Model of expressive behaviors for conversational agents

Catherine Pelachaud

CNRS – LTCI
Telecom ParisTech

catherine.pelachaud@telecom-paristech.fr
GRETA PLATFORM

GRETA TEAM
Greta platform

To create one robust architecture for the embodied conversational agent (ECA) that allows for different embodiments and agent representations.
Difficulties

Different embodiments/representations have different:

- communicative capabilities e.g. different modalities available (posture, facial expressions)
- body capabilities e.g. number of fingers
- physical constraints e.g. velocity of movements, joint angles, DOF
- parameterizations of the movements e.g. FAP/BAP
- animation creation mechanisms (interpolation)
Our solution

Intent Planner

Behavior Planner

Behavior Realizer

FML-APML

BML

audio

Instructions audio

Player

Web Player

Augmented Reality

NAO

virtual agent

robot
A pair:

- **Lexicon** – mapping between communicative intentions and signals
- Each communicative intention is associated with a **behavior set** $BS = (Name, Signals, Core, Rules)$:
  - *Name* - identification of the communicative intention,
  - *Signals* - the set of nonverbal signals,
  - *Core* - mandatory signals,
  - *Rules* - relations between signals.

- **Repositories** – signals in described with a symbolical language.
Agent Specification: lexicon

The same communicative intention = different nonverbal behaviors:
• to walk around limitations of different embodiments
• to define different agent tendencies for the behavior

- single signals are defined in repositories
- entry of lexicons may convey similar meaning but can correspond to different signal
Agent Specification: repository

The same entry identification = different signal (movements shapes)
- different degrees of freedom (joints),
- different body configurations,
- physical constraints e.g. crossing the hands, velocity of movement
FML-APML language:

• describes the agent's communicative intentions i.e. what the agent aims to communicate (its beliefs, goals, and emotions),
• not a standard yet,
• based on APML (De Carolis et al. 2002),
• intentions of speaker and/or listener,
• one common FML-APML script for all the embodiments and representations

```xml
<FML-Apml>
<BML>
  <speech id="s1">Hi there! <sync id="tm1"/> </text>
  <pitchaccent id="xpa1" type="Hstar" start="s1:tm1" end="s1:tm2"/>
  <boundary id="b1" type="HH" start="s1:tm1" end="1.0"/>
</speech>
</BML>

<performative id="p1" type="greeting" start="s1:tm1" end="s1:tm2"/>
</FML-Apml>
```
BML language:

- operates at signal level,
- different channels of (non)verbal communication: head movements, gaze, facial expressions, gestures, speech.

```xml
<bml>
  <face id="smile" start="2" end="5">
    <description level="1" type="agentbml"/>
  </face>
  <gesture id="greeting" start="2" end="5">
    <description level="1" type="agentbml"/>
  </gesture>
</bml>

Agent 1

```

```xml
<bml>
  <gesture id="greeting" start="2" end="5">
    <description level="1" type="naobml"/>
  </gesture>
</bml>

Agent n
```
NAO ROBOT

Quoc Anh Le
From video to gesture specification

Jean-Claude Martin, LIMSI

```xml
<?xml version="1.0" encoding="ISO-8859-1"?>
<!DOCTYPE bml SYSTEM "bml.dtd" []>

<bml>
  <gesture>
    <type>DEICTIC</type>
    <hand>BOTH</hand>
    <handshape>FIST</handshape>
    <extendedfingerdirection>UP</extendedfingerdirection>
    <palmdirection>POLAR</palmdirection>
    <trajectory>STRAIGHT</trajectory>
    <twohanded>MIRROR</twohanded>
    <location_vertical>HIGH</location_vertical>
    <location_horizontal>OUTWARD</location_horizontal>
    <location_distance>NEAR</location_distance>
  </gesture>
</bml>
```
Transform symbolic positions for each gesture phases to joint values

```xml
<robot_movement_realizer>
  <gesture type="DEICTIC" hand="BOTH" handshape="FIST" extendedfingerringdirection="UP" palmdirection="POLAR" trajectory="STRAIGHT" twohanded="MIRROR" location_vertical="HIGH" location_horizontal="OUTWARD" location_distance="NEAR">
    ...
  </gesture>
</robot_movement_realizer>

<table>
<thead>
<tr>
<th>Model</th>
<th>JointName</th>
<th>Sensor</th>
</tr>
</thead>
<tbody>
<tr>
<td>LShoulderRoll</td>
<td>0.061318</td>
<td></td>
</tr>
<tr>
<td>LElbowYaw</td>
<td>-1.552450</td>
<td></td>
</tr>
<tr>
<td>LElbowRoll</td>
<td>-1.512482</td>
<td></td>
</tr>
<tr>
<td>LWristYaw</td>
<td>0.009162</td>
<td></td>
</tr>
<tr>
<td>LHand</td>
<td>0.027663</td>
<td></td>
</tr>
<tr>
<td>RShoulderPitch</td>
<td>0.219404</td>
<td></td>
</tr>
<tr>
<td>RShoulderRoll</td>
<td>-0.069072</td>
<td></td>
</tr>
<tr>
<td>RElbowYaw</td>
<td>1.606056</td>
<td></td>
</tr>
<tr>
<td>RElbowRoll</td>
<td>1.560120</td>
<td></td>
</tr>
<tr>
<td>RWristYaw</td>
<td>-0.012314</td>
<td></td>
</tr>
<tr>
<td>RHand</td>
<td>0.012390</td>
<td></td>
</tr>
</tbody>
</table>
```
Lexicon-based animation

```xml
<?xml version="1.0" encoding="ISO-8859-1"?>
<!DOCTYPE bml SYSTEM "bml/bml.dtd" []>
<bml>
  <speech id="s1" start="0.0" type="application/wav" voice="acapela" language="en">
    <description level="1" type="gretahtml">
      <reference>tmp/from-fml-apml.pho</reference>
    </description>
    <tm id="tm1" value="Voilà">
      <tm id="tm2" value="bien">
        <tm id="tm3" value="longtemps,">
          <tm id="tm4" value="un soir">
            <tm id="tm5" value="de printemps,">
              <tm id="tm6" value="trois">
                <tm id="tm7" value="petits">
                  <tm id="tm8" value="morceaux"/>
                </tm>
                <speech id="un_soir" start="s1:tm2" end="s1:tm3" stroke="0.2">
                  <description level="1" type="gretahtml">
                    <reference>beat=RH_BEAT</reference>
                  </description>
                  <gesture id="un_soir" start="s1:tm2" end="s1:tm3" stroke="0.2">
                    <description level="1" type="gretahtml">
                      <reference>iconic=one_night</reference>
                    </description>
                    <gesture id="troi_morceaux" start="s1:tm2" end="s1:tm3" stroke="0.2">
                      <description level="1" type="gretahtml">
                        <reference>iconic=three_pieces</reference>
                      </description>
                    </gesture>
                  </speech>
                </tm>
              </tm>
            </tm>
          </tm>
        </tm>
      </tm>
    </tm>
  </speech>
</bml>
```

API.AngleInterpolation
(joints, values, times)

BML Realizer

BML Realizer

BML Realizer

BML Realizer
Behavior Animation and Expressivity

Jing Huang
Animation and Expressivity

**Animation:**
- From set of BML tags related to different modalities, compute whole body posture
- Interpolate between posture frames

**Behavior expressivity refers to the manner of execution of the behavior**
- Can be linked to emotional states
- Or behavior styles
Overview of our animation motor using kinematics Jing Huang

The animation pipeline takes as input sequences of symbolic gestures (BML definition), generates final animation frames with FK + IK + Expressive Parameters.

The animation pipeline takes as input sequences of symbolic gestures (BML definition), generates final animation frames with FK + IK + Expressive Parameters.
Posture Generations

- **Forward Kinematics:** Rotation (Quaternion/Matrix)
- **Inverse Kinematics:** Position->Rotation

- Retargets the gestures to certain positions with time and space constraints
  - Solving postures sequentially
    - Step 1
      - Torso generation using potential target of torso depend on hand gestures positions (horizontal and vertical) (PD controller based on hand positions)
    - Step 2
      - Arm gesture generation using mass spring IK solver (shoulder movements)
      - Head direction generation using analytical solution
Full Body IK

Results of Relative Motion
Gathering process for Full Body Posture

- **Complex (basic idea)**
- **Lazy (simplification)**
  - Level1: torso
  - Level2: arms, head
Expressivity

■ Expressive Posture: Volume Editing

- spatial parameter: wrist position along McNeill’ sectors
- Openness parameter: elbow swivel angle
- Power parameter: torso relative rotation varies with time and gesture target positions due to inertia

■ Expressive Animated Sequence: Sequential Editing

- “fluidity” and “tension” using TCB spline and noise functions(for trajectory)
- “Power”: acceleration simulation through slerp (frame interpolation) or trajectory interpolation: use of time variation functions (easing in out functions)
Expressive Trajectory Generation

- Joint rotation interpolation: use Slerp (spherical linear interpolation) with time warping: easing in out functions.

- Definition of trajectory parameters:
  - Various trajectory paths: line, circle, spiral, etc.
  - Expressivity: Kochanek Bartels splines (TCB splines)

\[ Tension(T), \text{ Continuity}(C), \text{ Overshooting}(B) \]

Using TCB spline to generate new curve path for hand gesture positions
Example

```xml
<gesture category="test" id="test">
  <phase type="STROKE-START">
    <hand distanceFixed="true" fingersOrientationRatio="1.0"
      handShapeFixed="true" horizontalFixed="true"
      palmOrientationRatio="0.0" side="Right" verticalFixed="true">
      <verticalLocation>YUpperC</verticalLocation>
      <horizontalLocation>XC</horizontalLocation>
      <locationDistance>ENear</locationDistance>
      <handShape>form_fist</handShape>
      <palmOrientation>INWARD</palmOrientation>
      <fingersOrientation>AWAY</fingersOrientation>
      <palmOrientationSupplementary>DOWN</palmOrientationSupplementary>
      <fingersOrientationSupplementary>AWAY</fingersOrientationSupplementary>
    </hand>
  </phase>
  <phase type="STROKE-END">
    <hand distanceFixed="true" fingersOrientationRatio="1.0"
      handShapeFixed="true" horizontalFixed="true"
      palmOrientationRatio="0.0" side="Right" verticalFixed="true">
      <verticalLocation>YCC</verticalLocation>
      <horizontalLocation>XP</horizontalLocation>
      <locationDistance>EFar</locationDistance>
      <handShape>form_Sapart</handShape>
      <palmOrientation>INWARD</palmOrientation>
      <fingersOrientation>AWAY</fingersOrientation>
      <palmOrientationSupplementary>DOWN</palmOrientationSupplementary>
      <fingersOrientationSupplementary>AWAY</fingersOrientationSupplementary>
    </hand>
  </phase>
</gesture>
```
IVA 2012

Expressive Body Animation Pipeline for Virtual Agent

Jing Huang
Catherine Pelachaud

Telecom ParisTech - CNRS
Many emotions are expressed by sequences (or combination) of multimodal signals rather than monomodal signals (e.g., static facial expressions).
Validation study

- **Different conditions**
  - Static image
  - Single signal
  - MSE

- **Study**
  - 48 participants
  - Recognition task
  - 8 emotions:
    - anger, anxiety, cheerful, embarrassed, panic, fear, pride, relief, tension

- **Higher recognition rate**
  - MSE
  - Except for Cheerfulness
Crowd sourcing method to collect behaviors directly created by users

- **Tasks**: create behaviors associated to different meanings
  - Smiles (amused, polite, embarrassed)
  - Attitudes (submissive/dominant)

- **Interactive selection** of morphological features
  - Multimodal
  - Expressivity
Demo: repertoire of attitudes

How should I look like when I deny something with a submissive attitude?

Now, can you help our character to deny something with a submissive attitude?

Select the appropriate parameters!

- Facial expression
- Head movement
- Voice
- Gesture
- Amplitude of gesture
- Voice
- Posture of body
- Head position
- Gaze

Are you satisfied with the result?

Not at all  ●  ●  ●  ●  ●  ●  Average  ●  ●  ●  ●  ●  ●  Great!

Next animation!
Challenges

- Work with Stacy Marsella from Geneviève Calbris’ theory (Elements of meaning in gesture)

- Repertoire:
  - Metaphor: Physico-semantic representation of gestures (Calbris)
  - Find mapping between gesture component and meaning
    - axis, movement type, hand shape...
  - Position in space
Challenges

■ Realization
  • Which gesture components vary? Which ones not?
  • Demarcative function:
    - Minimal change, but enough differentiation to ensure the interlocutor can perceive they are distinctive gestures: Indication of topic continuity
    - Large change: Indication of topic shifts

■ From meaning to gesture:
  • Single meaning level
  • Ideational unit level
Challenges

- Synchrony
  - Multimodal
  - Interrelational

- Rhythmic pattern:
  - gesture component patterns
Is it possible to implement some of the models of human speech-gesture production that have been proposed?

- with limitation inherent of computational models
- Highlight ‘missing’ information
- But use ECA platform as tools

Which parts of these models are still underspecified from a computational point of view?

- Instantiation of physical metaphors into gestures
- Demarcative function: going from one gesture specification to sequence of gestures within and outside ideational units
- Emergence of behavior patterns: shape and rhythm
Any questions?

Greta Team
CNRS LTCI – Télécom ParisTech - France