FML

FML (the Function Markup Language) encodes communicative and emotional functions the agent aims to transmit. Our version of this language, FML-APML is an XML-based markup language for representing the agent's communicative intentions and the text to be uttered by the agent. The communicative intentions of the agent correspond to what the agent aims to communicate to the user: e.g., its emotional states, beliefs and goals. It originates from the APML language (de Carolis et al., 2004) which uses Isabella Poggi's theory of communicative acts (Poggi, 2007). FML-APML uses a similar syntax as BML one. It has a flat structure and allows defining explicit duration for each communicative intention. In comparison with APML, the FML-APML language is simpler to use and tag nesting (i.e. tags had to be nested one in another; no partial overlap was allowed) is not required any more. The duration of each communicative intention can be specified explicitly (in seconds) or in relation to a speech act. The other novelty is the possibility to define not only the speaker's intentions but also the listener's ones.

Communicative intentions in FML-APML

Each FML-APML tag represents one communicative intention; different communicative intentions can overlap in time. We consider the following tags (taken from (Poggi, 2007)):

- certainty: is used to specify the degree of certainty the agent intends to express; possible values: certain, uncertain, certainly not, doubt;
- performative: represents the agent's performative e.g. suggest, approve, or disagree; possible values: planning, thinking, remembering;
- theme/rheme: represents the topic/comment of conversation; that is, respectively, the part of the discourse which is already known or new in the participants' conversation; Possible values: implore, order, suggest, propose, warn, approve, praise, recognize, disagree, agree, criticize, accept, advice, confirm, incite, refuse, question, ask, inform, request, announce, beg, greet;
- belief-relation: corresponds to the metadiscoursive goal, i.e. the goal of stating the relationship between different parts of the discourse; Possible values: gen-spec, cause-effect, solutionhood, suggestion, modifier, justification, contrast'
- turntaking: models the exchange of speaker turns; possible values: take, give;
- emotion: describes the emotional state of the agent. We can define simple emotions using emotional labels (e.g. anger or sadness) but also complex emotional states like masking (i.e.

the agent has a certain emotion but it hides it by showing another, fake, one) or superposition of two emotions;

- emphasis: is used to emphasize the agent's verbal or nonverbal message; possible values: low, medium, high;
- backchannel: Through backchannels the listener provides information about its communicative intentions, in particular about its will and ability to continue, perceive, understand the interaction and its attitude towards the speaker's speech (if it believes or not, likes or not, accepts or refuses what is being said) (Allwood et al., 1993);
- world: refers to objects of the world.

Attributes of FML-APML tags

The attributes of FML-APML tags are:

- name: the name of the tag, representing the communicative intention modeled by the tag. For example, the name performative represents a performative communicative intention;
- id: a unique identifier associated to the tag; it allows one to refer to it in an unambiguous way;
- type: this attribute specifies the communicative meaning of the tag. For example, a performative tag has many possible values for the type attribute e.g. suggest, propose, approve, etc. Depending on both the tag name (performative) and type (one of the above values), our Behavior Planning module determines the nonverbal behaviors the agent has to perform;
- start: starting time of the tag, in seconds. It can be absolute (time 0 corresponds to the start of the FML-APML message) or relative to another tag. It represents the point in time at which the intention specified by the tag starts to be communicated;
- end: duration of the tag. It can be a numeric value (in seconds) relative to the beginning of the tag or a reference to the beginning or end of another tag (or a mathematical expression involving hem). It represents the duration of the communicative intention modeled by the tag;
- importance: a value between 0 and 1 which represents the probability that the communicative intention encoded by the tag is communicated through nonverbal behavior;
- intensity: cerain communicative acts can be expressed with different intensities. The intensity of an emotional state is described by a value from the interval [0..1].

Emotion tag

Emotion has a central role in communication and ECAs should be able to communicate their emotional state in order to increase effectiveness of interaction with humans. In the FML-APML language we have introduced the emotion tag, which models the speaker's felt and expressed emotional state. The former is the emotional state the speaker is really experiencing (which can be caused by an event, a person, a situation, etc.) while the latter is the one the speaker wants to communicate to the others. These two emotional states can be completely different: for example, a person can produce a *polite smile* to his superior even if he is angry at him. In general, people can show their emotional state(the expressed state is the felt one), suppress (the felt state is expressed the least)or mask it (the expressed state is different from the felt one).

In FML-APML the emotion tag allows us to specify complex emotional states. We can for example model situations in which our agent is feeling a particular emotional state but

simulates another emotion, hiding the felt one. This is done by controlling the felt and expressed emotional states with the regulation attribute of the emotion tag. The possible values of the regulation attribute are:

- felt : this indicates that the tag refers to a felt emotion;
- fake: this indicates that the tag refers to a fake emotion, an emotion that the agent aims at simulating;
- inhibit : the emotion in the tag is felt by the agent but it aims at inhibiting it as much as possible;

Let us consider the following example:

```
<FML-APML>
<  emotion id="e1" type="anger" regulation="felt" start="0" end="3"/>
  <emotion id="e2" type="joy" regulation="fake" start="0" end="3"/>
</FML-APML>
```

The agent's real emotional state is anger (the regulation attribute of the emotion tag is set to felt) but it wants to hide it with a fake happiness (the regulation attribute of the emotion tag is set to fake).

Speech and synchronization

FML-APML tags can be attached and synchronized to the text spoken by the agent. This is modeled by including a special tag, called speech, in the BML syntax. Within this tag, we write the text to be spoken along with synchronization points (called time markers) which can be referred to by the other FML-APML tags. For example:

```
<FML-APML>
<bml> <speech id="s1">
<tm id="tm1"/>
what are you
<tm id="tm2"/>
doing
<tm id="tm3"/>
here
<tm id="tm4"/>
</speech>
</bml>
<fml> <emotion id="id3" type="anger" start="s1:tm2" end="s1:tm4"/>
</fml>
</FML-APML>
```

With the above code, we specify that the communicative intention of *emotion* starts in correspondence with the word *doing* and ends at the end of the word *here*.

In FML-APML each tag contains explicit timing data, similarly to BML tags. We also maintain coherence between the two languages defined inside the SAIBA framework. So, in FML-APML we can freely define the starting and ending time of each tag, or make tags referring to each other using symbolic labels. This also allows us to specify tags that are not linked to any spoken text. That is, with FML-APML we can define the communicative intention of non-speaking agents: for example we can represent the listener's communicative intention (e.g. the listener can have the intention to communicate that it is approving what the speaker says). FML-APML tags are used to model the agent's communicative intention. Each tag represents a communicative intention (to inform about something, to refer to a place/object/person, to express an emotional state, etc.) that lasts from a certain starting time, for a certain number of seconds.

The timing attributes start and end also allow us to model the synchronization of the FML-APML tags. They both can assume absolute or relative values. In the first case, the attributes are numeric non-negative values, considering time 0 as the beginning of the FML-APML command. E.g.

```
<emotion id="id3" type="anger" start="1.0" end="2.5"/>
```

defines communicative intention "emotion=anger" which starts after one seconds of animation and lasts 2.5 seconds.

In the second case we can specify the starting or ending time of other tags, or a mathematical operation involving them. Eg.

```
<emotion id="id3" type="anger" start="s1:tm2" end="s1:tm4"/>
```

See fml/fml-apml.dtd for details.

http://semaine.opendfki.de/wiki/FML

FML

The functional markup language FML is still under discussion. Its functionality being needed nevertheless, a working language FML-APML was created as a combination of the ideas of FML with the former Affective Presentation Markup Language APML. The following shows an example FML-APML document which contains the key elements. An `<fml-apml>` document contains a `<bml>` section in which the `<speech>` content contains `<ssml:mark>` markers identifying points in time in a symbolic way. An `<fml>` section then refers to those points in time to represent the fact, in this case, that an announcement is made and that the speaker herself is being referred to between marks `s1:tm2` and `s1:tm4`. This information can be used, for example, to generate relevant gestures when producing behaviour from the functional descriptions.

The representations in the `<fml>` section are provisional and are likely to change as consensus is formed in the community.

```
<fml-apml version="0.1">
   <bml xmlns="http://www.mindmakers.org/projects/BML" id="bml1">
      <speech id="s1" language="en-US" text="Hi, I'm Poppy."
          ssml:xmlns="http://www.w3.org/2001/10/synthesis">
         <ssml:mark name="s1:tm1"/>
            Hi,
            <ssml:mark name="s1:tm2"/>
            I'm
            <ssml:mark name="s1:tm3"/>
            Poppy.
            <ssml:mark name="s1:tm4"/>
      </speech>
   </bml>
   <fml xmlns="http://www.mindmakers.org/fml" id="fml1">
      <performative id="p2" type="announce" start="s1:tm1" end="s1:tm4"/>
      <world id="w1" ref_type="person" ref_id="self" start="s1:tm2"
end="s1:tm4"/>
   </fml>
</fml-apml>
```

For the conversion from FML to BML, information about pitch accents and boundaries is useful for the prediction of plausible behaviour time-aligned with the macro-structure of speech. In our current implementation, a speech preprocessor computes this information using TTS technology. The information is added to the end of the `<speech>` section as shown below. This is an ad hoc solution which should be reconsidered in the process of specifying FML.

```
<fml-apml version="0.1">
   <bml xmlns="http://www.mindmakers.org/projects/BML" id="bml1">
      <speech id="s1" language="en_US" text="Hi, I'm Poppy."
          ssml:xmlns="http://www.w3.org/2001/10/synthesis">
         <ssml:mark name="s1:tm1"/>
            Hi,
            <ssml:mark name="s1:tm2"/>
            I'm
            <ssml:mark name="s1:tm3"/>
            Poppy.
            <ssml:mark name="s1:tm4"/>
         <pitchaccent id="xpa1" start="s1:tm1" end="s1:tm2"/>
         <pitchaccent id="xpa2" start="s1:tm3" end="s1:tm4"/>
         <boundary id="b1" time="s1:tm4"/>
      </speech>
   </bml>
   <fml xmlns="http://www.mindmakers.org/fml" id="fml1">
      <performative id="p2" type="announce" start="s1:tm1" end="s1:tm4"/>
      <world id="w1" ref_type="person" ref_id="self" start="s1:tm2"
end="s1:tm4"/>
   </fml>
</fml-apml>
```