



Expressive Gesture Model for Storytelling Humanoid Agent

Le Quoc Anh, Catherine Pelachaud
Telecom ParisTech

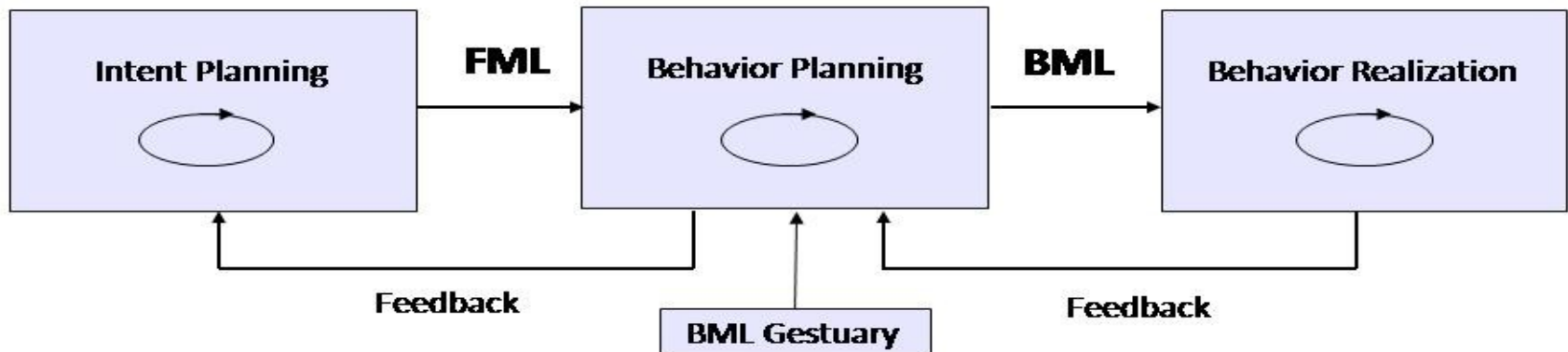


Overview

- **Objective:**
 - Build a model of expressive gestures
 - GVLEX project (Gesture and Voice for expressive reading):
 - Endow humanoid agents (NAO, GRETA) with gestures while reading a story to children.
 - Partners: Aldebaran, Acapela, LIMSI, Telecom ParisTech
- **Steps to be done:**
 - **Gesture lexicon:** Elaborate a repertoire (meaning, signals) based on gestural annotations from a storytelling video corpus.
 - **Gesture selection:** Based on extracted information of the story context to select gestures (to be realized) from the lexicon.
 - **Gesture realization:** Instantiate gestures animation in synchronization with the speech.

Method

- Use the platform of an existing virtual agent system, Greta
- Following SAIBA framework
- Two representation languages:
 - FML: Function Markup Language
 - BML: Behavior Markup Language



Affective Presentation Markup Language – FML-APML

- Describes the communicative functions
- Based on APML language (deCarolus et al)

```
<?xml version="1.0" encoding="ISO-8859-1"?>
<!DOCTYPE fml-apml SYSTEM "fml-apml.dtd" []>
<fml-apml>
  <bml>
    <speech id="s1" start="0.0" language="english" text="Hello world.">
      <description level="1" type="gretabml">
        <reference>tmp/from-fml-apml.pho</reference>
      </description>
      <tm id="tm1"/>
      Hello world!
      <tm id="tm2"/>
    </speech>
  </bml>
</fml>
```

unique name

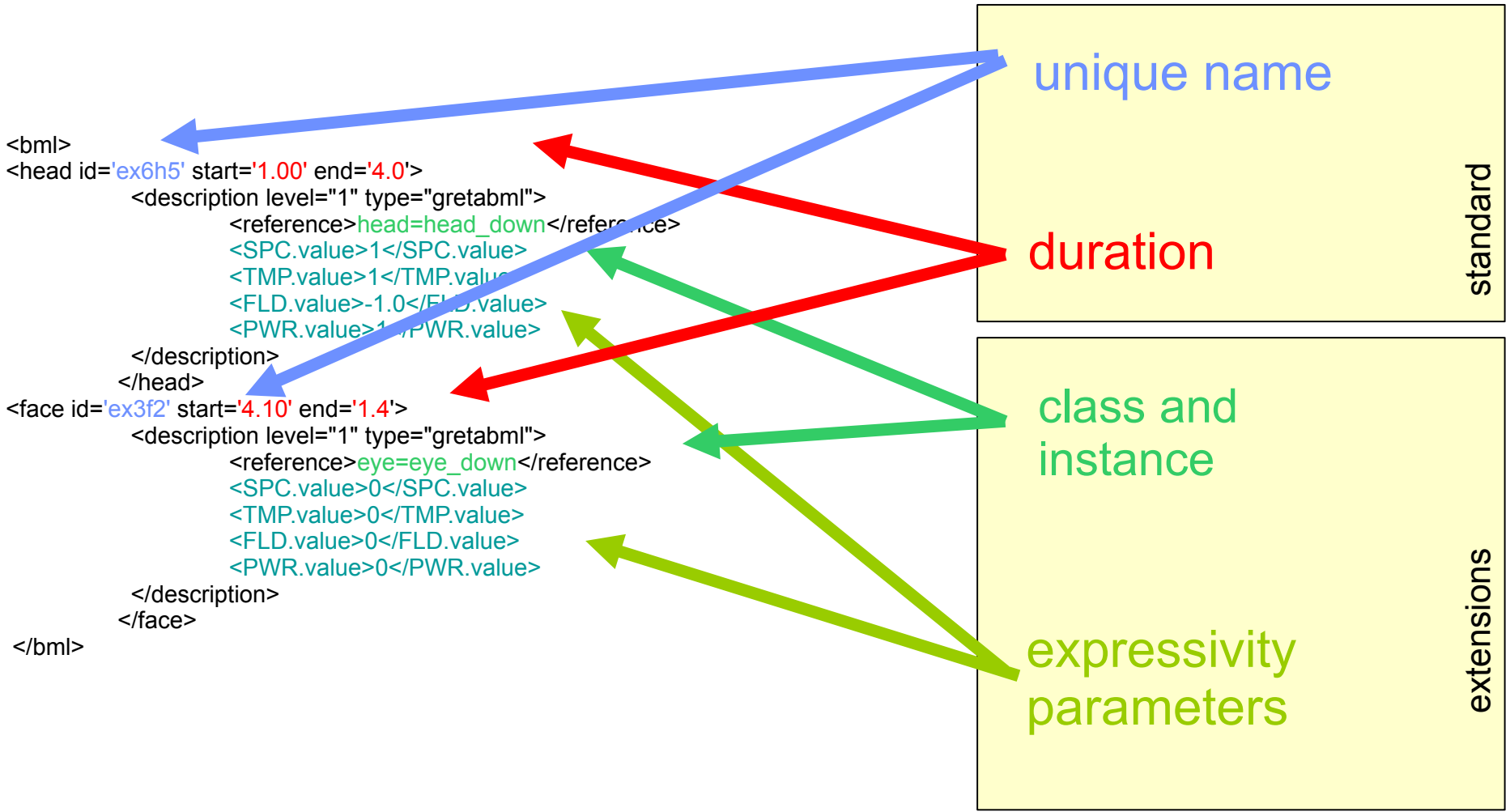
class and instance

duration

```
<performative id="p1" type="greet" start="s1:tm1" end="s1:tm2"/>
<emotion id="e1" type="joy" start="s1:tm1" end="s1:tm2"/>
<world id="w1" ref_type="place" ref_id="away" start="s1:tm1" end="s1:tm2"/>
```

```
</fml>
</fml-apml>
```

Behavior Markup Language



Robot vs. Greta

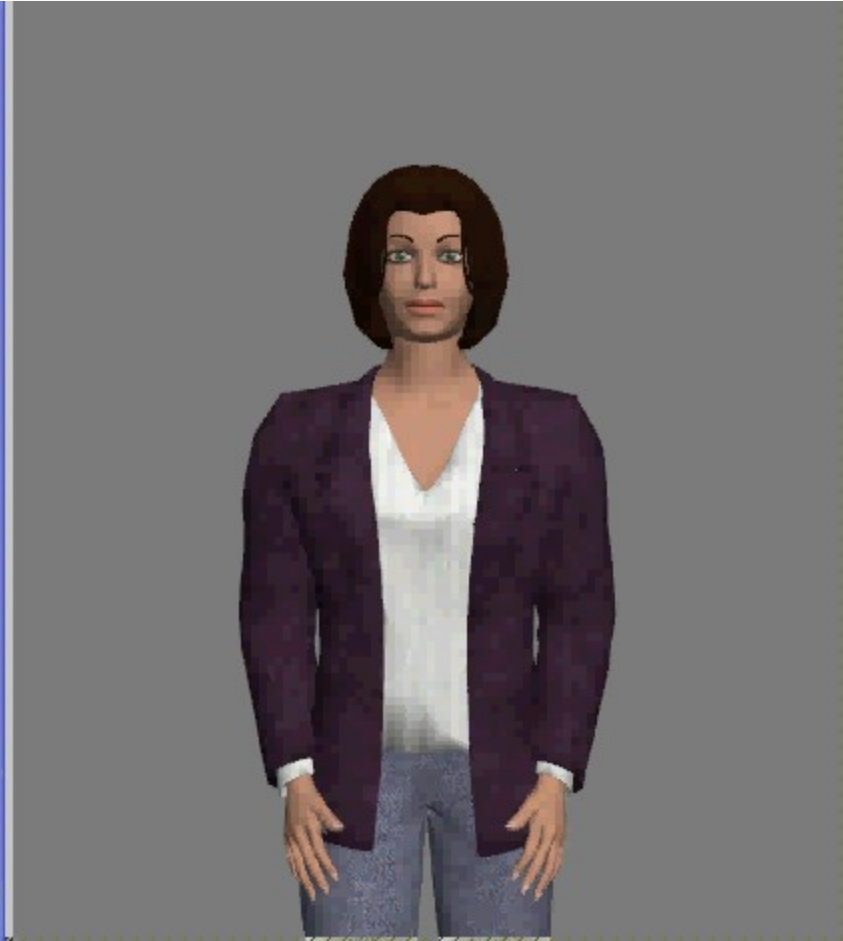
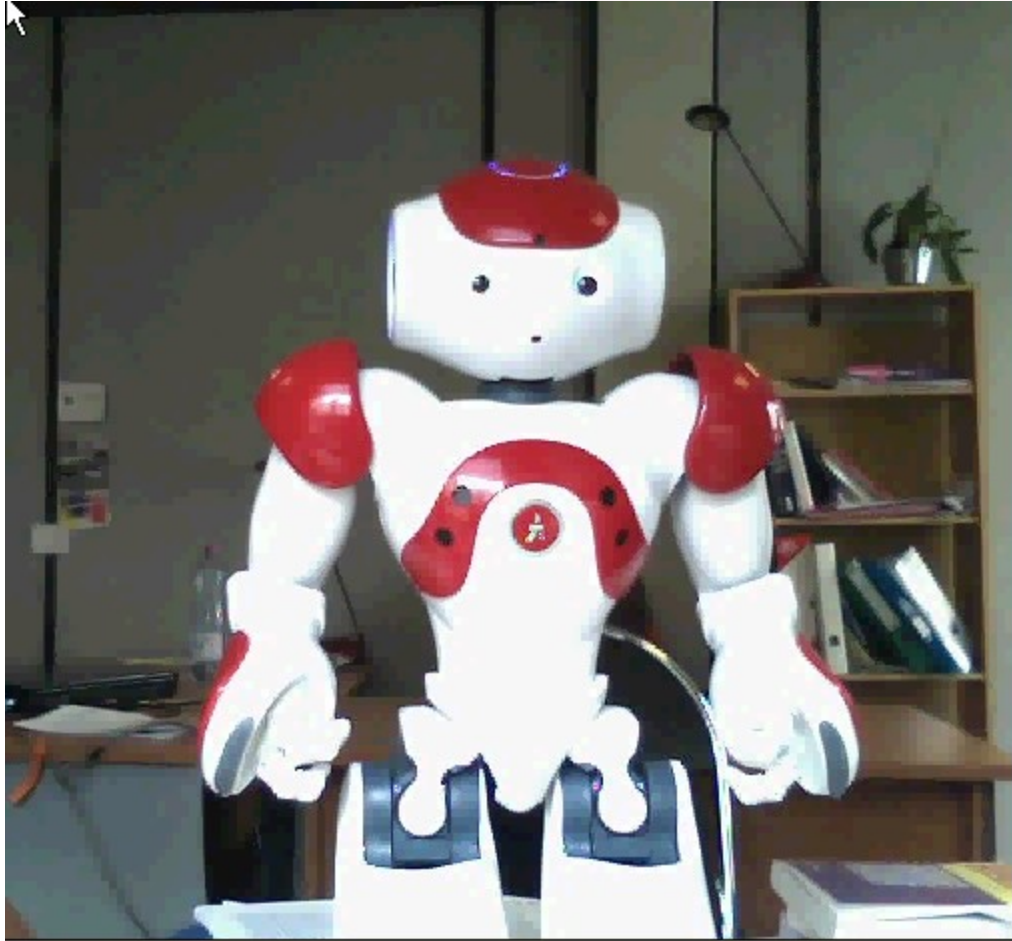
- Degree of freedoms
 - Not dynamic wrists
 - Three fingers that open or close together
 - Movement speed (>0.5 seconds)
 - Singular positions
- => Gestures may not be identical but should convey similar meaning



Gesture: Fall down



Gesture: Stop

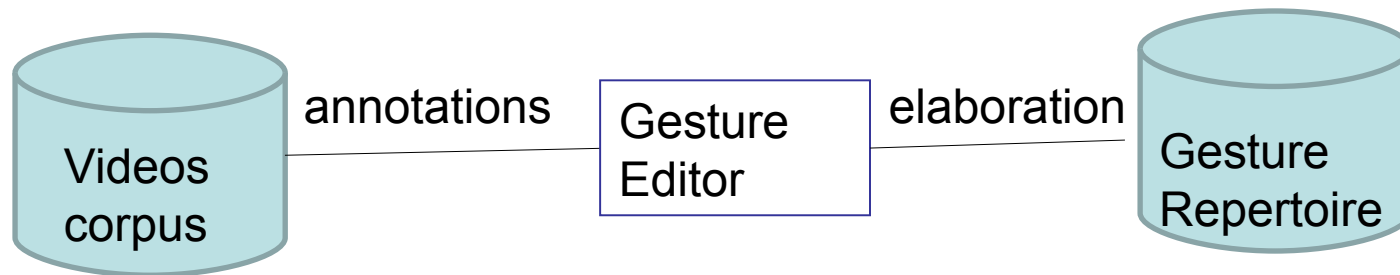


Gesture Lexicon

- Different degrees of freedom
- Variant of a gesture encompasses a family of gestures that shares
 - the same meaning (eg to stop someone)
 - a core signal (eg vertical flat hand toward the other)
- Gestures within a family may differ along the non-core signals they use
- Construction of a common lexicon with
 - Greta-Gestuary
 - Nao-Gestuary
- In the specific lexicon, variant shares similar meaning and signal-core.

Build Gesture Lexicon

- Goal: Collect expressive gestures of individuals in a specified context (story-tellers)
- Stages:
 1. Video collection
 2. Code schema and annotations
 3. Elaboration of symbolic gestures



Video collection

- 6 actors from an amateur troupe were videotaped
- Actors had received the script of the story beforehand
- The text was displayed during the session so that they could read it from time to time
- 2 digital cameras were used (front and side-view)
- Each actor was videotaped twice
 - 1st session as a training / warm-up session
 - the most expressive session can be kept for analysis

Video corpus

- Total duration: 80mn
- Average: 7 mn per story



Code schema and annotation

- Code schema
 - Goal: enable specification of gesture lexicons for Greta and Nao
 - Segmentation based on gesture phrases
 - Attributes
 - Handedness : Right hand / Left hand / 2 hands
 - Category: deictic, iconic, metaphoric, beat, emblem (McNeill 05, Kendon 04)
 - Lexicon: 47 different entries
- Annotations using Anvil tool (Kipp 01)
 - Current state: 125 gestures segmented for 1 actor
 - Rich in terms of gestures : 23 gestures per minutes for subject

Annotation

Anvil 5.0 beta 12

File Edit View Tools Bookmarks Analysis ?

Welcome to Anvil 5.0 beta 12
XML validation successful
Open ANVIL file: D:\Le Quoc Anh\Gvlex\S1-2-CM-DG-text
Loading video:
video codec: MJPG
screen size: 720x540
frame rate: 30.00029945373535fps
duration: 08:42:33 (15668 frames)
audio: LINEAR 32000.0Hz stereo

Current specification:
D:\Le Quoc Anh\Gvlex\gesture-affect.xml
07:19:60 modified frame 13188

Main Video: S1-2.avi (80%)

Emblem.Despair

Track: Speaker.Gesture

Track (primary): Speaker.Gesture
Time: 07:18:56 - 07:20:30 (52 frames)

Attributes
hand: 2H
category: emblem
lexicon-entry: Emblem.Despair
lexem: "Emblem.Despair"

Comment

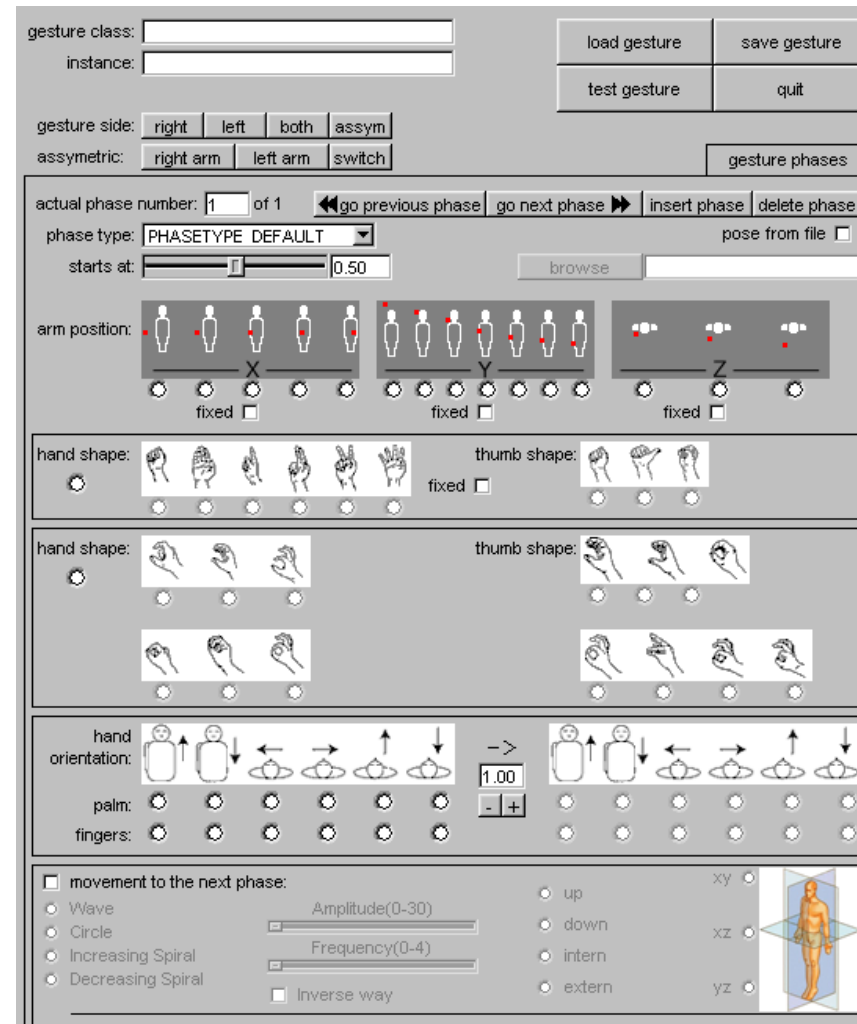
Start Edit Create & Edit Cut

Annotation: S1-2-CM-DG-texte.anvil

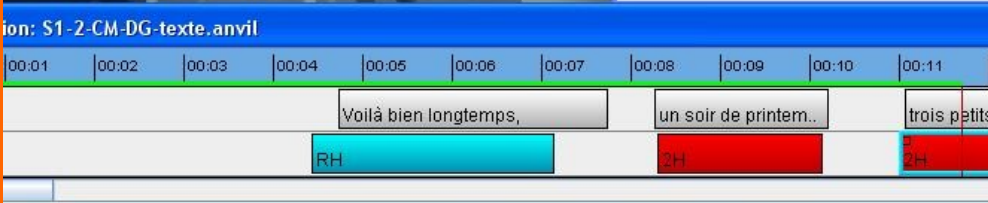
Speaker	speaking	Si le chat venait à passer, et les croquait dans leur sommeil ?	Ah là là ! se dit dame Souris.	La vie est bien compliquée.
	Gesture	Emblem Panic	Meta Chop	Emblem Strong
			Emblem Despair	Emblem So-What

Gesture Editor

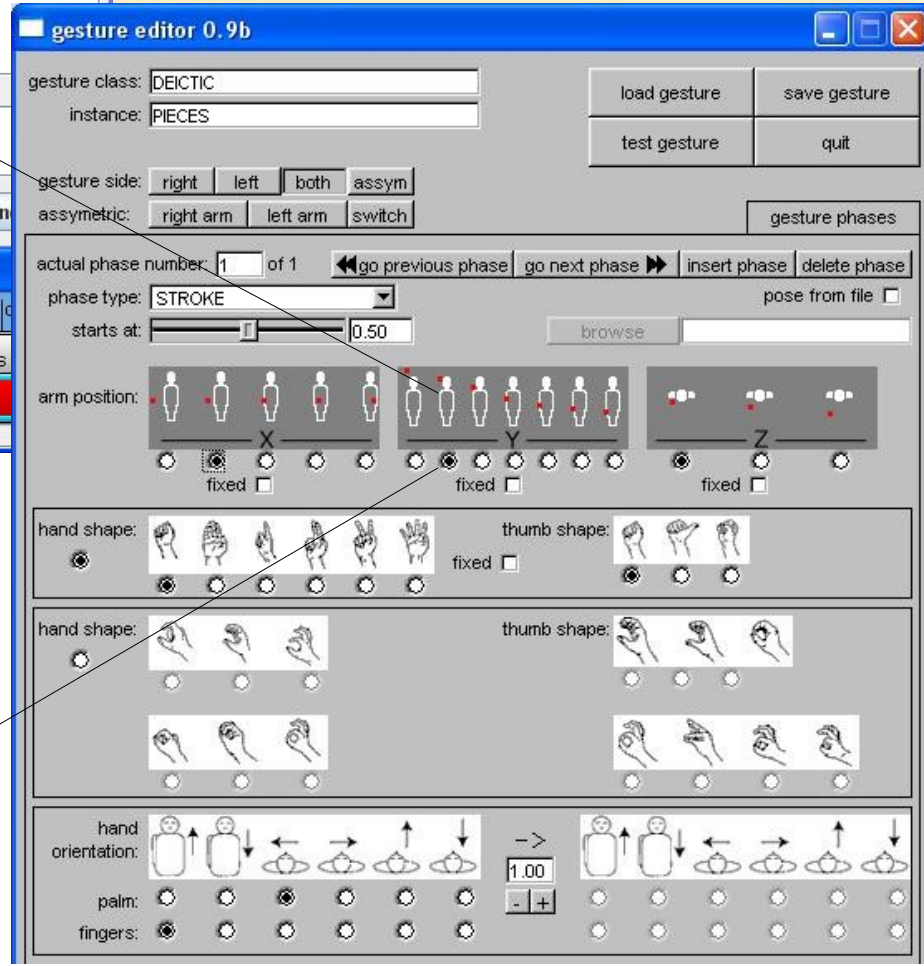
- Gesture described symbolically:
 - Gesture phases: preparation, stroke, hold, relaxation
 - Wrist position
 - Palm orientation
 - Finger orientation
 - Finger shape
 - Movement trajectory
 - Symmetry (one hand, two hand,..)



Gesture Editor



```
<?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>
```



Compilation

- Positions of hand
 - Pre-calculate joint values of all combinations of hand positions in 3D space (vertical, horizontal, distance) = (ShoulderRoll, ElbowYaw, ElbowRoll, WristYaw)
 - Current state: 105 positions corresponding to 7 vertical values, 5 horizontal values and 3 distance values
 - Replace symbolic positions by real joint values when compiling.
- Forms of hand
 - Open hand
 - Close hand

```
<?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>
```

Model	
JointName	Sensor
LShoulderRoll	0.061318
LElbowYaw	-1.552450
LElbowRoll	-1.512482
LWristYaw	0.009162
LHand	0.027663
RShoulderPitch	0.219404
RShoulderRoll	-0.069072
RElbowYaw	1.606056
RElbowRoll	1.560120
RWristYaw	-0.012314
RHand	0.012390

Reference to repertoire of gestures

```
<?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" langu
<description level="1" type="gretabml">
<reference>tmp/from-fml-apml.pho</reference>
</description>
<tm id="tm1"/> Voilà
<tm id="tm2"/> bien
<tm id="tm3"/> longtemps,
<tm id="tm4"/> un soir
<tm id="tm5"/> de printemps,
<tm id="tm6"/> trois
<tm id="tm7"/> petits
<tm id="tm8"/> morceaux
</speech>
```

BML Realizer

tm[1..8] = [2.9,3.1,3.2,3.4,3.6,3.7,3.9,4.2]

```
<gesture id="rhbeat" start="s1:tm2" end="s1:tm3" stroke="0.2">
<description level="1" type="gretabml">
<reference>beat=RH_BEAT</reference>
</description>
</gesture>
```

BML Realizer

```
<gesture id="un_soir" start="s1:tm6" end="s1:tm7" stroke="0.2">
<description level="1" type="gretabml">
<reference>iconic=one_night</reference>
</description>
</gesture>
```

BML Realizer

```
<gesture id="troi_morceaux" start="s1:tm9" end="s1:tm10" stroke="0.2">
<description level="1" type="gretabml">
<reference>iconic=three_pieces</reference>
</description>
</gesture>
```

BML Realizer

JointName	Sensor
LShoulderRoll	0.061318
LElbowYaw	-1.552450
LElbowRoll	-1.512482
LWristYaw	0.009162
LHand	0.027663
RShoulderPitch	0.219404
RShoulderRoll	-0.069072
RElbowYaw	1.606056
RElbowRoll	1.560120
RWristYaw	-0.012314
RHand	0.012390

JointName	Sensor
LShoulderRoll	0.061318
LElbowYaw	-1.552450
LElbowRoll	-1.512482
LWristYaw	0.009162
LHand	0.027663
RShoulderPitch	0.219404
RShoulderRoll	-0.069072
RElbowYaw	1.606056
RElbowRoll	1.560120
RWristYaw	-0.012314
RHand	0.012390

JointName	Sensor
LShoulderRoll	0.061318
LElbowYaw	-1.552450
LElbowRoll	-1.512482
LWristYaw	0.009162
LHand	0.027663
RShoulderPitch	0.219404
RShoulderRoll	-0.069072
RElbowYaw	1.606056
RElbowRoll	1.560120
RWristYaw	-0.012314
RHand	0.012390

API.AngleInterpolation(joints, values, times)



First result



- Voilà bien longtemps, un soir de printemps, trois petits morceaux de nuit se détachèrent du ciel et tombèrent sur Terre....

Future work

- **Lexicon Elaboration:**
 - Encode symbolic gestures in BML syntax.
 - Define invariant signification of gestures.
- **Gesture Realization:**
 - Improve synchronization mechanism to tie gestures to speech.
 - Add expressivity parameters for gesture implementation in real-time.