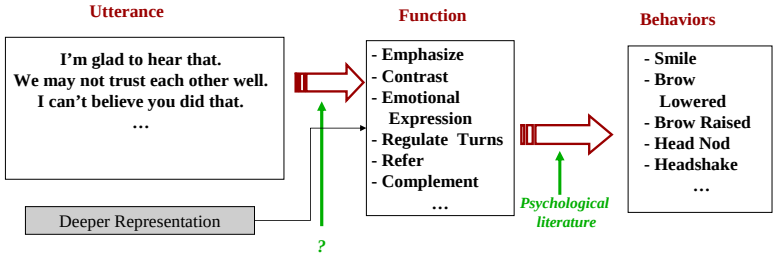
Nonverbal Behavior Generator for Virtual Humans

Jina Lee, Stacy Marsella
USC Information Sciences Institute
{jinal, marsella}@isi.edu

Introduction

Problem Description



How do we find the mapping between utterance + communicative function to behaviors?

Nonverbal Behavior Generator

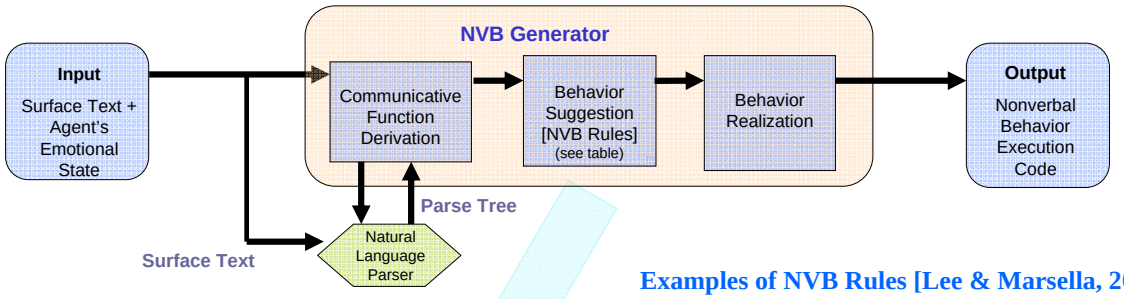
- Read in surface text and markup of virtual character's affective and intentional state and generate appropriate behaviors (e.g. facial expressions, body gestures, gaze)

Goal

- 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

Current Implementation

System Architecture



NVB Generator is informed by:

1. Analysis of Video Data

- Validate what's found in the literature
- Find out the dynamic properties of behaviors
  - speed, repetition, span of behaviors
- Explore the relation between the behavior and linguistic properties of the surface text
  - Guide rule construction
- Sensitive Artificial Listener [HUMAINE, 2004]

2. Study of Nonverbal Behaviors

- Literature on nonverbal behaviors and their communicative functions in human communication
  - Ekman, Heylen, Kendon, McClave

Examples of NVB Rules [Lee & Marsella, 2006]

- Constructed from the psychological literature and our video analysis.

Derivation	Function	Behavior
No, not, nothing, cannot, none	Negation	Head shakes on phrase
Really, very, quite, great, absolutely, gorgeous…	Intensification	Head nod and brow frown on word
Um/uh/well + interjection from parser	Word Search	Head tilt, brow raise, gaze away
But, however	Contrast	Head moved to side and brow raise
Everything, all, whole, several, plenty, full…	Inclusivity	Lateral head sweep and brow flash on word
Have to, need to, ought to	Obligation	Head nod once on phrase

Ongoing Research: Machine Learning for Gesture Models

- Start with head gesture model (e.g. head nods, head shakes, etc.), then extend to other gestures
- Gesture Corpus: AMI Meeting Corpus [Carletta et al., 2006]
  - Multimodal data set of meeting recordings
  - Contains video, audio, transcripts, dialog acts, topic segmentations, focus of attention, head gestures, hand gestures, etc.

Current Progress

Classifiers

- HMM used for visual recognition of arm gestures [Brand, Oliver, Pentland, 1996], sign languages [Assan & Groebel, 1997]
- CRF used for human motion activities [Sminchisescu et al., 2005]
- Variations of CRF (HCRF, LDCRF) used for head gesture recognition [Wang et al., 2006] [Morency et al., 2007]

Feature Selection

- Part of Speech
- Phrase boundary
- Dialog act
- Salient words (emotionally charged words, words used in current NVB rules)
- Speaker's role (speaker vs. listener)
- Theme / rheme

Future Work

- Explore ML techniques for other gesture models (arm, posture, etc.)
- Evaluation of the system and behaviors generated
- Modify and customize the current behavior generation for different gender, cultures, or personalities