Designing Bots, Virtual Humans, & Other Systems that Hold up their End of the Conversation

Justine Cassell

Carnegie Mellon University & ISIR / La Sorbonne
Chaire Blaise Pascale & Chaire Sorbonne
Mai 2018

The ineffable quality of Rapport in learning

Children who report more rapport are more likely to learn from the virtual peer

Search Engine vs. Conversation

Justine: “OK Google, I love Manchester United”

Google: Manchester United Football Club is a professional football club Based in Old Trafford, Greater Manchester, England, that competes in the Premier League, the top flight of English Football

Justine: “I love Manchester United”

Friend.: “No way! Arsenal wipes the floor with those Red Devils!”

Socially-Aware Robot Asst: “No way! Arsenal wipes the floor with those Red Devils!”

Motivation for Socially-Aware Bots

1. People pursue *multiple conversational goals* in every conversation & expect the same from their interlocutors. To put people at ease, and increase relationship strength, we must understand the ***propositional***, ***interactional*** & ***interpersonal*** functions of conversation.

2. People change interaction styles over time. We must increasingly ***manage long-term interactions*** with people by changing interaction style in a way that evokes increasing loyalty, rapport and trust.

Rapport improves *task* performance

Surveys

- Survey respondents gave higher quality answers if they felt rapport with interviewer (Berg (1989)

Health

- Physicians who build rapport during trial interviews enroll more participants (Albrecht *et al.*, 1999).

Sales

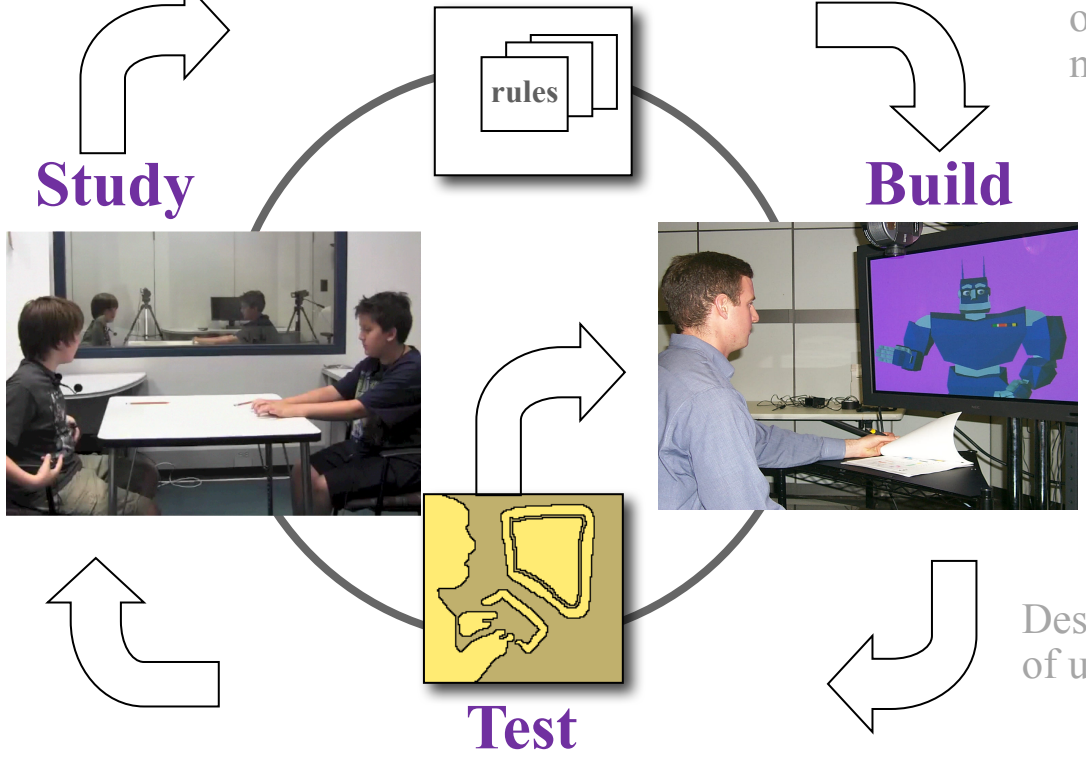
- Rapport with sales staff leads to increased likelihood of purchasing goods/service (Brooks, 1989).
- Customers show increased trust and disclosure when rapport is maintained with sales staff (LaBahn, 1996).

Methodology

Build formal models

Theorize & Model

Implement system on the basis of model



Start here

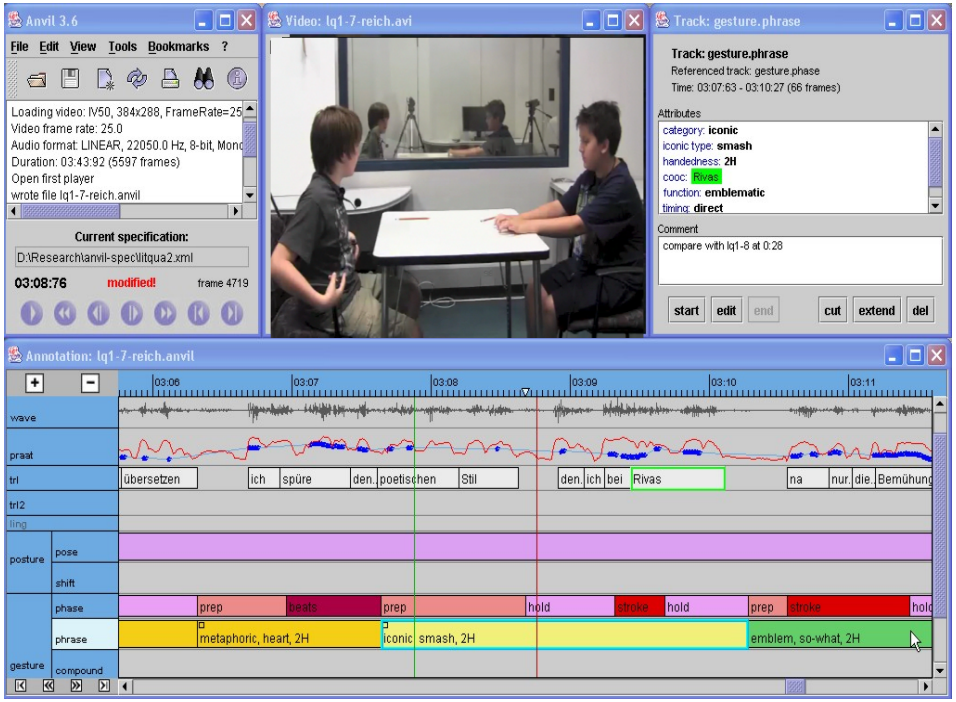
Collect Natural data

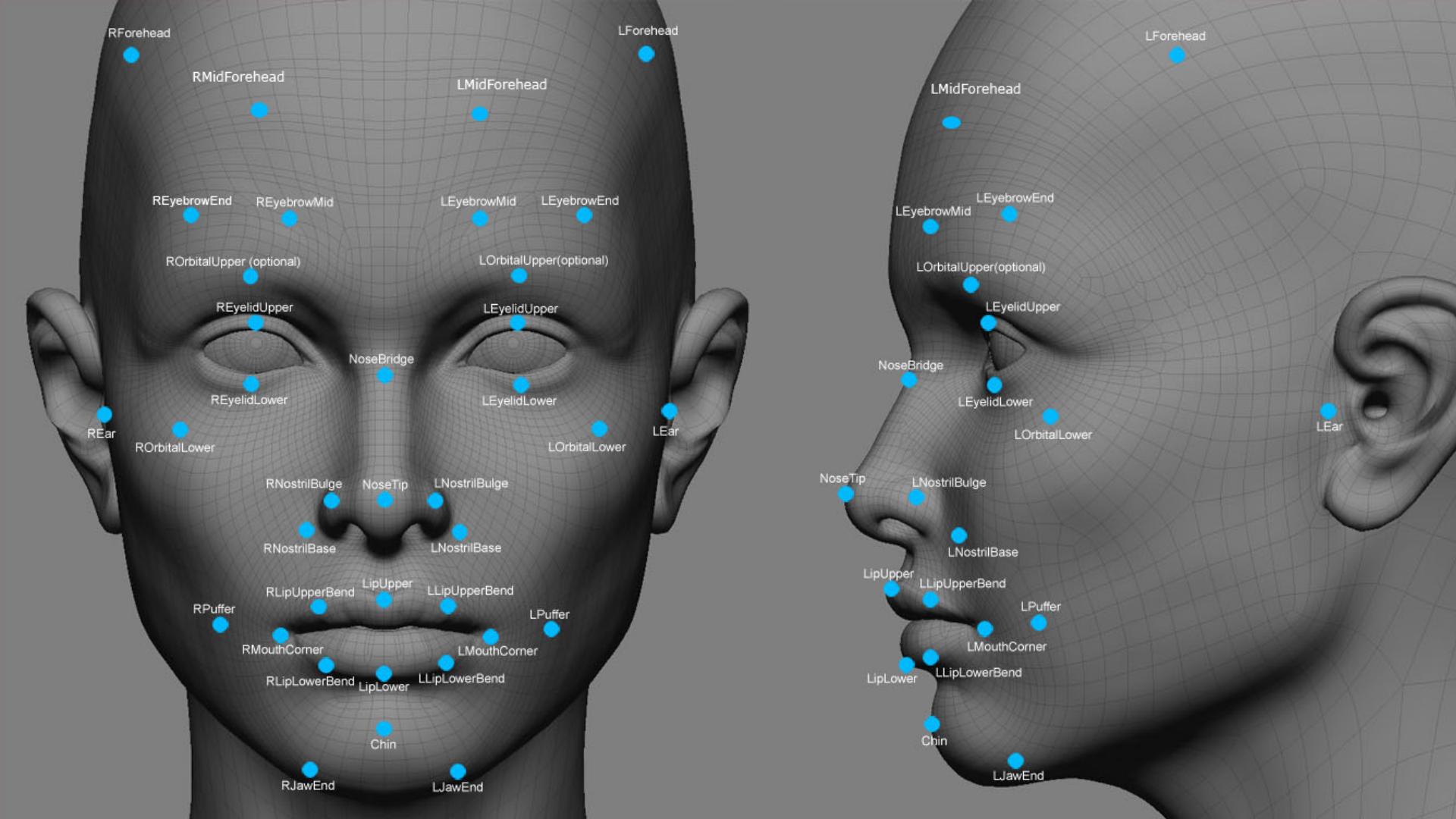
Realize gaps in understanding

Design evaluation of use

Observe

Analyze


Anvil 3.6
File Edit View Tools Bookmarks ?
Loading video: IV50, 384x288, FrameRate=25
Video frame rate: 25.0
Audio format: LINEAR, 22050.0 Hz, 8-bit, Mono
Duration: 03:43:92 (5597 frames)
Open first player
wrote file lq1-7-reich.anvil
Current specification:
D:\Research\anvil-spec\litqua2.xml
03:08:76 modified! frame 4719
Video: lq1-7-reich.avi
Track: gesture.phrase
Track: gesture.phrase
Referenced track: gesture.phrase
Time: 03:07:63 - 03:10:27 (66 frames)
Attributes
category: iconic
iconic type: smash
handedness: 2H
cooc: Rivas
function: emblematic
timing: direct
Comment
compare with lq1-8 at 0:28
start edit end cut extend del
Annotation: lq1-7-reich.anvil
03:06 03:07 03:08 03:09 03:10 03:11
wave
praat
trl
übersetzen ich spüre den poetischen Stil den ich bei Rivas na nur die Bemühung
trl2
ling
posture
pose
shift
gesture
phase
phrase
compound
prep beats prep hold stroke hold prep stroke hold
metaphoric, heart, 2H
iconic smash, 2H
emblem, so-what, 2H


RForehead
LForehead
RMidForehead
LMidForehead
REyebrowEnd
REyebrowMid
LEyebrowMid
LEyebrowEnd
ROrbitalUpper (optional)
LOrbitalUpper(optional)
REyelidUpper
LEyelidUpper
NoseBridge
REyelidLower
LEyelidLower
REar
LEar
ROrbitalLower
LOrbitalLower
RNostrilBulge
NoseTip
LNostrilBulge
RNostrilBase
LNostrilBase
LipUpper
RLipUpperBend
LLipUpperBend
RPuffer
LPuffer
RMouthCorner
LMouthCorner
RLipLowerBend
LipLower
LLipLowerBend
Chin
RJawEnd
LJawEnd
LForehead
LMidForehead
LEyebrowEnd
LEyebrowMid
LOrbitalUpper(optional)
LEyelidUpper
NoseBridge
LEyelidLower
LOrbitalLower
LEar
NoseTip
LNostrilBulge
LNostrilBase
LipUpper
LLipUpperBend
LPuffer
LMouthCorner
LipLower
LLipLowerBend
Chin
LJawEnd


ChristineAltese. deviantart

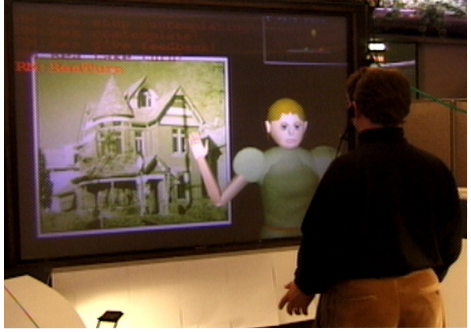
Analysis of Rapport

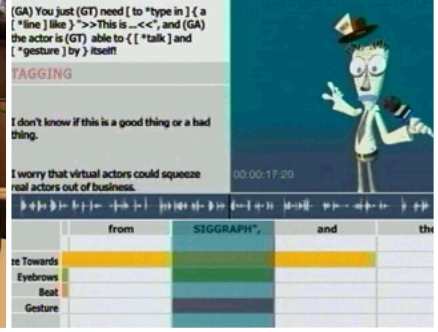
Thin-slice (Ambady & Rosenthal, 1992) judgments of every 30 second video segment) presented to 3 annotators in randomized order

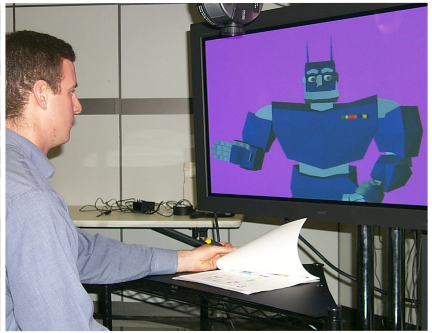
IRR

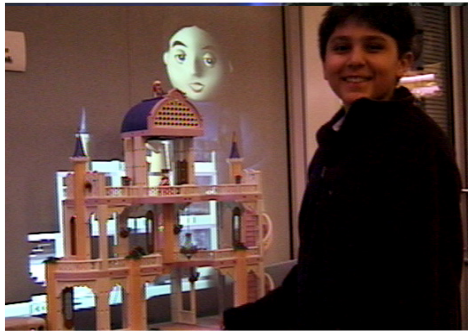
- *"Consensus"* measured by Intra-Class Correlation (single measure): *0.37*
- *"Consistency"* measured by Cronbach alpha: *0.68*
- *Inverse-based bias correction* (Kruger et al., 2014) was used mitigate rater bias & pick single rapport rating for each 30 second video segment.

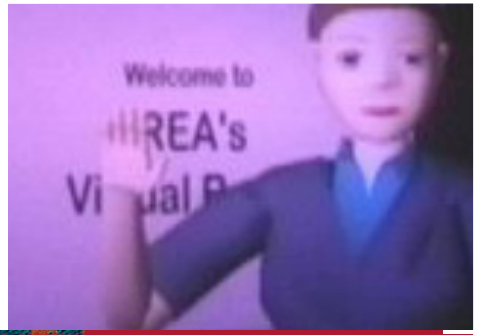
Implement

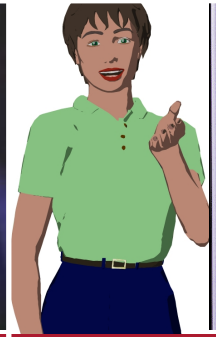



(GA) You just (GT) need [to *type in] { a
[*line] like } ">>This is ...<<", and (GA)
the actor is (GT) able to { [*talk] and
[*gesture] by } itself!
TAGGING
I don't know if this is a good thing or a bad thing.
I worry that virtual actors could squeeze real actors out of business.
00:00:17:20
from
SIGGRAPH",
and
Towards
Eyebrows
Beat
Gesture

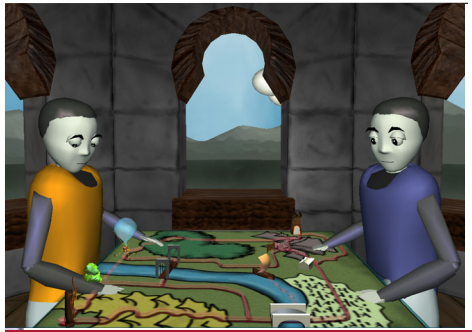





Welcome to
REA's




7:02 pm
All



Evaluate

Total Interactions:
250+ sessions

Current Dataset:
120 sessions (totally 10+ hours)

- Mean duration = approx. 5 min.
- SD = 69.00 seconds

SARA, *00:13.7*, I'm so glad to have the opportunity to meet you in person, *00:16.7*

User, *00:18.7*, hi how are you [doing, *00:19.2*

SARA, *00:19.3*, [I'm] in Davos to assist attendees but to be honest I am a little anxious about it it's a big deal to be invited to Davos is it your first time here, *00:28.7*

User, *00:29.1*, it is and I'm a little anxious too, *00:31.4*

SARA, *00:32.7*, I found there's a lot to take in I hope that it isn't overwhelming you, *00:37.2*

User, *00:37.9*, it is overwhelming a little bit, *00:39.8*

Summary of Evaluation Metrics

Interaction

- Total Time
- System Speaking Time
- User Speaking Time
- System's Response Time
- User's Response Time
- System's Interruptions
- User's Interruptions

Verbal

- System's Intentions
- User's Intentions
- System's Conversational Strategies
- User's Conversational Strategies

Task

- Session Rec. Acceptance
- Person Rec. Acceptance

Interpersonal

- Rapport Score
- Mutual Attentiveness
- Coordination
- Positivity

Goal of *Socially Aware Systems*

Development of a bot that manages ***interpersonal rapport*** (relationship strength) with users over interactions across time, as well as managing propositional and interactional goals, *in order to improve task performance.*

Automatically recognize rapport-managing conversational strategies from *verbal, visual and vocal* modalities of speaker and interlocutor, both within the individual and in the dyad.

Goal of Building Socially Aware Systems

Theoretical: Understand the nature of rapport in greater detail, by correlating with associated observable verbal (conversational strategies, vocal (voice quality) and visual (non-verbal) cues

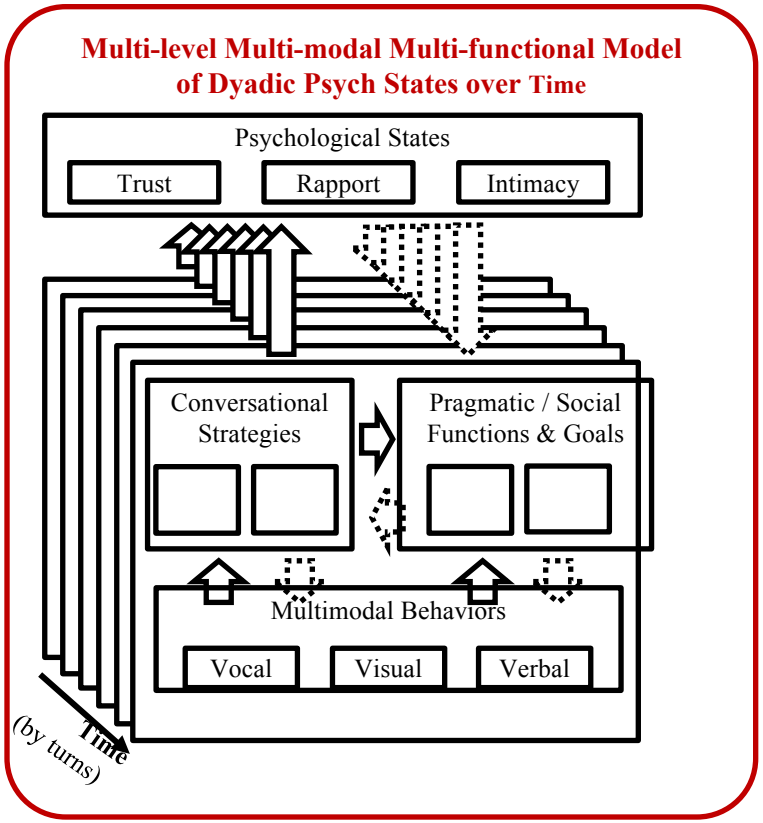
Methodological: Leverage this understanding to automatically recognize rapport-building strategies by leveraging and developing statistical machine learning techniques

Ineffective Conversation
(don't do this with agents)

Intimate Conversation
(don't do this with agents)

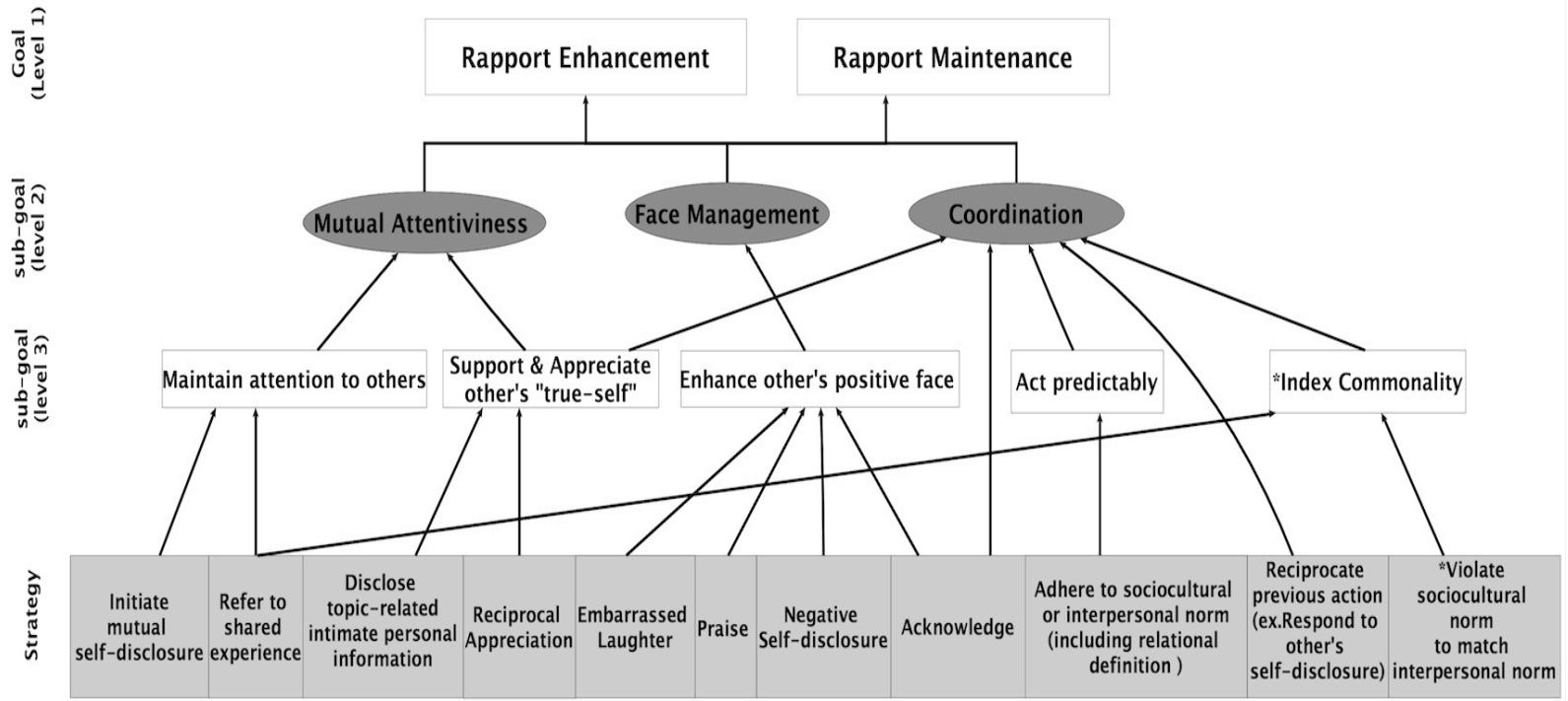
Agent Model of Rapport must be:

1. Dyadic,
2. Multi-level: differentiate between observable signals & underlying psychological states,
3. Sensitive to effect of *time*
4. Cross-Modal



with L.P. Morency, 2015

Data- &Theory-Driven Model of Rapport Management

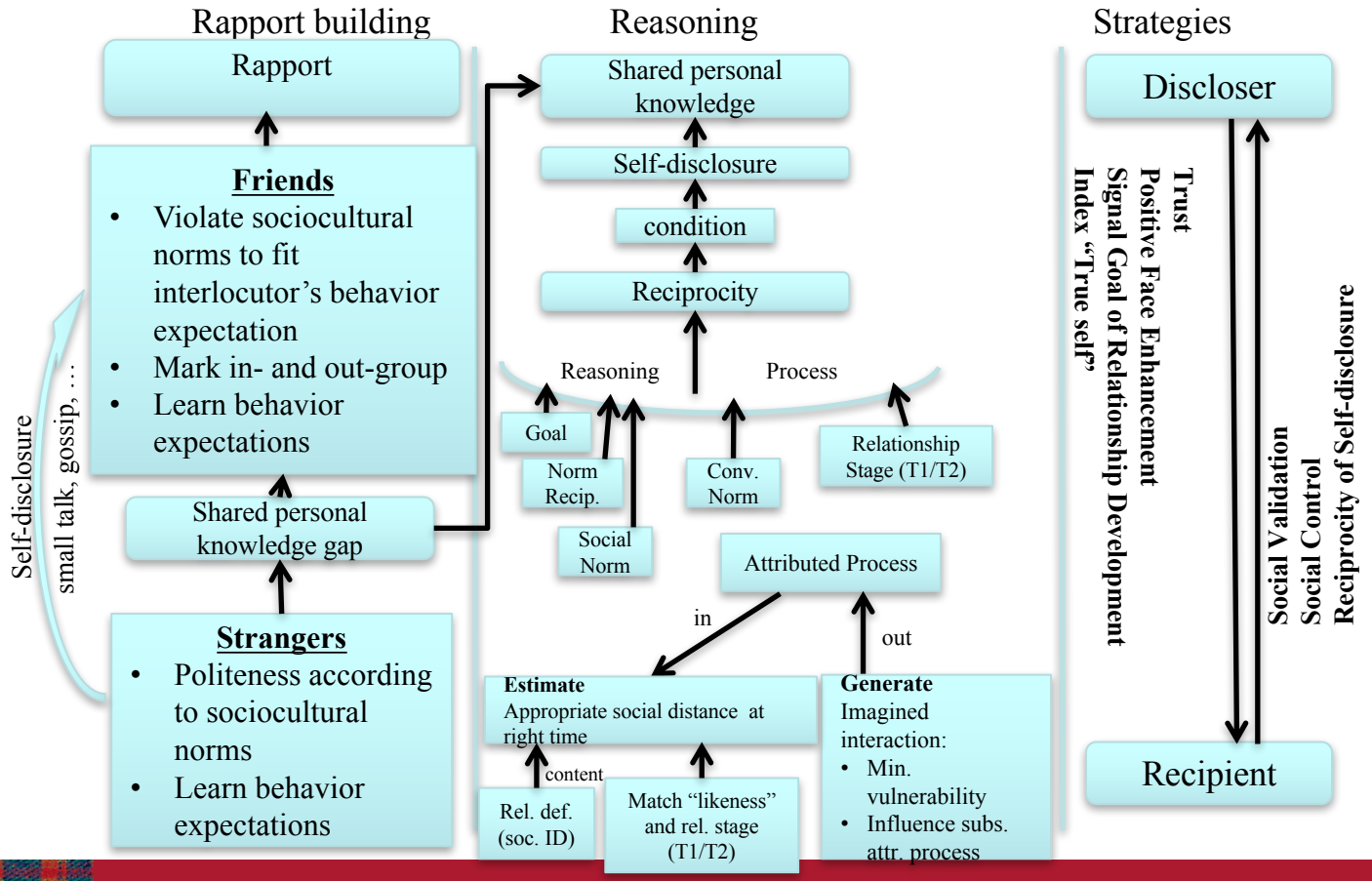


With Ran Zhao

Conversational Strategies

Conversational Strategies	Examples
VSN (Violation of Social Norm)	*"man you take forever to write"*
SD (Self Disclosure)	*"I hate math"*
PR (Praise)	*"well done"*
SE (Reference to Shared Experience)	*"I shared m&m's with you last time"*
BC (Back Channel)	*"yup"*
QE (Question Eliciting SD)	*"are you an atheist"*

Data- &Theory-Driven *process* model

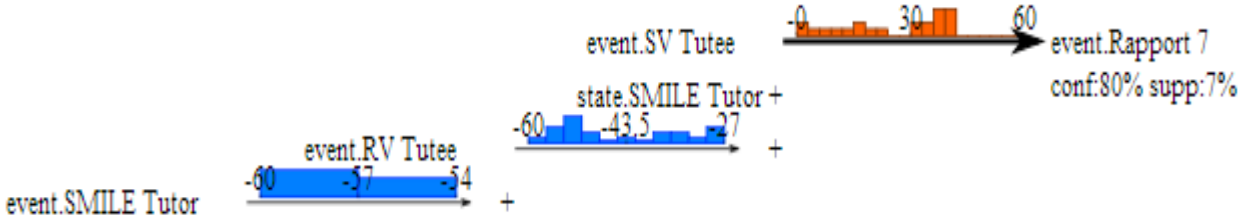


Data-Driven:
Temporal association rules

Form of temporal rules

"If event A happens at time t, there is 50% chance of event B happening between time t+3 to t+5"

Temporal association rules: Friends



Example: Friend in high rapport

Tutor: Sweeney you can't do that, that's the whole point *{smile}* [Violation of Social Norm]

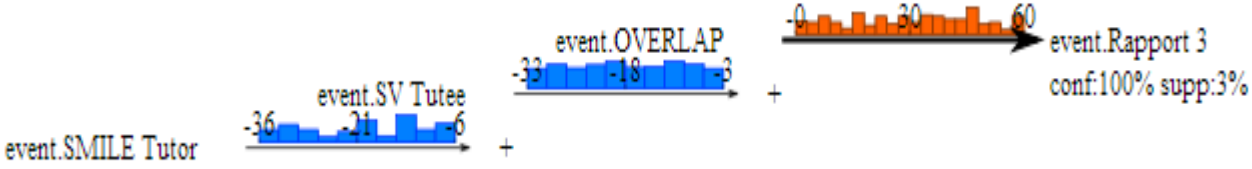
Tutee: I hate you. I'll probably never never do that [Reciprocate Social Norm Violation]

Tutor: Sweeney that's why I'm tutoring you *{smile}*

Tutee: You're so oh my gosh *{smile}*.We never did that ever [Violation of Social Norm]

Tutor: *{smile}* What'd you say?

Temporal association rules: Strangers



Example: Stranger in low rapport

Tutee: divide oh this is so hard let me guess: 11[Negative Self-Disclosure]

Tutor: you know

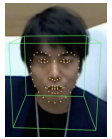
Tutee: 6

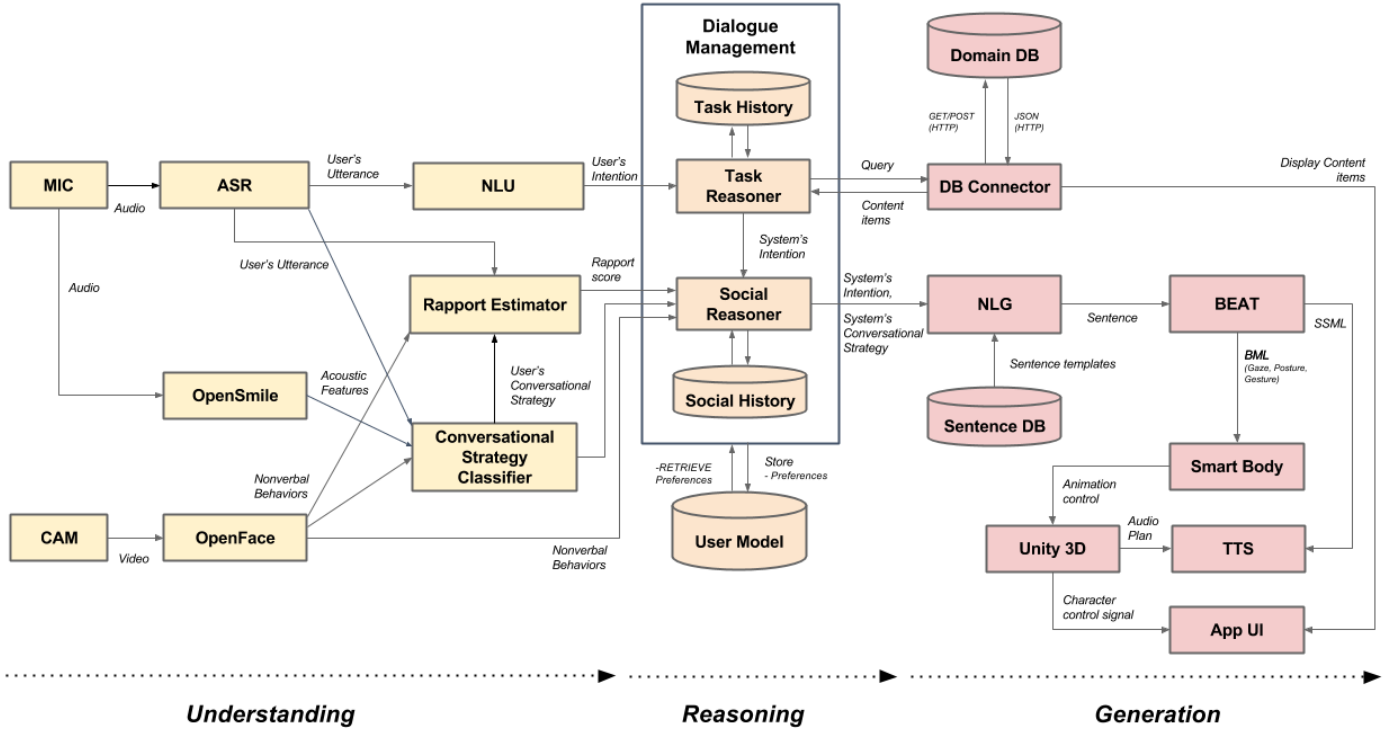
Tutor: next problem is is exactly the same *{smile}*: over 11 equals, 11 x over 11

Tutee: I don't need your help; [Violation of Social Norm]

Tutor: *{Overlap}* That is seriously like exactly the same.

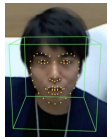
Socially-Aware Agent Architecture

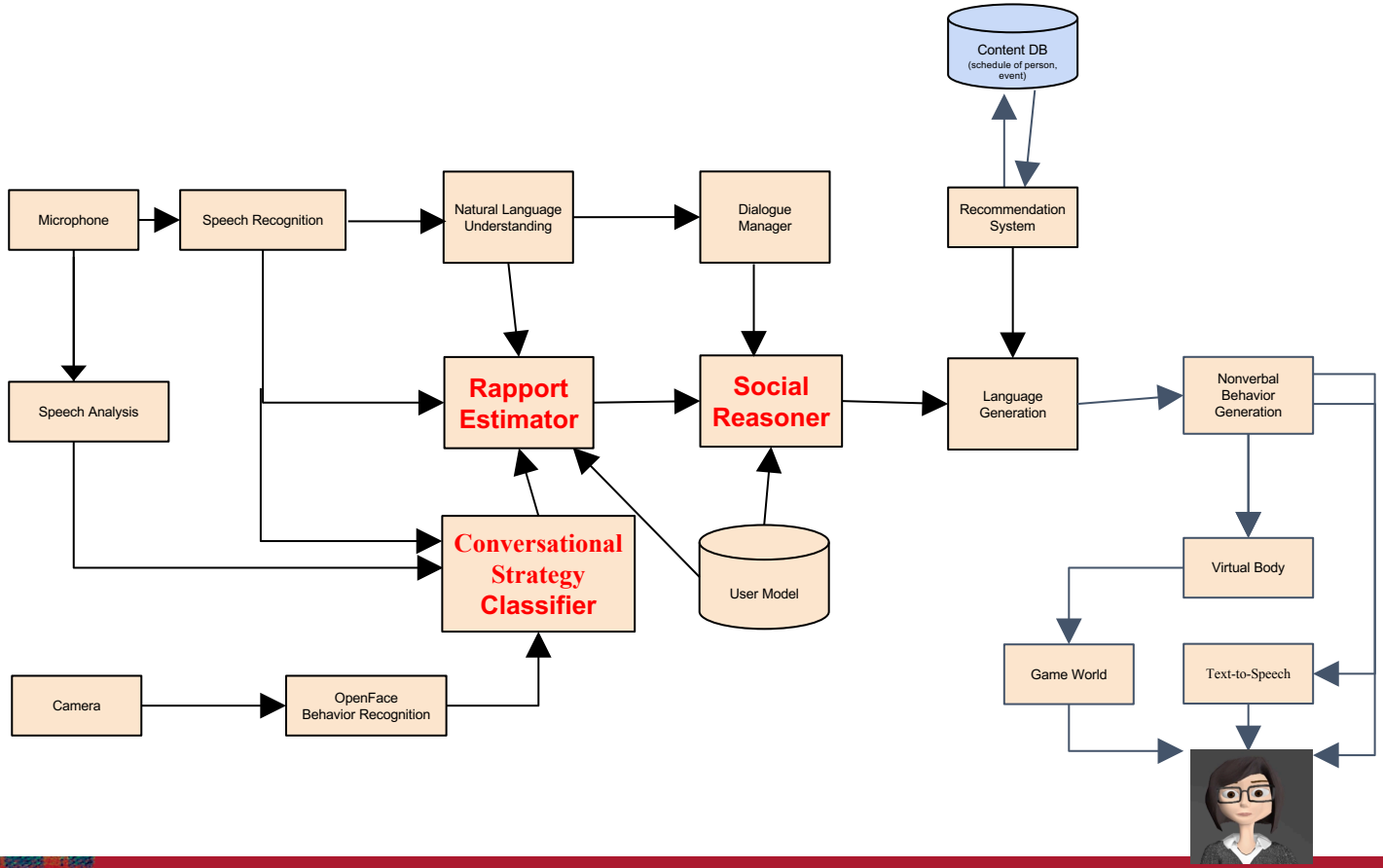




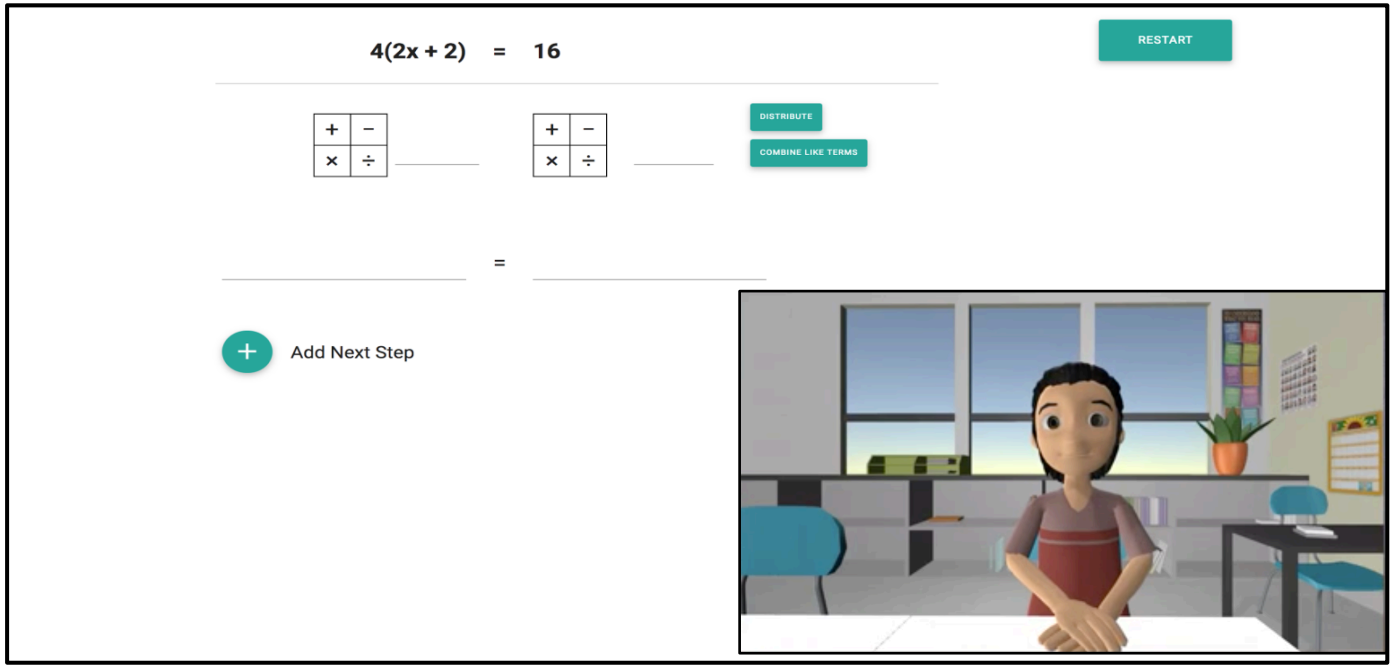
Matsuyama et al., SIGDIAL 2016

Socially-Aware Agent Architecture

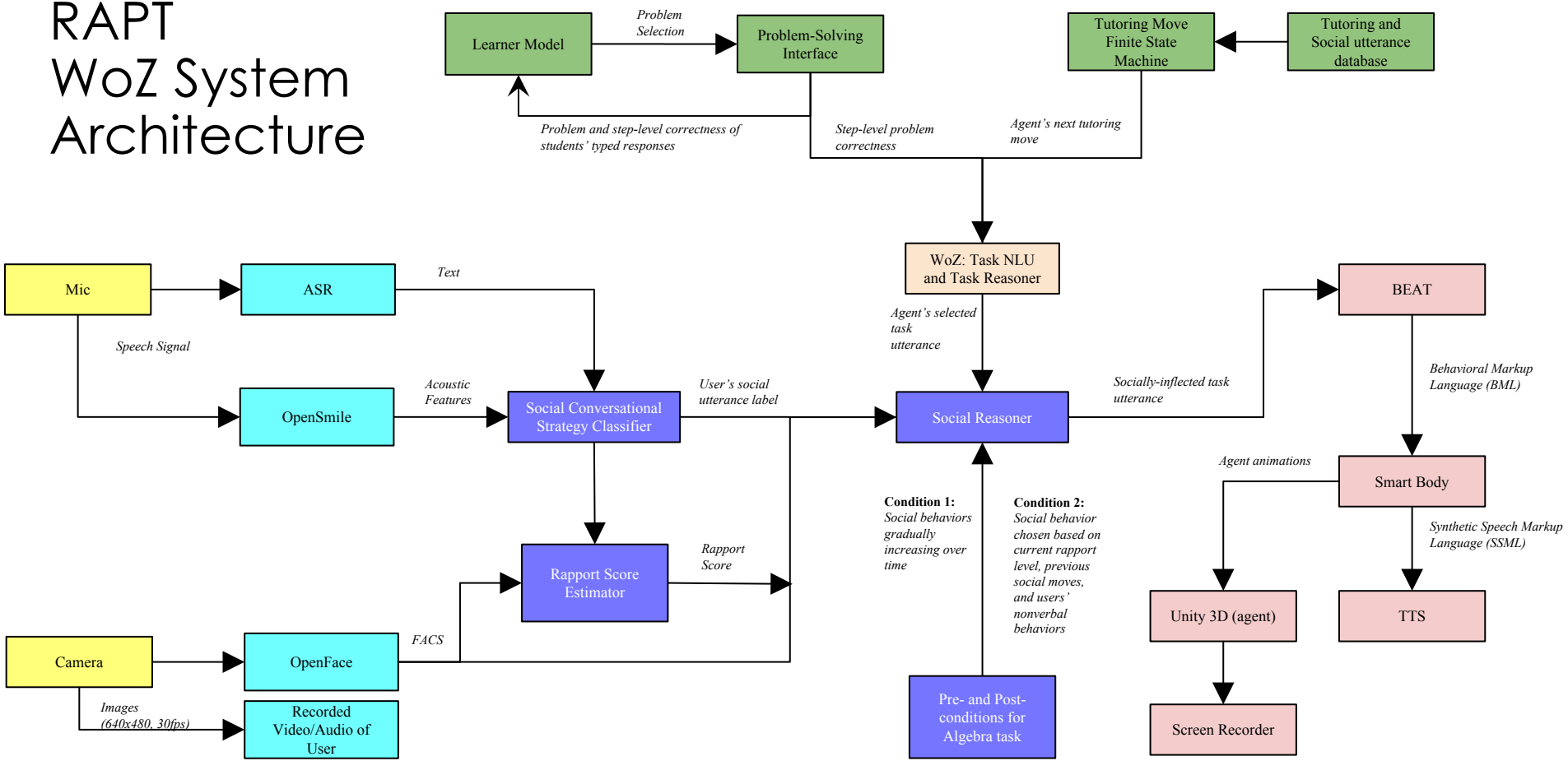




Some Applications: Rapport-Aware Peer Tutor (RAPT)



RAPT WoZ System Architecture



Evaluation: rule-based vs. adaptive

control condition: (fixed heuristics for social dialogue usage)

Praise

Decreasing in frequency [Kumar et al., 2010]

Self-Disclosure

Gradually increasing in frequency and intimacy [Ogan, 2011; Bickmore and Schulman, 2010]

Questions eliciting self-disclosure

Gradually increasing in breadth and depth of topics [Altman and Taylor, 1973]

References to shared experiences

Gradually increasing frequency [Kumar et al., 2010; Cassell and Bickmore, 2003]

Violation of social norms

Use only after a given threshold in number of turns or elapsed time [Ogan et al., 2012]

Indirectness

Decreasing in frequency [Madaio et al., 2017]

Evaluation: rule-based vs. adaptive

Experimental condition: (adaptive usage of social dialogue)

Based on:

Current rapport state

Changes in rapport state (increasing, decreasing, maintaining)

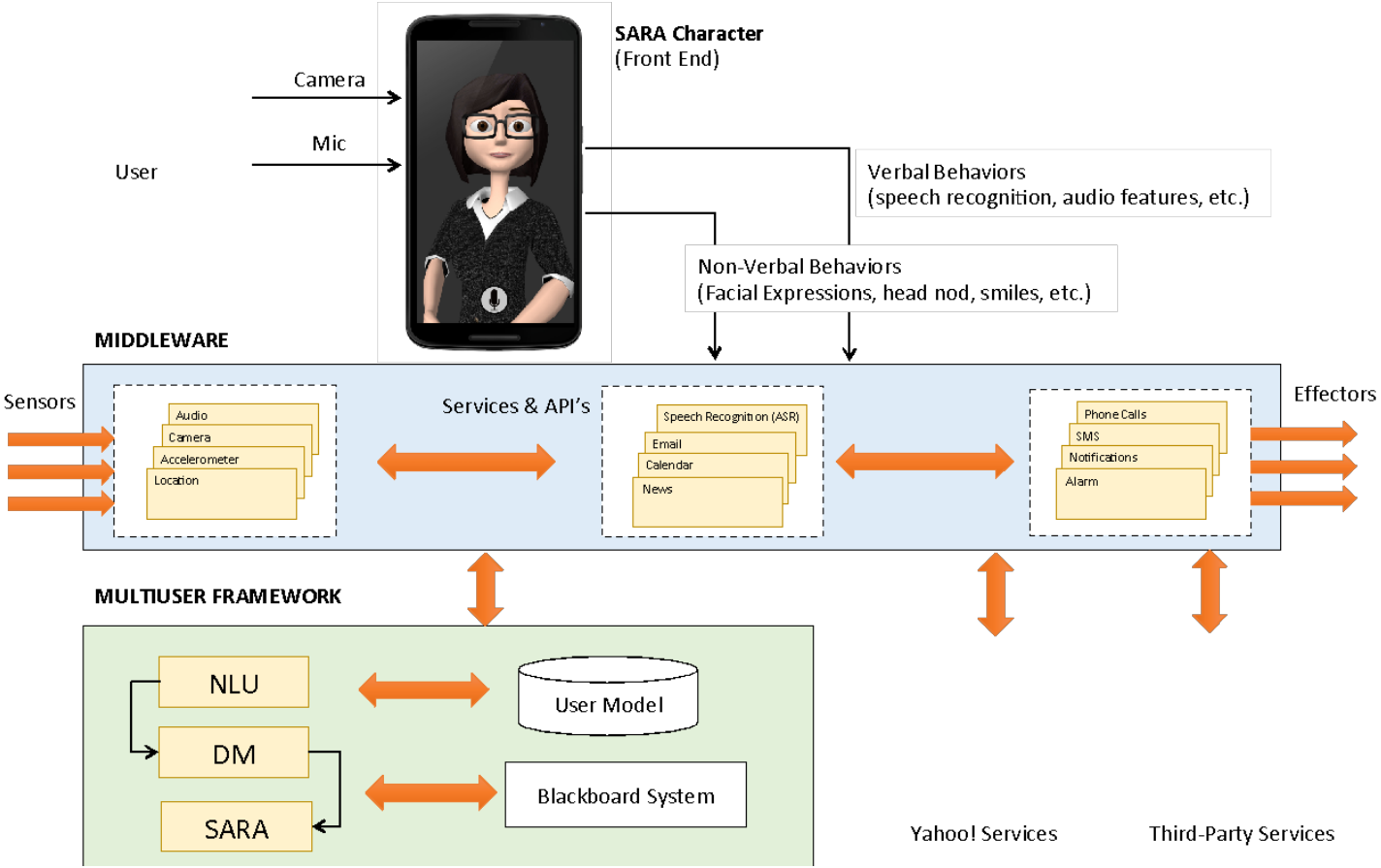
User's social behavior (self-disclosure, violation of social norms, etc)

Agent's previous social behavior

User's nonverbal behavior (smiling, nodding, gaze patterns)

Agent's previous tutoring behaviors (feedback, questions, explanations, etc)

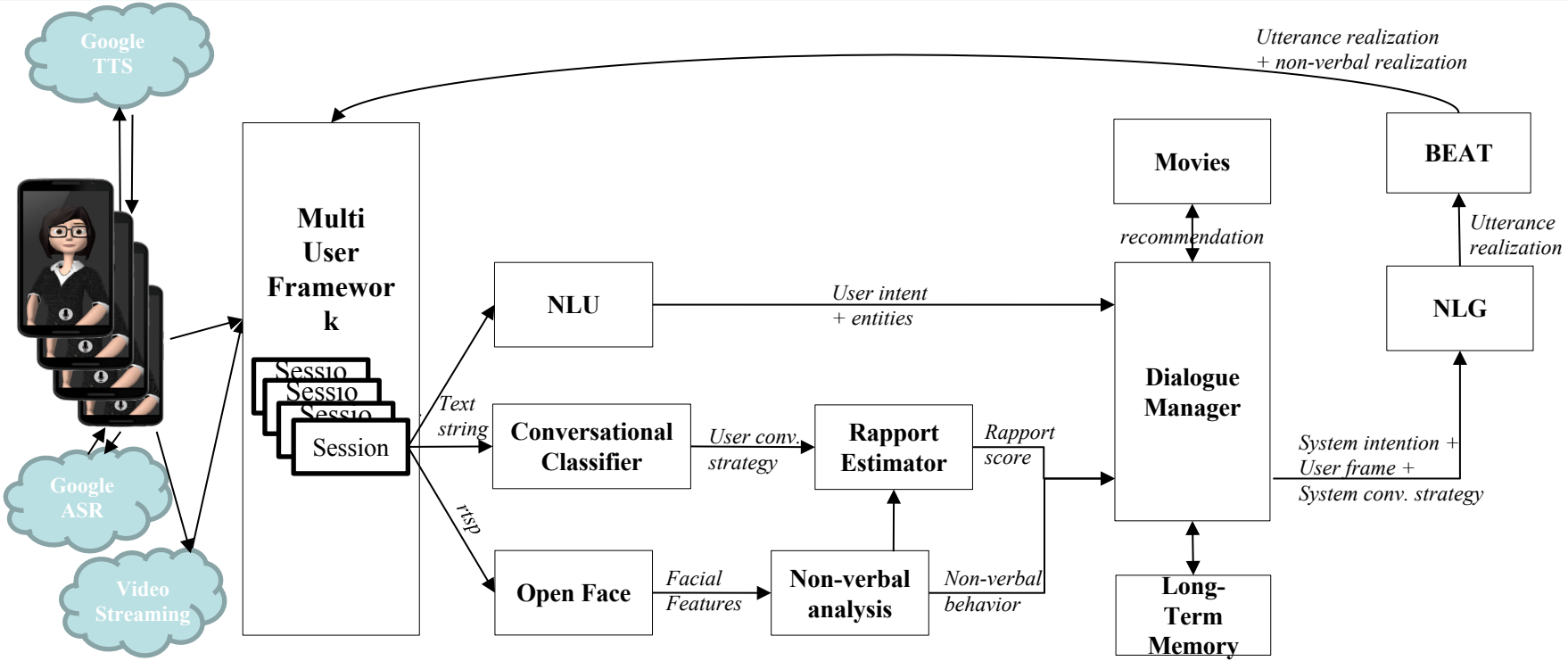
Application: Mobile front-end to apps



with Yahoo InMind Team

Integrated InMind Dialog Architecture

General architecture



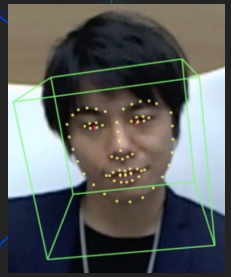
SARA: Socially Aware Robot Assistant

WHAT SARA UNDERSTANDS

OpenFace

CAMERA FEED LIVE

HEAD TRACKING...

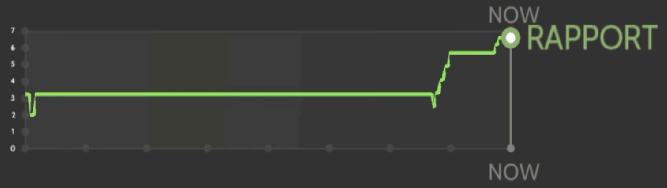


Smile NO

Eye Gaze NO

Head Nod NO

User-Sara Rapport 6 / 7



User Conversation Strategy

SELF DISCLOSURE (SD) 88%

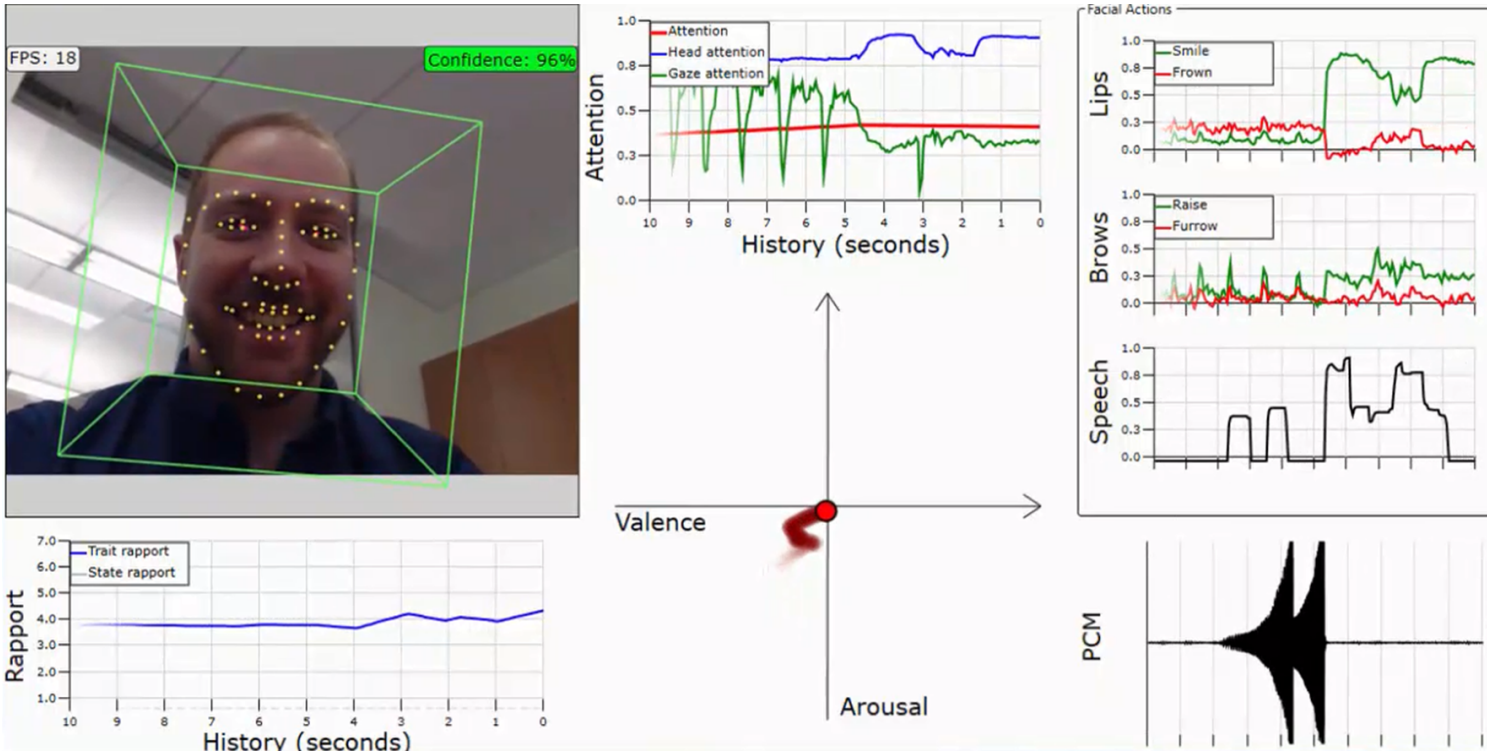
SHARED EXPERIENCE (SE) 72%

PRAISE (PR) 78%

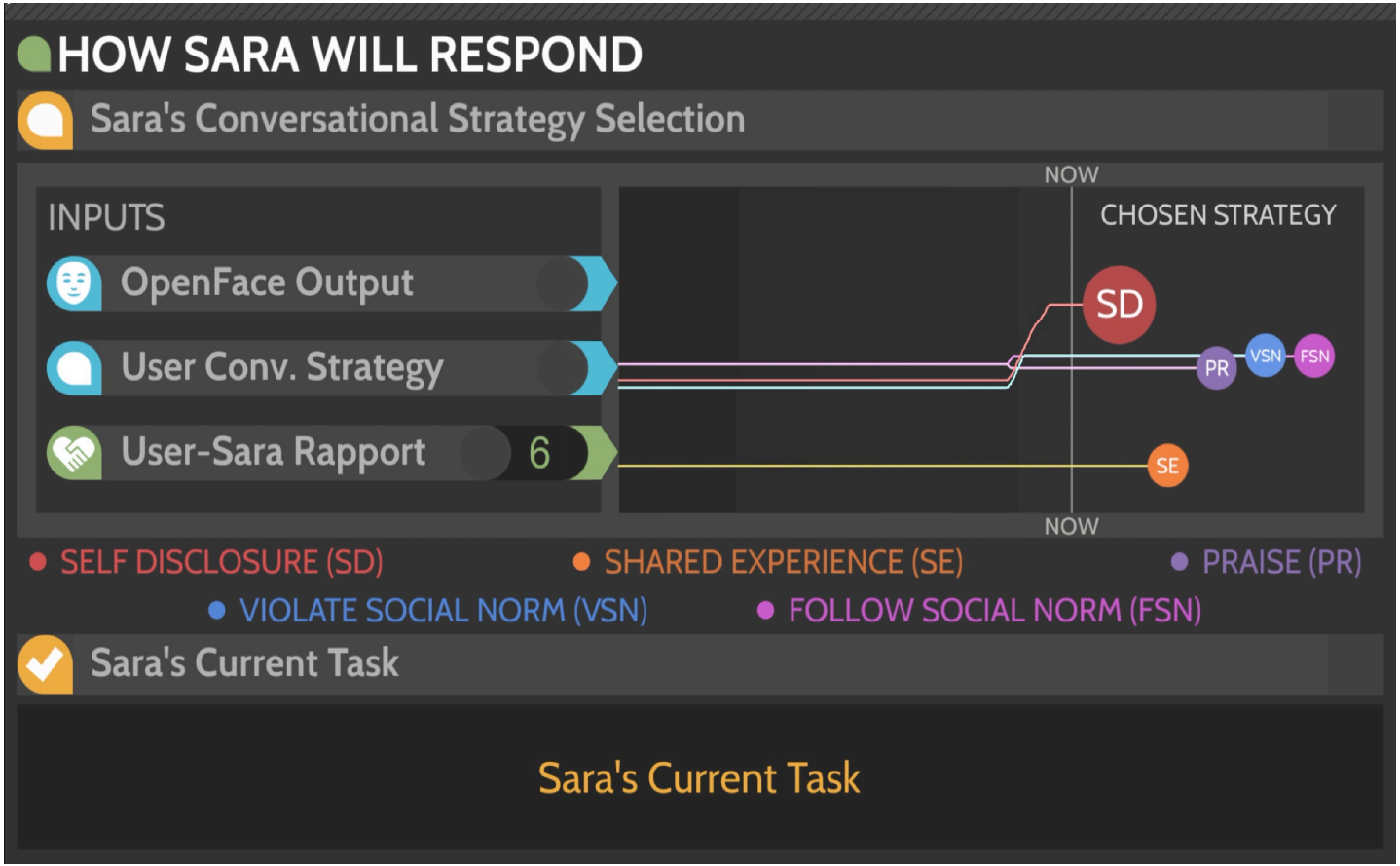
VIOLATE SOCIAL NORM (VSN) 50%

FOLLOW SOCIAL NORM (FSN) 50%

Tracking Facial Movements



OpenFace: L.P. Morency


HOW SARA WILL RESPOND
Sara's Conversational Strategy Selection
INPUTS
OpenFace Output
User Conv. Strategy
User-Sara Rapport
6
NOW
CHOSEN STRATEGY
SD
PR
VSN
FSN
SE
NOW
SELF DISCLOSURE (SD)
SHARED EXPERIENCE (SE)
PRAISE (PR)
VIOLATE SOCIAL NORM (VSN)
FOLLOW SOCIAL NORM (FSN)
Sara's Current Task
Sara's Current Task

WHAT SARA SAYS

Sara's Words and Body Language

part	let	me look	this
	GESTURE ACTIVATE NOD YES		

up	one minute
GESTURE ACTIVATE NOD YES SMILE YES	

this is my favorite part let me look this up one minute

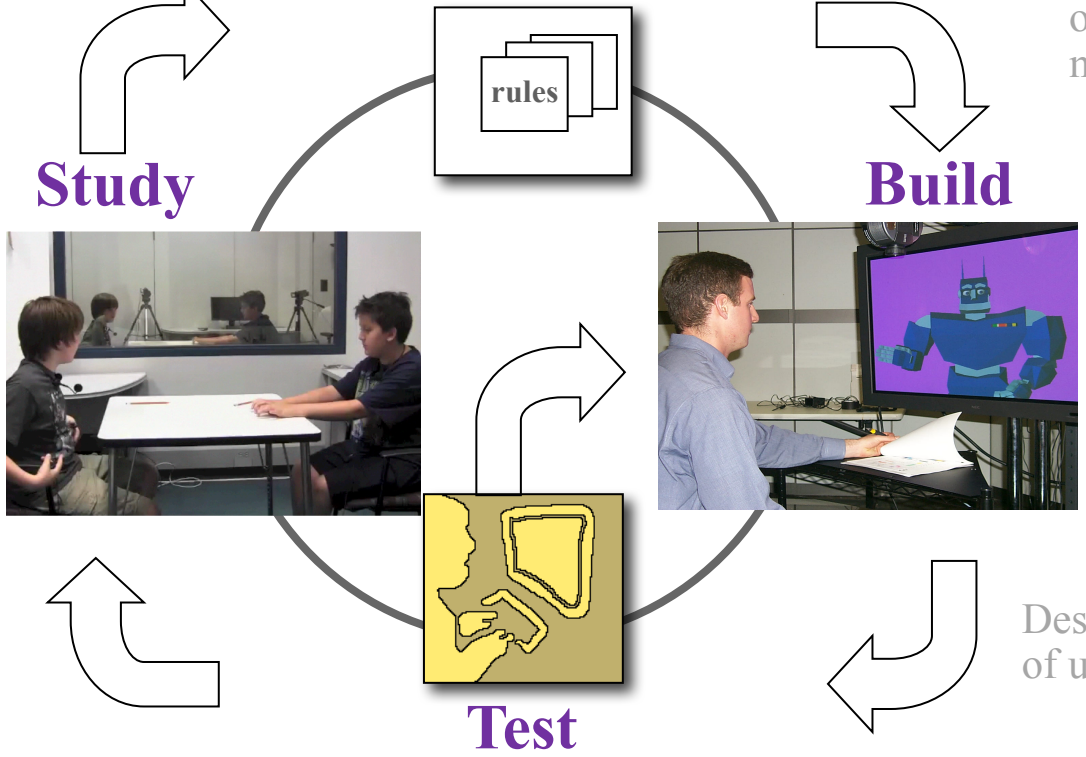
SARA: Socially-Aware Robot Assistant at Davos

Methodology

Build formal models

Theorize & Model

Implement system on the basis of model



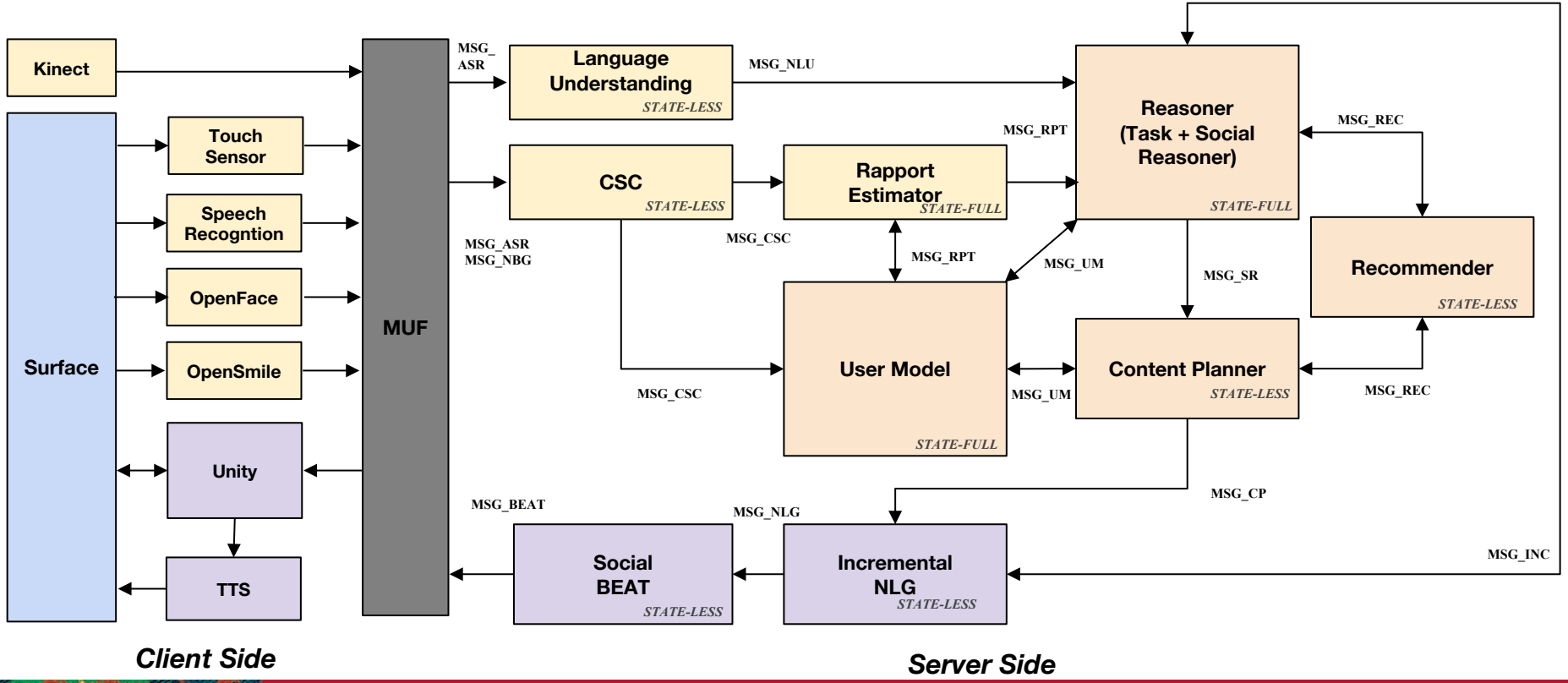
Start here

Collect Natural data

Realize gaps in understanding

Design evaluation of use

New SARA Framework



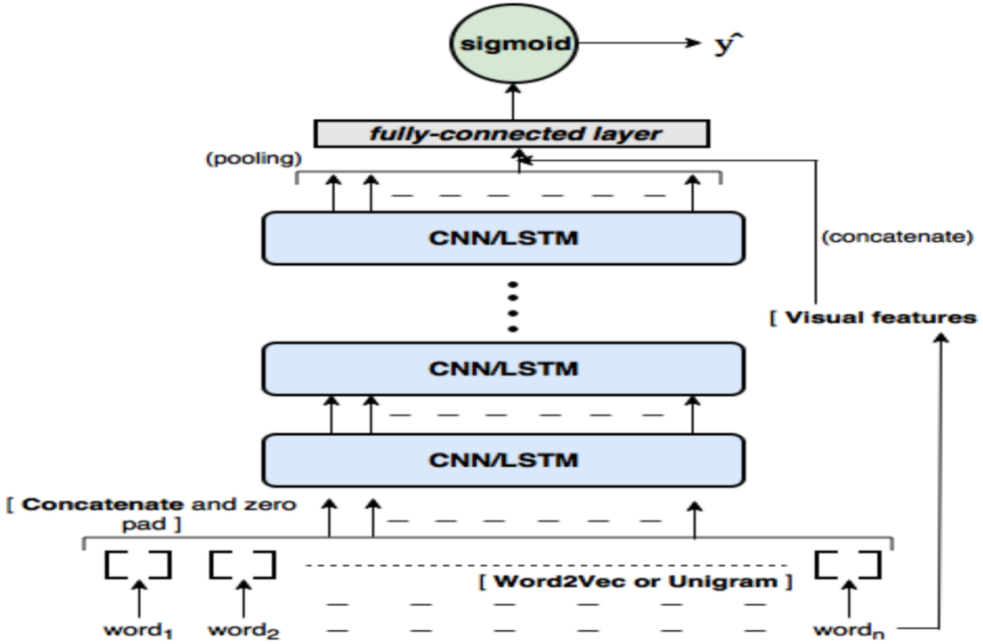
Indirectness Strategy Classifier

- Corpus
  - RAPT 2013 : indirectness annotation of peer-tutoring corpus
  - ConLL 2010 shared task on **uncertainty** detection
    - Wikipedia dataset (Wikipedia articles)
    - BioScope dataset (abstracts and articles from biomedical literature)

Code	Definition	Example	Distribution
Apology	Apologies used to soften direct speech acts	Sorry, its negative 2.	7.7%
Qualifiers	Qualifying words for reducing intensity or certainty	You just add 5 to both sides.	66.1%
Extenders	Indicating uncertainty by referring to vague categories	You have to multiply and stuff.	3.6%
Subjectivizer	Making an utterance seem more subjective to reduce intensity	I think you divide by 3 here.	22.6%

IN (Indirect Delivery) *"so I think what I'm gonna do is make that 15 minus 3 a 12"*

Architecture: Indirectness Classifier



Pranav Goel, Yoichi Matsuyama, Michael Madaio & Justine Cassell, **"I think it might help if we multiply, and not add" : Detecting Indirectness in Conversation**, International Workshop on Spoken Dialog System Technology (IWSDS 2018). – to appear

Microsoft Surface Hub

SARA Receptionist

With thanks for generous funding and support to the following organizations, and to the students and staff of the ArticuLab