International Journal of Engineering Science Invention
ISSN (Online): 2319 – 6734, ISSN (Print): 2319 – 6726
www.ijesi.org ||Volume 5 Issue 11|| November 2016 || PP. 126-138

ProFi's Process for Multi-modal Fission System

Nibedita Sahoo1, Jagannath Ray2, Narendra Kumar Rout3
[1,2]*Associate Professor, Department of Computer Science Engineering, Gandhi Institute For Technology (GIFT), Bhubaneswar*
[3] *Assistant Professor, Department of Computer Science Engineering, Gandhi Engineering College, Bhubaneswar*

ABSTRACT:*Humanbeingscontinuouslyadapttheirwayofcommunicationtotheirsurroundingsandtheircommunicationpartner.Althoughcontextawareubiquitoussystemsgatheralotofinformationtomaximizetheirfunctionality,theypredominantlyusestaticwaystocommunicate.Inordertofulfilltheuser'scommunicationneedsanddemands,thesensor'sdiverseandsometimesuncertaininformationmustalsobeusedtodynamicallyadapttheuserinterface.Inthisarticlewepresent ProFi,asystemforProbabilisticFission,designedtoreasononadaptiveandmultimodaloutputbasedonuncertainorambiguousdata.Inaddition,wepresentasystemarchitectureaswellasanewmetamodelformultimodalinteractivesystems.BasedonthismetamodelwedescribeProFi'sprocessofmultimodalfissionalongwithourcurrentimplementation.*

Categories and Subject Descriptors
*H.5.2[**InformationInterfacesandPresentation**]:UserInterfaces—Theoryandmethods,Userinterfacemanagementsystems(UIMS);D.2.11[**SoftwareEngineering**]:SoftwareArchitectures*
General Terms
Algorithms,Design,Theory
Keywords*:probabilisticmultimodalfission,modalityarbitration,adaptiveuserinterface,multimodalinteraction*

I. Introduction

Sinceeverybehaviorisakindofcommunication,"onecannotnotcommunicate"[21].Thepowertocommunicateandthecompetencetoadaptthestyleofcommunicationineachpossiblesituationisoneofthemostremarkablehumanabilities.Butwhatcapabilitiesdoourdigitalgadgets encompass? The attribute smart seems to be restricted to these devices' underlying function. Until now, however, they are not very smart in how they offer and communicate their function. We – as human beings – own the ability to reason about the way we express ourselves. We inspect our surroundings, judge the information to be communicated, and monitor our communicative counterpart. We gather lots of information from our surroundings, which we permanently interpret to adapt our verbal and non-verbal communication. In doing so, we try to meet our communication partners' needs, our surroundings' demands, and last but not least we take into account the constraints of the information which shall be communicated.

1.1 Motivation

With respect to the evolution chain from distributed computing to mobile computing through to ubiquitous computing [19], it becomes obvious, that up to now, technical systems focus on providing a maximum of functionality to the user. Although advances of multimodal systems and multimodal interaction are manifold (see [5, 16, 18]), the aforementioned adaption of communication that is carried out by humans is still not very prominent. Research into Companion Systems is about to change this. These cognitive technical systems are continually available and attempt to adapt their behavior to the users' preferences, needs, capabilities, emotional state, and situation. In comparison to the paradigm of ubiquitous computing there are three criteria that set a companion system apart: intention-awareness, artificial intelligence planning, and adaption by learning (cf. Figure 1).

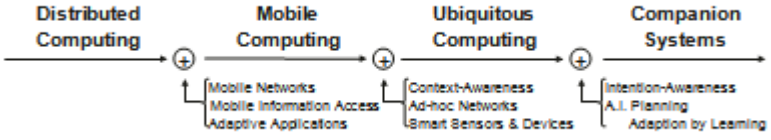

Figure 1: The evolution chain towards Companion Systems **inspired by [19]**

Within our collaborative research center we developed technical components to realize a multi-sensor companion system. A simplified architecture scheme is depicted in Figure 2. Such a system is aware of its user

and its surroundings as described in [22]. Companion systems are able to support the user by intention recognition and A.I.-based task planning. The system shall assist its user in any given situation. With the need of permanent availability, such a system has to offer a very flexible UI concept, to be able to realize a proper UI via diverse device components in any given situation. Focusing on the system's adaptive communication concept, the user's preferences, needs, capabilities, emotional state, and situation must be taken into account. All of this information can be affected by uncertainty. The process of delivering an interface to the user through the available modalities can no longer be hard wired. A flexible reasoning strategy of modality arbitration even with ambiguous input data is necessary. That is why we present a new approach of probabilistic fission.

Based on findings in neuroscience, we can adopt some concepts of neuronal information processing to the domain of modality arbitration. As stated by Paramythis and Weibelzahl, or Strnad et al. [17, 20], the fission process can be seen as a chain of afference, inference, and efference. The system's sensors and fusion layers deliver the necessary information. Each information feature is stored in and offered via a central knowledge base. This step stands for afference. Within our system different tasks from various domains are concerned with inference, e.g. knowledge processing, task planning, dialogue management, and last but not least the process of multimodal fission.

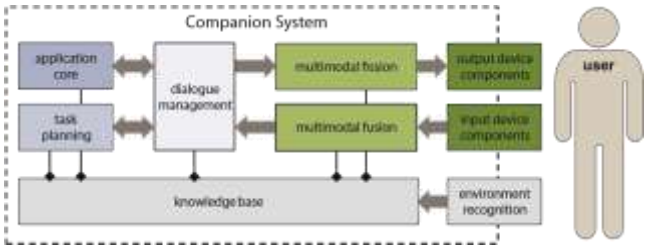

Figure 2: Architectural overview of our system in the human computer interaction loop

The fission's process of modality arbitration goes beyond the inference tasks. In terms of efference, the fission process is responsible to deliver dedicated output signals to specific system output components. Each output component, in turn, is in charge to render its attributed information. On this level again, adaption could take place. Imagine a list of selection items. It may depend on the number of items how the selection will be rendered (e.g. a group of radio buttons vs. a pull-down list or a scrollable list).

All in all, the fission shall provide adequate output via multiple modalities. But how does the fission's process look like? And what requirements have to be met when working with uncertain, fuzzy or ambiguous real world data?

1.2 Research Contributions

In the remainder we describe a meta model for interactive multimodal systems. We motivate to solve the problem of multimodal fission via a two level fission approach. The early fission stands for information partitioning and semantically combining of cross-modal information representation on an abstract level. The late fission is concerned with modality arbitration and the mapping from abstract information to concrete UI components. We present the ProFi system, which enables real time reasoning for modality arbitration with uncertain or ambiguous data in a modelbased approach. ProFi's probabilistic rule based reasoning concept can easily be evaluated and is designed to easily integrate new reasoning knowledge at runtime.

II. Multimodal Output Generation

According to Horchani et al. [12] each system in the field of HCI can be assigned to two different types of systems. On the one hand, a system can be used as an interactive tool. This type of system remains rather passive. On the other hand, a system can act as an active partner to reach a goal in a collaborative dialogue situation. Graphic tools or traditional office applications belong to the first category. An example for the second category might be a dialogue system which assists the user while booking a flight in a classical wizard structure. In the remainder we focus on the latter category.

As described by Foster [6], the main tasks in fission are concerned with (1) content selection and structuring, (2) modality selection, and (3) output coordination. But also current work, like [1, 5, 9] and [18] is concerned with the problem of determining the right device combination for output. It is still seen as a challenging task to design a system which adapts appropriately to a constantly changing interaction context. According to Costa and Duarte [3] novel systems shall not force the user to interact in a pre-defined way. Systems have to adapt their user interface to the users' needs. That is why many current approaches focus on model driven UI concepts. Starting with a basic abstract UI description, the later UI can be determined at runtime in a context-aware adaptive manner. This UI evolution from an abstract and modality independent to a concrete and final modality specific UI is described in [2], and is also known as the CAMELEON reference framework. The logical continuation of this idea concerns the refinement of information. As we will see, in the

same way as the UI, communicable information has to exist on different levels of abstraction; in a modality independent manner as well as in a modality specific form.
As requirements for contemporary context aware systems for multimodal interaction, our approach shall meet the following objectives:

- allow for a general model driven UI generation without being bound by a specific and domain-dependent scenario
- support context-aware UI adaption and evolution
- enable the integration of uncertain sensor knowledge for modality arbitration
- use an easy-to-understand reasoning methodology for modality arbitration with transparency and traceability of the results
- modular system architecture shall be the basis to easily enable future extensions or changes of the reasoning and context knowledge

2.1Architectural Meta-Model for Multimodal Interaction

Our approach focuses on adaptive multimodal interaction. The System will obtain context data from diverse sensors via different channels and shall provide its multimodal interface through multiple devices as described in [22]. As motivated in [10, 11] we use a flexible dialogue management as link to an independent functional application core. The modality independent dialogue management allows us to provide an independent interaction concept for our system. The devices, environmental and user statuses affect the final user interface at runtime which map the logical to the physical interaction and vice versa.

To meet these requirements we present a new meta model for adaptive multimodal interactive systems. Our stormy tree model (cf. Figure 3) is based on a modified Arch/Slinky meta-model [8] and the principles and architecture described in [5]. The forked branches on the right side represent different concepts of multimodal interaction, where the blue lines mark important information transitions. For the input stream these transitions are known as early fusion at the feature level and late fusion at the semantic level, as described in [16].

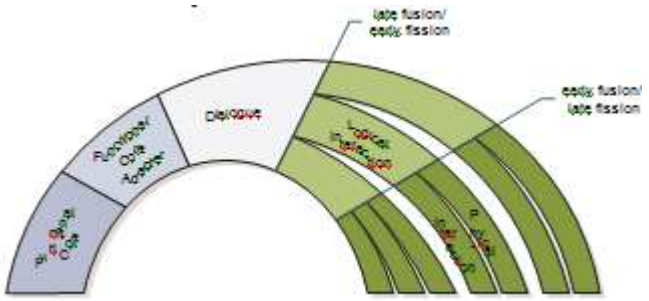

Figure 3: The stormy tree **meta model for multimodal interaction with early and late fusion and fission**

We endorse this sub-division concept and adopt it for the process of multimodal fission. By early fission we subsume tasks like information partitioning and semantically combining for cross-modal information presentation on an abstract level. These tasks are related to Foster's first task of fission: "content selection and structuring" [6] or the "semantic fission" as named by Rousseau et al. [18].

With a given set of abstract information, the next step is to determine the proper and concrete output representation for each abstract information item. We call this step modality arbitration or late fission. Within this step, the set output information items will be arranged in a spatial and temporal way. These steps are covered by the remaining two important steps of fission [6]: "modality selection", and "output coordination". Others may name this step modality election, allocation or arbitration (cf. Rousseau et al. [18]). The remainder of this article focuses on the process of late fission (modality arbitration) with uncertain or ambiguous decision knowledge.

2.2 State of the Art in Fission

According to Costa and Duarte [3], systems, which combine different output modalities like text and speech evolved since the early nineties. The allocation of output modalities of the early multimodal systems was rather hard-coded than based on intelligent algorithms.

In 2002 Foster summarized the state of the art concerning the fission task for the COMIC project [6]. She proposed the aforementioned three important tasks in fission. She further named several systems and classified these different approaches into four categories. At first, the composition approaches, where different UI primitives got ranked and assembled together to form a coherent and more complex UI. At second – and according to Foster the most popular ones – the rule based approaches. Here different pre-defined rules act as

processable design guide to identify the adequate UI in a given situation. The third category is dedicated to the plan-based approaches. These systems apply different rendering strategies based on varying preconditions of a given feature set. In comparison with the second category, one single plan-based strategy can embrace the logic of multiple rules. As a fourth category, Foster named the competing and cooperative agents. Here a hierarchy of agents strives for a UI, which meets some given requirements. The first attempt that satisfies the given requirements will be realized. Based on that Review, Foster and White described their plan-based fission approach for the COMIC project in 2005 [7]. COMIC is used to realize an intelligent bathroom designer. It can realize output via a GUI and a virtual talking head. The head can underpin the verbal utterances by facial expressions or support deictic references by gaze shifts. They use a user model along with the dialogue history to reason about an adequate output. An information ontology provides the communicable information.

In 2006 Rousseau et al. presented the ELOQUENCE platform and introduced another important aspect: the user interface's temporal evolution [18]. According to this, they motivate the WWHT-tasks to answer four questions which will occur during the presentation life-cycle of a certain information: (1) What is the information to present? (2) Which modalities should be used to present this information? (3) How to present the information using these modalities? (4) and Then, how to handle the evolution of the resulting presentation? Similar to the Stormy-Tree meta model, their architecture is based on an ARCH concept. ELOQUENCE makes use of a modality independent semantic information definition at the functional core. They use a dialogue component to assemble the information fragments and communicate them via different output modules. The early fission is pre-set by a human designer to build the semantic structure. For modality allocation they propose the use of rules, automats, or Petri networks. For realization they use a rulebased composition approach in the ELOQUENCE system.

In 2009 Dumas et al. gave an important survey on multimodal interfaces, principles, models, and frameworks [5]. Concerning the fission process, they do not impart further knowledge about other systems or fission concepts, than mentioned here before. But Dumas et al. mentioned another interesting idea: machine learning for multimodal interaction. They gave examples for machine learning in multimodal fusion, but these techniques might be appropriate for fission approaches, too.

In 2011 Costa and Duarte motivated a system which uses multimodal fission to infer an adequate user interface for elderly and differently impaired users [3]. The work was developed in the scope of the European GUIDE project. They came up with the fact, that "there is not much research done on fission of output modalities". Referred to Costa and Duarte, this is because most applications use only few different output modalities, and therefore use simple and direct output mechanisms. Their fission process is also oriented towards Rousseau's WWHT tasks. As recommended by Calvary et al. [2] their UI generation process starts with an abstract information which is mapped to a final UI element as motivated by the CAMELEON process. They plan to realize a rule-based composition approach.

Another interesting approach was presented by Hina et al. in 2011 [9]. They present a multi-agent system, where the interaction history is taken into account to reason about the new output. They recommend a machine learning approach for case based reasoning. For unknown cases the user can decide about the final UI and the system stores this decision for future tasks. They also make use of rules and priority rankings to determinate the final media devices.

Currently rule-based approaches can be seen as established practice. Recent work goes together with modeldriven UI generation as described in the CAMELEON reference framework. It does not seem that any of the aforementioned approaches use probabilistic models to enable uncertain knowledge as basis for the reasoning process.

III. Own Approach

Before going into detail, we want to give a short overview of our process from an abstract dialogue description to a final UI. In doing so we will describe our realized system along with ProFi's probabilistic rule based reasoning approach for modality arbitration. Figure 4 shows our current approach and processing pipeline. In an exemplary scenario a companion system assists a user with the home theater set-up. We start with a set of abstract information references provided by the dialogue management as modality independent dialogue output (cf. Listing 2). In an optional first step (1) the information items can be semantically and temporal rearranged in an early fission step. This decision depends on the context knowledge. In this example we proceed with an unmodified dialogue output. Late fission takes place in step (2): the probabilistic reasoning of the output modalities. We will focus on this process in the remainder of this article. Step (3) can be seen as a post-processing decorator pattern. If, for example, private information shall be communicated, but only public devices are available, the information can be obfuscated in this step. Step (4) takes place on the realizing device component. Here, the final UI-widget will be assigned for each information item. For instance, depending on the user model a male or female voice will be used. Visual selection items will be presented as buttons or as pull-down list depending on the available space and the fission's output specifications. This step can also be used to

adapt the UI to a given style guide with the use of so called beautifications before it will be rendered. The context information which we use is provided by our consortium's central knowledge base. The offered information is based on real world observations as described in [22].

3.1 Reasoning Algebra

To dissociate the current approach from static examples we use this subsection to motivate a general algebra to introduce the solution for multimodal fission with uncertain or ambiguous data.

3.1.1 Knowledge Declaration

In accordance with Dey and Abowd's definition, we understand context as follows [4]: "Context is any information that can be used to characterize the situation of an entity. An entity is a person, place, or object that is considered relevant to the interaction between a user and an application, including the user and applications themselves." That is why we belief that the one and ultimate context-model for any purpose does not exist. Instead we provide an algebra which can be used with any context model.

Let Φ be a set of possible probability values and S a set of statements that:

$\Phi := \{\varphi \in R \mid \varphi \in [0,1]\}$

$S := \{s \in (S_b \cup S_c \cup S_v)\}$,

where S_b is a statement set of Boolean values, S_c is a value set of pairwise distinct and classifiable textual statements, and S_v is the set of real valued variables in a way that:

$S_b := \{s_b \in \{true,false\}\}$

$S_c := \{s_c \in C \mid C \text{ is a set of classifiers}\}$

and

$S_v := \{s_v \in R\}$.

Let K be a knowledge item as a set of probable statements that:

$$K := \{k \in \{S \times \Phi\} \; with \sum_k \phi_k \leq 1 \qquad \forall k \in K\}$$

.

If the sum $\sum_k \phi_k$ is less than 1, the assigned knowledge item expresses incompleteness. For instance, by inference of other knowledge a user's preference for the tactile channel can be assumed to be false by a probability of 0.2, true by 0.1, and remains undetermined by 0.7. Finally we define K3):as knowledge set of different knowledge items K_i (cf.Listing

$$\mathcal{K} := \bigcup K_i$$

.

Our context knowledge encompasses the models as depicted in Figure 4. The surroundings model provides information about the current situation, like the level of noise or lighting, the demands of the environment (like silence in a library), to name a view. The user model provides information about the user's handicaps, preferences, etc. The device models provide information about each device and its input and output components. Each component provides knowledge about its supported decoder resp. encoder concepts. The information model provides the mapping from abstract to concrete information items (cf. Listing 1). An additional relations set is used to store relations like distance distributions for different users and device components. Each knowledge item from the user model, surroundings model, as well as their distance relations, can be represented by flexible value distributions to enable the modeling of uncertain, ambiguous, or unknown knowledge. By unknown knowledge we understand assignable but unspecified or missing knowledge.

3.1.2 Information Declaration

Our paradigm of adaptive multimodal communication is designed to use abstract information, and to map this abstract information to concrete information. This could be done, for instance, by representing the abstract information "book" via a concrete picture showing a book, via the concrete letters book on screen, or even via the concrete verbal utterance: "book". Therefore letformation items: I be a set of abstract in-

$I := \{i \mid i \text{ is an abstract information item}\}$

In our current realization, each abstract information item i can be mapped on up to six encoder mediums: picture, text, text for text to speech synthesis (TTS text), a grammar for Automatic text or Speech Recognition

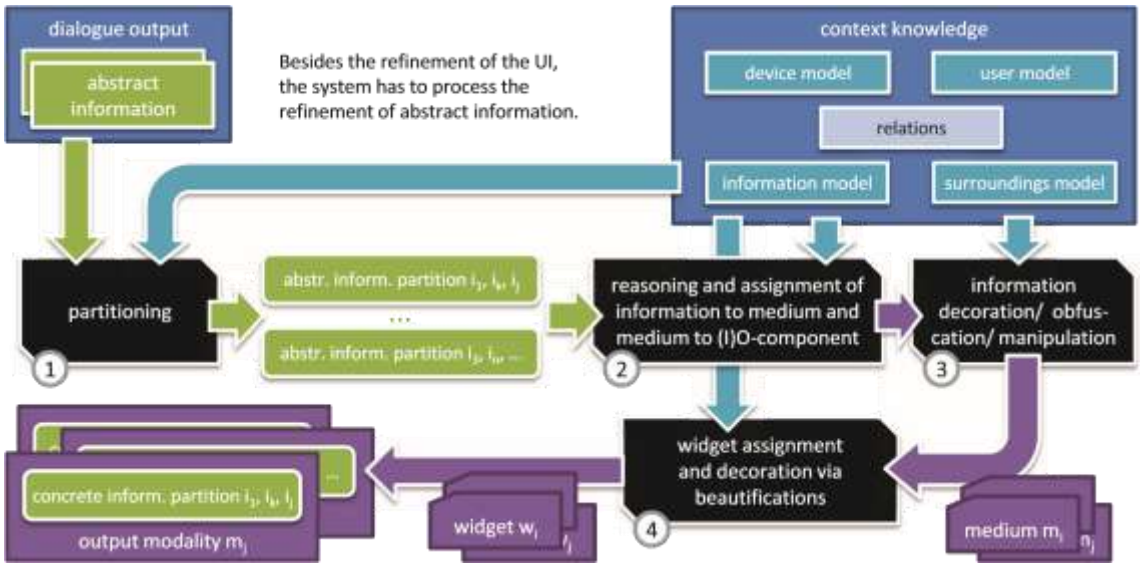

Figure 4: The fission's process to derive context aware outputs from abstract information and dialogue data.

(ASR grammar), audio representation, and video representation (cf. Listing 1). Thus, let I be a set of concrete information items with:

$I = I_1 \times ... \times I_6$

$:= \{(i_1,...,i_6) \mid i_j \in I_j \vee i_j = nil \; j = 1,...,6\}$

where

$I_P = I_1 := \{i_1 \in P \mid P \text{ is a set of pictures}\}$

$I_T = I_2 := \{i_2 \in T \mid T \text{ is a set of texts}\}$

$I_{TTS} = I_3 := \{i_3 \in TTS \mid TTS \text{ is a set of TTS texts}\}$

$I_{ASR} = I_4 := \{i_4 \in ASR \mid ASR \text{ is a set of ASR grammars}\}$

$I_A = I_5 := \{i_5 \in A \mid A \text{ is a set of audio samples}\}$ $I_V = I_6 := \{i_6 \in V \mid V \text{ is a set of videos}\}$.

Concluding Remarks.

By the term $i_j = nil$ we understand that there is no concrete information i_j within the set of I_j which can be assigned to an abstract information i.

Further characteristics of concrete information (e.g. for tactile signals) can simply be introduced by additional information classes I_j.

A possible difference between two or more concrete information fragments in terms of different information contents (e.g. a difference between picture and text) shall be ignored here.

3.2 Probabilistic Modality Arbitration

This subsection will describe and focus on reasoning and assignment (box (2), Figure 4); ProFi's probabilistic reasoning in late fission. Due to the fact that we cannot rely on nonambiguous information for modality arbitration we present a probabilistic approach to handle uncertain or ambiguous data. The presented approach consists of two main steps. First, we identify all possible output configurations and their combinations. Then we evaluate each configuration with a given rule set to identify the best output modali-

```
<informationSet>
<informationinformationID="yes_trigger">
<text>yes</text>
<ttsText>yes</ttsText>
<recognitionChoices>yes|okay|of course</recognitionChoices>
</information>
<!-- other information -->
</informationSet>
```

Listing 1: Excerpt from our set of concrete information**, as used in our system. The** abstract **information** yes_trigger **can be mapped onto text, a verbal utterance or onto recognition choices as a grammar for ASR. The concrete information can be used by different widgets to realize a** yes **trigger (e.g. button or speech dialog).**ties. This second step is designed to be computed in parallel using multi-core CPUs.

3.2.1 Exploring the Output Possibilities

To judge a potential output configuration to communicate an abstract information item we need some knowledge about the later output. We want to inspect and judge its properties and encoding type. For that reason we define an output possibility o as combination of:

- The item's later use (information item/ recognition item/ trigger item)
- A reference to a device and component model
- The abstract information's concrete encoder medium in terms of $I_1,...,I_6$ (cf. section 3.1.2) • Optional desires and dislikes specified by the designer or the user concerning the channel, a special device or component, as well as the favored encoder medium

To judge even multimodal output combinations we build the power set of all potential output possibilities with:
$P(O) := \{U \mid U \subseteq O\}$ where
$O := \{o \mid o \text{ is a specific output possibility}\}$.

So each $o \in O$ represents one concrete output representation for a given abstract information with respect to the given act type (e.g. here: trigger) including references to the knowledge about the device and component that will render the information as well as the encoder medium in terms of $I_1,...,I_6$ (cf. section 3.1.2). As an example, the abstract yes_trigger from Listing 1 may be realized as: o_1: trigger{device:PDA, component:Screen, text:yes [1] } o_2: trigger{device:PC, component:Screen, text:yes} o_3: trigger{device:PC, component:TTS, ttsText:yes[2]}
o_4: trigger{device:PC, component:ASR,okay|yea|of course|sure}
recognitionChoices:yes|

For these exemplary n = 4 output possibilities our power set will carry $2n-1 = 24-1 = 15$ items (without the empty set). The power set may be even larger if we add additional parameters (e.g. font size, volume, ...). Thus each output multimodal) output combination to realize a concrete outputpermutation item $p \in P(O)$ represents a valid (unimodal or for the given abstract information. These permutation items include each possible output combination, and each can be categorized by different glossary concepts, like the CASE[3]or CARE[4]properties [1, 5, 15]. Even if those attributes originally were used to describe fusion concepts, they can also be applied to modality constellations in fission, where this is considered meaningful.
The next step will be to evaluate each of the possible output permutation to identify the best one.

3.2.2 Rating the Output

To identify the best output solution we integrated design rules in our process to infer the proper solution in terms of an expert system. Beyond that, we can use uncertain and ambiguous knowledge within our reasoning process to meet the requirements of real word sensory systems. Our reasoning component for modality arbitration relies on a set of design rules R with:
$R := \{r \mid r \text{ is a rule for modality arbitration}\}$.

These rules are used to judge each potential output configuration which is able to communicate a given abstract information. Each rule comes with an activation signature and a pre-defined reward value. If a rule is activated by a certain knowledge, its positive or negative reward is assigned to its currently activating output configuration. In our implementation we use rewards from -100 up to 100. The assigned reward is biased by the activation knowledge's probability φK. Therefore each rule can be written as function
$r : (p,K) \rightarrow [-100 * \varphi K, 100 * \varphi K]$,
with p being a certain output permutation item, and K representing knowledge as described above.
The next step is to evaluate each permutation item p in

P(O) by each rule r and its biased reward. Each rule r can be activated by certain knowledge items K in combination with pre-defined parameters of each output permutation item. In our current implementation the utilized context knowledge can be taken from (a) the user model, (b) from the different device and component models, (c) from the surroundings model, or (d) from relations set up by different items of these models[5](cf. Figure 4). We identify the most appropriate output permutation(s) with the use of the maximum reward over all rated permutation items where:

$$p_{max} = \max_{p_j} \sum_i r_i : (p_j, \mathcal{K}) \qquad \forall p_j \in \mathcal{P}(O), \forall r_i \in \mathcal{R}$$

.

No matter if there is more than one maximum – they are all rated equal and we can choose an arbitrary one to realize the output.

[1] Later the simple word "yes" can be rendered as a button.
[2] The simple utterance "yes" will be extended to a complete sentence at runtime.
[3] Concurrent, Alternate, Synergistic, Exclusive
[4] Complementarity, Assignment, Redundancy, Equivalence
[5] The distances from each device to each user, for example, is such a relation.

3.2.3 Combining Probabilities

To calculate the bias for each rule's reward, we have to calculate the overall probability of each rule's activation knowledge. The activation knowledge can stem from different domains (e.g. user and surroundings knowledge). Imagine the rule: "If the surroundings demands silence and the user has a hearing impairment it is good to avoid aural output, but use visual output." Then this rule will only be activated if both knowledge items' statements are set. In addition, the current output permutation item has to fulfill the two requirements: no aural, but visual information encoding, in order to activate the rule. If all preconditions are met, the knowledge's overall probability has to bias the rule's reward, which will then be assigned to the current output permutation item.

For mutual independent knowledge items K_i (more precisely their probable statements k_j) with probability φ_{kj} we use the combined probability:

$$\phi_K := \prod_i \phi_{k_j} \quad \text{with } k_j \in K_i \in K,$$

where K is the rule's activation knowledge.

Concluding Remarks.

At this point we assume that each K_i from a rule's activation sequence is mutually independent from the others. This might not be true, if the knowledge items themselves have been inferred from the same data and, therefore, might be dependent [14].

3.3 Information Decoration and Obfuscation

Under some circumstances it might happen that even the best rated output permutation might have drawbacks in terms of privacy. There could even be a rule with a negative reward stating: "It is the worst case to communicate private information via public device components" (reward = −100). Imagine an abstract information which is marked with the optional private-flag, but the only solution would be to communicate this information via a component which can be perceived by the public, because there is no other device available.

Compared to suppressing the output by a simple abort we can alter the concrete information to obfuscate the private content (cf. Figure 4, step 3). This is possible because we know which part of the act's information is the private one. We also own the knowledge, that it is a non-private device component when examining our device model. The surroundings model gives us the information that there are other people in the scene.

In this case we advise to apply an obfuscation strategy with each encoder medium of the concrete information set I. For example, blurring an image or replacing original text with stars. Another option offered by this process step is the ability to add decorations to the concrete information to meet the requirements of a corporate identity style guide (e.g. a certain color filter applied to photos).

3.4 Widget Assignment

When the mapping from the abstract to the concrete information is done, and the reasoned device components are assigned to each information item, this concrete interaction output is sent to the referenced device components. The specified output has to be rendered via the named components (cf. ClientRuntime in Figure 5). It is time to realize the specified information encoding via a certain widget[6] or UI concept. In our ubiquitous system it is up to the client devices to render the received output information via their designated components.

As an example, imagine a selection task over n selectable items. The prior process might lead to an output where each selection item is designated to be encoded via text. The question is: how shall these selectable text items be realized by their assigned device component? The items could be rendered as a group of buttons, checkboxes, radio buttons, as a list and so on.

The problem can be solved with a similar algorithm like the one which solved the problem of modality arbitration. Merely the rules may be more design specific. We want to note, that both, the refinement of the information (cf. Figure 4, stage 3) as well as the refinement of the user interface (cf. Figure 4, stage 4), are part of the process of multimodal fission.

3.5 Realization

We implemented the proposed process as an ubiquitous system based to the architecture depicted in Figure 2. The respective components use a message oriented middleware to communicate their data. The fundamental software components for the presented approach are the fission module with the rule engine, the knowledge manager, the information manager, and the client runtime to render the output on different devices (see Figure 5).

[6] We use the term *widget* as metaphor for any channel: a visual widget, an aural widget, or a tactile widget.

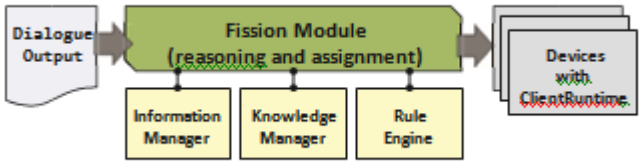

Figure 5: Part of the current implementation. The fission module receives a dialogue input and distributes the interaction output for different devices.

The whole information flow in our system is based on XML messages. The fission module receives a dialogue output with abstract output description as shown in Listing 2. In Addition, each abstract information item can be associated with one of the following act types: information, listen or trigger. The information type thereby represents system output-only information, whereas the listen type stands for active listening with no output but system input. A trigger (e.g. each information item within a selection) represents both: output and input; output information which can trigger an input. Then the process mentioned in section 3.2 inspects the dialogue output and reasons the adequate output for each abstract information item. For better layout results and to structure the interaction logic we added further containers (e.g. a selection environment) on different hierarchical layers which wrap the abstract information objects.

```
<?xmlversion="1.0"encoding="utf-8"?>
<dialogueOutputdialogueID="connection_selection">
<topic>
<abstractInformationobjectID="topic_widget"informationID="connection_selection_topic"/>
</topic>
<dialogueAct>
<desiredOutputChannel>aural</desiredOutputChannel>
<desiredOutputChannel>visual</desiredOutputChannel>
<desiredInputDecoderMedium>speech</desiredInputDecoderMedium><desiredInputDecoderMedium>touch</
desiredInputDecoderMedium>
<selectionobjectID="selection_container"informationID="connection_selection_information">
<abstractInformationobjectID="item_1"informationID="bluray_and_amplifier"/>
<abstractInformationobjectID="item_2"informationID="receiver_and_amplifier"/>
<abstractInformationobjectID="item_3"informationID="tv_and_amplifier"/>
</selection></dialogueAct>
<listen>
<abstractInformationobjectID="selection_assistance"informationID="help"/>
</listen>
</dialogueOutput>
```

Listing 2: An exemplary dialogue output defining a topic as well as a selection as main part of the dialogue act. In this case, the dialogue manager added additional desires which shall be considered by the fission process.

During the process of modality arbitration the fission module in combination with the information manager inspects the current known information set (cf. Listing 1) as well as the device model to build up the set of valid output permutation items. Then each possible output permutation gets evaluated by each rule. Whether or not a rule's reward is assigned depends on the particular output permutation item and the current context knowledge (cf. Listing 3). The later one is provided by the knowledge manager. Due to their logical behavior the rules are implemented as classes and are stored in an auxiliary library. Each rule is based on a common marker interface and gets integrated by the rule engine at runtime by reflection. This principle offers the potential to integrate new rules or replace existing ones without the need to stop and rebuild the whole system.

For a better validation each output permutation gets annotated with its activated rules and the context dependent percentage reward bias. For the selection item tv_and_amplifier (see Listing 2 and Figure 6) the fission's evaluation output is shown in Listing 4. Application developers can

```
<personid="p5"isUser="true">
<!-- other knowledge -->
<handicapHearing>
<pBoolvalue="true"probability="0.087" />
<pBoolvalue="false"probability="0.913" />
```

```
</handicapHearing>
<handicapSpeaking>
<pBoolvalue="false"probability="1" />
</handicapSpeaking>
<preferenceChannelAural>
<pBoolvalue="true"probability="0.1119" />
<pBoolvalue="false"probability="0.8881" />
</preferenceChannelAural>
<!-- further knowledge -->
</person>
```

Listing 3: Excerpt from a person knowledge set Kp**, as used in our system, with three exemplary Boolean** knowledge items **(hearing and speaking handicap as well as the preference for the aural channel). Each item is represented by the probability distribution of its** statements**.**

use these information to identify the need of further rules in cases where the results are not satisfied with the inferred output.

At the end of the reasoning process, when the best rated output configuration for each information item is determined, the output specification is passed to the responsible devices. The devices' client runtime is in charge to render the output via the specified device components.

Figure 6: The visual output representation of an abstract dialogue output as defined in Listing 2. The abstract information of each selection item is realized in a redundant multimodal way (picture plus text).

As an example, we process the given dialogue output from Listing 2. A set of specific context parameters will result in a multimodal user interface as shown in Figure 6. Besides that, as desired (cf. the desires in Listing 2), an additional speech interface is realized in parallel to offer the selection via speech. But with the given context knowledge not all desires have been satisfied. As shown in Listing 4, the rule Rule_a_doc matched, but contributes its reward by only 50 % for the listed, and best rated output permutation. It cannot contribute its full reward, because of the desired (cf. Listing 2), but not supported aural output channel. In this case, the absence of any aural output is influenced by the user's dislike of the aural channel (cf. Listing 3). That is also the reason why the output permutation with the additional TTS output was rated worse than the listed"winning one" without the audio output.

The final permutation for this information item is depicted in Figure 7. ProFi's inspection tool offers such a representation for each communicable information item. Thereby the bright blocks represent possible but unused devices, components, or encoding concepts (here: verbal utterances via TTS). The corresponding evaluation output for this selection item item 3 is shown in Listing 4. The corresponding final UI is depicted in figure 6.

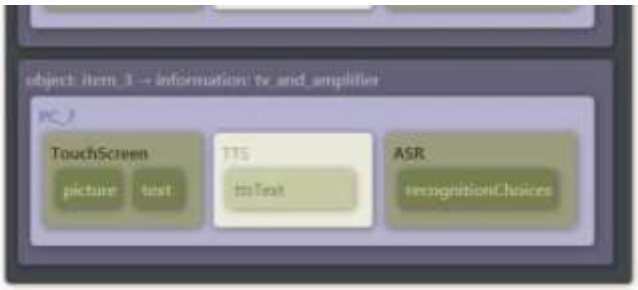

Figure 7: Excerpt of the fission's reasoning inspection tool. The clipped screen shot represents the final rendering allocation for the exemplary selection item with ObjectID item 3 **(cf. the given dialogue output from Listing 2). The output will be realized via the device** PC 7**. The used device components are** TouchScreen **and** ASR**.**

As we can see in Listing 4, the probabilistic reasoning approach offers the ability to handle conflict of objectives while reasoning. When inspecting the matched rules, we see the the two rules Rule_u_10_i_n and Rule_u_10_i_p. The first one is associated with a positive reward. The second one influenced the overall rating with a negative reward. Both rules common activation knowledge item is the Boolean value distribution which describes the preference for the aural channel (cf. Listing 3). This is one further benefit, which cannot be achieved by conventional non-probabilistic approaches, since the common Boolean flag would be either true or false, but nothing in between. In combination with the different rules, the probabilistic approach led to a compromise output which respects the designer's desires as well as the preferences and demands of the user and the environment.

With different context parameters the user interface for the same dialogue output might be realized in a diverse way. Due to our fully model driven approach it can be realized, for example, as a solely unimodal speech dialogue or in a multimodal way, even on multiple devices.

3.6 Complexity and Scalability

We can solve the problem of modality arbitration, which is a problem of exponential complexity. We have to handle this problem in a domain where time is a crucial factor. According to [23] a system's feedback time shall be about 200 ms, and should be communicated in at least 500 ms

The selection item "item_3" will be communicated with the following OutputPermutation:
rating: 222,30794124 (best of 15 possible output permutations)
PermutationItem[deviceID=PC_7, ComponentID=TouchScreen, ObjectID=item_3, InformationID=tv_and_amplifier, InformationType=picture]
PermutationItem[deviceID=PC_7, ComponentID=TouchScreen, ObjectID=item_3, InformationID=tv_and_amplifier, InformationType=text] PermutationItem[deviceID=PC_7, ComponentID=ASR, ObjectID=item_3, InformationID=tv_and_amplifier, InformationType=recognitionChoices]
100,00 %: [Rule_a_didm] It is good that the permutation supports the desired input decoder mediums.
86,64 %: [Rule_su_1] The surrounding indicates a high noise level, so it is good to provide visual output.
50,00 %: [Rule_a_doc] It is good that the permutation supports the desired output channels.
100,00 %: [Rule_s_1] Better offer multimodal input possibilities (tactile and aural) for selection items.
100,00 %: [Rule_s_II] Better use multiple redundant outputs (text + picture) for a desired visual selection offer.
8,70 %: [Rule_u_1_a] The user has a hearing impairment, so better provide visual output. 8,70 %: [Rule_u_11_p] If the user has a hearing handicap, it is good to add text to pictures if there is no supporting TTS output.
92,40 %: [Rule_u_9_o_p] The user prefers the visual channel. It is good to offer the information via this one.
11,19 %: [Rule_u_10_i_p] The user prefers the aural channel. It is good to offer this input channel.
13,95 %: [Rule_u_14_ach_f_p] It is good to avoid aural output only, if the user is not able to listen right now (availability).
7,60 %: [Rule_u_9_o_n] The user does not prefer the visual channel. It is bad to use this output channel.
88,81 %: [Rule_u_10_i_n] The user does not prefer the aural channel. It is bad to use this input channel.
5,11 %: [Rule_u_15_acs_f_n] It is a worse case to rely on visual output only, if the user is not able to watch right now (availability).

Listing 4: The fission's evaluation output for the exemplary selection item item_3 **(cf. Listing 2), representing the tv and amplifier connection. The best rated output permutation includes three of the four possible permutation items (cf. Figure 7). The finally realized UI element is the lower button in Figure 6.**

[13]. To handle the complexity we use algorithm concepts like branch and bound to evaluate the rules with a positive reward before the ones with a negative reward. We use the negative ones, only if necessary. We also apply case based reasoning, since without any change in the context models, similar information items will always result in the same configuration of output permutations (e.g. each of our three exemplary selection items).

Due to the high dynamic complexity (number and complexity of rules, different parameter sets, number of available device components) of each output task, it is hard to evaluate the process and name an exact duration. We will now explore the exponential complexity of $O(2^n)$, when evaluating each possible output permutation for our exemplary selection item item 3 (cf. Table 1).

number of devices	**number of components**	**possible output permutations for item_3**	**avg. reasoning time for complete dialogue output**
1	3	15	47 ms
2	5	127	91 ms
3	8	2047	298 ms

Table 1: Comparison of the increasing reasoning time (for the complete dialogue output) and the exponential increase of possible output permutations (for selection item item 3**) by increasing the number of available output components.**

We did several runs to realize the exemplary dialogue output from Listing 2. All runs were computed on a CPU with 2.8 GHz on a PC with 4 GB RAM. We used diverse randomized user and surroundings models with a total number of 18 probabilistic knowledge items. For one device, with three components for GUI, TTS, and ASR, the exemplary item 3 can be realized in 15 different ways. The fission needs an averaged time of 47 ms to reason the whole dialogue, with all its six items. By adding a further device with an additional GUI and TTS component the reasoning time increases by 44 ms. Providing a further device with GUI, TTS, and ASR will offer 2047 different ways to realize item 3. The reasoning will take an averaged duration of 298 ms. This might represent a scenario with a tablet device, smart phone and desktop computer.

IV. Conclusions

The task of multimodal fission is a complex problem. Besides the task of UI refinement, the fission process also has to refine and combine each information item which shall be communicated.

Based on the new Stormy Tree meta model for multimodal interactive systems we described the process of probabilistic fission for adaptive multimodal interaction along with our current implementation. Our model driven realization extends the often theoretical approaches of the state of the art and introduces probabilistic reasoning to integrate uncertain and ambiguous knowledge. With respect to the reasoning time, the presented approach is qualified to be used in interactive systems with real-time requirements.

The quality of the realized output is hard to evaluate, because each assessor evaluates the UI with its own and subjective conceivability. Nevertheless we did several evaluation runs with different subjects in diverse context situations and identified 77 rules which lead to very pleasing results with broad acceptance. ProFi provides insights into the reasoning process by providing information as shown in Figure 7 and Listing 4 to get hints for reasoning optimization by modifying the rule set. In doing so we thereby met all of our objectives, stated in section 2.

In the near future, we will address the research topics of temporal reasoning and cross-modal output creation. We plan to solve these problems by broaden our current concept of early fission. Further work will be done to integrate supervised learning with our rule engine. We plan to offer the user the ability to intervene, and suggest a better form of output constellation. We are going to inspect the user's interaction history to infer additional knowledge to further improve ProFi's reasoning results.

As written earlier the fission process can be seen as a chain of afference, inference, and efference. If we refer to the neuroscience we can use the so-called efference copy as internal feedback and can pass it to the fusion process. Fusion concepts can use this knowledge about the output to resolve pointing references, as the system understands its own output. The efference copy can also be used to track the reasoning history and to support the supervised learning approach.

References

[1]. M. Blumendorf, D. Roscher, and S. Albayrak. Dynamic user interface distribution for flexible multimodal interaction. In International Conference on Multimodal Interfaces and the Workshop on Machine Learning for Multimodal Interaction, ICMI-MLMI '10, pages 20:1–20:8, New York, NY, USA, November 2010. ACM.

[2]. G. Calvary, J. Coutaz, D. Thevenin, Q. Limbourg, L. Bouillon, and J. Vanderdonckt. A unifying reference framework for multi-target user interfaces. Interacting with Computers, 15(3):289–308, 2003.

[3]. D. Costa and C. Duarte. Adapting multimodal fission to user's abilities. In Proceedings of the 6th international conference on Universal access in human-computer interaction: design for all and eInclusion - Volume Part I, UAHCI'11, pages 347–356, Berlin, Heidelberg, 2011. Springer-Verlag.
[4]. A. K. Dey and G. D. Abowd. Towards a better understanding of context and context-awareness. In In HUC'99: Proceedings of the 1st international symposium on Handheld and Ubiquitous Computing, pages 304–307. Springer-Verlag, 1999.
[5]. B. Dumas, D. Lalanne, and S. Oviatt. Multimodal Interfaces: A Survey of Principles, Models and
[6]. Frameworks. In D. Lalanne and J. Kohlas, editors,
[7]. Human Machine Interaction – Research Results of the MMI Program, volume 5440/2009 of Lecture Notes in Computer Science, chapter 1, pages 3–26.
[8]. Springer-Verlag, Berlin, Heidelberg, March 2009.
[9]. M. E. Foster. State of the art review: Multimodal fission. Public Deliverable 6.1, University of
[10]. Edinburgh, September 2002. COMIC Project.
[11]. M. E. Foster and M. White. Assessing the impact of adaptive generation in the comic multimodal dialogue system. In In Proceedings of the IJCAI 2005 Workshop on Knowledge and Reasoning in Practical Dialogue Systems, 2005.
[12]. C. Gram and G. Cockton. Design principles for interactive software. Chapman & Hall, Ltd., London, UK, 1997.
[13]. M. D. Hina, C. Tadj, A. Ramdane-Cherif, and N. Levy. A Multi-Agent based Multimodal System Adaptive to the User's Interaction Context. In Multi-Agent Systems – Modeling, Interactions, Simulations and Case Studies, chapter 2, pages 29–56. InTech, 2011.
[14]. F. Honold, M. Poguntke, F. Schussel, and M. Weber.¨ Adaptive Dialogue Management and UIDL-based Interactive Applications. In A. Coyette, D. Faure, J. Gonzalez, and J. Vanderdonckt, editors, Proceedings of the International Workshop on Software Support for User Interface Description Language (UIDL 2011), page 4, Paris, September 2011. Thales Research and Technology France. Interact 2011, Workshop.
[15]. F. Honold, F. Schussel, M. Weber, G. Bertrand,¨ F. Nothdurft, and W. Minker. Ein goal-basierter ansatz fur adaptive multimodale systeme. In M. Eibl,¨ editor, Mensch & Computer 2011: uberMEDIEN¨ |UBERmorgen¨ , pages 357–360, Munchen,¨ 2011. Oldenbourg Verlag.
[16]. M. Horchani, L. Nigay, and F. Panaget. A platform for output dialogic strategies in natural multimodal dialogue systems. In IUI '07: Proceedings of the 12th international conference on Intelligent user interfaces, pages 206–215, New York, NY, USA, 2007. ACM.
[17]. C. Kohrs, N. Angenstein, H. Scheich, and A. Brechmann. Human striatum is differentially activated by delayed, omitted, and immediate registering feedback. Frontiers in Human Neuroscience, 6(00243), 2012.
[18]. D. Koller and N. Friedman. Probabilistic graphical models: Principles and techniques. The MIT Press, 2009.
[19]. L. Nigay and J. Coutaz. Multifeature systems: The care properties and their impact on software design. In Multimedia Interfaces: Research and Applications, chapter 9. AAAI Press, 1995.
[20]. S. Oviatt. Multimodal Interfaces. In A. Sears and J. A. Jacko, editors, The Human-Computer Interaction Handbook, pages 413–432. CRC Press, 2007.
[21]. R. Paramythis and S. Weibelzahl. A Decomposition Model for the Layered Evaluation of Interactive
[22]. Adaptive Systems. In Proceedings of the 10th International Conference on User Modeling, LNAI 3538, pages 438–442. Springer, 2005.
[23]. C. Rousseau, Y. Bellik, F. Vernier, and D. Bazalgette. A framework for the intelligent multimodal presentation of information. Signal Process., 86(12):3696–3713, 2006.
[24]. T. Strang and C. Linnhoff-Popien. A Context
[25]. Modeling Survey. In Workshop on Advanced Context
[26]. Modelling, Reasoning and Management, UbiComp 2004 – The Sixth Int. Conf. on Ubiquitous Computing, Nottingham/England, page 8, 2004.
[27]. O. Strnad, A. Felic, and A. Schmidt. Context Management for Self-adaptive User Interfaces in the Project MyUI, pages 261–272. VDE, Springer, 2012.
[28]. P. Watzlawick, J. B. Bavelas, and D. D. Jackson. Pragmatics of Human Communication: A Study of Interactional Patterns, Pathologies and Paradoxes. Norton, New York, 1967. Some Tentative Axioms of Communication.
[29]. A. Wendemuth and S. Biundo. A companion technology for cognitive technical systems. In A. Esposito, A. Vinciarelli, R. Hoffman, and V. C. Muller, editors,¨Proceedings of the EUCogII-SSPNET-COST2102 International Conference (2011), LNCS Proceedings on Cognitive Behavioural Systems, Dresden, 2012. Springer LNCS [in press or to appear 2012].
[30]. S. Wolff, C. Kohrs, H. Scheich, and A. Brechmann. Temporal contingency and prosodic modulation of feedback in human-computer interaction: Effects on brain activation and performance in cognitive tasks. In Informatik 2011, GI-Jahrestagung 2011, Berlin, volume 192 of LNI, page 238, 2011.

Nibedita Sahoo "ProFi's Process for Multi-modal Fission System" International Journal of Engineering Science Invention (IJESI), Vol. 05, No. 11, 2016, PP 126-138.