A generic formal model for fission of modalities in output multi-modal interactive systems

Linda Mohand Oussaïd
Centre de Recherche sur l’Information Scientifique et Technique
LSI/Université des Sciences Technologiques Houari Boumediene
Algiers, Algeria
www.cerist.dz
lmohand@cerist.dz

Yamine Ait Ameur
Laboratoire d’Informatique Scientifique et Industrielle /
Ecole Nationale Supérieure de Mécanique et d’Aérotechnique
Téléport 2- 1 Avenue Clément Ader BP 40109 86961 Futuroscope Chasseneuil cedex
http://www.lisi.ensma.fr/
yamine@ensma.fr

Mohamed Ahmed Nacer
Laboratoire des Systèmes Informatiques
Université des Sciences Technologiques Houari Boumediene
Algiers, Algeria
http://wwwlsi.usthb.dz/
anacer@mail.cerist.dz

Abstract

Output multi-modal human-machine interfaces combine semantically output medias and modalities in order to increase interaction machine capabilities. The aim of this paper is to present a formal model supporting formal specifications of output multi-modal interactive systems. As a consequence, the expected usability and robustness properties can be expressed and checked. This paper proposes a generic model which permits to specify the output multi-modal interfaces following the CASE design space.

Keywords: Multi-modal interaction, fission of modalities, formal design, formal verification.

1. INTRODUCTION

The emergence of new interaction devices (touch-screens, haptic devices ...) and the use of machines in several situations, with diverse modes (visual, auditory ...) by different users made possible the emergence of several man-machine interaction capabilities. Setting up such interactions, namely multi-modal interactions, increased the user interactions capabilities. Indeed, parallel, concomitant, synergistic, etc. interactions became possible. Moreover, these new interactions capabilities opened several applications for disabled people, in critical systems, in games developments, etc.

The use of multi-modal interfaces increases interactive capabilities of machines, but the sequential and parallel combination of modalities, increases the complexity of the information representation. Therefore, it leads to more complex development and validation processes. Moreover, the introduction of computing interactive systems in critical systems requires a high safety level and needs rigorous design approaches which guarantee the satisfaction of the expressed specifications.

Currently, the traditional approaches used for the specification of the functional core of a system are actually used for the development of interfaces. These traditional approaches are not well adapted to such interactive systems. The main drawback is their incapability to handle the user needs and requirements, from interactive
A generic formal model for fission of modalities in output multi-modal interactive systems

point of view, in the description of the specification. Indeed, usability requirements like tasks analysis are not available in such specifications. Several work has been proposed by the interactive systems community in order to handle such requirements [1], [2]. But few of work has addressed the formal specification of the output multi-modal interaction.

The formal design of input multi-modality interaction has been studied in [3] and [4]. A generic formal model supporting the expression of input multi-modal system interactions has been described. This work focussed on the fusion of input modalities. Our proposal consists in defining such a generic system for the output multi-modal interaction and particularly for the fission of output modalities. In this paper we present a generic and formal model handling output multi-modal system interaction specification.

Our paper is structured as follows. After a brief definition of the main concepts of output multi-modality in section 2 and of most important related work in section 3, we describe, in section 4 the proposed formal model formalizing the specification of an output multi-modal man-machine interface. It includes a fission model refined by an allocation model. Section 5 instantiates this model by the output interaction case study of the Smartkom System [7]. Finally, we conclude and give some perspectives to this work.

2. PRELIMINARIES (OUTPUT MULTI-MODALITY CONCEPTS)

Before describing our formal model, we propose to fix the multi-modality terminology we use in this paper. There are many definitions in the literature [9] according to the point of view (user/technology) or to the abstraction level where definitions are provided. Below we give these definitions and some examples related to output multimodality.

**Mode:** a state or a way in which an action is performed during an output interaction. It determines its interpretation. Examples of output modes are **visual, auditory and tactile**.

**Media:** a physical device in a computing system. It delivers the information to the user. Examples are a **monitor, a speaker**.

**Modality:** is the information structure as it is perceived by the user. For example, **text, alarm, vibration**. A modality combines a mode and a media.

**Statement:** an information produced by an elementary function of the functional core. Function and functional core are elements of the manipulation.

**Interactive task:** elementary or complex informational exchange from machine to user. It presents a statement of the functional core to the external environment.

**Output multi-modal interaction:** interaction in which the information produced by the functional core is split and returned to the user via different output modalities according to the interaction context. For example, when the user asks the system, about the trains going from the city A to the city B, the system answers by speech synthesis: “your query can be satisfied by the four following trains”, and, at the same moment, it displays the detailed trains list. For this example, the text and speech synthesis modalities are used. The text modality is obtained by combining the visual mode to the media **screen** and the speech synthesis is obtained by combining the **auditory** mode to the media **speaker**.

3. OUTPUT MULTI-MODAL INTERACTION : A BRIEF STATE OF THE ART

Output multi-modal systems have been developed in many areas. Examples of such systems are: COMET [5] designed for the automatic generation of maintenance diagnosis of portable military radio. It combines visual modalities (text and graphics). Magic [6] was designed for the generation of cardiovascular postoperative briefings using visual and auditory modalities. Smartkom [7] supports the administration of diverse services: address book, booking hotels, ordering appliances. It is a symmetrical multi-modal system using the same modalities (speech, gesture, facial expression) for input and output interactions.

Multi-modality in general, and output multi-modality in particular, may be defined according to different design processes. Several work has been devoted to the characterization of multi-modal interfaces in order...
A generic formal model for fission of modalities in output multi-modal interactive systems

Semantic fission

UIE$_1$ UIE$_2$ UIE$_n$

PME$_1$ PME$_2$ PME$_n$

Allocation

Instanciation

Information

Dialog controller

Fonctional core

FIGURE 1: The WWHT model

to assess and compare them. The most well known design space is CASE [8]. It characterizes: temporality, statements scheduling and their involvement in the interactive task. The CASE design space suggests classifying multi-modality according to two axes: the composition of medias (sequential or parallel) and the relations between information (combined or independent). This classification leads to four multi-modal interaction types: concurrent (parallel and independent), alternate (sequential and combined), synergistic (parallel and combined) and exclusive (sequential and independent).

Moreover, other work has focussed on modelling the design process of an output multi-modal interaction. For example, the SRM (Standard Reference Model) [10] builds the output interaction starting from the target goal. It involves five layers.

- Control layer to select the next goal to achieve (output interactive task).
- Content layer to first, refine the goal into more specialized subgoals, and second, for each elementary goal, to select the adequate pair (modality, media) and the presentation content.
- Design layer to set the morphological presentation attributes (for example, used font size) and spatial-temporal attributes (timing and layout on the interface).
- Realization layer, responsible of the generation of the effective presentation.
- Presentation display layer whose role is to distribute the different components of the presentation to the appropriate media and coordinate the various components to construct the global presentation.

The WWHT model (What, Which, How, Then) [11] (see Figure 1.) builds the output multi-modal interface by answering to the four following questions:

- What: what is the provided information?
- Which: what is the chosen multi-modal presentation?
- How: how this presentation is instantiated?
- Then: how this presentation evolves?

Answers to these questions leads to follow the multi-modal output presentation architectural design process shown in Figure 1. It includes four steps.
A generic formal model for fission of modalities in output multi-modal interactive systems

- **Semantic fission**: it is a semantic decomposition of the information produced by the functional core into elementary information units that may be processed for output presentation purposes. It answers the question : What?
- **Presentation allocation**: for each elementary information unit, it selects the appropriate multi-modal presentation associated to the current state of the interaction context and consolidates the selected multi-modal presentations into a global one. It answers the question : Which?
- **Instantiation**: determines for the presentation modalities, lexico-syntactic content and morphological attributes according to the interaction context. It answers the question : How?
- **Evolution**: defines the evolution of the multi-modal presentation according to the interaction context change. This evolution can reset the presentation design, according to the interaction context change, either to the allocation phase, or to the instantiation phase. It answers the question : Then?

From this brief overview of the state of the art of output multi-modal interface design, we can state that few work for defining development cycles for such interfaces has been performed. Moreover, we can assert that no formal modelling of the overviewed design spaces is currently available. In practice, only informal design processes are put into practice. The previously overviewed models do not offer a formal framework ensuring a sure design according to rigorous specifications or the possibility to verify functional nor usability properties. Our work proposes a generic and formal model for output multi-modal interaction design and validation based on the WWHT model.

4. THE FORMAL MODEL PROPOSED

The formal model we suggest, formally specifies a multi-modal output interactive system. It models the two first phases of the multi-modal presentation generation process (Figure 1), based on the WWHT model (semantic fission and allocation) in a formal framework and enriches it with additional useful operators.

This model formalizes the successive representations of output information throughout the refinement process of the information generated by the functional core induced by the four steps of the WWHT design model. This refinement leads to the multi-modal presentation. Therefore, two formal models: the fission model and the allocation model compose our formal global model. Each model is defined by its syntax and static and dynamic semantics.

4.1. The Fission model

The fission model expresses the semantic fission or decomposition of the information generated by the functional core into elementary information units delivered to the user. The description of the semantic fission model includes the description of the syntax and the static and dynamic semantics. The objective of fission is to describe the correct basic information unit composition (static semantics) and its temporal occurrence (dynamic semantics); a fission description parameterized for the CASE design space is introduced. Notice that other design spaces could have been defined.

4.1.1. Syntax

Let $I$ be the set of continuous\(^1\) information to fission, and $UIE$ be the set of elementary information units. The description of the fissioned elements is given by the following BNF\(^2\) rules.

$$I ::= UIE \mid (op_{temp}, op_{sem})(I, I) \mid It(n, I) \text{ where } n \in \mathbb{N}$$

Where

- $op_{temp}$ is a binary temporal operator belonging to $TEMP = \{An, Sq, Ct, Cd, Pl, Ch, In\}$.
- $op_{sem}$ is a binary semantic operator belonging to $SEM = \{Cc, Cp, Cr, Pr, Tr\}$.
- $It$ is a binary temporal operator expressing iteration.

\(^1\)Information whose restitution to the user takes a significant time.
\(^2\)Backus Naur Form.
The temporal and semantic binary operators are defined on traces of events that express the production of the information $i \in I$ resulting from the fission description. Their signatures are: $op_{temp} : I \times I \rightarrow I$, $op_{sem} : I \times I \rightarrow I$ and $IT : \mathbb{N} \times I \rightarrow I$.

To define the meaning of the introduced temporal operators, let $i_i, i_j$ be two information elements of $I$, then:

- $i_i$, $S \in \{ i_j \}$ for $i_j$ anachronic to $i_i$, i.e. $i_j$ occurs after an interval of time following the end of $i_i$.
- $i_i$, $S \in \{ i_j \}$ for $i_j$ sequential to $i_i$, i.e. $i_j$ occurs immediately when $i_i$ ends.
- $i_i$, $C \in \{ i_j \}$ for $i_j$ concomitant to $i_i$, i.e. $i_j$ occurs after the beginning of $i_i$ and ends after $i_i$ ends.
- $i_i$, $C \in \{ i_j \}$ for $i_j$ coincident with $i_i$, i.e. $i_j$ occurs after the beginning of $i_i$ and ends before $i_i$ ends.
- $i_i$, $P \in \{ i_j \}$ for $i_j$ parallel to $i_i$, i.e. $i_i$ and $i_j$ begin and end at the same moment.
- $i_i$, $C \in \{ i_j \}$ for choice between $i_i$ and $i_j$, i.e. deterministic choice between $i_i$ and $i_j$.
- $i_i$, $I \in \{ i_j \}$ for independent order between $i_i$ and $i_j$, i.e. the temporal relation between $i_i$ and $i_j$ is unknown.

From the previous definition, we notice that the duration of events is not null. The definition of the semantic operators relies on the semantics of the information elements manipulated by the interface and produced by the functional core. Since our approach is generic, we assume that an interpretation function $int$ interpreting the information elements over a semantic domain $D$ equipped with the relevant operators is available. It is a parameter of the generic approach we propose. The definition of this function together with the semantic domain are given when the interface output information are specified.

Consider the interpretation function $int$ which associates to an information, its semantic interpretation characterizing the multi-modality. Its signature is $int : I \rightarrow D$. $D$ is defined according to the interface output information. The interpretation function together with the semantic domain are given when the interface output information are specified.

Let $int(i) = \{ i_1, \ldots, i_n \}$, then:

- $i_i$, $C \in \{ i_j \}$ for $i_j$, concurrent to $i_i$, $int(i_i)$ and $int(i_j)$ are independent.
- $i_i$, $C \in \{ i_j \}$ for $i_j$, complementary to $i_i$, $int(i_i)$ and $int(i_j)$ are complementary without any redundancy.
- $i_i$, $C \in \{ i_j \}$ for $i_j$, complementary and redundant to $i_i$, $int(i_i)$ and $int(i_j)$ are complementary and a part of their interpretations is redundant.
- $i_i$, $P \in \{ i_j \}$ for $i_j$, partially redundant to $i_i$, $int(i_i)$ is completely included in $int(i_j)$ or $int(i_j)$ is completely included in $int(i_i)$.
- $i_i$, $T \in \{ i_j \}$ for $i_j$, totally redundant to $i_i$, $int(i_i)$ and $int(i_j)$ are equivalent.

Notice that independence, complementarity, inclusion, equivalence, etc. are operators defined on the domain $D$. Given an information $i$ in $I$, and $n$ an integer, $It(n, i)$ expresses that the information $i$ occurs $n$ times.

4.1.2. Static Semantics

The first part of the static semantics has been defined by introducing the $int$ interpretation function and the definition of the semantic operators of the $SEM$ set. The second part is related to the definition of the static properties related to the duration of the information event occurrences. More precisely, it defines the duration of restitution of an information $i$. It is expressed for each elementary information unit by introducing time boundaries using the temporal relation $T$.

Consider $Time = \{ t_1 \}$ the set of discrete time events and the functions $start$ and $end$ defined on $I$ as follows:

$$\forall i \in I \rightarrow Time \ \forall i \in I \ \text{start}(i)$$ is the start of $i$ event occurrence.

$$\forall i \in I \rightarrow Time \ \forall i \in I \ \text{end}(i)$$ is the end of $i$ event occurrence.

$T$ is the temporal relation that combines information with its start and end event occurrence.

$$T : I \rightarrow Time \times Time$$

$\forall i \in I \ \exists (\text{start}(i), \text{end}(i)) \in Time \times Time$ such that $\text{start}(i) < \text{end}(i)$ and $T(i) = (\text{start}(i), \text{end}(i))$.
4.1.3. Dynamic Semantics

The dynamic semantics of the fission model addresses the temporal interleaving of event occurrence. It defines the temporal operators by means of the temporal relation $T$.

$$\forall op_{temp} \in \{An, Sq, Ct, Cd, Pl, Ch, In\} \forall i, j \in I, \text{ with } T(i) = (\text{start}(i), \text{end}(i)),$$

$$T(i) = (\text{start}(i), \text{end}(i)) \text{ iff } op_{temp} = An \text{ and } \text{end}(i) < \text{start}(i)$$

and

$$T(i) = (\text{start}(i), \text{end}(i)) \text{ iff } op_{temp} = Sq \text{ and } \text{end}(i) = \text{start}(i)$$

$$T(i) = (\text{start}(i), \text{end}(i)) \text{ iff } op_{temp} = Ct \text{ and } \text{start}(i) < \text{end}(i) < \text{end}(i)$$

$$T(i) = (\text{start}(i), \text{end}(i)) \text{ iff } op_{temp} = Cd \text{ and } \text{start}(i) < \text{end}(i) < \text{end}(i)$$

$$T(i) = (\text{start}(i), \text{end}(i)) \text{ iff } op_{temp} = Pl \text{ and } \text{start}(i) = \text{end}(i) = \text{start}(i)$$

$$i_k = i_i \lor i_k = i_j \text{ iff } op_{temp} = Ch$$

$$T(i) = (\text{start}(i), \text{end}(i)) \text{ iff } op_{temp} = In$$

Finally, the temporal binary operator $\text{iter}$ is defined. Given $i$, an information in $I$, $n$ a natural integer greater than or equal to 1

$$\text{iter}(n, i) = (\ldots ((i, Sq i) Sq i) \ldots Sq i) \text{ where } n \text{ times}$$

4.1.4. Modelling of CASE space

The CASE formal model is obtained from the formal model previously described, by defining the syntax of the four multi-modality types of the CASE space. We start by describing the temporal and semantic operators allowed for each of the two values of the two space axes (use of media and relations between information). Then, the syntax grammar of the four types resulting from crossing the two axes (Concurrent, alternate, synergistic and exclusive) is provided.

1. **Use of media.** This axis refers to time scheduling. Thus, it concerns the use of temporal operators.
   - Sequential: it constrains the temporal operators excluding the use of parallelism: anachronistic, sequential, choice and iteration.
   - Parallel: it constrains the temporal operators to use types of parallelism: concomitant, coincident and parallel.

2. **Link between information.** This axis refers to the semantic relationship between information and thereby it determines the semantic operators.
   - Combined: restricts the semantic operators to the use of: complementary, complementary and redundant, partially redundant and totally redundant operators.
   - Independent: it restricts semantic operators to the use of concurrent operator.

Thus, the four types of multi-modality resulting from the previous constraints are defined as follows:

**Concurrent type** : $I ::= UIE \mid (op_{temp}, op_{sem})(I, I)$ where $op_{temp} \in \{Ct, Cd, Pl\}$ and $op_{sem} \in \{Cc\}$

**Alternate type** : $I ::= UIE \mid (op_{temp}, op_{sem})(I, I) \mid It(n, I)$ where $n \in \mathbb{N}$, $op_{temp} \in \{An, Sq, Ch\}$ and $op_{sem} \in \{Cp, Cr, Pr, Tr\}$

**Synergistic type** : $I ::= UIE \mid (op_{temp}, op_{sem})(I, I)$ where $op_{temp} \in \{Ct, Cd, Pl\}$ and $op_{sem} \in \{Cp, Cr, Pr, Tr\}$

**Exclusive type** : $I ::= UIE \mid (op_{temp}, op_{sem})(I, I) \mid It(n, I)$ where $n \in \mathbb{N}$, $op_{temp} \in \{An, Sq, Ch\}$ and $op_{sem} \in \{Cc\}$

3rd International Workshop on Verification and Evaluation of Computer and Communication Systems
4.2. Allocation model

Once the information crosses the fission process, the allocation process takes place. The proposed allocation model is based on the second phase of WWHT refinement process. It includes the operators devoted to allocation and proposed in the WWHT model: complementary and redundant, and two other operators we propose: choice and iteration.

The formal allocation model we introduce formalizes the allocation for each elementary information unit resulting from the fission process. They correspond to (modality, media) pairs, combined with the: complementary, redundant, choice and iteration operators in order to apply the usability choices for the output multi-modal interface. These usability choices are provided by the interface designer. Again, we follow the same structure, as for fission, to define the allocation model by describing syntax and semantics.

4.2.1. Syntax

The allocation of output information is described according to two BNF rules: the first one defines the multi-modal presentation pm for an information i, and the second one defines the elementary multi-modal presentation pme related to an elementary information unit uie. Here, the refinement process introduced the concrete presentation of the information issued from fission.

Consider PM as the set of multi-modal presentations, PME the set of elementary multi-modal presentations. We define the multi-modal presentation pm corresponding to the information i obtained after fission as the combination of the different elementary multi-modal presentations pme corresponding to the elementary information unit uie, also issued from fission process, which compose i. This correspondence is formalized as a morphism that maps information (i), information units (uie) together with the semantic (opsem) and temporal (optemp) operators, to multi-modal presentations of the sets PM and PME and the corresponding semantic (op' sem) and temporal (op' temp) operators acting a multi-modal presentation. The multi-modal presentation syntax PM is defined as follows:

\[ PM := PME \mid (op'_temp, op'_sem)(PM, PM) \mid It'(n, PM) \text{ where } n \in \mathbb{N} \]

where:

- op'_temp is a binary temporal operator belonging to TEMP' = \{An', Sq', Ct', Cd', Pl', Ch', Ln'\}.
- op'_sem is a binary semantic operator belonging to SEM' = \{Cc', Cp', Cr', Pr', Tr'\}.
- It' is a binary temporal operator expressing iteration.

The binary temporal and semantic operators are defined on events traces expressing the production of the multi-modal presentations pm ∈ PM. Their signature is given by:

\[ op'_temp : PM \times PM \rightarrow PM; \quad op'_sem : PM \times PM \rightarrow PM; \quad It' : \mathbb{N} \times PM \rightarrow PM \]

Given pm_i, pm_j two multi-modal presentations in PM, then the different operations are described as follows:

- pm_i, An' pm_j for pm_i anachronic to pm_j ie pm_j occurs after an interval of time following the end of pm_i.
- pm_i, Sq' pm_j for pm_i sequential to pm_j ie pm_j occurs immediately after pm_i ends.
- pm_i, Ct' pm_j for pm_i concomitant to pm_j ie pm_j occurs after the beginning of pm_i ends and ends after pm_i ends.
- pm_i, Cd' pm_j for pm_i coincident to pm_j ie pm_j occurs after the beginning of pm_i and ends before pm_j ends.
- pm_i, Pl' pm_j for pm_i parallel to pm_j ie pm_i and pm_j begin and end at the same moment.
- pm_i, Ch' pm_j for choice between pm_i and pm_j ie the deterministic choice between pm_i and pm_j.
- pm_i, In' pm_j for independent order between pm_i and pm_j ie the temporal relation between pm_i and pm_j is unknown.
Analogously to the definition of interpretation function \( \text{int} \) introduced for fission, the interpretation function \( \text{int}' : \text{PM} \to \text{D}' \) is defined for multi-modal presentations. \( \text{D}' \) is the interpretation domain of multi-modal presentations, it is also defined according to the interface context or designer. The semantic operators are:

- \( \text{pm}_i \), \( \text{Cc}' \ \text{pm}_i \) for \( \text{pm}_j \) concurrent to \( \text{pm}_j \) ie \( \text{int}'(\text{pm}_i) \) and \( \text{int}'(\text{pm}_j) \) are independents.
- \( \text{pm}_i \), \( \text{Cp}' \ \text{pm}_j \) for \( \text{pm}_m \) complementary to \( \text{pm}_j \) ie \( \text{int}'(\text{pm}_i) \) and \( \text{int}'(\text{pm}_m) \) are complementary without any redundancy.
- \( \text{pm}_i \), \( \text{Cr}' \ \text{pm}_j \) for \( \text{pm}_n \) complementary and redundant to \( \text{pm}_j \) ie \( \text{int}'(\text{pm}_i) \) and \( \text{int}'(\text{pm}_n) \) are complementary and that a part of their interpretations is redundant.
- \( \text{pm}_i \), \( \text{Pr}' \ \text{pm}_j \) for \( \text{pm}_m \) partially redundant to \( \text{pm}_j \) ie \( \text{int}'(\text{pm}_i) \) is completely included in \( \text{int}'(\text{pm}_m) \) or that \( \text{int}'(\text{pm}_j) \) is completely included in \( \text{int}'(\text{pm}_m) \).
- \( \text{pm}_i \), \( \text{Tr}' \ \text{pm}_j \) for \( \text{pm}_n \) totally redundant to \( \text{pm}_j \) ie \( \text{int}'(\text{pm}_i) \) and \( \text{int}'(\text{pm}_n) \) are equivalent.

Other specific sets defining modalities and medias are required to define the allocation.

Consider \( \text{MOD} \) to be the set of output modalities, \( \text{MED} \) the set of output medias, and the relation \( \text{rest} \) indicating whether a modality can be returned by a media or not; \( \text{rest} : \text{MOD} \times \text{MED} \to \{\text{true}, \text{false}\} \). \( \forall \text{mod}_i \in \text{MOD}, \forall \text{med}_j \in \text{MED}, \text{mod}_i \text{rest med}_j \) expresses that \( \text{mod}_i \) can be returned by \( \text{med}_j \).

Consider \( \text{ITEM} \) the set of pairs (modality, media), such that the modality can be returned by the media. \( \text{ITEM} = \{((\text{mod}_i, \text{med}_j) \text{ such as } \text{mod}_i \in \text{MOD} \wedge \text{med}_j \in \text{MED} \wedge \text{mod}_j \text{rest med}_j)\} \)

And finally, the function \( \text{affect} : \text{UIE} \to \text{ITEM} \) which associates to each elementary information unit \( \text{uie} \), a pair (modality, media) in \( \text{ITEM} \) set;

Then, the syntax for \( \text{PME} \) is described by the following BNF grammar:

\[
\text{PME} ::= \text{affect} (\text{UIE}) \mid \text{iter} (n, \text{PME}) \mid \text{affect} (\text{UIE}) \text{ compl } \text{PME} \mid \text{affect} (\text{UIE}) \text{ redon } \text{PME} \mid \text{affect} (\text{UIE}) \text{ choix } \text{PME} \text{ with } n \in \mathbb{N}
\]

\( \text{Iter} \) expresses \( n \) iterations of an elementary multi-modal presentation \( \text{pme} \).

\[ \text{Iter} : \mathbb{N} \times \text{PME} \to \text{PME} \]

Once, the multi-modal presentations are defined (through the previous BNF rules), the \( \text{repres} \) function \( \text{repres} : \text{PME} \to \text{DR} \) is introduced to associate representation in a representation domain \( \text{DR} \). Here, \( \text{DR} \) is the representational interpretation domain associated to multi-modal presentations.

Hence, \( \text{Compl}, \text{redon} \) and \( \text{choice} \) operators expressing:

- \( \text{Compl} \): representational complementarity between two elementary multi-modal presentations to return an elementary information unit \( \text{uie} \); \( \text{compl} : \text{PME} \times \text{PME} \to \text{PME} \).
- \( \text{redon} \): representational redundancy between two elementary multi-modal presentations to return an elementary information unit \( \text{uie} \); \( \text{redon} : \text{PME} \times \text{PME} \to \text{PME} \).
- \( \text{choice} \): representational choice between two elementary multi-modal presentations to return an elementary information unit \( \text{uie} \); \( \text{choice} : \text{PME} \times \text{PME} \to \text{PME} \).

4.2.2. Static semantics

The static semantics of the allocation model expresses on the one hand, the duration of multimodal presentations \( \text{pm} \) and elementary multi-modal presentations \( \text{pme} \), by the definition of their temporal boundaries, respectively, through the temporal functions \( T' \) and \( T'' \), and on the other hand, a set of properties defined on the syntactic elements of the model. These expressions make it possible to describe the usability properties of the interface.

Consider the \( \text{start}' \) and \( \text{end}' \) functions defined on \( \text{PM} \) as follows:

\[
\text{start}' : \text{PM} \to \text{Time} \ \forall \text{pm}_i \in \text{PM} \ \text{start}'(\text{pm}_i) \text{ is the beginning instant of } \text{pm}_i
\]
Representing two modalities in the same temporal window, on a non-shareable media such as presenting, at the same time, a tone with a sentence produced by speech synthesis.

Consider the \(\text{start}^*\) and \(\text{end}^*\) functions defined on \(\text{PME}\) as follows:

\[
\text{start}^* : \text{PME} \rightarrow \text{Time} \forall \text{pme}_i \in \text{PME} \ \text{start}^*(\text{pme}_i) \text{ is the beginning instant of } \text{pme}_i
\]

\[
\text{end}^* : \text{PME} \rightarrow \text{Time} \forall \text{pme}_i \in \text{PME} \ \text{end}^*(\text{pme}_i) \text{ is the ending instant of } \text{pme}_i
\]

and \(T'\) the temporal relation associating to each elementary multi-modal presentation its beginning and ending instant.

\[
T' : \text{PME} \rightarrow \text{Time} \times \text{Time} \forall \text{pme}_i \in \text{PME}
\]

\[
\exists (\text{start}^*(\text{pme}_i), \text{end}^*(\text{pme}_i)) \in \text{Time} \times \text{Time} \text{ such that } \text{start}^*(\text{pme}_i) < \text{end}^*(\text{pme}_i)
\]

\[
\text{and } T'(\text{pme}_i) = ((\text{start}^*(\text{pme}_i), \text{end}^*(\text{pme}_i))
\]

There are some usability constraints of the output interface that can be relative to the environment (for example, noisy environment compromises the use of auditory mode), the system (for example, using a PDA pleads for the graphic modalities rather than textual ones) and user (a hearing impaired user favours the visual mode to the auditory mode). In order to describe these constraints, we introduce the following static properties:

- The used mode by a media: \(\text{mode} : \text{MED} \rightarrow \{\text{visual, auditory, tactile}\}. \) For example, \(\text{mode}(<\text{speaker}>\) = auditory, \(\text{mode}(<\text{histogram}>\) = visual.

- Media shareability which expresses whether it is possible or not to share a media between more than one modality of one or different types, in order to avoid collisions\(^3\) on a media: \(\text{share} : \text{MED} \rightarrow \{\text{true, false}\}. \) For example, \(\text{share}(<\text{screen}>\) = true, \(\text{share}(<\text{speaker}>\) = false.

- Priorities between modalities: \(\text{priority} : \text{MOD} \rightarrow \mathbb{N}. \) For example, \(\text{priority}(<\text{speech}>\) > \text{priority}(<\text{text}>).\n
- Priorities between modes: \(\text{priority}' : \text{MODE} \rightarrow \mathbb{N}. \) For example, \(\text{priority}'(<\text{visual}>\) > \text{priority}'(<\text{auditory}>). \) (a hearing impaired user).

4.2.3. Dynamic semantics

The dynamic semantics of the allocation model describes the temporal scheduling and relationships of \(\text{pme}_i\). It defines the \text{iter} and \text{choice} operators.

The binary temporal operator \(\text{iter} : \text{PME} \times \mathbb{N} \rightarrow \text{PME}\) is defined using the sequential operator \(\text{Seq}\). Let \(\text{pme}_i\) an elementary multi-modal presentation in \(\text{PME}\) and \(n\) a natural integer greater than or equal to 1.

\[
\text{iter}(n, \text{pme}_i) = (\ldots (\text{pme}_i\text{Seqpme}_i)\text{Seqpme}_i)\ldots \text{Seqpme}_i)
\]

The sequential operator \(\text{Seq}\) is defined by the temporal relation \(T'\) as follows:

\[
\forall \text{pme}_i, \text{pme}_j \in \text{PME} \times \text{PME} \text{ With } T'(\text{pme}_i) = (\text{start}^*(\text{pme}_i), \text{end}^*(\text{pme}_i)),
\]

\[
T'(\text{pme}_j) = (\text{start}^*(\text{pme}_j), \text{end}^*(\text{pme}_j)) \text{ and } \text{end}^*(\text{pme}_i) = \text{start}^*(\text{pme}_j)
\]

\(^3\)Representing two modalities in the same temporal window, on a non-shareable media such as presenting, at the same time, a tone with a sentence produced by speech synthesis.
The choice operator choice : PME × PME → PME is defined as follows:

∀ pmei, pmej ∈ PME × PME ∃ pmek ∈ PME such as pmek = pmei Choice pmej ⇒ pmek = pmei ∨ pmek = pmej

4.3. Properties expression and verification

Once, the output interface is formally modelled using the model described above, it is possible to verify some usability and robustness properties. For that, it is necessary to express these properties in the same formal language that we used to model the interface, in order to compare the two descriptions. Inclusion and or simulation relations between these descriptions permit to assert (or not) the satisfaction of these properties.

We propose to verify the robustness property consisting in establishing the absence of collisions. A collision corresponds to a parallel use of a non-sharable media to present two modalities. Thus, the generation of a collision implies first, a parallel combination of two elementary multi-modal presentations (pmi and pmj) using one of the three parallel temporal operators (concomitant, coincident and parallel); second, the use of the same non-sharable media in these two elementary multi-modal presentations.

Formally, a collision is produced if the multi-modal presentation is described by the following expression:

\[ pm = (C/t_1 | C/t_2 | P', op_{sem})(pm_i, pm_j) \]

where \( pm_i = affect(uie_i) \odot pm'_i \) and \( pm_j = affect(uie_j) \odot pm'_j \)

and \( affect(uie_i) = (mod_i, med) \) and \( affect(uie_j) = (mod_j, med) \)

and \( \neg share(med) \)

with: \( \{ pm_i, pm_j, pm'_i, pm'_j \} \subset PM; \{ uie_i, uie_j \} \subset UIE; \{ mod_i, mod_j \} \subset MOD; med \in MED \)

We currently study the possible concretization of this verification in verifying tools such as the Promela/Spin model checkers or B event proof method.

5. CASE STUDY

To illustrate the formal model described above, we use an output interaction scenario based on the SmartKom system [7]. Smartkom is a system managing various applications related to communicative services (phone, fax, email ...), access to computing devices and positioning, navigation and road information services. The system SmartKom is a symmetrical multi-modal system. Output multi-modality is supported by the Smartakus conversational agent which has a collection of postures and facial expressions.

The modelled output multi-modal interaction scenario consists of a dialogue between the user and Smartakus. In reaction to the user request about Heidelberg city map, the conversational agent Smartakus answers to the request by means of speech synthesis: "Here you can see the map of the city" and displays at the same time the map of Heidelberg.

5.1. Modelling the output interaction

The interface which responds to the user request is modelled according to the scenario described above. We consider, I as the set of output information containing the singleton \( i = \text{"Here you can see the map of the city"} \) combined with the map of the city of Heidelberg, UIE as the set of elementary information units consisting of \( uie_1 = \text{"Here you can see the map of the city"} \) and \( uie_2 = \text{map of the city of Heidelberg} \). Finally, we also consider the sets: MOD = \{speech, facial expression, picture\} the set of output modes; MED = \{screen, speaker\} the set of medias used in Smartkom; ITEM =
A generic formal model for fission of modalities in output multi-modal interactive systems

\{(speech, speaker), (facial expression, screen), (picture, screen)\} the set of the possible combinations (modality, media).

5.1.1. Semantic Fission

The information \(i\) resulting from fission is expressed by combining the temporal parallel and the semantic complementary operators of the elementary information units \(uie_1\) and \(uie_2\) as follows:

\[ i = (Pl, cp)(uie_1, uie_2) \]

From this expression, we can deduce that the interface is designed according to the synergistic multi-modality type.

5.1.2. Allocation

Elementary multi-modal presentations \(pme_1\) and \(pme_2\) related respectively to the elementary information units \(uie_1\) and \(uie_2\) are:

The elementary information unit \(uie_1\) is expressed by the modality speech on the speaker completed by facial expressions on screen (construction of the conversational agent discourse).

\[ pme_1 = (speech, speaker)(uie_1) \text{ compl} (facial expression, screen)(uie_1) \]

The elementary information unit \(uie_2\) is expressed by the picture modality on the screen (restitution of the map) with the expression \(pme_2 = (picture, screen)(uie_2)\)

Then, the multi-modal presentation \(pm\) corresponding to \(i\) is the combination of temporal parallel and complementary semantic elementary multi-modal presentations \(pme_1\) and \(pme_2\) can be obtained by the expression :

\[ pm = (Pl', C p')( (speech, speaker)(uie_1) \text{ compl} (facial expression, screen)(uie_1)), (picture, screen)(uie_2)) \]

5.2. Properties expression and verification

To ensure that the interaction described above does not produce any collision, the \(pm\) expression which specifies the interaction is checked. The multi-modal presentation combines the two elementary multi-modal presentations using the parallel temporal operator.

\[ pm = (Pl', C p')(pme_1, pme_2) \text{ where} \]

\[ pme_1 = (speech, speaker)(uie_1) \text{ compl} (facial expression, screen)(uie_1) \]

\[ pme_2 = (picture, screen)(uie_2) \]

The multi-modal presentation uses simultaneously the media screen trough the elementary multi-modal presentations \(pme_1\) and \(pme_2\). The screen is a shareable media: \(share(screen) = true\). Hence, the multi-modal presentation \(pm\) does not present any collision.

6. CONCLUSION AND PERSPECTIVES

Multi-modal interfaces are increasingly available in real life systems and in critical systems. Like for the functional core of systems, they require the use and application of rigorous design techniques. However, few work has addressed the modelling of multi-modal interactive systems and mainly no work has addressed the formal modelling of output multi-modal interactive systems. The Available models are still informal, based on empirical rules. There does not exist any a priori approach supporting formal validation.

This paper addressed the formal design of output multi-modal interactive systems. The interests of the proposed approach is twofold. On the one hand, it masters the development process complexity of the
A generic formal model for fission of modalities in output multi-modal interactive systems

interface. This complexity is accentuated by the introduction of synchronization primitives and semantic references between information because of the multi-modal nature of the output interface. The introduction of refinement permitted to decompose the development of the output multi-modal interaction. On the other hand, once the formal model is established, the approach supports the a priori expression of usability properties related to the described interface. In addition, the proposed approach allows the designer to progressively and iteratively define the output multi-modal interface by introducing different characteristics throughout the modelling and refinement process. The proposed approach offers a generic design model and a parameterized design model according to the CASE space.

This work can be pursued in different directions. We mention two of them. Regarding the interface specification and in order to make the proposed model as generic as possible, it seems important to define other instantiations of the proposed model for other design spaces or other multi-modal interactive systems description approaches. The second direction relates to the verification of the expressed properties. We expect to define generic properties like collision (defined in this paper) or shareability qualifying the quality of the modelled output multi-modal interfaces. As a second step, we project to propose safe transformations of the specified output multi-modal system together with its properties, that preserve the semantics of the modelled system, into a target formal technique that supports the verification of desired properties. We are currently working on the definition of such transformations for the Promela/Spin model checker and the event B proof based formal method.

REFERENCES