Pragmatics and Human Factors for Intelligent Multimedia Presentation: A Synthesis and a Set of Principles

Frédéric Landragin

Multimodal Output Generation Symposium

April 4th, 2008, Aberdeen

Objectives

Context: *providing early specifications for multimodal systems (for a private company)*

1. **Clarifying the set of parameters that intervene when presenting**
   - how do human factors (attention, salience) intervene?
   - what are the inputs for the presentation module?
2. **Defining communicative acts for IMMPS**
   - are they the same than in interpretation?
   - how does the dialogue manager exploit them?
3. **Specifying the treatment steps from the output of the dialogue manager to the user interface**

First Classifications, Principles and Architectural Concerns

Roles of an IMMPS

- **WH-** = the dialogue manager takes the **decisions** on
  1. who = to whom the information has to be presented
  2. what = what is the information to present
  3. which = which part of the information has to be valorized
  4. where = where can the information be displayed (= on which devices)
  5. when = when and for how long must the information be presented

- **HOW** = the IMMPS **realizes** this decision by
  - choosing the way of valorizing the related piece of information (cf. item 3)
  - choosing the modality(ies) and then the device(s) to exploit (cf. item 4)
  - dividing the information to determine for each modality the related pieces of information (cf. item 4)
  - dividing the information to spread its presentation over time (cf. item 5)
  - managing a HMI or GUI that is specific to the presentation (navigation buttons)

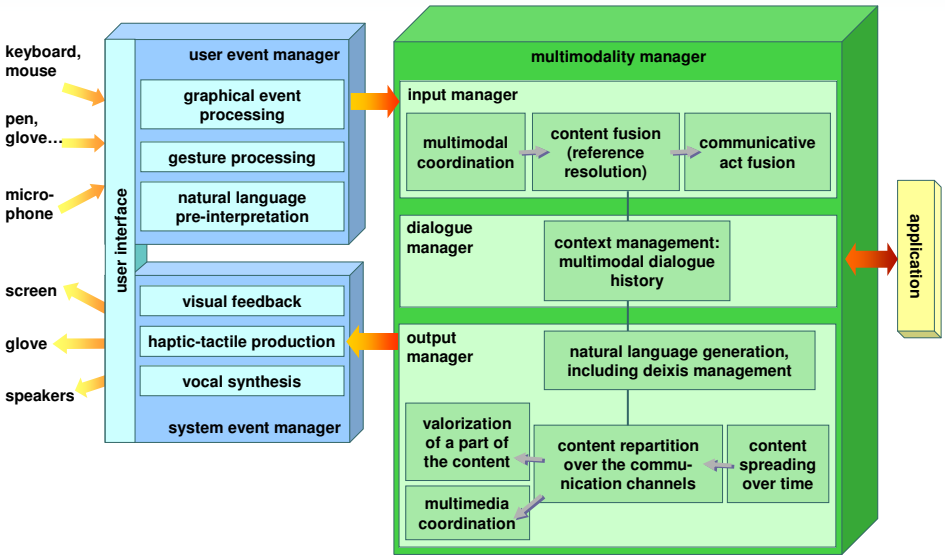
Nine principles for natural IMMPS

- Taking into account the characteristics of the information (or message) in its context (dialogue history for instance)
  1. *with a better repartition of information over the communication channels*
  2. *with a natural rendering and valorization of the information on a channel*
  3. *with a better exploitation of the semantic content of the message*
  4. *maintaining better cohesion and coherence with previous messages*
- Taking into account the characteristics of the terminal, and of the physical and situational environment
  5. *with a more refined exploitation of presentation means*
  6. *with a more refined exploitation of presentation conditions*
- Taking into account the user's physical and cognitive abilities, roles, preferences
  7. *with a better exploitation of the addressee's expectations*
  8. *for a better perception of the message by the addressee*
  9. *for more relevant further reactions from the addressee*

Presentation and adaptation

- Adaptation to the **terminal**
  intervenes during the presentation because the presentation method depends on the terminal characteristics
- Adaptation to the **physical environment**
  intervenes during the presentation because criteria such as background noise level consist of parameters for IMMPS
- Adaptation to the **user's preferences**
  intervenes during the presentation
  (it is of course in the interest of IMMPS to follow the display preferences)
- Adaptation to the **user's roles** (or user task)
  intervenes during the presentation because IMMPS can exploit its knowledge of the user's roles in order to emphasize a piece of information
- Adaptation to the **user's access rights** (or user's prerogatives or profile)
  intervenes before the presentation because the dialogue manager has to filter the information that the user must not know

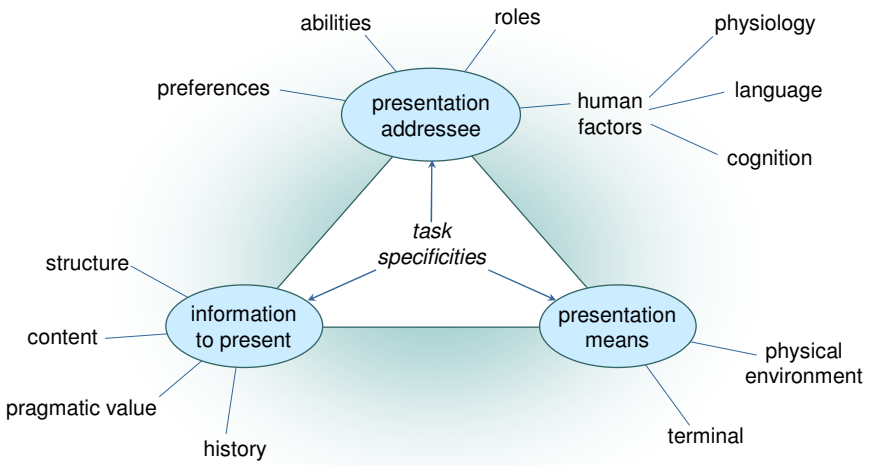
Architecture



Input Parameters for IMMPS

Three poles

A “cartography” for IMMPS parameters



First pole: information itself

- Content of the message to present
  - level of criticality
  - level of urgency
  - information complexity (data structuring, size, number of items)
  - information constitution (discrete versus continuous, e.g.: list of items; timetable)
  - information scope (whole information + zoom in on one element)
  - presentation constraints that are inherent to the information (visual constraint for cartography, no constraint for a linguistic message or for data that can either be displayed or verbalized)
- Message illocutionary and perlocutionary values and forces
  - communicative act(s) that is/are determined by the dialogue manager
  - expected reaction from the user (feedback or not, immediate action or not)
- Interaction history
  - displayed data stack (in order to allow the mention of an already displayed data)
  - displaying action history (to allow the mention of an already executed action)

Second pole: presentation means

- Characteristics of the terminal
  - terminal availability
  - dimensions constraints (screen size)
  - constraints on the processing delays
  - constraints and preferences on output modalities
- Presentation environment
  - **epistemic** constraints (*learning from the environment*):
    picking up and taking into account the ambient noise and the ambient luminosity
  - **ergotic** constraints (*transforming, changing the state of the environment*):
    thresholds for ambient noise and ambient luminosity, that have to not be overstepped in order to not bother the environment
  - **semiotic** constraints (*communicating meaningful information toward the environment*):
    quantity and quality of speech delivery
    (e.g.: too loud or too rapid considering the environment)

Third pole: addressee (or user)

- User's abilities
  - constraints on the communication channels ways of working (e.g.: handicap)
  - constraints and preferences on the communication channels exploitation levels (e.g.: auditory channel already monopolized)
- User's roles
  - constraints on the access rights and bans that come from the user profile
  - constraints and preferences that come from the ongoing user task
- User's preferences (= *individual preferences*)
  - preferences on the linguistic terms and on the presentation metaphors, that were expressed before by the user
  - preferences on the dialogue management, that are detected and exploited by the dialogue manager (e.g.: the user prefers short answers to long answers; the user always prefers to conclude a sub-dialogue before going back to the main dialogue)
- Human factors (= *universal preferences*)
  - **physiological** preferences (color theory; Gestalt theory; salience)
  - **linguistic** preferences (informational structure; salience; Grice's maxims)
  - **cognitive** preferences (short-term memory; focal attention; way of reasoning)

Human factors: physiology

- Sound modality (beep, horn)
  - the more strong the sound is, the more powerful it is (but the more stressful it is)
  - high pitch is more strident than low pitch
- Visual modality
  - color theories: red is perceived much quicker than blue or yellow, and therefore is more often exploited for visual alerts; blue can be perceived much easier in dark environment than in luminous environment; etc.
  - the center of the visual field (corresponding to the fovea) is a privileged place
  - the "good form" notion from Gestalt (e.g.: perfect circle): the more a visual message is close to it, the better the message is perceived
- Whatever the modality, exploitation of saliency and pregnancy
  - a **salient** element, i.e., an element that can be distinguished by singular properties (e.g.: the only red element), is more easily perceived
  - a **pregnant** element, i.e., an element that has been the object of previous repetitions so that it impregnates the user's memory, is more easily perceived

Human factors: language

- Lexical and syntactic levels
  - using the terms and syntactic constructions from the user
  - using simple words and simple syntactic constructions…
- Semantic and pragmatic levels
  - exploiting the Grice's maxims when determining the message to present
  - minimizing the risks of ambiguities (no anaphora if several potential antecedents are possible)
  - avoiding indirect and composite speech acts
- Stylistic level
  - exploiting the **informational (or communicative) structure** in order to put a message element forward (e.g.: putting into salience (or saliencing) by choosing the relevant grammatical function, thematic role, theme, focus, etc.)
  - exploiting the **coherence** (generating a message with a logical link with previous ones)
  - exploiting the **cohesion** (generating a message whose form is in direct continuity with the form of previous messages)

Human factors: cognition

- Lower cognitive processes (perception, attention, memory)
  - short-term memory size: taking its capacity into account (from 5 to 7 independent elements), and limiting the flow of new information
  - exploiting attention: a message can have the purpose to capture selective attention (e.g.: alerts) or to request an important amount of persistent attention for a thorough treatment (presentation of an important information)
  - giving no opportunity for **selective attention** to be captured in various directions
  - giving time to **persistent attention**
- Upper cognitive processes (mental representation, judgment, decision)
  - each message leads to a representation process whose complexity depends on the complexity of the information in its canonical form ⇒ stay inside reasonable limits
  - some pieces of information require a judgment ⇒ do not multiply in the same presentation act such pieces of information
  - because of their visual characteristics, some pieces of information have an influence on actions that can be done on them ⇒ manage such *affordances* in a relevant way
- In a general manner, exploiting all that has already worked well

  (e.g.: if the system noticed that a particular visual message had a positive and efficient influence on the user, it can decide to use it again in similar sensitive situations)

Statement

- From the applicative domain and the user task and the user profile
  - levels of criticality and urgency
  - self-descriptive information (**structured**/**organized** and **quantified** information)
  - presentation constraints and preferences that are specific to the task or task type
- Computed by the dialogue manager
  - pragmatic forces and other labels on the message (e.g.: emotion to render)
  - cohesion and coherence indications
  - linguistic valorizations
  - constraints and preferences on linguistic terms and dialogue management
- Determined by IMMPS on the basis of the constraints from the previous items
  - information ordering (e.g.: depending only on urgency levels)
  - way of dissociating an information into several presentation phases
  - way of dissociating an information over the communication channels
  - for each piece of information, level of valorization (e.g.: depending only on criticality)
  - way of valorizing a piece of information
  - way of exploiting the preferences (in particular if they contradict each other)

Communicative Acts for IMMPS

Illocutionary force for output multimodality

- When interpreting as well as when generating, to the message content is added an illocutionary force that expresses the act that is realized by the enunciation, and that depends on an underlying intention
  - **say that** = the speaker expresses an **assertion** in order to **make** the addressee **know** something
  - **tell to** = the speaker expresses a **demand** in order to **make** the addressee **do** something
  - **ask** = the speaker expresses a **question** in order to **know** something from the addressee (two cases: the close question, i.e., ask-if, whose answer is yes or no, and the open question, i.e., ask-WH, whose answer is an information)
- More precisely on multimedia presentation
  - the way of presenting (for instance an alert) depends on the illocutionary force: if we just want to **inform** (say that), we can use a certain way of presenting that totally differs from the way of presenting we use to **encourage to act** (tell to)
  - the dialogue system may need a confirmation of the message reception: then we can distinguish a "say that without feedback" from a "say that with a mandatory feedback" (it is the use of the "OK" or "OK/Cancel" dialogue boxes)
  - the two previous items can be modeled using composite acts:
    - alert = "**say that** problem" + "**tell to** react to it"
    - tell that with feedback = "**say that** information" + "**tell to**/**ask if** feedback"

Perlocutionary force for output multimodality

- Each message has the aim to produce an effect on its addressee, whatever this effect is (just taking the message content into account, or realizing something precise)
  - it is to the dialogue manager to manage the perlocutionary force, in particular the expected reaction following an order from itself (next state in the user task model)
  - it is to the multimedia presenter to correctly convey the perlocutionary aim, for instance by making obvious a waiting attitude from itself
- Example
  - following the detection of an inconsistency in the database, an alert can:
    - inform the user ("be careful, there is an inconsistency!")
    - encourage the user to give the information he is susceptible to know and that he has not yet passed on
  - first solution: supposing that informing will be sufficient
  - second solution : informing + opening a text box window (equivalent of a "ask-WH")
  - third solution (case of an animated agent): informing + displaying an attitude that clearly conveys an expectation on the user's behavior

Perlocutionary force for a GUI

- The choice of the GUI elements has an influence on the user's further actions
- Classical examples
  - we push buttons
  - we try to modify the content of the cells of a table
  - we try to write some text inside each element that looks like a text box
  - in a general manner we know that each displayed element has a function, and if we don't know this function we try to identify it
- Consequences
  - the presenter must know the functions of all the GUI elements that it may present
  - it must take these functions into account during the various phases of the presentation
  - for each GUI element, it must be aware of the input interaction possibilities (it must inform the input events manager, and indeed the fusion module)
- Example: presenting a table of numeric values can be made using several ways depending on whether the values can be modified or not
  - displaying the cells with a particular color or rendering (grey tint if not modifiable)
  - each cell is accompanied with a text box that makes obvious the possibility to modify

Perlocutionary force for a vocal interface

- Considering the difficulties of speech recognition, several recognition grammars can be specified depending on the type of expected input utterances right after a multimedia presentation
- Classical examples
  - a very **general grammar**, which by consequence is not very precise, is used when the system has to detect the theme in order to launch the related application
  - a **command grammar** is used when the user has the dialogue initiative
  - a **grammar** that is **specific** to numbers, dates, and numeric values, is used when the user must answer to a question from the system that deals with such data
- Consequences
  - during all the presentation phases, the presenter must take into account the type of vocal feedback that is susceptible to follow a presentation act
  - if the dialogue manager does not, the presenter must inform the recognition module
- Example: using the general grammar as a default grammar
  - activating the command grammar further to an inform-like presentation
  - activating a specific grammar further to a question-like presentation

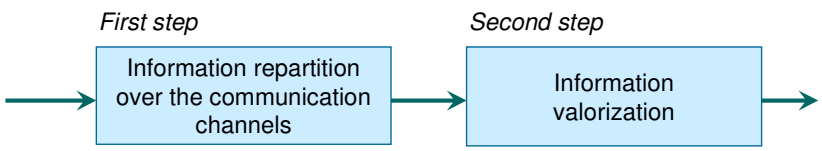
Perlocutionary force for natural language output

- The way of presenting multimedia information has an influence on the user's further linguistic choices
- Examples
  - displaying pieces of information following an obvious order (arrival order or visual organization order) will favor **mentional expressions** such as "the first", "the next one", "the last one"
  - displaying pieces of information that are obviously dissociated will favor **quantified expressions** such as "each", "all the"
- Consequences for the understanding of a linguistic message that follows a multimedia presentation
  - during all the presentation phases, the presenter must take into account the type of vocal feedback that is susceptible to follow a presentation act, and must inform not only the recognition module, but also the module dedicated to natural language understanding
  - the presenter must be aware that it may have to make obvious an ordering that was not expressed by the dialogue manager (e.g. the default occidental visual order)
  - the presenter must be aware of the way it sticks together or not pieces of information

Statement on communicative acts

- Presentation intentions
  - inform without feedback
  - inform with mandatory feedback
  - encourage to react
  - question
- Corresponding communicative acts classification
  - **"inform"** (equivalent to the "say that" speech act, with no feedback required)
  - **"feedback inform"** (equivalent to the "say that" + "tell to"/"ask if" composite speech act)
  - **"demand"** (equivalent to the "tell to" speech act)
  - **"question"** (equivalent to the corresponding speech act, with the distinction between open question and close question)
- Rules with the nature of the information to present as main parameter
  - the communicative act depends on the level of criticality:
    "feedback inform" for a high level and "inform" for a lower level
  - the act also depends on the level of urgency:
    "demand" for a high level and "inform" for a lower level

A Two-Steps Process for IMMPS



1st step: information repartition over the communication channels

- First, take into account the **constraints** that are inherent to the information (visual modality for a map), the **constraints** that are linked to the terminal (no vocal synthesis), the **constraints** that are linked to the presentation environment (strong ambient noise), and the **constraints** that are linked to the user's abilities and roles (handicap, profile)
- Second, take into account a set of rules that relies on:
  - the message content (exploitation of the channel that fits at best the information constitution; exploitation of several channels when the information is very complex)
  - the communicative act (one single channel is favored for a simple act, two channels are favored for a composite act)
  - the interaction history (the channel that is already exploited is favored)
  - the user's preferences (exploitation of one channel if that corresponds to a preference)
  - human factors (displaying a very huge information, whose reading requires an important amount of persistent attention, is spread out over time)
- **To reinforce** the message (high urgency level), duplicate the information over two or all communication channels (= exploit redundancy)
- **To make the link** between several distributed information, indicate with a modality that another part of the message is conveyed using another modality (vocal messages: "on the currently displayed map, you can see…", "flight 102 is the one that flashes")

Link between several pieces of information (deixis)

"Flight 102 is the one that flashes" corresponds to the realization of

1. the dialogue manager produces the following presentation request: make_obvious_to_the_user (flight_102)
2. the multimedia presenter chooses a both visual and vocal realization with the generation of a deixis, so that the user brings the two realizations together
3. the multimedia presenter asks the natural language generation module for materializing the inter-modal deixis (it indicates the nature of the display)
4. the natural language generation module produces the expression "the one that flashes"
5. the multimedia presenter produces "Flight 102 is the one that flashes" and activates the visual flashing rendering (explicit inter-modal deixis)

4. the natural language generation module produces the expression "here is"
5. the multimedia presenter produces "Here is the flight 102" and activates the visual flashing (implicit inter-modal deixis)

Comparison with multimodal fusion and fission

The repartition phase corresponds to the multimodal fission problem

- **at the signal level:** considering its nature, sending information to the correct channel (e.g.: sending the sound track of a video to the auditory channel and the visual track to the visual channel)
  → fission that corresponds to the constraint-based repartition
- **at a semantic level:** dissociating the information content over several modalities in order to better manage its complexity and to simplify the resulting monomodal messages (e.g.: displaying the part of the information that requires an important amount of persistent attention, and verbalizing the part whose only aim is to capture selective attention)
  → fission that corresponds to the preference-based repartition
- **at a pragmatic level:** dissociating the message illocutionary force over several modalities in order to simplify the illocutionary force of each resulting monomodal message (e.g.: 'feedback inform' where 'inform' is verbalized and 'feedback' is required via a text box)
  → not frequent in the literature, but seems to correspond to a kind of fission
- **the 3 previous levels correspond to the 3 multimodal fusion steps:**
  - signal level = multimodal coordination ↔ **multimedia coordination**
  - semantic level = content fusion ↔ **content fission**
  - pragmatic level = event fusion ↔ **presentation act fission**

For and against redundancy

- For redundancy
  - if a channel does not work well, the other one makes up
  - the more information is emitted, the more chances has the addressee to receive it
  - the more information is presented again, the more chances has the addressee to become imbued with it
- Against redundancy
  - too many messages don't encourage the addressee to maintain his persistent attention
  - too many messages increase the processing time and then the reaction delay
  - an air crash example: "– why didn't you answer to the control tower who indicated you that your landing gear was not out?
    – because I had a klaxon that was resonant in my ears!
    – it's incredible! that signal precisely indicated you that your landing gear was not out!"
- Statement
  - exploit redundancy only if the addressee should be able to make the link between the various emissions of the same information (i.e., if he can notice that it is redundancy)
  - don't exploit redundancy in the same communication channel (e.g.: sound + voice)
  - when the message is so urgent or important that it cannot be ignored, be careful that redundancy does not introduce any perturbation

2nd step: information valorization

- First, take into account the **constraints** that are inherent to the information (numbers of lines and columns of a table), the **constraints** that are linked to the terminal (screen size), and the **constraints** that are linked to the presentation environment (threshold for the ambient noise)
- Second, take into account a set of rules that rely on:
  - the message content (display rules linked to data structure)
  - the communicative act (strong intensity for a 'demand' act)
  - the user's preferences (displaying with font size 16 if it is a preference)
  - human factors (exploiting the red color for an alert)
- **To optimize** the content within a modality, spread out the information to the limits of the terminal (when displaying a picture, take all the available space)
- **To emphasize** a content, exploit salience (adjust the communicative structure for putting one element into salience)
- **To render an emotion** on a content, exploit the prosody and if necessary the multiple possibilities of a conversational animated agent
- **To have a hold over user's attention**, take into account the distinction between selective attention (that is captured by a transient verbalization or display) and persistent attention (that requires a persistent or permanent display)

Conclusion and Future Work

Conclusion and future work

- Propositions for the design of a task-oriented IMMPS
  - that integrate adaptation to the terminal, the environment, and the user
  - that consist of three processing phases (filtering, repartitioning, valorizing)
  - that are based on speech act theory
  - that take human factors into account
- Future work
  - clarification of the sets of rules for repartition and valorization phases
  - case study in order to demonstrate the proposal interest and feasibility (e.g.: interactive support for cooperative decision making in the domain of air traffic management)
  - formalization of the propositions