the 6th international conference on Multimodal interfaces. ACM Press, State College (2004)
17. Sottet, J.-S., et al.: A Model-Driven Engineering Approach for the Usability of Plastic User Interfaces. In: Engineering Interactive Systems, Salamanca, Spain (2007)
18. Sweller, J., Van Merrienboer, J.J.G., Paas, F.G.W.C.: Cognitive Architecture and Instructional Design. Educational Psychology Review 10(3), 251–296 (1998)