SCHOOL OF INTERACTIVE ARTS + TECHNOLOGY

SFU SIMON FRASER UNIVERSITY
ENGAGING THE WORLD

How to Build an Embodied Conversational Agent

Ö. Nilay Yalçın, PhD.
School of Interactive Arts and Technology
Simon Fraser University
oyalcin@sfu.ca

digital health circle

iVizLab
EXPRESSION BASED INTELLIGENT VISUALS AND VISUALIZATION LAB

Outline

- What is an Embodied Conversational Agent?
- Behavior Realizer
  - BML, SSML, AIML
  - Smartbody
  - Unity
- Inputs/Percepts
  - Visual
  - Speech
- Decision Making
  - Dialogue Manager
- Discussion

Embodied Conversational Agents (ECAs)

… are **agents** that can interact with users with a **multimodal**, **situated** (and often anthropomorphic), and **real-time interaction** to emulate a similar experience of ***human-to-human conversational interaction*** (Cassell, Bickmore, Campbell, & Vilhjálmsson, 2000).

What do you need for an Embodied Conversational Agents?

Behavior Realizer

SMARTBODY : http://smartbody.ict.usc.edu/

SmartBody is a Behavioral Markup Language (BML) realization engine that transforms BML behavior descriptions into real-time animations (keyframes).

iVizLab Upcoming Project:

3D Model Creation

Model rigging and animation

Behavior Synchronization

BML - Behavior Markup Language

http://www.mindmakers.org/projects/bml-1-0/wiki

A Standard Markup Language for Behavior Realizers. An example:

```
<bml>
    <speech id="speech0" type="application/ssml+xml">
        <mark name="T0" />My <mark name="T1" />morning<mark name="T2" />started
        <mark name="T3" />off<mark name="T4" />amazing <mark name="T5" />
    </speech>
    <gesture name="ChrBrad@Idle01_BeatLowLf02" stroke="speech0:T4" />
    <face amount="0.7" au="102" ready="speech0:T0"  relax="speech0:T5" type="facs"/>
</bml>
```

Behavior Block

BML - Behavior Markup Language

http://www.mindmakers.org/projects/bml-1-0/wiki

A Standard Markup Language for Behavior Realizers. An example:

```
<bml>
    <speech id="speech0" type="application/ssml+xml">
        <mark name="T0" />My <mark name="T1" />morning<mark name="T2" />started
        <mark name="T3" />off<mark name="T4" />amazing <mark name="T5" />
    </speech>
    <gesture name="ChrBrad@Idle01_BeatLowLf02" stroke="speech0:T4" />
    <face amount="0.7" au="102" ready="speech0:T0"  relax="speech0:T5" type="facs"/>
</bml>
```

Behaviors

Behavior Block

BML - Behavior Markup Language

http://www.mindmakers.org/projects/bml-1-0/wiki

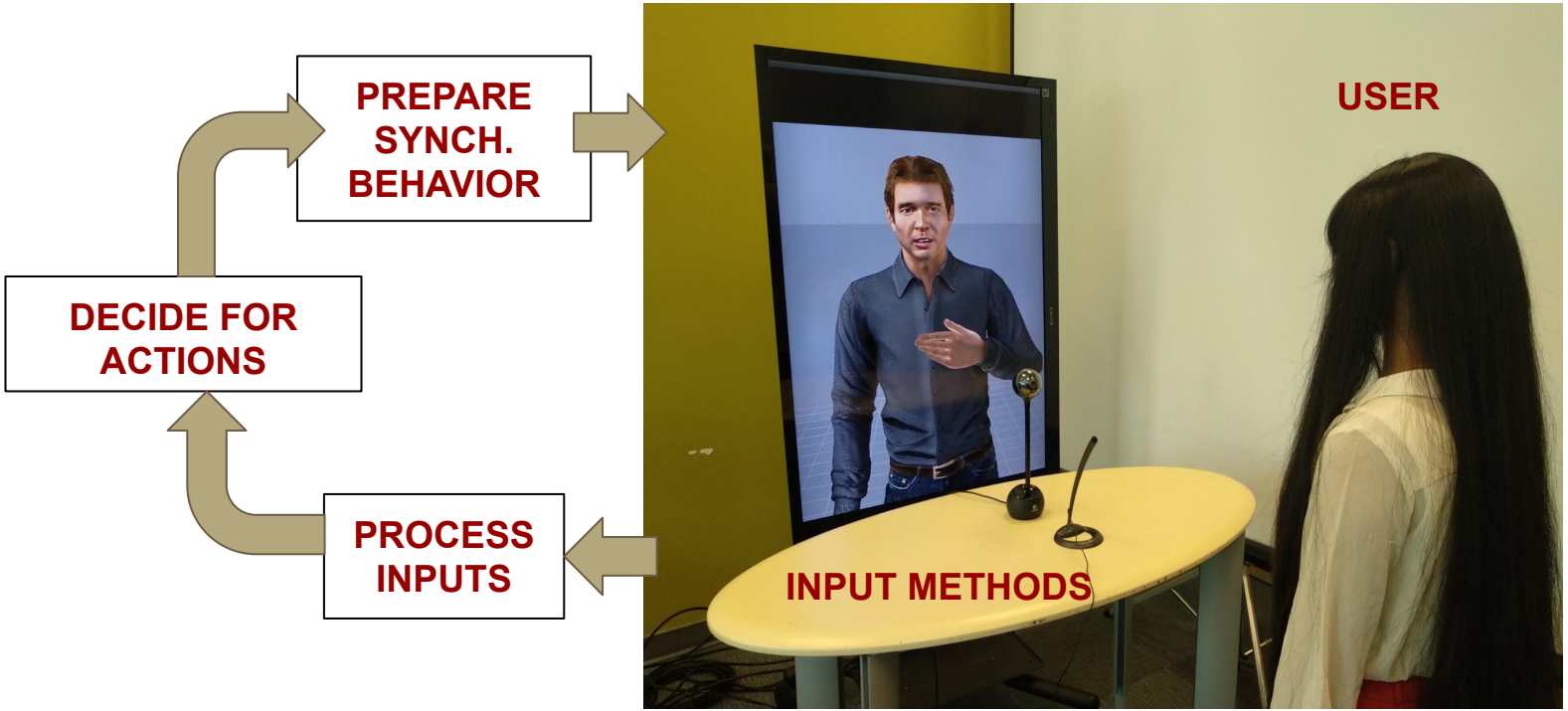
A Standard Markup Language for Behavior Realizers. An example:

```
<bml>
    <speech id="speech0" type="application/ssml+xml">
        <mark name="T0" />My <mark name="T1" />morning<mark name="T2" />started
        <mark name="T3" />off<mark name="T4" />amazing <mark name="T5" />
    </speech>
    <gesture name="ChrBrad@Idle01_BeatLowLf02" stroke="speech0:T4" />
    <face amount="0.7" au="102" ready="speech0:T0"  relax="speech0:T5" type="facs"/>
</bml>
```

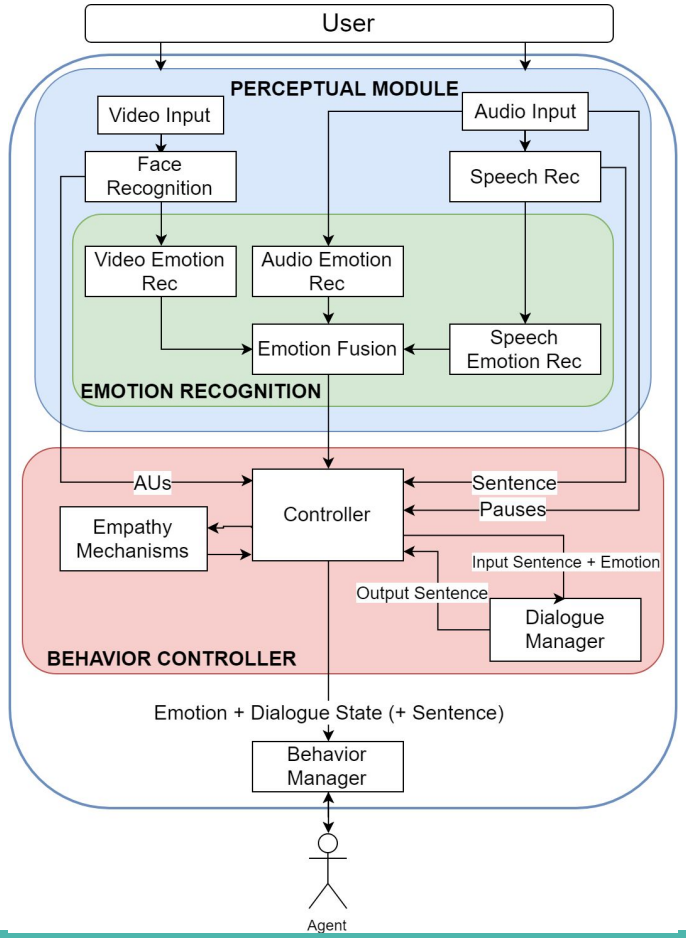
Behaviors — **Behavior Block** — **Synchronization constraint**

```
<bml>
    <speech id="speech0" type="application/ssml+xml">
        <mark name="T0" />My <mark name="T1" />morning<mark name="T2" />started
        <mark name="T3" />off<mark name="T4" />amazing <mark name="T5" />
    </speech>
    <gesture name="ChrBrad@Idle01_BeatLowLf02" stroke="speech0:T4" />
    <face amount="0.7" au="102" ready="speech0:T0"  relax="speech0:T5" type="facs"/>
</bml>
```

Creating the AI Framework



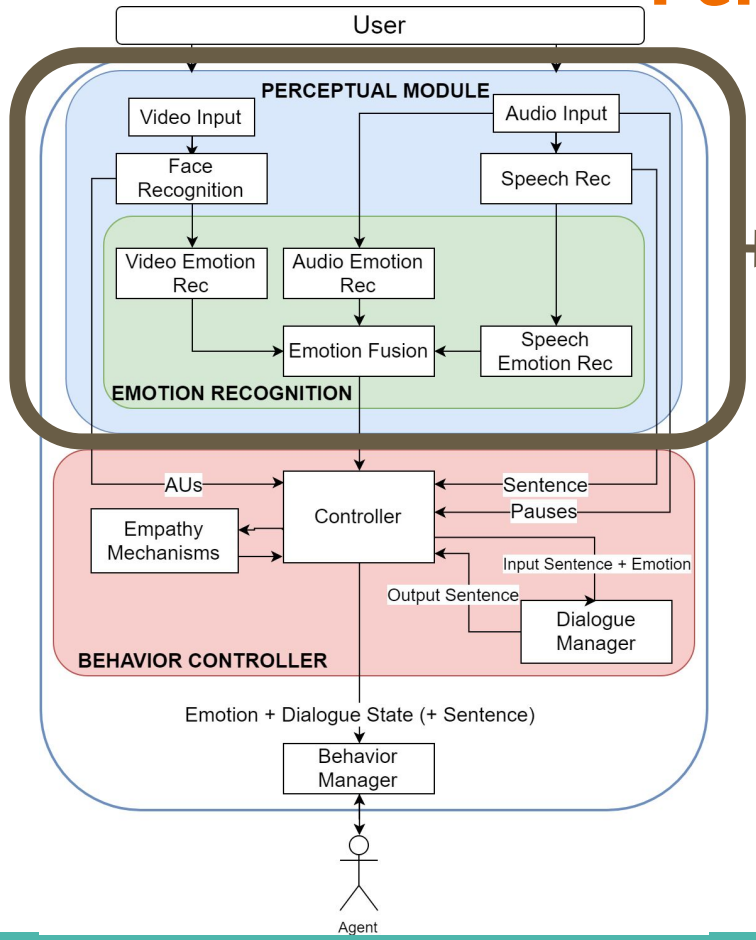
Framework

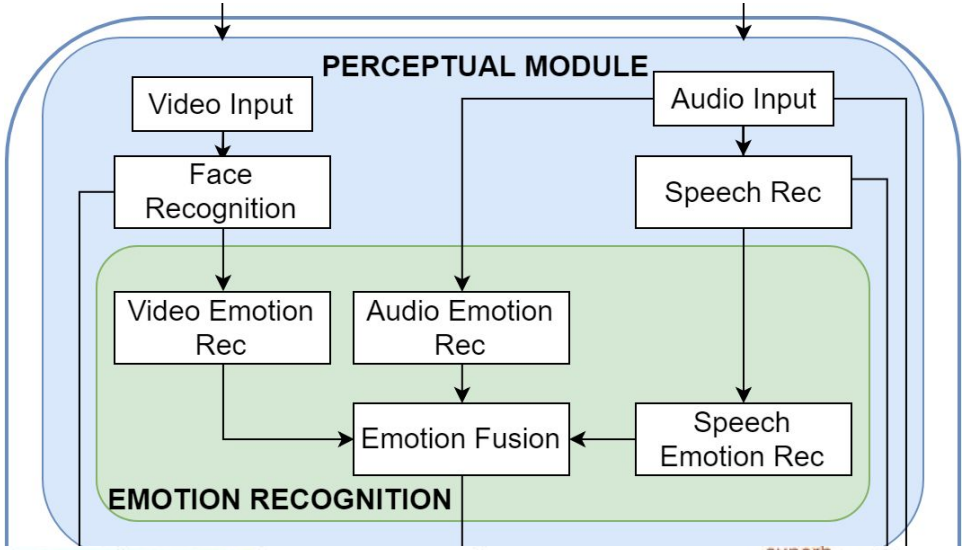


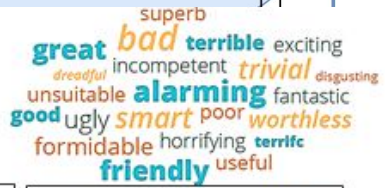
Yalçın, Ö. N. (2019). Empathy Framework for Embodied Conversational Agents. *Cognitive Systems Research.* https://doi.org/10.1016/j.cogsys.2019.09.016

- Code available at github: https://github.com/onyalcin/M-PATH/
- Data generated available at SFU Radar: https://researchdata.sfu.ca/islandora/object/islandora%3A10691
- Python 3.5, Windows platform

Perceptual Module







- OpenCV
- CK + Dataset
- CNN – Keras with Tensorflow backend

- PyAudio
- Ravdess Dataset
- CNN – Keras with Tensorflow

- Google Cloud API
- Python NLTK
- SO-CAL + Vader
- NRC Dataset

Speech Recognition

Tools:

- https://pypi.org/project/SpeechRecognition/
  - Google API: https://cloud.google.com/speech-to-text/
  - CMU Sphinx: https://cmusphinx.github.io/wiki/
  - PocketSphinx: https://github.com/bambocher/pocketsphinx-python
  - Microsoft: https://azure.microsoft.com/en-us/services/cognitive-services/speech-services/

Datasets:

- http://www.openslr.org/resources.php
- LibriSpeech audiobook : http://www.openslr.org/12/
- Speech commands : https://ai.googleblog.com/2017/08/launching-speech-commands-dataset.html
- VoxForge accented English : http://www.voxforge.org/

Facial Emotion Recognition (FER)

Tools:

- https://github.com/onyalcin/face_emotion_recognition
  - OpenCV + Keras (CNN)
- OpenFace : https://cmusatyalab.github.io/openface/

7 basic emotions = neutral, happy, sad, angry, surprised, disgust, fear

Datasets:

- CK+ http://www.consortium.ri.cmu.edu/ckagree/
- Microsoft FER+ https://github.com/Microsoft/FERPlus

1st stage: Echo Bot

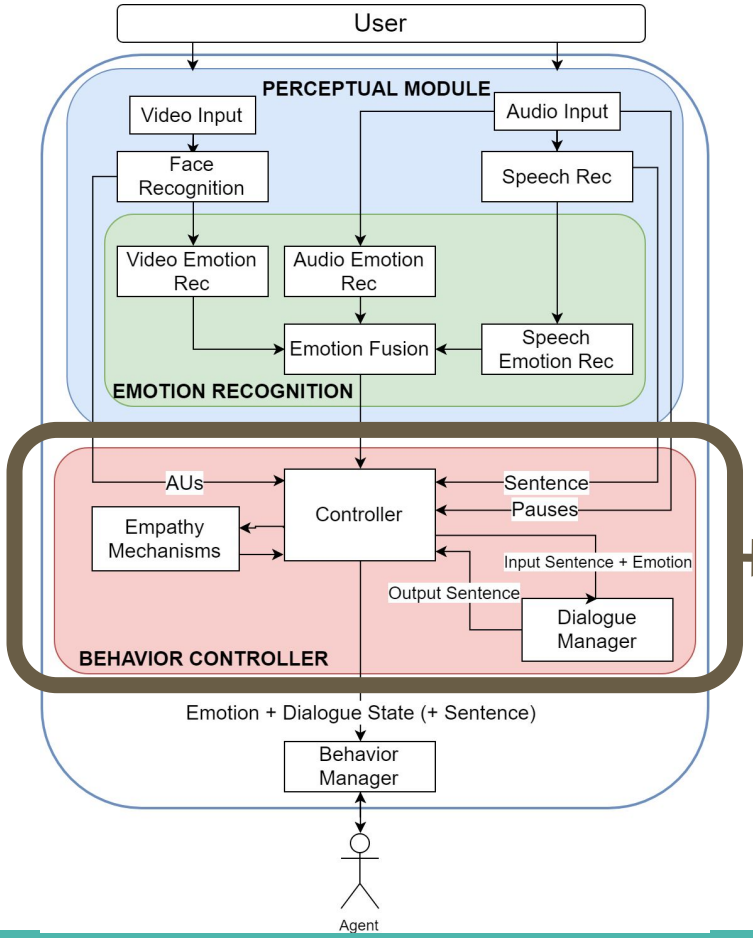
INPUTS:

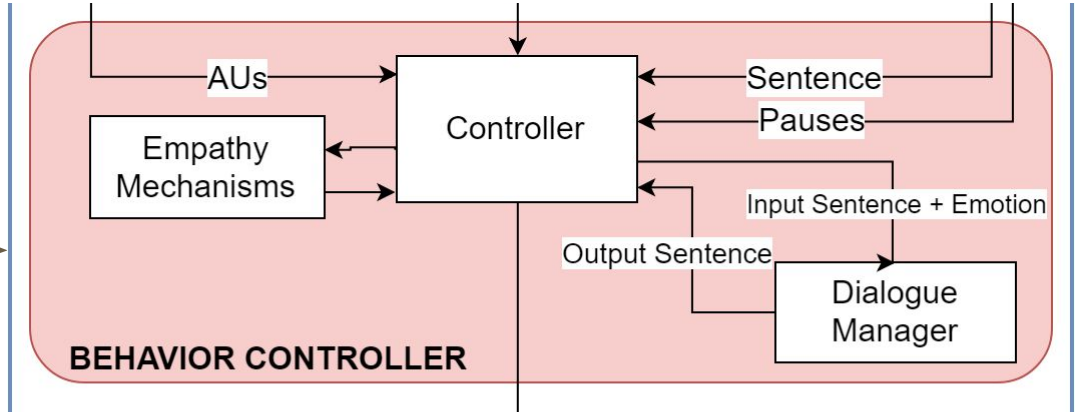
OUTPUTS:

BML(emotion categories + speech text)

https://github.com/onyalcin/echo_bot

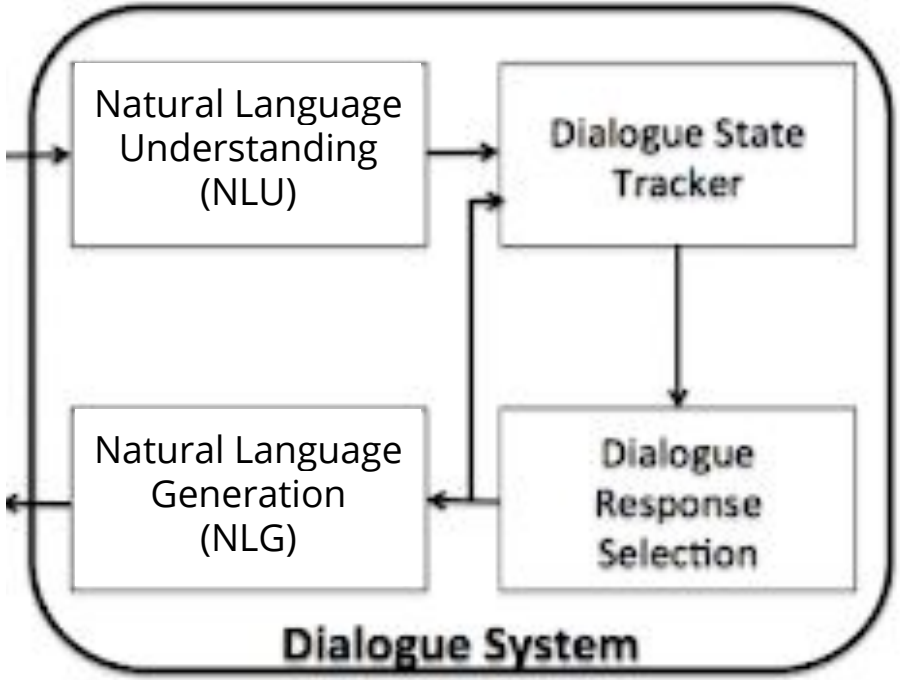
Controller


User
PERCEPTUAL MODULE
Video Input
Audio Input
Face Recognition
Speech Rec
Video Emotion Rec
Audio Emotion Rec
Emotion Fusion
Speech Emotion Rec
EMOTION RECOGNITION
AUs
Controller
Sentence
Pauses
Empathy Mechanisms
Input Sentence + Emotion
Output Sentence
Dialogue Manager
BEHAVIOR CONTROLLER
Emotion + Dialogue State (+ Sentence)
Behavior Manager
Agent


AUs
Controller
Sentence
Pauses
Empathy Mechanisms
Input Sentence + Emotion
Output Sentence
Dialogue Manager
BEHAVIOR CONTROLLER

Dialogue Management

USER INPUT
"What is the weather tomorrow?"



SYSTEM RESPONSE
"Where are you based?"

```
intent: weather_query, 93%
entities:
  date: tomorrow, 98%
```

```
next_action:
  location_question, 80%
   weather_api_call, 10%
```

Your database, backend or CRM

Dialogue Management

Tools:

- RASA : https://rasa.com/docs/getting-started/
  - https://github.com/RasaHQ/rasa
- DialogFlow (https://dialogflow.com/), Azure (https://azure.microsoft.com/en-ca/services/bot-service/), Watson Assistant (https://www.ibm.com/cloud/watson-assistant/)

Datasets:

- https://breakend.github.io/DialogDatasets/

2nd Stage: Conversational Bot

INPUTS:

emotion categories + speech text

REASONING:

speech response

OUTPUTS:

BML(emotion categories + response text)

Further implementations for a rainy day…

Dialogue:

- https://github.com/SenticNet/conv-emotion
- https://github.com/deepmipt/DeepPavlov
- https://tutorials.botsfloor.com/dialog-management-799c20a39aad
- Xiaoice: https://arxiv.org/abs/1812.08989

Audio Emotion Recognition:

- https://github.com/marcogdepinto/Emotion-Classification-Ravdess
- RAVDESS : https://smartlaboratory.org/ravdess/

Sentiment Analysis:

- https://github.com/cjhutto/vaderSentiment
- http://saifmohammad.com/WebPages/lexicons.html

Thank you!
Questions?

http://www.sfu.ca/~oyalcin/

oyalcin@sfu.ca

https://ivizlab.org/

https://github.com/onyalcin/