Representing Communicative Functions in SAIBA with a Unified Function Markup Language

Angelo Cafaro[1], Hannes Högni Vilhjálmsson[2], Timothy Bickmore[3], Dirk Heylen[4], and Catherine Pelachaud[1]

[1] CNRS-LTCI, Telecom ParisTech, France
{angelo.cafaro,catherine.pelachaud}@telecom-paristech.fr
[2] Center for Analysis and Design of Intelligent Agents, Reykjavik University, Iceland
hannes@ru.is
[3] College of Computer and Information Science, Northeastern University, USA
bickmore@ccs.neu.edu
[4] Human Media Interaction, University of Twente, The Netherlands
d.k.j.heylen@utwente.nl

Abstract. The SAIBA framework proposes two interface languages to represent separately an intelligent agent's communicative functions (or intents) and the multimodal behavior determining how the functions are accomplished with a particular multimodal realization. For the functional level, the Function Markup Language (FML) has been proposed. In this paper we summarize the current status of FML as discussed by the SAIBA community, we underline the major issues that need to be addressed to obtain a unified FML specification, we suggest further issues that we identified and we propose a new unified FML specification that addresses many of these issues.

Keywords: function markup language, communicative function, multimodal communication, embodied conversational agents.

1 Introduction

Over the past decade many Embodied Conversational Agent (ECA) systems [1,2,3,4,5,6] adopted an abstraction approach to multimodal behavior generation that separates communicative function or intent from its behavioral realization. This design and the need for sharing working components motivated the SAIBA framework [7]. SAIBA supports this separation through two interface languages named **Function Markup Language** (FML) [8] and **Behavior Markup Language** (BML) [7,9] respectively. A first version of BML has been adopted internationally [7,9], but a common unified FML specification has not yet emerged, although several specialized contributions exist [6,10,11]. There is an ongoing discussion about several issues that FML needs to address [8]. In this paper we summarize the status of FML based on discussions within the SAIBA community, we underline the major issues that need to be addressed to obtain a unified FML specification and suggest further issues that need to be tackled. Finally we propose a new unified FML specification that addresses many of these issues.

[1] Author conducted this research at CADIA, Reykjavik University.

T. Bickmore et al. (Eds.): IVA 2014, LNAI 8637, pp. 81–94, 2014.

2 Current Status of the FML Discussion

The first attempts to define the FML standard were based on various existing ECA systems. This included the REA system [1] and the systems that followed it: BEAT [3] and Spark [4]. Also the Multimodal Utterance Representation Markup Language (MURML) [12] used in the MAX system and the FML-APML mark-up language [11] developed for the Greta framework [2] provided inspiration. Other important systems that featured in the various discussions on FML were The Tactical Language and Culture Training System (TLCTS) [10] and the Virtual Human Toolkit [5] developed at the Institute of Creative Technologies (ICT) which uses FML-like concepts in the *Non-Verbal Behavior Generator* module. These systems attempted to adopt a clear separation between communicative function representation and corresponding behavior that would accomplish those functions. However, these systems focused on domain specific issues such as representation of emotions [11], cultural and other contextual information [6,10] and subsets of communicative functions [5]. In this paper we propose a new unified FML specification that builds on these earlier contributions and the current status of the FML discussion.

There has been an ongoing discussion about FML through a series of targeted workshops[1]. The work in [8] summarizes the discussions to date, and outlines the most important components of FML, including the following.

Contextual information and **person characteristics**. The former includes cultural and social setting, environmental information (e.g. time of the day), history of interactions and topics discussed. The latter, referring to a participant performing communicative functions, are organized in two main dimensions: person information (e.g. identifier, name, gender, role) and personality.

Communicative actions include dialogue acts, grounding actions and turn taking. Formal (logical) languages have been proposed to represent **propositional content** and a certain organization of propositions has emerged at both *sentence* level (emphasis, given/new information, theme/rheme) and *discourse* level (topics and rhetorical relations between different parts of the discourse). It is assumed that extra-linguistic or certain non-linguistic actions, such as picking up a glass of water, can also perform certain communicative functions.

Emotional and **mental states** are believed to contribute to the motivation of a communicative intent. Emotions are divided between *felt*, *faked* and *leaked*. Mental states are defined as cognitive processes such as *planning*, *thinking* or *remembering*.

Social psychological aspects and **relational goals** are also considered. The concepts of *interpersonal framing* (e.g. showing empathy in comforting interactions) and *relational stance* (e.g. warmth) functions are introduced to affect the behavior produced by an agent with those goals.

3 Outline of Important FML Issues

Defining and Separating Contextual Information. Some contextual information may be necessary, but how much and how should it be represented? Do we need a new

[1] At Reykjavik in 2005, AAMAS 2008, AAMAS 2009, ICT and Paris in 2010.

language to represent this information (e.g. CML as proposed in [10])? In addition to the two dimensions of person characteristics proposed earlier (*person information* and *personality*), what do we mean by enough context? The contextual parameters have been shown to be important for the generation of behavior, for example, with respect to the *environmental context* (greetings depending on the time of the day), *cultural background* or *socio-relational* goals. It is also important to consider how context could affect the planning of functions.

Defining and Classifying Functions. A communication function might arise from an action that does not have propositional content, these functions have been classified more generally as *communicative actions*. The main concern is what to consider as a communicative action. Choosing a classification scheme that embraces all prevailing perspectives on communicative function is not easy, but it will aid the designers of ECAs to use FML at different levels. At a higher level it will be possible to obtain a general outline of the human communicative capacity of a system by noting what general kinds of function specification are available. At a lower level, a designer can expect that functions belonging to the same category will share some specification characteristics and parametrization [13]. The main question that arises is how many groups and categories of functions are needed.

Characterizing and Separating Conscious vs. Unconscious Intents. Contextual information does not represent the only determinant for generating communicative functions. A broader distinction needs to be made between consciously planned intents and communicative functions resulting from unconscious determinants such as mental and emotional states. We may also want to support a direct path from perception to the realization of behavior as in FML-APML [11], but this raises the issue of possible conflict between unconscious/reactive intents and conscious planned intents.

Defining Temporal Constraints and a Prioritization Scheme. Assigning timing information to communicative functions and supporting the temporal coordination among them are important issues to consider as suggested by [11]. [13] suggested that temporal constraints at the functional level of description should be much more coarse-grained than those at lower levels since it becomes hard to specify exactly how long it will take to accomplish a specific function. In addition to supporting the specification of temporal constraints, an advantage of cutting FML in *"smaller chunks"* [14] or *"chunk plans"* [13] is that they could be processed separately allowing faster generation of corresponding BML compared to a larger FML input (i.e. containing several communicative functions). When dealing with real time reactions (e.g. back channel feedback) a large amount of communicative functions processed as a whole could create an unacceptable delay that would slow down the system's response and make the whole interaction feel unnatural from the user's point of view. However, assuming that FML chunks are adopted, several issues need to be resolved. First we have to come up with a precise definition of an FML chunk, keeping in mind what constitutes a useful semantic unit. Secondly, what timing primitives will suffice to temporally coordinate chunks? Finally, what kind of conflict resolution scheme is needed when chunks collide? For example, FML-APML proposed an *"importance"* attribute to help with this.

Defining an FML Representation Structure. Previous representation languages mainly adopted an XML-like syntax and assigned a nested structure to the specified set of tags (cf. [11,10]). While it is tempting to adopt a similar structure for a unified FML representation, one may wonder whether this is still a valid solution, and if so, what are the rules that govern the embedding of a set of tags into others. At the current stage of the work, the discussion has been kept on a theoretical level, but this is an important issue when it comes to practically defining a structure for an FML representation.

Single or Multiple Agents? An instance of FML could refer to a single or multiple ECAs. Dealing with individual ECAs might offer scalability and improved performance through distributed processing while mixing ECAs may require central processing, which does not scale well. But the latter makes it easier to solve complex highly coordinated interactions.

Multiple Interaction Floors and Roles of Participants. A person may be engaged in more than one conversation at the same time and assume different roles, including bystander. This moves from dyadic settings towards more complex scenarios. Specifying the configuration of interactions at the functional level ensures that behaviors can take this into account. Some examples of configurations might be simple *1-to-1* interactions (for example dyadic), *1-to-many* (for example when describing a public speech) and *many-to-many* (two groups interacting as a whole with each other).

4 A Unified FML Specification

This section summarizes the complete proposed specification [2]. First some key terms:

Participant An entity (e.g. virtual agent or user) participating in an interaction and carrying out or being affected by communicative functions.

Floor A participant can be engaged in several interactions with other participants that we name *floors*. A metaphor for the social contract that binds participants together in the common purpose of interacting.

FML chunk The smallest unit of FML functions associated with a single participant that is ready to be turned into supporting BML-specified behavior.

4.1 Overview

Representation Structure and Target. A single FML representation instance includes functions that several participants want to accomplish. It is divided in two main sections: a *declaration* (described in Section 4.2) and a *body* (Section 4.3), as illustrated in Figure 1. The declarations incorporate contextual information, whereas the body includes all participants' generated functions grouped in FML chunks and belonging to

[2] See `http://secom.ru.is/fml/` for full specification.

three different tracks (named interactional, performative and mental state) as a result of our functions categorization described below.

Contextual Information and Multiple Interaction Floors. We divided contextual information in two components: a *static* component describing participants information (e.g. gender, age, personality, etc…) and a *dynamic* component providing information about the active floors (e.g. participants in each floor and their attitudes). *Participant information* is labeled as *static* since it is meant to endure over time. It affects all active floors in which the participant is involved. This specification supports the co-existence of multiple active floors for each participant and each floor involving one or more participants. The *floors information* supports the specification of the active floors that the FML instance describes. It is labeled *dynamic* since the information included is meant to be temporarily associated with a particular floor.

Functions Categorization and Body Tracks. The body of an FML representation is divided into three sub-sections or *"tracks"*. This design reflects the choice of categorizing the communicative functions as suggested by [15]. The first category of functions (named **interactional**) deals with establishing, maintaining and closing the communication channel, instantiated with a floor, between participants. The second category (named **performative**) covers the actual content that gets exchanged across the communication channel. The third category deals with functions describing mental states and emotions (for simplicity it has been named **mental state**).

Temporal Constraints and FML Chunks. Splitting the body up into separate tracks requires an overall orchestration of the functions in relation to each other. The order of appearance of functions in the FML instance does not necessarily imply delivery time. Coarse-grained temporal constraints (described in Section 4.3) allow synchronization and relative timing among chunks across all the tracks.

Unconscious Intents. The mental state track assumes a particular meaning that addresses the issue of representing functions that are not deliberately planned by a participant. Every participant has a *ground state* that comprises his mental and emotional states (mood could be considered as well). Only functions in the mental state track can change the participant's ground state for a limited or unlimited time depending on the particular temporal constraint adopted. In essence, the ground state provides additional contextual information about the participant that can affect the generation and realization of multimodal behavior in the later stages of the SAIBA generation process.

4.2 FML Representation: Declaration

The declaration section stores contextual information in two separate sub-sections for participant's information and floors configurations as shown in Figure 2.

The identikits tag contains an `<identikit>` for each participant including person characteristics (e.g. a human readable *name* and *gender*). Each tag supports the inclusion of embedded information about the participant. We provide `<personality>` and `<relationship>` as examples. The former is based on the Big 5 [16] model dimensions (other models can be supported). The latter specifies a relationship level with other participants. The example scenario described in Section 4.4 shows the use of

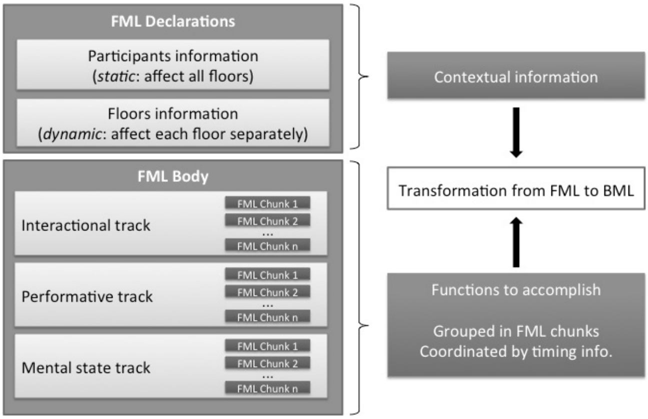

Fig. 1. An overview of our proposed FML specification. A representation instance is divided in declarations and body sections, respectively, for contextual information and communicative functions. Contextual information can affect all communication floors (participant information) or selected ones (floors information). The communicative functions are temporally coordinated in FML chunks.

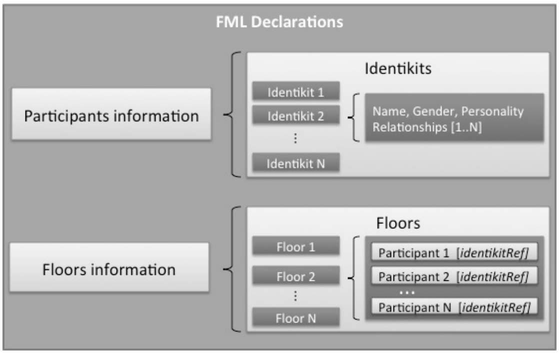

Fig. 2. The declarations section of an FML instance stores contextual information divided in participants information (identikits) and floors configurations (floors).

these tags. This part of the declaration section can be created once and then cached for later usage since it contains static information.

The floors tag describes each active floor in the FML instance. Each floor described has a *floor-cfg* attribute that specifies its configuration. We identified four possible configurations: *individual*, *unicast*, *broadcast* and *multicast* (naming inpsired by network protocols). An *individual* configuration describes a single entity (i.e. participant), *unicast* represents the classical dyadic interaction, *broadcast* describes an individual entity interacting with a group and *multicast* characterizes two groups, as a whole, interacting with each other.

A **<floor>** can include one or more **<participant>** tags depending on the number of participants involved. These tags have an *entity* attribute describing whether in the given floor configuration the participant is an *individual* or a *group*. A *role* attribute specifies the role assumed by the participant in the given floor, currently inpsired by Goffman's participation framework [17]. According to Goffman, participants can have a **speaker** or a **hearer** role. Two types of hearers are identified in this framework: *ratified* (official) and *unratified* (unofficial) participants. Ratified participants are subdivided into *addressed* and *unaddressed* recipients, and unratified participants or bystanders are subdivided into *eavesdroppers* and *overhearers*, based on their intent and degree of interest.

Furthermore, a **<participant>** tag can embed some contextual information. As an example, we defined **<attitude>** tags to specify the attitude that the participant has toward another participant in any given floor according to Argyle's status and affiliation model [18] (see the full example in Section 4.4).

4.3 FML Representation: Body

The body of an FML instance is divided in three tracks. Each track includes **FML Chunks** that are timed with relative **Temporal Constraints**. **FML Chunk tag:** Each **<fml-chunk>** refers to a single participant's identikit with the *participantRef* attribute. The first element within a chunk can be a single occurrence of a **<timing>** tag followed by any number of functions defined for the track in which the chunk appears. The **<timing>** tag temporally constrains the whole chunk relative to other chunks.

Temporal Constraints: Temporal constraints work on a chunk level with the following design principles: (1) the chunks' order of appearance in the body of an FML instance is not meaningful, (2) unless specified by the **<timing>** element an FML chunk should be scheduled for later processing (i.e. transformation to BML) "as soon as possible" and (3) the order of appearance of functions within a chunk is not meaningful and they will be considered in arbitrary order at later stages. The **<timing>** tag has a *primitive* attribute to specify the temporal relationship between the current chunk and the referenced one. Possible values are: *immediately*, *must_end_before*, *execute_anytime_during*, *start_immediately_after*, *start_sometime_after*, *start_together*.

FML Functions Specification: All FML functions tags have in common a unique identifier and a *floorID* attribute for referencing the floor in which the communicative function is meant to be accomplished.

Interactional track functions. This track supports the specification of a category of communicative functions that serve to coordinate a multimodal interaction. Table 1 shows the possible functions that can appear within an FML chunk in this track. The first column on the left side represents a broad category of interactional functions and it is also the name adopted for the corresponding tag. These tags have a common attribute named *type* that narrows down the specification of functions within the category. Some of the functions (marked with a "*") require the specification of the *addressee* attribute indicating the participant to which the function is addressed.

The initiation and closing categories describe the communicative functions, respectively, to manage the initial and termination phases of the interaction. In particular, the different available types of initiation and closing functions are based on the stages of a greeting encounter as suggested by Kendon's greeting model [19]. As starting point, the turn-taking, speech-act and grounding functions have type attribute values

Table 1. Interactional functions: suggested tag names on the left and possible type attribute values on the right. Functions marked with "*" have an *addressee* attribute.

Function Category	Type Attribute
initiation*	*react, recognize, salute-distant, salute-close, initiate*
closing*	*break-away, farewell*
turn-taking*	*take, give, keep, request, accept*
speech-act	*inform, ask, request*
grounding	*request-ack, ack, repair, cancel*

following the suggestions in [15]. All tags in this track can be linked to others in another track (e.g. speech acts linked with a performative tag) by using the temporal constraints.

Performative track functions. The various functions in this category can be divided across different organizational levels, from the largest organizational structure of a discourse down to the specification of each proposition. In our proposal, the performative track acts as place holder for further embedded extensions of FML specifically targeted to describe performative functions. Therefore, chunks in this track can host one or more `<performative-extension>` tags.

This tag is merely a stub and the description of an extension that will handle its contents is out of the scope of this paper, though we foresee an extension mechanism similar to BML's[3]. Following the recommendations in [15], Table 2 suggests a set of possible function categories and their specific types that could be included within this tag. Similarly to interactional functions, an *addressee* attribute specifies to which participant the included performative act is directed to.

Table 2. Performative functions: suggested tag names and type values

Function Category	Type Attribute
discourse-structure	*topic, segment, ...*
rhetorical-structure	*elaborate, summarize, clarify, contrast, emphasize, ...*
information-structure	*rheme, theme, given, new, ...*
proposition	*any formal notation (e.g. "own(A,B)")*

Mental State Track Functions. Functions in this track are the only ones capable of changing the *ground state* of a participant. The concept of ground state is kept at abstract level in this proposal, but the idea is that it may affect the manner in which other functions get realized, thus modeling the unconscious side of a participant. We do not

[3] `http://www.mindmakers.org/projects/bml-1-0/wiki#Extensions`

specify how the ground state should be modeled and how it should affect the behavior generation and realization in later stages. However, we provide several design ideas as a starting point, we propose functions describing mental states and emotions (as shown in Table 3).

First, multiple functions can occur simultaneously in this track. We propose a *weight-Factor* attribute ranging from 0 to 1 to establish the impact that each single one has on the ground state. Secondly, we propose that every function appearing in this track gets sustained by default and unless specified with a temporal constraint (e.g. *must_end_before*), it changes the ground state permanently. However, reverting to a previous state or voiding the effect of a sustained emotion will be possible by specifying the same function again with the same *weightFactor* as it was before or zeroing it. Finally, by using the temporal constraints it is possible to sustain mental states or emotions only during the accomplishment of a function in another track.

Table 3. Mental and emotional state functions: suggested tag names and possible types

Function Category	Type Attribute
cognitive-process	*remember, infer, decide, idle ...*
emotion	*anger, disgust, embarrassment, fear, happiness, sadness, surprise, shame ...*

We based the specification of the **<emotion>** tag on FML-APML [11]. So each **<emotion>** has two attributes that specify the *intensity* and *regulation* of the emotion. The regulation can be: *felt* (a felt emotion), *fake* (an emotion that the participant aims at simulating) and *inhibit* (the emotion is felt by the participant but is inhibited as much as possible).

4.4 Example Scenario

We now use the proposed specification to describe the communicative functions of an example scenario about ordering a cheeseburger in a diner, a scenario that has been subject of discussion in earlier FML workshops. The following declarations describe a two floors interaction among three individuals: Gilda, Pete and George. Gilda is a customer, Pete is the cashier taking orders and George makes the burgers.

We assume that Gilda, after having approached the cashier, has already placed her order. Thus, the FML describes a floor where Pete acknowledges the order just placed by Gilda and another floor where Pete requests George to make a cheeseburger. Gilda has a friendly attitude towards Pete and acts as by-stander in the second floor between Pete and George.

Declaration Section. First we show the declaration section in Listing 1.1. Contextual information appears in the participants' identikits. In particular, Pete's `personality` is defined as *LOW* for the `extraversion` trait and his `relationships` with other participants are specified. Both Pete and George work in the same place, therefore we assumed that they are friends. The relationship information of the other two participants is left out but can be specified similarly.

As for the floors, they describe a *unicast* configuration. In `floor1` both Pete assumes the role of *speaker* and Gilda has *addressed-hearer* role. Furthermore, Gilda has a *FRIENDLY* attitude towards Pete. In `floor2`, Pete is the *speaker*, George is an *addressed-hearer* while Gilda is an *unaddressed-hearer*. They are all represented as *individual* entities in the two floors described.

```
<declarations>

<!— Participants identikits —>
 <identikits>

  <identikit id="PET" name="Pete" gender="male">
   <personality extraversion="LOW"/>
   <relationships>
    <relationship level="STRANGER" with="GIL" />
    <relationship level="FRIEND" with="GEO" />
   </relationships>
  </identikit>

  <identikit id="GIL" name="Gilda" gender="female" />
  <identikit id="GEO" name="George" gender="male" />

 </identikits>

<!— Floors configuration —>
 <floors>

  <!— Floor1 is between Pete and Gilda —>
   <floor floorID="floor1" floor-cfg="unicast">
    <participant identikitRef="PET" role="speaker" entity="individual" />
    <participant identikitRef="GIL" role="addressed-hearer" entity="individual"
         >
     <attitude affiliation="FRIENDLY" status="NEUTRAL" towards="PET" />
    </participant>
   </floor>

   <!— Floor2 is between Pete and George with Gilda as by-stander  —>
   <floor floorID="floor2" floor-cfg="unicast">
   <participant identikitRef="PET" role="speaker" entity="individual"/>
   <participant identikitRef="GEO" role="addressed-hearer" entity="individual"/
        >
   <participant identikitRef="GIL" role="unaddressed-hearer" entity="individual
        "/>
   </floor>

 </floors>

</declarations>
```

Listing 1.1. The **<declarations>** section of the cheeseburger example

Body Section. Listing 1.2 shows the body section of our example. Pete *acknowledges* the order just placed by Gilda with a `grounding` function, as can be seen in the chunk at line 6. This *must_end_before* the beginning of the second chunk described at line 11. Within this second chunk, Pete switches to the floor with George, he *takes* the `turn` and performs a *speech act* in the form of a *request*.

Starting *immediately_after*, Pete tells George to make a cheeseburger as described in the `performative` track (see 26). The two chunks in the mental state track accomplish this function with a *fake* emotional state of *anger* (see at line 39). This Pete's emotional state is sustained only for the duration of the performative act, afterwards it gets voided as we can see at line 45. Finally, *immediately_after* that Pete requests George to make a cheeseburger, Pete *gives* the `turn` away as shown at line 17.

```
1 <body>
2
3 <!-- Interactional track -->
4  <interactional>
5
6   <fml-chunk actID="ACT01" participantRef="PET" >
7    <timing primitive="must_end_before" actRef="ACT02" />
8     <grounding floorID="floor1" id="id1" type="ack" />
9   </fml-chunk>
10
11   <fml-chunk actID="ACT02" participantRef="PET" >
12    <timing primitive="start_sometime_after" actRef="ACT01" />
13     <turn-taking floorID="floor2" id="id2" type="take" />
14     <speech-act floorID="floor2" id="id3" type="request"/>
15   </fml-chunk>
16
17   <fml-chunk actID="ACT03" participantRef="PET"  >
18    <timing primitive="start_immediately_after" actRef="ACT04" />
19     <turn-taking floorID="floor2" id="id4" type="give" />
20   </fml-chunk>
21
22  </interactional>
23
24 <!-- Performative track -->
25  <performative>
26   <fml-chunk actID="ACT04" participantRef="PET" >
27    <timing primitive="start_immediately_after" actRef="ACT02"  />
28     <performative-extension id="id5" floorID="floor2" addressee="GEO">
29      <discourse-structure type="topic">
30      George make a <rhetorical-structure type="emphasis">cheesburger</
           rhetorical-structure>
31      </discourse-structure>
32     </performative-extension>
33   </fml-chunk>
34  </performative>
35
36 <!-- Mental state track -->
37  <mental-state>
38
39   <fml-chunk actID="ACT05" participantRef="PET" >
40    <timing primitive="start_together" actRef="ACT03"/>
41     <emotion floorID="floor2" id="id6" type="anger" regulation="fake"
42                intensity="0.7" weightFactor="1.0" />
43   </fml-chunk>
44
45   <fml-chunk actID="ACT06" participantRef="PET">
46    <timing primitive="start_immediately_after" actRef="ACT03"/>
47     <emotion floorID="floor2" id="id7" type="anger" regulation="fake"
48                weightFactor="0.0" />
49   </fml-chunk>
50
51  </mental-state>
52
53 </body>
```

Listing 1.2. The **<body>** section of the cheeseburger example

5 Conclusions and Future Work

In this paper we outlined the issues that an FML representation should address and we proposed a unified specification within the SAIBA framework. A preliminary interpreter for a subset of this specification has been implemented used to generate ECA behavior for a few sample scenarios [20].

The proposed FML specification is preliminary and has many limitations. The contextual information needs the inclusion of other important determinants discussed earlier, such as participant's culture and socio-relational goals. We think that the logical separation we have made in the declaration section will easily allow the inclusion of such information, for example culture and age could be part of the identikit, while socio-relational goals can be specified per floor basis.

We merely introduced the concept of *ground state* and we have suggested a simple mechanism (i.e. mental state track functions) to affect this state. However, where the ground state information is stored and the format needs to be defined. We introduced a simple prioritization schema for mental state functions with the *weightFactor* attribute, however an overall prioritization across the three tracks also needs to be defined.

The process of analyzing all the issues to address and the design of this specification led us to some final important considerations. First, modeling functions and categorizing them, separating and defining contextual information, and in general, dealing with all the aspects of human communicative functions when shaping this proposal required the adoption of a theoretical stance. For example, we adopted specific models of personality (Big 5) and interpersonal attitude (Argyle) to define contextual information. We also assumed that a communicative function can arise either from a consciously planned communicative intent that the participant aims to accomplish or unconsciously, for example, due to the participant's mental-emotional state. In either case (i.e. intentionally or unintentionally planned) our assumption is that a communicative function represents a goal to achieve in multimodal interaction and based on this assumption we designed our FML representation. These aspects certainly need agreement among the community, considering also the alternatives (for example other personality models) and the advantages of adopting specific models rather than others.

Secondly, we underlined that contextual information (or ground state of a participant) can have impact across different stages of the SAIBA framework. At functional level they can impact the production of functions (i.e. FML), at behavioral level they can impact the generation of multimodal behavior (i.e. BML) and how this behavior is realized (i.e. realization parameters). For this proposal we have chosen to deal with the last two when transforming from FML to BML. However, there seems to be a demand for inclusion in the SAIBA framework of an external standardized mechanism to handle this transformation and also a specification that goes beyond the mere representation of the two interface languages (FML and BML) is needed. In general, our recommendation is that SAIBA should not only provide standardized interface languages but also techniques and best practices that enable proper transfer between SAIBA components.

In conclusion, the FML specification proposed with this paper needs community feedback as part of an iterative process aimed at validating and improving it with further suggestions. We plan to keep working on top of this concrete specification and (1) add the missing tags to represent a wider set of communicative functions, (2) complete the specification of the *ground state* concept in the *mental state* track and possibly adopt a wider standard to express emotions (e.g. W3C EmotionML), and (3) provide a more detailed ontology to describe contextual information (e.g. incorporating participant's mood and environmental information to be used in case of iconic gestures).

Acknowledgements. This work was conducted at CADIA with support from the School of Computer Science at Reykjavik University and the Icelandic Research Fund (Learning Icelandic Language and Culture in Virtual Reykjavik). Further support provided by the EC FP7 (FP7/2007-2013) project VERVE and the Dutch national program COMMIT.

References

1. Cassell, J., Bickmore, T., Billinghurst, M., Campbell, L., Chang, K., Vilhjálmsson, H., Yan, H.: Embodiment in conversational interfaces: Rea. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, CHI 1999, pp. 520–527. ACM (1999)
2. Niewiadomski, R., Bevacqua, E., Mancini, M., Pelachaud, C.: Greta: an interactive expressive eca system. In: Proceedings of the 8th International Conference on Autonomous Agents and Multiagent Systems, vol. 2, pp. 1399–1400 (2009)
3. Cassell, J., Vilhjálmsson, H.H., Bickmore, T.: Beat: the behavior expression animation toolkit. In: Proceedings of the 28th Annual Conference on Computer Graphics and Interactive Techniques, SIGGRAPH 2001, pp. 477–486. ACM (2001)
4. Vilhjálmsson, H.H.: Augmenting online conversation through automated discourse tagging. In: Proceedings of the 38th Annual Hawaii International Conference on System Sciences (HICSS 2005) - Track 4, vol. 04, p. 109.1. IEEE Computer Society (2005)
5. Hartholt, A., Traum, D., Marsella, S.C., Shapiro, A., Stratou, G., Leuski, A., Morency, L.-P., Gratch, J.: All together now - introducing the virtual human toolkit. In: Aylett, R., Krenn, B., Pelachaud, C., Shimodaira, H. (eds.) IVA 2013. LNCS, vol. 8108, pp. 368–381. Springer, Heidelberg (2013)
6. van Oijen, J.: A framework to support the influence of culture on nonverbal behavior generation in embodied conversational agents. Master's thesis, University of Twente (2007)
7. Kopp, S., Krenn, B., Marsella, S.C., Marshall, A.N., Pelachaud, C., Pirker, H., Thórisson, K.R., Vilhjálmsson, H.H.: Towards a common framework for multimodal generation: The behavior markup language. In: Gratch, J., Young, M., Aylett, R.S., Ballin, D., Olivier, P. (eds.) IVA 2006. LNCS (LNAI), vol. 4133, pp. 205–217. Springer, Heidelberg (2006)
8. Heylen, D., Kopp, S., Marsella, S.C., Pelachaud, C., Vilhjálmsson, H.H.: The next step towards a function markup language. In: Prendinger, H., Lester, J.C., Ishizuka, M. (eds.) IVA 2008. LNCS (LNAI), vol. 5208, pp. 270–280. Springer, Heidelberg (2008)
9. Vilhjálmsson, H.H., et al.: The behavior markup language: Recent developments and challenges. In: Pelachaud, C., Martin, J.-C., André, E., Chollet, G., Karpouzis, K., Pelé, D. (eds.) IVA 2007. LNCS (LNAI), vol. 4722, pp. 99–111. Springer, Heidelberg (2007)
10. Samtani, P., Valente, A., Johnson, W.L.: Applying the saiba framework to the tactical language and culture training system. In: Workshop on Functional Representations for Generating Conversational Agent Behavior at AAMAS (2008)
11. Mancini, M., Pelachaud, C.: The fml-apml language. In: Workshop on Functional Representations for Generating Conversational Agents Behavior at AAMAS (2008)
12. Kranstedt, A., Kopp, S., Wachsmuth, I.: Murml: A multimodal utterance representation markup language for conversational agents. In: Proceedings of the AAMAS Workshop on Embodied Conversational Agents Let's Specify and Evaluate Them! (2002)

13. Thórisson, K.R., Vilhjálmsson, H.H.: Functional description of multimodal acts: A proposal. In: Proceedings of the 2nd Function Markup Language Workshop "Towards a Standard Markup Langauge for Embodied Dialogue Acts" at AAMAS (2009)
14. Bevacqua, E., Prepin, K., de Sevin, E., Niewiadomski, R., Pelachaud, C.: Reactive behaviors in saiba architecture. In: Proceedings of the 2nd Function Markup Language Workshop "Towards a Standard Markup Langauge for Embodied Dialogue Acts" at AAMAS (2009)
15. Vilhjálmsson, H.H.: Representing communicative function and behavior in multimodal communication. In: Esposito, A., Hussain, A., Marinaro, M., Martone, R. (eds.) COST Action 2102. LNCS, vol. 5398, pp. 47–59. Springer, Heidelberg (2009)
16. McCrae, R.R., Costa Jr., P.T.: Personality trait structure as a human universal. American Psychologist 52(5), 509–516 (1997)
17. Goffman, E.: Forms of Talk. University of Pennsylvania Press, Philadelphia (1981)
18. Argyle, M.: Bodily communication, 2nd edn. Methuen, New York (1988)
19. Kendon, A.: Conducting Interaction: Patterns of Behavior in Focused Encounters (Studies in Interactional Sociolinguistics). Cambridge University Press, New York (1990)
20. Cafaro, A.: First Impressions in Human-Agent Virtual Encounters. PhD thesis, Center for Analysis and Design of Intelligent Agents, Reykjavik University, Iceland (2014)