Nonverbal Behavior Generator for Embodied Conversational Agents

Jina Lee, Ki-young Jang
{jinal, kjang}@usc.edu
University of Southern California

April 2, 2007
Virtual Humans

• Goal: Virtual humans that act like real humans

• Behaviors not pre-scripted
  – Behave induced by understanding and reasoning about the current situation

• Communicate in Natural Language
  – Language Recognition & Generation

• Understand social situation

• Respond emotionally to situation
Virtual Human Project at USC - SmartBody

• Joint work between Institute for Creative Technologies and Information Sciences Institute

• Related Research Topics
  - Emotion modeling
  - Multi-party dialogue model
  - Speech recognition in noisy environments
  - Natural language pragmatics
  - Social reasoning
  - Negotiation about tasks
Virtual Human Body

- **Capabilities:**
  - Basic Physical Behavior
  - Walking, grasping

- **Nonverbal, expressive behavior**
  - Gestures, facial expressions, gaze

- **Requirements:**
  - Spontaneous, interactive
    - Behaviors on the fly
    - Responsiveness to events

April 2, 2007

Jina Lee, Ki-young Jang
Nonverbal Behaviors

• Nonverbal communication
  – All the messages other than words that people exchange in interactive contexts. (Hecht, DeVito, and Guerrero)

• Nonverbal Behavior (NVB)
  – Behaviors people make that convey communicative functions
  – Gestures, facial expressions, gaze, etc.

• Nonverbal behaviors serve various functions
Problem

- **Utterance**
  - I'm glad to hear that.
  - We may not trust each other well.
  - I can't believe you did that.
  - ...

- **Function**
  - Emphasize
  - Contrast
  - Emotional Expression
  - Regulate Turns
  - Refer
  - Complement
  - ...

- **Behaviors**
  - Smile
  - Brow Lowered
  - Brow Raise
  - Head Nod
  - Headshake
  - ...

- **Psychological literature**

- **Challenge**
  - To find the mapping between utterance and function
  - To model the nonverbal behavior generation for ECA using this mapping without a rich markup

April 2, 2007

Jina Lee, Ki-young Jang
Goals for NVB Generator

• Robust NVB Generation that can use markup of communicative function if provided, but can also extract/infer it if not

• Extraction that leverages syntactic and semantic analysis of text

• Use open-source tools

• Use evolving standards for markup
  – SAIBA framework (FML & BML)
  – Clear distinction of function and behavior

April 2, 2007

Jina Lee, Ki-young Jang
**Surface Text:**
Yes, Prudence. Many times. I actually quite like you

**Function:**
- **Affirmation**
- **Intensification**

**Behavior:**
- **Head nods**
- **Head nod & brow frown on word**
SAIBA Framework

- SAIBA [Kopp et al., 2006] – Situation, Agent, Intention, Behavior, and Animation
- A distinction between communicative function and communicative behavior

Function
- Emphasize
- Contrast
- Express Emotion
- Refer
- Regulate Turns
- complement ...

Behaviors
- Smile
- Brow Lowered
- Brow Raise
- Head Nod
- Headshake ...

Function Markup Language (FML)
- Specifies the communicative and expressive intent of the agent.
  - AFFECT, INTENT, TURN
  - Persistent Features:
    PERSONALITY, CULTURE, GENDER

Behavior Markup Language (BML)
- Elements roughly correspond to the parts involved in the behavior
  - BODY, GESTURE, HEAD, FACE, GAZE, LIPS, SPEECH
Virtual Human Body Embodiment

VHuman Brain
- Reasoning
- Language
- Emotion

VHuman Behavior Subsystem

Intentions & Emotions

Behavior Generation

Behavior Markup

SmartBody

Integrated Arch.


Realization: How to animate behaviors? How to schedule & compose them?
NVB Generator System Architecture

Input
Surface Text + Agent’s Emotional State (FML)

NVBGenerator
Communicative Function Derivation
Behavior Suggestion
Behavior Realization

Function Rules
NVB Rules
Behavior Description

Output
Nonverbal Behavior Execution Code (BML)

Cache
Parse Tree
Natural Language Parser (Charniak Parser)
Possible Approaches

- How do we extract the communicative function from linguistic features?
  - Information from the natural language generator (comm. Intent / affect)
    - e.g. Multi-modal NLG [Krenn et al., 2002]
  - Machine learning techniques using a gesture corpora
  - Top down analysis of video data
**Psychological Literature**

- Literature on NVB
  - e.g. Ekman, Hadar, Kendon, McClave, etc.

<table>
<thead>
<tr>
<th>Function</th>
<th>Behavior</th>
</tr>
</thead>
<tbody>
<tr>
<td>Signs of affirmation</td>
<td>Head nods</td>
</tr>
<tr>
<td>Backchannel (response) requests</td>
<td>Head nod</td>
</tr>
<tr>
<td>Self correction</td>
<td>Head shake</td>
</tr>
<tr>
<td>Concepts of inclusivity</td>
<td>Lateral sweep or head shake</td>
</tr>
<tr>
<td>(i.e. everyone, all)</td>
<td></td>
</tr>
<tr>
<td>Listing</td>
<td>Head moves with succeeding items</td>
</tr>
<tr>
<td>Uncertainty (I guess, I think…)</td>
<td>Lateral shakes</td>
</tr>
<tr>
<td>Negative expression</td>
<td>Head shake</td>
</tr>
<tr>
<td>Superlative or intensified expression</td>
<td>Head shake</td>
</tr>
<tr>
<td>(i.e. very, really)</td>
<td>Brow frown</td>
</tr>
<tr>
<td>Mark Contrast</td>
<td>Head movement</td>
</tr>
</tbody>
</table>
Analysis of Video Data

• To validate what’s found in the literature

• Find out the dynamic properties of behaviors
  – speed, repetition, span of behaviors
    (word/phrase, cross-syntactic boundaries)

• To see what the actual NVB look like
  – Do head nods across different functions appear differently?

• Relation between the behavior and linguistic properties of the surface text
  – Guide rule construction

• Sensitive Artificial Listener [HUMAINE, 2004]
Nonverbal Behaviors Observed

- **Head**
  - Nod, shake, tilt, moved to the side, pulled back, pulled down

- **Eyebrow**
  - Raised, frowned (lowered), flash

- **Eyes / Gaze**
  - Look up, look down, look away, squinted, squeezed, rolled

- **Others**
  - Shoulder shrug, mouth pulled on one side

### Breakdown of the number of utterances with corresponding function

<table>
<thead>
<tr>
<th>Function</th>
<th># of utterances (out of 223)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Negation</td>
<td>62</td>
</tr>
<tr>
<td>Intensification</td>
<td>62</td>
</tr>
<tr>
<td>Affirmation</td>
<td>36</td>
</tr>
<tr>
<td>Assumption</td>
<td>28</td>
</tr>
<tr>
<td>Word Search</td>
<td>23</td>
</tr>
<tr>
<td>Contrast</td>
<td>23</td>
</tr>
<tr>
<td>Interjection</td>
<td>13</td>
</tr>
<tr>
<td>Response Request</td>
<td>10</td>
</tr>
<tr>
<td>Listing</td>
<td>9</td>
</tr>
<tr>
<td>Obligation</td>
<td>9</td>
</tr>
<tr>
<td>Inclusivity</td>
<td>7</td>
</tr>
</tbody>
</table>
Interesting Features from Video Analysis

• Interjection
  – *Yes, no*: labeled as interjection
    • big nod once on word
  – e.g. No, you’re not.
    Yes, please.

• Word Search
  – *Well, um, uh* labeled as interjection

• Intensification
  – Literature: head shake and lowered brows on intensifying word
  – Video: big head nod and lowered brows on intensifying word
## Nonverbal Behavior Rules

<table>
<thead>
<tr>
<th>Derivation</th>
<th>Function</th>
<th>Behavior</th>
</tr>
</thead>
<tbody>
<tr>
<td>No, not, nothing, cannot, none</td>
<td>Negation</td>
<td>Head shakes on phrase</td>
</tr>
<tr>
<td>Really, very, quite, great, absolutely, gorgeous…</td>
<td>Intensification</td>
<td>Head nod and brow frown on word</td>
</tr>
<tr>
<td>Yes, yeah, I do, We have, It’s true, OK</td>
<td>Affirmation</td>
<td>Head nods and brow raise on phrase</td>
</tr>
<tr>
<td>I guess, I suppose, I think, maybe, probably, perhaps, could</td>
<td>Assumption / Possibility</td>
<td>Head nods on phrase</td>
</tr>
<tr>
<td>Um/uh/well + interjection from parser</td>
<td>Word Search</td>
<td>Head tilt, brow raise, gaze away</td>
</tr>
<tr>
<td>But, however</td>
<td>Contrast</td>
<td>Head moved to side and brow raise</td>
</tr>
<tr>
<td>Derivation</td>
<td>Function</td>
<td>Behavior</td>
</tr>
<tr>
<td>------------------------------------------------</td>
<td>----------------</td>
<td>---------------------------------------------------</td>
</tr>
<tr>
<td>Yes, no + interjection from parser</td>
<td>Interjection</td>
<td>Head nod on word</td>
</tr>
<tr>
<td>You know</td>
<td>Response Request</td>
<td>Head move to side and brow raise on word</td>
</tr>
<tr>
<td>X and Y</td>
<td>Listing</td>
<td>Head moved to one side and to the other on word</td>
</tr>
<tr>
<td>Everything, all, whole, several, plenty, full...</td>
<td>Inclusivity</td>
<td>Lateral head sweep and brow flash on word</td>
</tr>
<tr>
<td>Have to, need to, ought to</td>
<td>Obligation</td>
<td>Head nod once on phrase</td>
</tr>
</tbody>
</table>
Priorities of NVB Rules

Example
She wouldn’t be really happy.
- Head shakes over the whole sentence
- negation overrides intensification

Algorithm
1. Find all the utterances where two or more rules co-occur
2. Mark which rule overrides the other (looking at the behavior) in the matrix of rules and count the frequencies of these cases

Result
1. Interjection
2. Negation
3. Affirmation
4. Assumption/possibility, obligation
5. Contrast, word search, response request
6. Intensification, inclusivity, listing
Surface Text:

Yes, Prudence. Many times. I actually quite like you

Function Rules:

Affirmation Rule
Priority 3

Interjection Rule
Priority 1

Intensification Rule
Priority 6

Behavior Rules:

Head nods on phrase

Head nod on word

Head nod & brow frown on word

April 2, 2007

Jina Lee, Ki-young Jang
Status

Work is at a preliminary phase, but...

**SASO**¹: A leadership and negotiation skills training application.

**ELECT**¹: A cultural training application.

¹ Developed at USC Institute for Creative Technologies

April 2, 2007

Jina Lee, Ki-young Jang
Conclusion and Future Work

- A framework for NVB generator that extracts the communicative function from the input text and generates appropriate NVB

- Designed for easy modification and extension of the rules

- Module was incorporated into several applications

- Evaluation of the system and behaviors generated needed

- Machine learning techniques to aid us in the process of behavior generation
  - Requirement: large gesture corpora

- Modify and customize the current behavior generation for different gender, cultures, or personalities

- Model the affective state of the user interacting with the ECA