Interacting with Socially Interactive Agents

Catherine Pelachaud

CNRS – ISIR – Sorbonne Université

catherine.pelachaud@upmc.fr
Social Interactive Agents

Virtual characters that can:

• simulate cognitive and expressive human capabilities

• communicate using verbal and nonverbal means

• display a wide range of socio-emotional behaviors

• be socially aware and emotionally competent

• be capable of holding multi-modal social interactions
Nonverbal Behaviors

Nonverbal behaviors:

- have a communicative function
- have different meanings; context dependent
- are socially shared
- may be intentional or not (scratching one’s head because it hitches is not a NVB)

They are dynamic, multimodal, transient

- Facial expression, Gesture
- Gaze, Posture
- Voice, Touch
Metaphoric gestures

GESTURE GENERATION FROM IMAGE SCHEMAS

BRIAN RAVENET, CHOE CLAVEL, CATHERINE PELACHAUD
Metaphoric Gesture

Metaphoric gestures: convey abstract concepts through the physical behavior of a gesture, its form and motion.

Communication of message

- Build from concrete elements, the properties of those elements and actions on them.

Eg:

* Anything = set of items

Hands and head $\rightarrow$ delineate space of items
Metaphoric Gestures

Autonomously generate meaningful and coordinated verbal and nonverbal behaviors:

◦ From the textual surface discourse of the agent augmented with prosodic information (e.g. pitch accents), compute:
  ◦ timing (when to place a gesture)
  ◦ shape (which gesture form and movement)

◦ Capture mental imagery from text and map it into gesture
Image Schemas

Image Schemas: allows for manipulation of spatial, temporal and compositional concepts (container vs object and whole vs split for instance).

- Image Schemas
  - can be used to drive gesture (Cienki, 2005)

- Examples
  - UP, DOWN, FRONT, BACK, LEFT, RIGHT, NEAR, FAR, INTERVAL, BIG, SMALL, GROWING, REDUCING, CONTAINER, IN, OUT, SURFACE, FULL, EMPTY, ENABLEMENT, ATTRACTION, SPLIT, WHOLE, LINK, OBJECT.
Computational Model

Theoretical model

- Cognitive processes
- Image Schemas
  - Speech channel
    - Verbal content and prosody
      - Observed correlations
    - Nonverbal channel
      - Gestures, facial expressions...

Computational model

- Image Schemas
  - Speech channel
    - Verbal content and prosodic markers
  - Nonverbal channel
    - Prosodic markers
    - Gestures
Metaphoric Gesture - Example

Idea = concrete object with physical properties, such as size, location or weight

Example:
- Important idea = an object big in size
- Ideas can be thrown away
- Ideas can be held tightly
- Ideas can be made visible
Extract Image Schemas from the text

“idea to rise to the surface”
Extract Image Schemas from the text

"idea to rise to the surface"

OBJECT PATH UP

Textual speech of the agent
Extract Image Schemas
Associate Image Schemas to Gesture Primitives
Combine primitives into ideational units
Produce co-verbal metaphorical gestures

Brian Ravenet
Associate Image Schemas to gesture primitives

Based on Geneviève Calbris
- Object

- Path
- UP

Textual speech of the agent
Extract Image Schemas
Associate Image Schemas to Gesture Primitives
Combine primitives into ideational units
Produce co-verbal metaphoric gestures
Associate Image Schemas to gesture primitives

« idea OBJECT to rise PATH to the surface »

Ideational Units (Xu et al. 2014)

Textual speech of the agent
Extract Image Schemas
Associate Image Schemas to Gesture Primitives
Combine primitives into ideational units
Produce co-verbal metaphoric gestures

Brian Ravenet
Combine the primitives into ideational units

Textual speech of the agent ➔ Extract Image Schemas ➔ Associate Image Schemas to Gesture Primitives ➔ Combine primitives into ideational units ➔ Produce co-verbal metaphoric gestures

Brian Ravenet
Interaction

Build socially interactive agents:

- Able to manage the impression it gives on users
- Optimize engagement
- Manage interpersonal relationship

In order to enhance user’s perception of SIAs

Focus on adaptation mechanisms
Adaptation

During interaction, several mechanisms of adaptation occur to:

- Favor engagement
- Enhance user’s experience
- Signal interpersonal relationship
- Increase rapport
- Establish trust
- ...

Can be signaled through:

- Imitation, backchannel, verbal alignment, synchronization
Adaptation

How an agent able to perceive user’s state and to adapt to her state affects user’s perception?

How does it impact user’s perception of her interaction with the agent?

Conduct studies where agent adapts to user to enhance engagement or impression

Studies

- Adaptation of body movement
- Adaptation at nonverbal behaviors level
- Adaptation at conversational strategies level
- Adaptation at cues level
Laughter

Laughter is an essential social signal in human-human communication. It is a social vocalization universal across cultures and languages. It can convey various functions:

- Feedbacks to humorous stimuli or praised statements
- Mask embarrassment
- Reinforce bonding
- Social indicator of in-group belonging [Adelsward 1989]
- Speech regulator during conversation [Provine 1996]
- Elicit interlocutor’s laughter [Provine 1996]
LoL – Laugh out Loud
Pecune & Mancini

Study impact of laughter on human experience during a human-agent interaction

Effect of dynamic coupling between interactants → enhances sense of engagement (Prepin et al., 2012)

Behavior expressivity conveys affective content (Castellano et al., 2012)

Copying paradigm: copying the expressivity dynamically as it evolves in human's performance.

AIM: Study how a virtual agent able to copy and to adapt its laughing and expressive behaviors on the fly participates in enhancing user's experience in the interaction.
Condition: No Copying
Agent does not copy user’s laugh behavior quality
LoL – Evaluation study
Biancardi

Context
- 32 participants listen to funny music first without then with a virtual agent.
- Conditions: The agent can laugh performing a prefixed behavior, or copying user’s laughter intensity.
- Evaluate user’s perception of music funniness
- Evaluate user’s mood during the experience
- Evaluate social and spatial presence, and believability of the virtual agent

Results
- Participants perceived the music as **funnier** when the agent was present and copied user’s laughter intensity, than when they listened to the music alone;
- Participants’ mood was **more positive** when the agent was present and copied user’s laughter intensity, than when they listened to the music alone;
Overall Architecture
Mancini, Wang

User’s analysis

Dialog model

Agent’s behavior

Adaptation mechanism
Agent = virtual guide of an exhibit on video games
Location: Museum of Sciences and Industry at La Villette, Paris
Participants: museum visitors

3 stages

- Pre-questionnaire: NARS. apriori attitude of participants towards the agent
- Interaction with agent
- Post-questionnaires:
  - Perception of agent
  - Interaction quality
Scenario
Biancardi
Study 1: Adapting agent’s behavior according to user’s impressions
Biancardi, Mancini, Wang

Goal: Let agent learn the best combinations of nonverbal behaviors according to its goal to be perceived as warm or socially competent

Detection of user’s impressions of agent’s warmth and competence → from the analysis of their facial expressions. (C. Wang et al, 2019)

Agent’s impression manager: Flipper, dialog manager + Reinforcement learning

→ verbal and non-verbal behaviors agent displays next
→ rewards: detected impressions

Communicative Intention
◦ Types of gestures
◦ Arm rest position
◦ Smile
◦ Verbal
Experiment
Biancardi

Interaction with a virtual guide at Museum of Sciences and Industry at La Villette,

- 71 participants
- 3 conditions:
  - **Warmth**, when the agent adapted its behaviors according to user’s warmth impressions, with the goal to maximize its warmth;
  - **Competence**, when the agent adapted its behaviors according to user’s competence impressions, with the goal to maximize its competence;
  - **Random**, when the model was not exploited and the agent randomly chose its behavior, without considering user’s reactions.

pre-questionnaire: NARS
post-questionnaires:

- Perception of Warmth & Competence: using 4 adjectives for each
- Perception of interaction
Results

Biancardi

Adaption at behavior level to maximize participant’s impression

Capture behaviors that play a role in impression formation

Significant results for the *competence* condition

But did not manage to change user’s negative apriori (NARS) on agent’s warmth

But

No assured behaviors coherency across turns

No consideration of conversational strategy
Study 2: Adapting agent’s strategies

Biancardi

Goal: Let agent learn the best combinations of conversational strategies according to its goal to engage user.

Conversational strategies linked to warmth and competence

Reinforcement learning:
- Conversational strategy
- Reward function: user’s engagement
Conversational Strategies

Conversational strategies:

use of you- and we-words, the level of formality of the language, the length of the sentences.

Example:

Intention: explain a topic

**Supplication:** “I think that while you play there are captors that measure tons of stuffs!” accompanied by smiling and beat gestures

**Intimidation:** “While you play at video games, several captors measure your physiological signals.” accompanied by ideational gestures without smiling

Jones & Pittman, Pennebaker, Callejas
Experiment
Biancardi

Interaction with a virtual guide at Museum of Sciences and Industry at La Villette, Paris

- 75 participants
- 6 conditions: adaptation, random, ingratiation, supplication, self-promotion and intimidation

Results:

Primacy of warmth dimension:

- Supplication, ingratiation: agent appears warmer
- Self-promotion: same level of warmth as supplication and ingratiation → halo effect with competence

Stronger impact of negative impression over positive one (Peeters and Czapinski, 1990)

- Intimidation: agent appears colder

Interaction quality: adaptive agent increases user’s engagement and user’s preference
Study 3: Adapting agent’s behavior to user’s cues

Dermouche

Aim: develop a computational model that

- Captures the adaptation of interactants at the cues level
- Predicts agent’s behavior taken into account user’s cues
- Conveys agent’s communicative intents

Steps:

- Analyze behavior adaptation in human-human interaction
- Predicts agent’s behaviors from user’s ones
- Merge agent’s behaviors from its intentions with the predicted behaviors
Study 3: Architecture
Dermouche

Adaptation mechanism

- IL-LSTM: predict behavior from learned model
  - Input:
    - user’s detected behavior
      - smile, head and gaze
    - Agent’s previous behavior
  - Output:
    - Agent’s predicted behavior
    - Merge predicted behavior and communicative behavior of agent
Prediction Model

Dermouche

Corpus: NoXi
- Smile activation and intensity
- Head rotation
- Gaze direction
- Conversational state: who speaks

Prediction Model: IL-LSTM

Input:
- Time window of 20 frames of both humans A’s and B’s behaviors

Output:
- Next human B’s behaviors
Prediction Model
Dermouche

Input:
- Time window of 20 frames of user’s and agent’s behaviors

Model: IL-LSTM

Output: next agent’s frame
- \((f_0, ..., f_{20}) \rightarrow f_{21}\)
- \((f_1, ..., f_{21}) \rightarrow f_{22}\)
- \((f_2, ..., f_{22}) \rightarrow f_{23}\)
Evaluation
Dermouche

Same scenario
Same location

5 conditions:

- **REF**: agent does not adapt its behavior.
- **HEAD**: agent adapts its head rotation according to the user’s behavior.
- **SMILE**: agent adapts its smile according to the user’s behavior.
- **GAZE**: agent adapts its gaze according to the user’s behavior.
- **ALL**: agent adapts its head rotation, smile, and gaze according to the user’s behavior.
Experiment
Dermouche

101 participants (50 F)

Results: unpaired t-tests

Compared to REF condition, agent in SMILE condition is evaluated as:

- more friendly ($p = .01$)
- more involved ($p < .01$),
- less distant ($p < .01$)
- more satisfied ($p = .01$)

Idem for agent in ALL condition

Agent in HEAD and EYE condition: not validated

Effect of user’s apriori (NARS)
Conclusion

Validation for SMILE

For other modalities: users stared at screen, so they did not move their head and gaze

→ agent adapted its head and gaze behaviors to user’s ones

→ agent did not move head and gaze

Future:

need to validate with measure

° user’s behaviors

° synchronization between interactants

° behaviors coupling
Adaptation - Conclusion

Presentation of studies that model different aspects of adaptation

- Body movement
- Behaviors
- Strategies
- Cues

Need to model complex phenomena such as rapport building, empathy...
Social Touch
Boucaud - Thouvenin

- Study touch in human-agent interaction in virtual reality
- Functions of social touch
  - Attract attention
  - Manage turn taking
  - Backchannel
  - Emotional emphasis
  - Encourage
  - Comfort
  - Calm
- Develop a decision model to trigger when
  - Agent can touch human
  - When and with which touch
- Evaluate decision model and touch acceptability
Touch communicates distinct emotions. (Hertenstein et al. (2006 & 2009))

Relationship between toucher and touchee important for better results (Thompson & Hampton, 2011) and touch facilitates bonding. (Montagu, 1971) (Chatel-Goldman et al., 2014)
Virtual Social Touch
Boucaud - Thouvenin

To what extent can touch help an ECA to express emotions and bond with a human in virtual reality?

How can we determine when and how to touch?

Important to monitor touch acceptability
Framework
Boucaud - Thouvenin

Mental state: emotion/intention

Touch Production
- VR Tracking

Decision: emotion/intention

Touch Reception
- Colliders

SOFTLY

Touch Production
Touch Perception
Boucaud - Thouvenin

Virtual receptors and proximity zones by Nguyen & Wachsmuth (2009)

Tactile cells of our agent (back, arms, shoulders)
DECISION MODEL

Boucaud - Thouvenin

- Human sensing:
  - Gaze direction: agent’s eyes/head/body
  - Proxemics: intimate, personal, social, public
  - Touch: body part (head, arm, body), type of touch

- Agent’s emotion state: FaTiMa (Dias et al, 14)

- Decision model:
  - Agent’s emotional state
  - Agent’s perception of human’s emotional state
  - Agent’s perception of rapport (emotion valence, human attentiveness)
  - Dialog state

- Human touch avoidance: low, medium, high

- Output: verbal and nonverbal behavior including touch
DECISION MODEL
Boucaud - Thouvenin

Functions of touch:
- Attention getting: try to grab the attention
- Turn management: taking or giving the floor
- Emotional emphasis: display emotion
- Supporting touch: comfort, calm

Types of touch:
- Hit
- Tap
- Caress
- Sustained touch

Example: initial inputs:
- Attentiveness(H) = 2, Mood(H) = 5 and StaticTouchAvoidance(H) = Medium.

→ Rapport(H) = 60 → Speak(Step1, Step1, Inform, Gesture)

→ Failure to task → Mood(H) = 4, Attentiveness(H) = 1

→ Rapport(H) < 60 ; Mood(A) decreases → Speak(Fail1, Fail1, GetAttention, Touch)
Evaluation
Boucaud - Thouvenin

Scenario: discussion on experience of lockdown
- Autonomous virtual agent (dark blue shirt) driven by the decision model
- Avatar of human (turquoise shirt) driven by human

VEED.IO
The pandemic is really hard with all the confinements, the impossibility
Any questions?

AGENT PLATFORM: GRETA AVAILABLE AT HTTPS://GITHUB.COM/ISIR/GRETA