Generation of Multimodal Behaviors in the Greta platform
Michele Grimaldi, Catherine Pelachaud

To cite this version:
Michele Grimaldi, Catherine Pelachaud. Generation of Multimodal Behaviors in the Greta platform. 21st ACM International Conference on Intelligent Virtual Agents, Sep 2021, Kyoto (virtual), Japan. 10.1145/3472306.3478368. hal-03428896

HAL Id: hal-03428896
https://hal.archives-ouvertes.fr/hal-03428896
Submitted on 15 Nov 2021

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L’archive ouverte pluridisciplinaire HAL, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d’enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.
Generation of Multimodal Behaviors in the Greta platform

Michele Grimaldi
ISIR, Sorbonne University
Paris, France
michele.grimaldi@isir.upmc.fr

Catherine Pelachaud
CNRS, ISIR, Sorbonne University
Paris, France
catherine.pelachaud@upmc.fr

KEYWORDS
virtual agent, FML, image schema, ideational unit, multimodal behavior

1 INTRODUCTION
One of the main challenges when developing Embodied Conversational Agents ECAs is to give them the ability to autonomously produce meaningful and coordinated verbal and nonverbal behaviors. Those behaviors concern facial expressions, gestures, head movement, posture, etc. These multimodal behaviors may have different communicative functions; they can stress a word, expression an emotion, indicate a point in space, to name a few functions. They are aligned with speech, showing a high degree of synchronization. There exist many computational models that can compute multimodal behaviors for virtual agent [2, 4, 6, 9, 12, 14]. Our goal is to merge some to have a large range of behaviors to create expressive ECAs.

2 GRETA PLATFORM
Greta [13] is a real-time three dimensional embodied conversational agent with a 3D model of a human-like appearance compliant with MPEG-4 animation standard. It is able to communicate using a rich palette of verbal and nonverbal behaviors. Greta can talk and simultaneously show facial expressions, gestures, gaze, and head movements. It relies on procedural animation.

Two standard XML languages FML [7] and BML [8] are used to encode, respectively, its communicative intentions and behaviors based on the standard SAIBA architecture. Our current goal is, given communicative intentions specified in an FML file, to generate multimodal behaviors relying on three different computational models.

In our earlier model, Greta standard treatment instantiates the communicative intentions defined in an FML file into multimodal behaviors. The instantiation is based on a lexicon that contains pairs of the type (intention, multimodal behaviors). The instantiated behaviors are then combined and synchronized with speech; they are sent to the Behavior Realizer module that produces the final animation of the virtual agent.

That process is at the basis of all Greta’s behaviors but is limited in terms of behaviors complexity.

We present now the two other modules we have added. We first introduce the concept of Ideational Unit.

3 IDEATIONAL UNIT
Ideational Units are units of meaning that give rhythm to the discourse of a person and during which gestures show similar properties [16]. A gesture shape can have invariant shape properties that are critical for its meaning [16]. For instance, a gesture representing an ascension would probably have an upward movement but the hand shape may not be of particular relevance (for conveying the ascension meaning).

Within an Ideational Unit, successive gestures need to show significant changes to be distinguished. However, invariant properties of a gesture can be transferred to the variant properties of the surrounding gestures [16]. As such within an Ideational Unit, successive gestures may share common conformational features while still being distinguishable by other features.

4 MEANING MINER MODULE

The Meaning Miner module takes as input the FML file marked with prosodic and Ideational Unit [3]. Then it finds the corresponding image schemas and builds the corresponding gesture [??]. The Image Schemas Extraction module has the task of identifying the Image Schemas from the surface text of the agent’s speech and to align them with the spoken utterance (for future gesture alignment).

After obtaining a list of aligned Image Schemas for a sequence of spoken text, the gesture modeler module builds the corresponding gestures. The first step is to retrieve the gesture invariants to build the final gestures. Gestures are described by several features, namely hand shape, wrist orientation, movement and position in gesture space [10]. As explained in Section 3, gesture invariants are conformational features of gesture (e.g. a hand shape) that need not to be altered to express a given meaning and that remains invariant in successive gestures within a same temporal unit [16].

For each Image Schema we search which features are needed to express its meaning and how it is expressed using a dictionary that maps each Image Schema into its corresponding invariants.

The Meaning Miner module has to co-articulate gestures within an Ideational Unit, that structures speech, by computing either a hold or an intermediate relaxed pose between successive gestures (instead of returning to a rest pose); it also has to transfer properties of the main gesture onto the variant properties of the other gestures of the same Ideational Unit. Doing so allows us to ensure that a
meaning expressed through an invariant is carried on the same hand throughout an Ideational Unit [17].

5 NVBG
To extend the range of behaviors that Greta’s agent can assume, the NVBG (Non Verbal Behavior Generator) [11] module has been integrated in the Greta platform. Greta launches NVBG module and its chartnik parser [5]. It sends the text to be said to the NVBG module as a vrExpress Message. In the NVBG module, gestures are distinguished by:

- animation: that concerns the whole body movement except for the head
- head: that involves head movement such as nod or shake

In NVBG, an animation is a macro-type that conveys communicative acts such as: contemplate, negation, contrast, response, request, listening and so on. If NVBG finds an animation or head movement that is linked a such a communicative act, it will send back the encrypted response of the treatment.

The animation tag, the macro-type and the name of the gesture that NVBG outputs are directly not recognized in the Greta platform. The animations computed with the NVBG module are motion capture data. For that reason the response of the NVBG module is treated to suit the XML elements and gesture names understandable by Greta. Two mapping files are used to map the name of NVBG’s gesture and types to Greta’s gesture and types.

Each animation outputted by the NVBG module is an XML line that has some attributes. Firstly the mapping files allow us to change the NVBG gesture type and name with the corresponding gesture in Greta. Then the attributes that are not useful for Greta are removed or modified.

The NVBG module defines only the start of a gesture at a certain time marker without defining when it should finish (as it is defined by the corresponding motion capture animation file). Thus, when translating the output of the NVBG module to Greta, as Greta uses procedural animation, an end attribute is added for each animation line. The full treated response is a series of XML lines with tags linked to APML-FML entries (eg deictic, iconic, performative) that are understandable by Greta [16].

Now the Behavior Planner module can understand the gestures computed from NVBG. It then needs to translate them into BML signals. The whole process can be summarized as a transformation of mocap animation into BML signals that are defined by phases (such as stroke-start and stroke-end phases [8]. Finally the Behavior Realizer module receives all the signals and computes the animation via the MPEG-4 player.

6 INTEGRATION
The integration that has been done allows Greta to use Meaning Miner and NVBG models independently and to combine the results of their treatments with the standard treatment that is done by default. The two modules are fully integrated within the Greta platform.

6.1 Architecture
The Greta architecture has been updated to allow it to access the treatment of both modules, NVBG and Meaning Miner. The two models are added as external modules. The FML module reads an FML file and interacts sequentially with the three behavior generation modules. The gestures that are computed by each of them are sent to the Behavior Planner modules that orders and synchronizes them.

Figure 1: Updated treatment within the Behavior Planner module

The Behavior Planner module is re-designed to contain the NVBG and the Meaning Miner models. The remaining of the animation computation process, involving translation of the gesture into signals and then signals into video-audio animation, is not affected. The Behavior Planner module now contains all the treatments. Even if one of them does not find gestures this does not affect the rest of the process.

7 CONCLUSION AND FUTURE WORKS
The redesigned architecture integrates three modules to compute multimodal behaviors to convey communicative intentions. It allows a virtual agent to display a larger spectrum of behaviors. In the foreseeable future we intend to add another model to the agent architecture relying on a machine learning approach, namely the Sequence-to-Sequence Predictive Model [18]. It will allow Greta to extend its range of gesture and improve gestural performances for life-like virtual characters. Examples of animation can be found in the demo video at:

https://user-images.githubusercontent.com/49474878/117296170-08fae880-ae75-11eb-941d-d36012a870fc.mp4

So far, we are using in-house procedural animation. We are currently integrating the Greta platform within Unity3D to be able to use the animation capabilities of the game engine.
8 ACKNOWLEDGMENT

We thank Philippe Gauthier for his help and support. We are very greatful to Stacy Marsella and Jinah Lee for giving us access to the code of NVBG.

REFERENCES


