Embodied Conversational Agent

Catherine Pelachaud

CNRS – ISIR – Sorbonne Université

catherine.pelachaud@upmc.fr

dépasser les frontières

INSTITUT DES SYSTÈMES INTELLIGENTS ET DE ROBOTIQUE

Outline

1. Introduction
2. Nonverbal behaviors
3. ECA Architecture
4. Communicative gestures
5. Turn-taking system
6. Expressions of emotions
7. Evaluation studies
8. Believability and Uncanny Valley

Embodied Conversational Agents : an introduction

GRETA - CNRS

ECA : human-machine interaction

Virtual characters that are :

- interactive
- expressive verbally and nonverbally
- autonomous

Autonomous : they can plan what to say, know when to start a conversation, when to answer and when to take the conversation turn → complex process

Embodied Conversational Agents : aims

Create 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

Embodied Conversational Agents: an introduction

Required capabilities for an ECA :

Perception

- Pay attention to its environment (object, user, context…)

Interaction

- be an active speaker and an active listener
- *emits* and *perceives* continuously signals from the other
- *reacts* emotionally and be *socially* awareness

Generation

- synchronized visual and acoustic behaviors.
- emotional behaviors

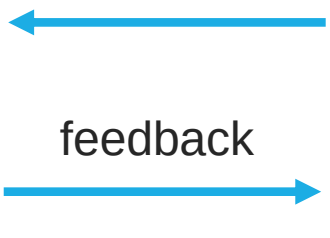
Embodied Conversational Agents: an introduction

General overview

- cognitive processes
- reactive processes

ECA's decision

perception



feedback

Embodied Conversational Agents: an introduction

General overview : example

- cognitive processes:
  « human shows me her friendship »
- reactive processes :
  mirror smile

ECA's decision:
smile

ECA perceives human is smiling

ECA smiles to human

Embodied Conversational Agents: Application

Serious games:

- Training: job interview, intercultural communication, announcing bad news
- Social ability: bullying, shyness

Health & wellbeing

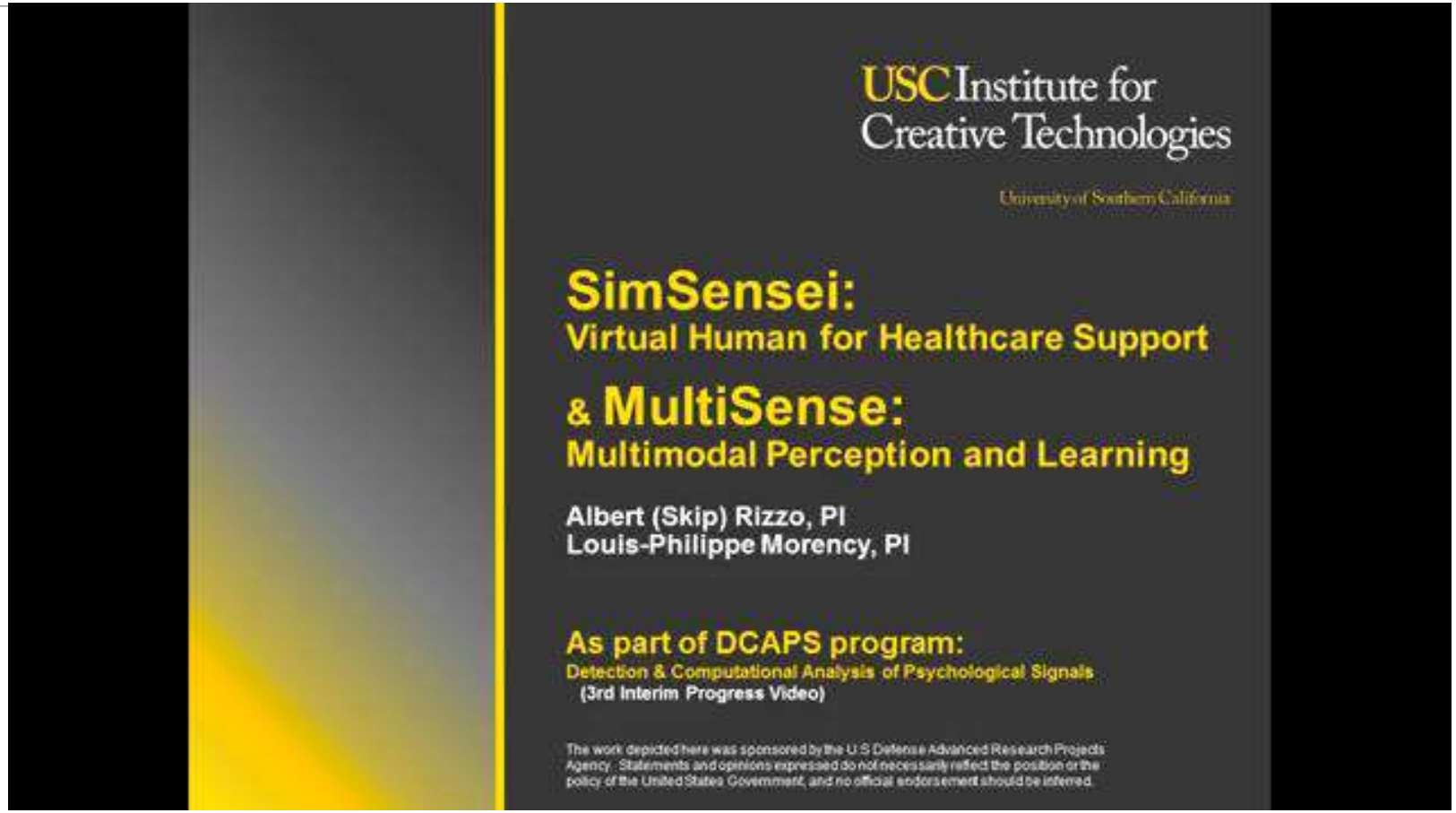
- Depression
- Post-traumatic stress disorder (PTSD)
- Autism

Education: virtual tutors

Rehabilitation: speech therapy

Companions

Embodied Conversational Agents : examples - SimSenSei



Nonverbal Behaviors

Communication

Definition of communication (Allwood, 06)

- not simply a transfer function
- sharing anything between arbitrary entities where all entities are active, interacting with each other and within a social and interelational context

Involve different processes: sense, perceive, adapt, generate, plan

Communication involves:

- Verbal
- Nonverbal

Definition

Nonverbal behavior

- Also called body language

Burgoon, 1991: “those behaviors other than words themselves that form a socially shared coding system; i.e., they are typically sent with intent, typically interpreted as intentional, used with regularity among members of a speech community, and have consensually recognizable interpretations.”

- Burgoon: ‘message orientation’
- Has a communicative function
- Requires: Sender and Receiver
- They are dynamic
- May be intentional or not (scratching one’s head because it hitches is not a NVB)
- NVB have different meanings; context dependent
- NVB are socially shared rather than to correspond to idiosyncratic behavior patterns

Communicative behaviors

Example from Marsella

Example from Calbris

Functions of NVB

- Communicative intent
  - Production and reception of messages: aid to create a message and to understand it
  - Carry propositional content
  - Manage interaction: begining and ending of conversation, turn-taking…
  - Express emotion: display of felt, masked, fake emotion
- Internal regulation
  - Attention
  - Cognitive load
  - Emotion regulation
  - Expression of felt emotion
- External regulation
  - Impression formation and management
  - Social attitude
    - Dominant, friendly
  - Emotion
  - Interpersonal relationship
    - Show rapport, support, confort…
    - Conflict management
  - Expression of real or desired self, linked to
    - personality, socio-economical status, gender orientation, age, culture…
    - expectancy to conform to some social norms
  - Deceive other

NVB channels

Involves:

- Facial expression
- Gesture
- Gaze
- Posture
- Head movement
- Voice
- Touch
- …

NVB Role

Links between Verbal and NonVerbal Behavior (Ekman & Friensen, 1969)

- Redundancy (diplucate verbal message)
- Substitution (replace verbal message)
- Complementation (add information to verbal message)
- Emphasis (underline verbal message)
- Contradiction (provide opposite meaning to verbal message)

But NVB is not secondary to VB

Mapping Mind - Body

Not a single mapping

- A behavior can have different functions
  - Raised eyebrow may signal surprise, emphasis, question mark, suggestion…
  - Smile may express happiness, be a polite greeting, be a backchannel signal…
- A function can be expressed by different behaviors
  - Greeting: word, smile, hand wave, eyebrow flashing…

Interpretation of nonverbal behavior:

- Context dependent
- Interactant

Mapping Mind - Body

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

Marsella & Gratch

Smiles

Important social signals

Bear many meanings: more than 50 functions (Ekman; Hess)

- Positive/negative emotion (joy, embarrassment)
- Affiliation (social bonding, rapport)
- Attitude (dominance, irony)
- Greeting
- Politeness
- …
- Not all related to emotions

Differences in expression:

- Change in facial and body cues
- Symmetry
- Dynamic

Joy

Nervous/
Embarrased

Polite

(Cohn et al)

Ironic smile

Facial feedback hypothesis

One own expression can affect our emotion

- Participants rated a movie while holding a pencil with their
  - Lips → pout
  - Teeth → smile

Change in:

- Perception of stimuli
- Mood

(Strack, Martin, Stepper, 1988)

Similar results with

- Association to smiling avatar
- Manipulation of own voice (more smiling voice)

Embodied cognition

Embodied mental representation

Shared representation in body and mind

Body that perceives and acts on object

Mental representation based on sensori-motor experience

Perception of object: emergence of partial reaction of sensori-motor experience triggered from previous interactions

Emotion, knowledge: embodied in motoric system

Seeing expression of emotion in other, hearing emotional voice or even saying emotional word: trigger expression of emotion in self

- Spontaneous imitation
- May be different from empathy

Imitation

Imitation, mirroring, mimicry, entrainment, alignment

- Smile
- Posture
- Gaze
- Linguistic alignment
- Paralinguistic
- …

Meltzoff&Moore, 1977

Imitation

Linked to

- Empathy: the ability and tendency to share and understand others' internal states (Zaki & Ochsner, 2016)
  - Affective empathy: "experience sharing, or the tendency of perceivers (individuals focusing on someone else) to take on the sensorimotor, visceral, and affective states of targets (individuals on whom perceivers focus" (Zaki & Ochsner, 2016)

    → theory of mind
  - Cognitive empathy: "mentalizing, describes perceivers' explicit reasoning about targets' internal states using lay "theories" about how situations produce internal states (Gopnik & Wellman, 1992)" (Zaki & Ochsner, 2016)

    → people "draw inferences about targets' underlying emotions, intentions, and beliefs" from targets' NVB
  - Prosocial motivation: sharing and understanding states of the others

    → tendency to feel sympathy, show concern, be cooperative

Imitation

Engagement:

- “the process by which two (or more) participants establish, maintain and end their perceived connection during interactions they jointly undertake” (Sidner. et al, 2004)
- “the value that a participant in an interaction attributes to the goal of being together with the other participant(s) and of continuing the interaction” (Poggi, 2007)

Chameleon effect : «nonconscious mimicry of the postures, mannerisms, facial expressions, and other behaviors of one's interaction partners” (Chartrand&Bargh, 1999) → increase of engagement

Building ECAs Systems

Building ECAs Systems

Objective (example) : I want to model an ECA that smiles in appropriate situations

Big steps :

Corpus constitution : how people smile ? When do they smile ?

Model : my agent will display a facial expression x in the situation y

Evaluation :

- my agent is judged as believable by human observers (or not…)
- my agent behaves similarly to a human (or not…)

Methodology:
From Human-Human to Human-Agent

To approach research issues:

1. Study human-human interaction
2. Acquire database
3. Multilevel analysis
4. Build computational model for agent
5. Evaluate model

Building ECAs Systems

- Working group on “ECAs and human-human interaction” Dagstuhl Seminar 2004, from J. Cassell ’s work methodology

Model Building

Model-based Implementation

Data Collection (Corpus)

Model

Refine model

Implemented Model

Filling data gaps

Experimental Design

Evaluation

BUILDING ECAS SYSTEMS : BUILDING A CORPUS

How to Acquire Data

Elaborate video corpus:

- Acted data
- Naturalistic data
- Induced data
- TV shows, films

Different settings:

- Controlled lab situations
- Hidden camera (spontaneous, rare !)

Create animation:

- Motion capture
- From scratch (animators)

Corpora

NoXi

Tardis

Ilhaire

Emilya

Multimodal Repertoire

Multimodal Signals

Need two information to characterize multimodal signals:

- Their meaning
- Their visual action

Need to encode two types of information

- Meaning: communicative function representation - FML
- Signals: behavior representation - BML

Taxonomy of Communicative Functions

A semantic topology based on the information to be conveyed has been defined by I. Poggi:

- Information on the Speaker's Identity
  - Age, culture, gender…
- Information on the World
  - Description of object (iconic)
  - Reference to object (deictic)
- Information on the Speaker's Mind
  - Information about Speaker's belief (degree of certainty, belief relation)
  - Information about Speaker's intention (performative, turn-taking, emphasis)
  - Information about Speaker's affective state
  - Meta-cognitive information about Speaker's mental action

Gesture

Definition: hand and arm movements used to communicate an intention, a belief, an emotion

Synchronized with speech

Speech independent gestures:

- Often occur as a single gesture with no speech
- Consciously made
- Can occur as sequence of gesture: ‘come on in; sit down’
- Emblems: direct verbal translation

Gesture

Speech-related gesture, also called illustrators

meaning tied with speech

Gestures can:

- Relate to speaker's referent
- Indicate speaker's relationship to referent
- Punctuate speech
- Regulate and organize dialogue between interactants

Gesture

Communicative gesture: Semiotic taxonomy (Kendon, McNeill)

- Deictic: indicates a (concrete or abstract) point in space
- Iconic: describes the physical property of item
- Metaphoric: relates to abstract idea
- Beat: rhythms the speech

Non-communicative gesture:

- adaptator

Iconic gesture

Deictic

Metaphoric

Attention

[HF=Forefinger, Orient=FU]

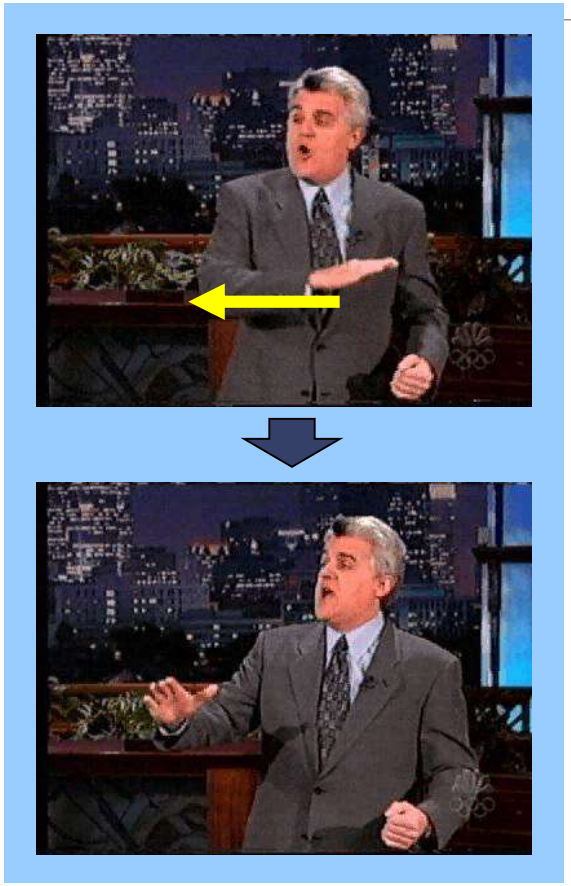
→ (Efron 1941)
→ (Calbris 1990)
→ (Saitz/Cervenka 1972)
...

Progress

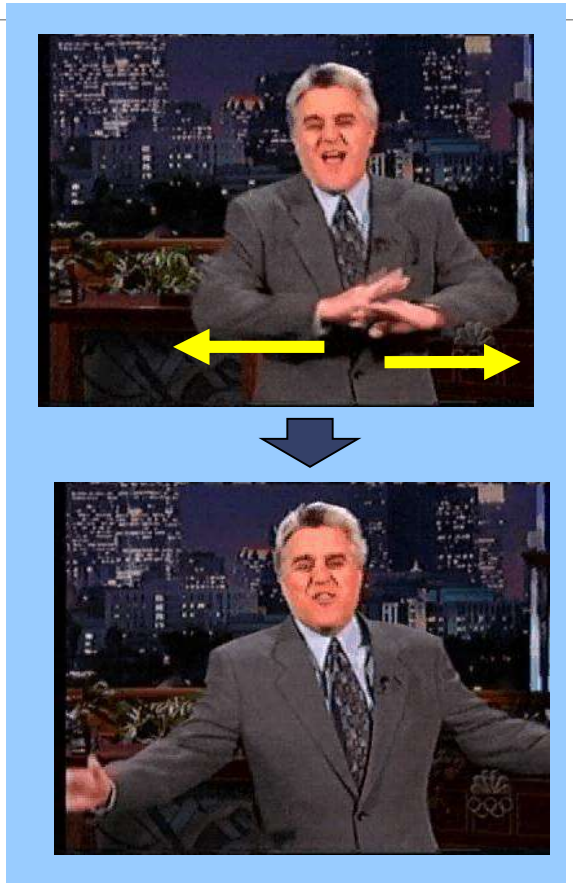
[Move=circ/alt/progressive]

→ (McNeill 1992)
→ (Payrató 1993)
→ (Webb 1997)
...

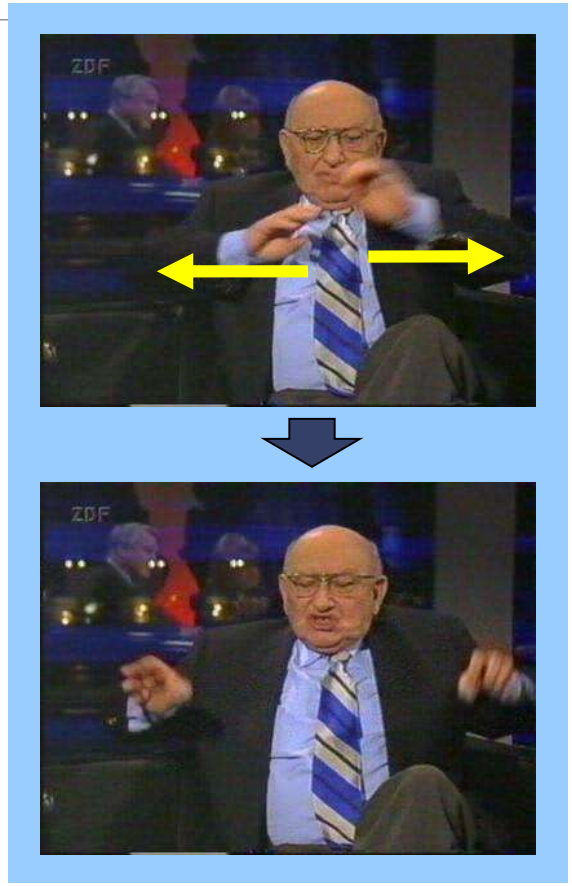
Example: Lexeme “Wipe”



JL, RH



JL, 2H



MR, 2H

Kipp, 07

Gesture Configuration

Arm:

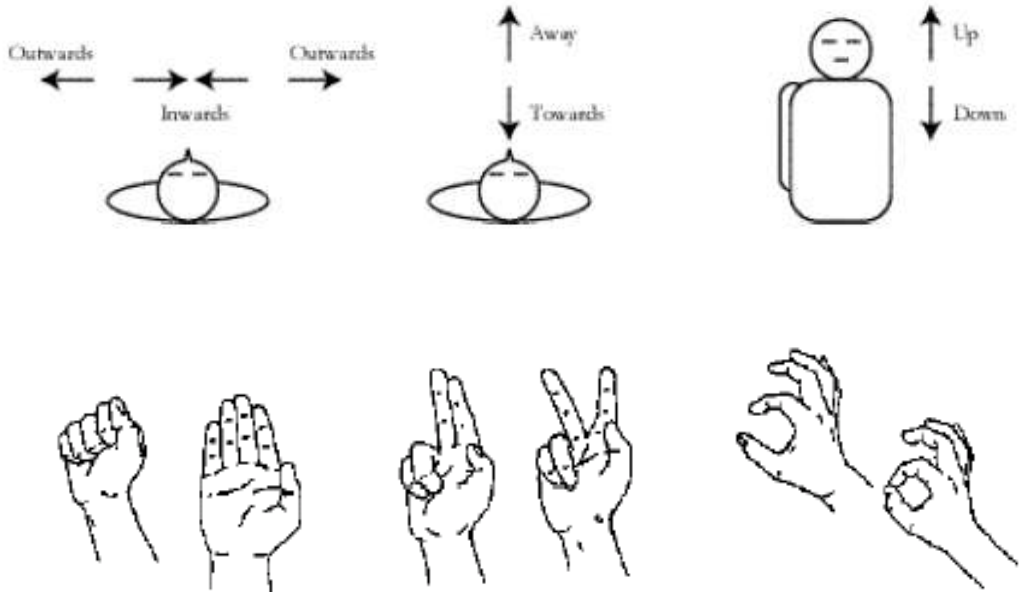
- height
- distance
- radial orientation
- arm swivel

Wrist orientation:

- Palm direction
- Finger direction

Hand shape

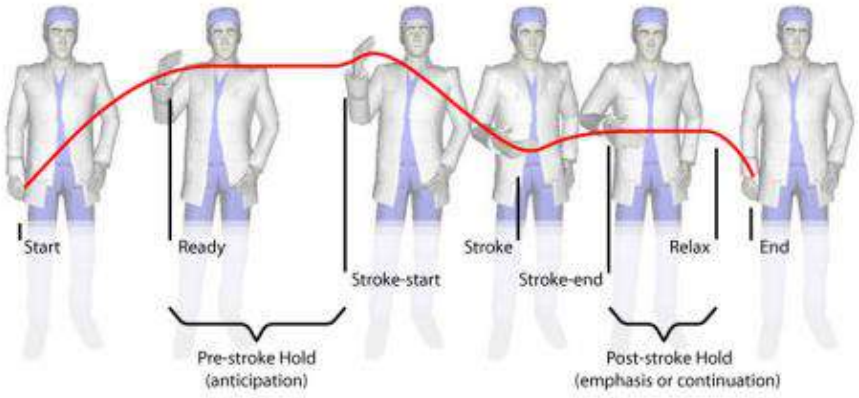
- 4 fingers shape
- Thumb modifier



Gesture

Temporal structure:

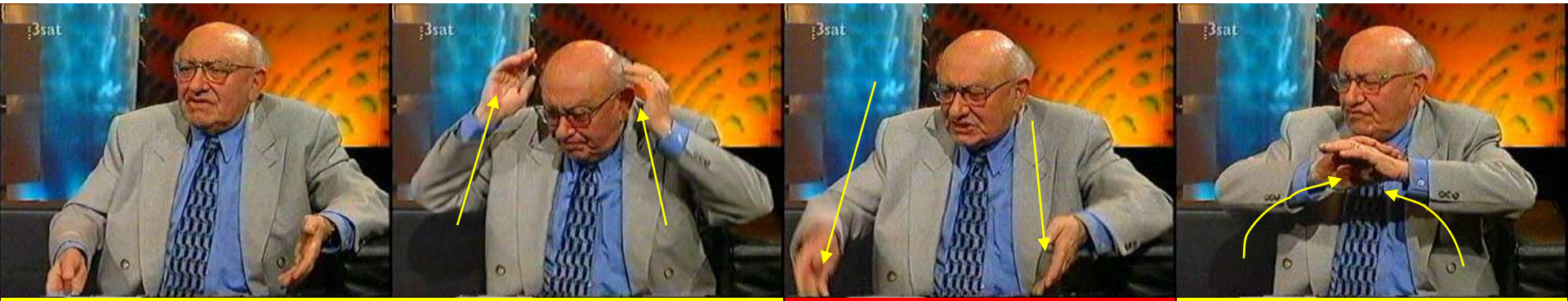
- Gesture phases: preparation, strole, hold, retraction
- Stroke:
  - carry meaning of gesture
  - Often coincide with emphasized syllable



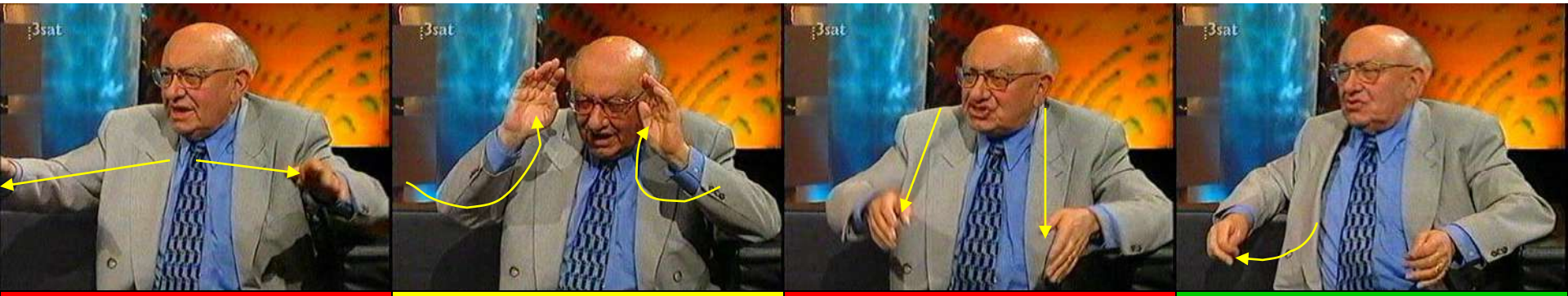
Kopp et. al 2006

- Gesture unit: coarticulated gestures between retraction phases

Temporal Structure



preparation	stroke	preparation
Fling		Wipe



stroke	preparation	stroke	retraction
	Dismiss		

Facial Coding Scheme

Facial Action Coding System FACS: Paul Ekman and Wallace Friesen, 1978 – 2002

FACS allows expert coders to manually measure facial expressions by breaking them down into component movements of individual facial muscles (Action Unit).

Reference : FACS manual (500 pages)

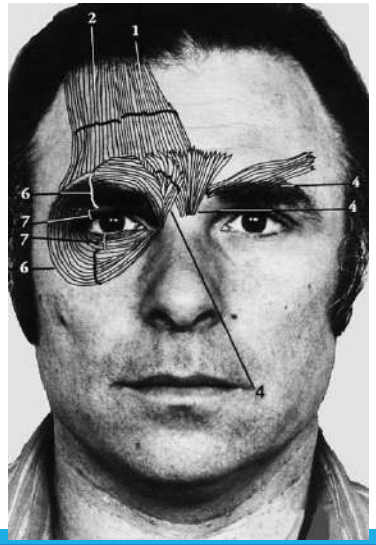
FACS

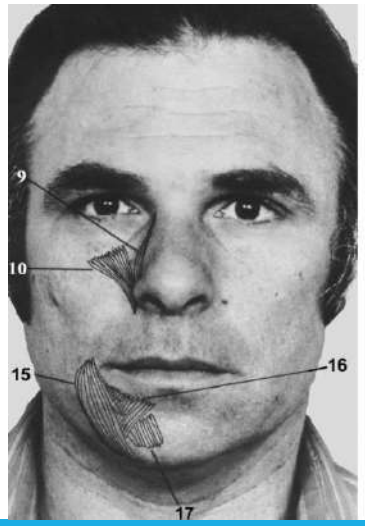
Upper face action units (brows, blink…)

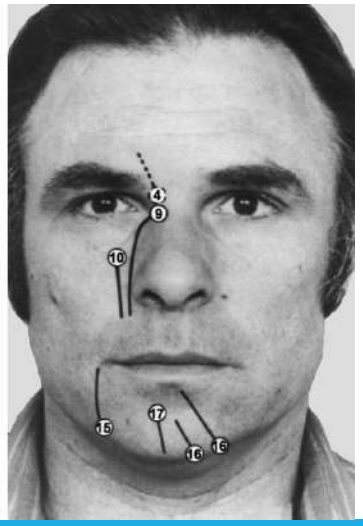
Lower face action units (lips, nasolabial)

Head and eye positions (eye turn left, head up…)

Miscellaneous actions (jaw, bite, tongue show…)



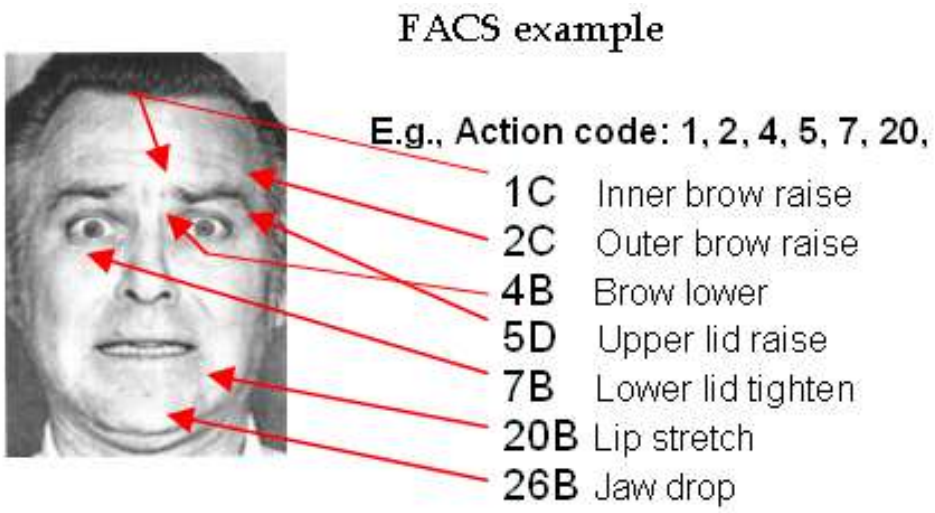




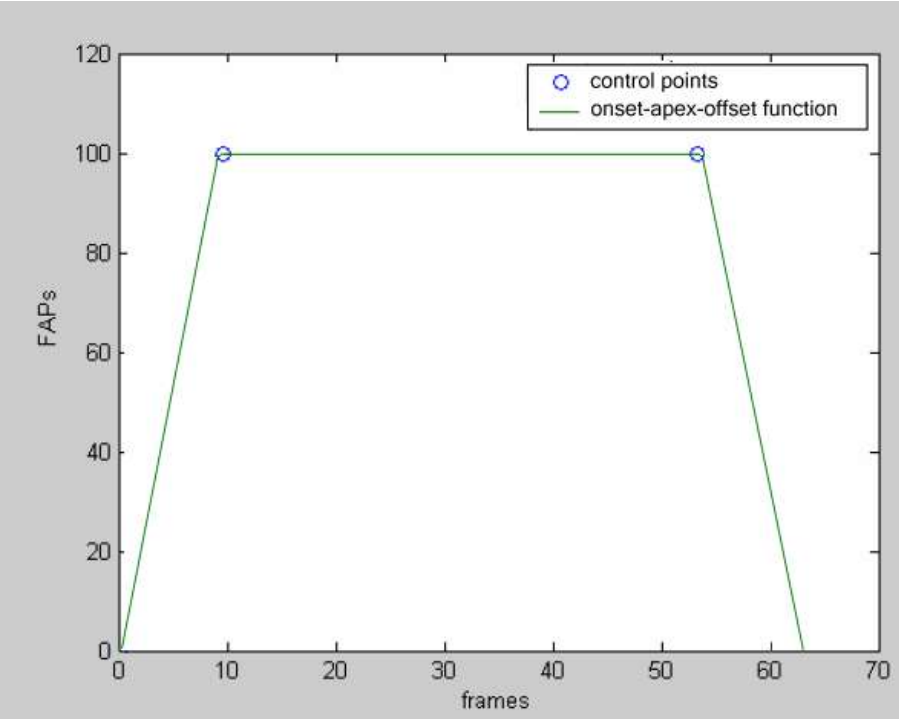
FACS

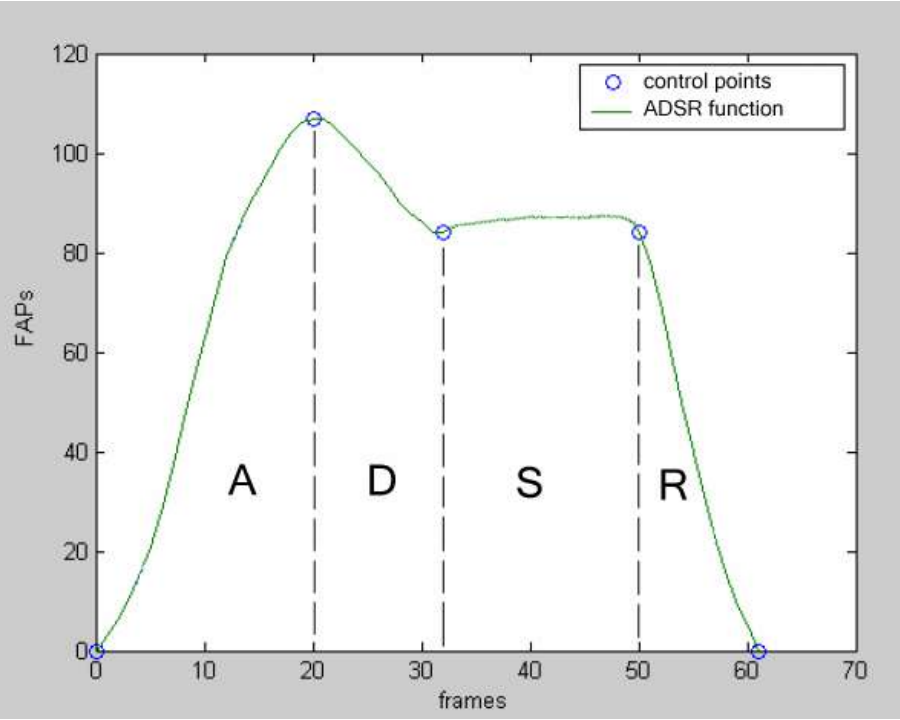
Upper Face Action Units					
AU 1	AU 2	AU 4	AU 5	AU 6	AU 7
					
Inner Brow Raiser	Outer Brow Raiser	Brow Lowerer	Upper Lid Raiser	Cheek Raiser	Lid Tightener
*AU 41	*AU 42	*AU 43	AU 44	AU 45	AU 46
					
Lid Droop	Slit	Eyes Closed	Squint	Blink	Wink

FACS


FACS example
E.g., Action code: 1, 2, 4, 5, 7, 20,
1C Inner brow raise
2C Outer brow raise
4B Brow lower
5D Upper lid raise
7B Lower lid tighten
20B Lip stretch
26B Jaw drop

Facial Temporal Course





ONSET – APEX - OFFSET

ATTACK – DECAY – SUSTAIN - RELEASE

Gaze

Gaze is linked to

- intonational pattern
- Emotional state
- Communicative function (performative, remembering, …)

It is used to

- Regulate flow of conversation
- Process information during social interactions
- Show interest
- Signal search for feedback
- Request information
- Express emotion
- Influence another person's behavior

Example from Belfast Corpus

Eye Contact

Establish relationship and communication with others

Increase with degree of intimacy and friendship between interactants

Decrease during lies, difficulties in organizing speech

Mutual gaze

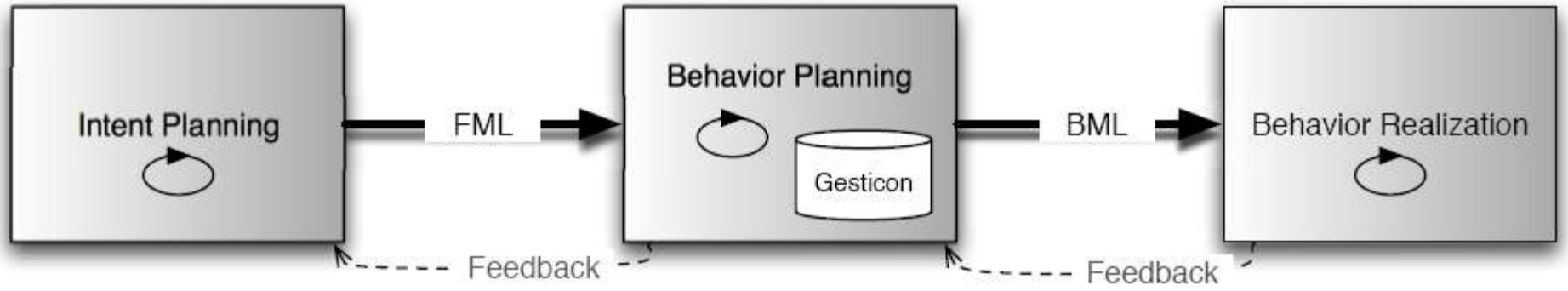
ECA System Architecture

SAIBA Framework

S. Kopp, B. Krenn, J.-C. Martin, S. Marsella, C. Pelachaud, H. Pirker, K. Thórisson, H. Vilhjálmsson

Towards a Common Framework for Multimodal Generation

SAIBA: Situation, Agent, Intention, Behavior, Animation



Function Markup Language

- FML Function Markup Language
  - encodes communicative intentions the agent aims to transmit to the user: eg, its emotional states, beliefs and goals different communicative intentions can overlap in time
  - is still at a very early age of specification
  - is an XML-based markup language

FML - APML

FML-APML (Affective Presentation Markup Language)

- **certainty**: degree of certainty the agent intends to express
- **performative**: e.g. suggest, approve, or disagree
- **theme/rheme**: represents the topic/comment of conversation;
- **belief-relation**: goal of stating the relationship between different parts of the discourse
- **turntaking**: models the exchange of speaker turns

FML - APML

FML-APML (Affective Presentation Markup Language)

- **emotion**: emotional state of the agent. Simple emotion (e.g. anger or sadness), masking (fake emotion) or superposition of two emotions
- **emphasis**: emphasize the agent's verbal or nonverbal message
- **backchannel**: listener's communicative intentions, i.e. its will and ability to continue, perceive, understand the interaction and its attitude towards the speaker's speech
- **world**: refers to an object of the world, used to communicate something about the object

FML – APML example

```
<fml-apml version="0.1">
  <bml xmlns="http://www.mindmakers.org/projects/BML" id="bml1">
    <speech id="s1" language="en-US" text="Hi, I'm Poppy."
        ssml:xmlns="http://www.w3.org/2001/10/synthesis">
         <ssml:mark name="s1:tm1"/>
           Hi,
           <ssml:mark name="s1:tm2"/>
           I'm
           <ssml:mark name="s1:tm3"/>
           Poppy.
           <ssml:mark name="s1:tm4"/>
    </speech>
  </bml>
  <fml xmlns="http://www.mindmakers.org/fml" id="fml1”>
    <performative id="p2" type="announce" start="s1:tm1" end="s1:tm4"/>
    <world id="w1" ref_type="person" ref_id="self" start="s1:tm2" end="s1:tm4"/>
  </fml>
</fml-apml>
```

duration

Behavior Markup Language

- BML operates at signal level
- Different communication channels

e.g. : head movements, gaze, facial expressions, gestures, speech

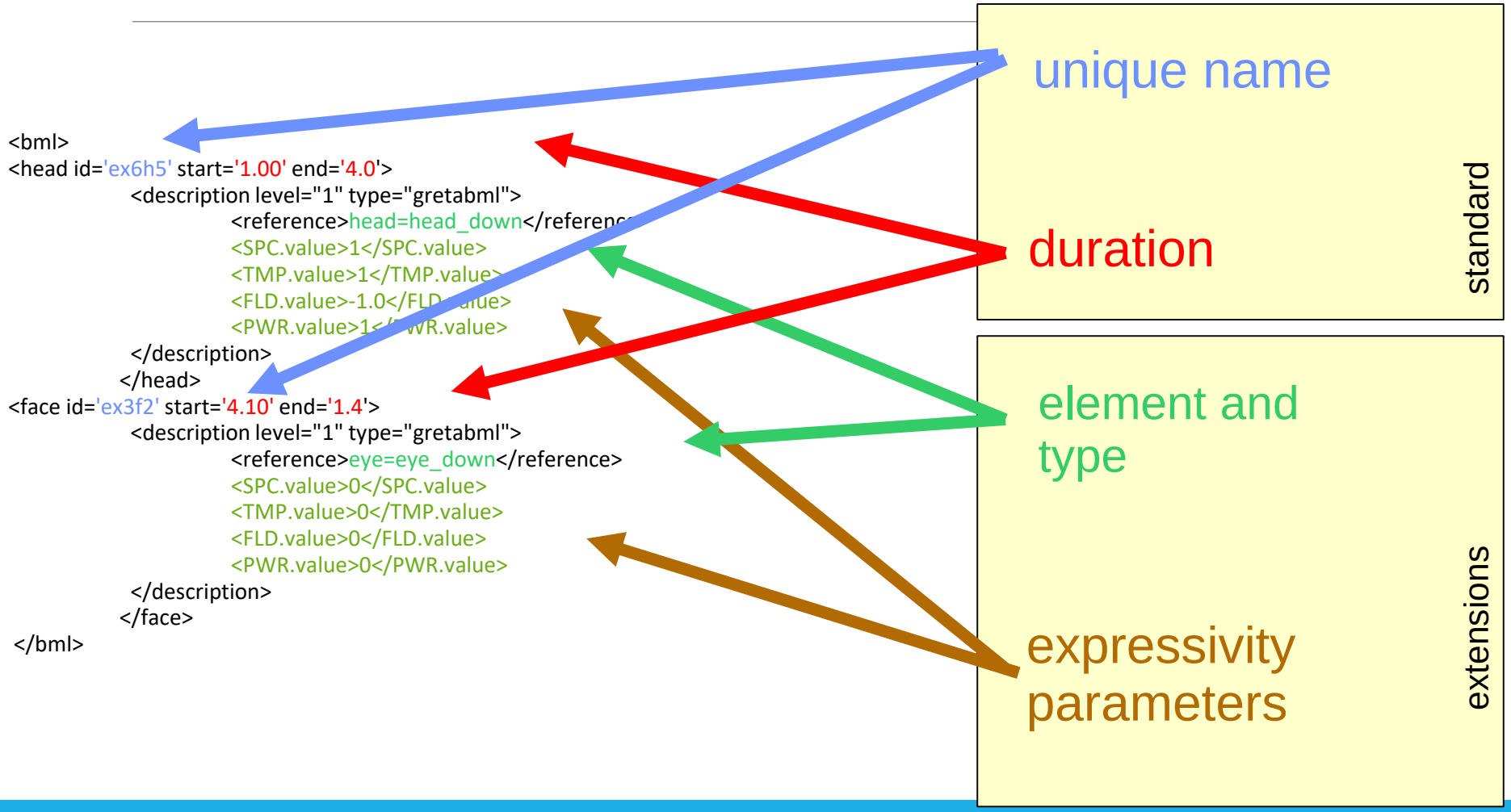
- Start time, end time specification
- Expressivity parameters

e.g. : quickly or slowly, with more or less energy

⇨ specify which signals to perform and how

Behavior Markup Language

```
<bml>
<head id='ex6h5' start='1.00' end='4.0'>
        <description level="1" type="gretabml">
                <reference>head=head_down</reference>
                <SPC.value>1</SPC.value>
                <TMP.value>1</TMP.value>
                <FLD.value>-1.0</FLD.value>
                <PWR.value>1</PWR.value>
        </description>
      </head>
<face id='ex3f2' start='4.10' end='1.4'>
        <description level="1" type="gretabml">
                <reference>eye=eye_down</reference>
                <SPC.value>0</SPC.value>
                <TMP.value>0</TMP.value>
                <FLD.value>0</FLD.value>
                <PWR.value>0</PWR.value>
        </description>
      </face>
</bml>
```



Behavior Markup Language

Synchronization of head movement and gaze with a gesture

```
<bml>
 <gesture id=”g1” type=”beat”/>
 <head type=”nod” stroke=”g1:stroke”/>
 <gaze target=”object1” start=”g1:ready” end=”g1:relax”/>
</bml>
```

Kopp et. al 2006

- « Nod » stroke occurs at the same time as « beat » stroke
- Gaze starts when « beat » is in ready phase, end when « beat » is in relax phrase

Lexicon creation: FML → BML

Behavior set:

BS = (*name; Sigs;Core; Implications*);

name of the corresponding communicative function; eg *refuse*, we set the *name* of the set to *refuse*.

Sigs, containing the name of the signals produced on single modalities (like head, face, gaze...)

Core: list of signals that are mandatory to communicate the function

- to communicate *refuse*, the agent *MUST* shake its head

Implications: a set of logic rules like *if A then B* that allow us to conditionally constrain the presence of a signal of the Sigs set depending on the presence of the other signals.

Lexicon creation

Behavior set:

BS = (*name; Sigs;Core; Implications*);

```
<behavior-set name=«preformative-refuse">
<signals>
<signal id="1" name="shake"
 modality="head"/>
<signal id="3" name="at" modality="gaze"/>
<signal id="2" name="frown"
 modality="face">
<alternative name="lip tension"
 probability="0.3"/>
<alternative name="frown+lip tension"
 probability="0.4"/>
</signal>
<signal id="4" name="no"
 modality="gesture"/>
</signals>
<constraints>
  <core>
    <item id="1"/>
  </core>
  <rules>
    <implication>
      <ifpresent id="4"/>
      <thenpresent id="2"/>
    </implication>
  <rules>
</constraints>
</behavior-set>
```

Greta

ECA System Architecture : Summary

3 main stages :

1. perception
2. reasoning/reactive processes
3. behavior realization

FML-APML : communicate intention

BML : specify how to realize behaviors

Computational Models of Communicative Gesture

Predicting gestures

Gestures stroke either precedes or coincides with speech prosody (pitch accent)

It rarely follow speech

→ difficulty to predict in real-time in an incremental manner where to place a gesture

There is not a direct mapping with speech/word/prosody and gestures

→ highly non-derministic

Link between information conveyed by speech and by gesture:

- Redundant
- Complementary
- Substitution
- …

Gesture models

Existing methods

- Ruled-based
- Statistically-based
- ML

Goal:

- Where to place a gesture
- Which gesture form
- Capture gesturing style

Rule-based

Rule-based

Rules defined from the literature, corpus annotation

Capture links with linguistic structures including

- prosody
- Iconic meaning
- Metaphor
- Deictic

Embodied Conversational Agents : where we started

Autonomous agents designed to be:

-Speaker

-Listener

With multimodal behaviors:

-Gesture

-Facial expression

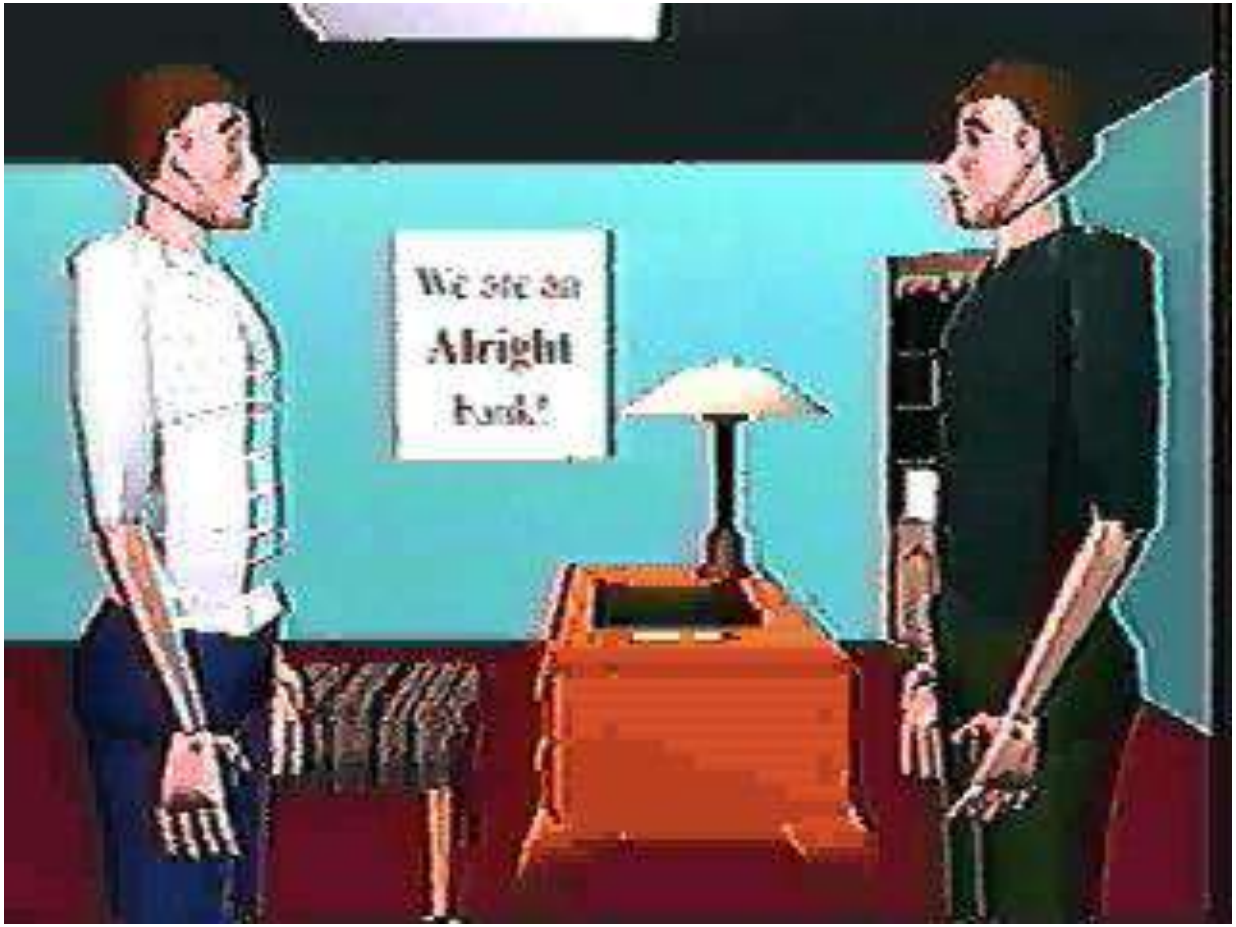
-Gaze

-Lip movement

Cassell et al, 94

Gilbert and George at the Bank (Upenn)

Cassell, Pelachaud, Badler, Steedman… Suggraph'94



Cassell et al, 94

Synchrony tool - BEAT

Based on linguistic and contextual information extracted from text

- Theme, rheme
- Contrast
- Word newness

Decomposition of text into theme and rheme

Linked to WordNet

Computation of:

- intonation
- gaze
- gesture

Cassell et al

Statistically based

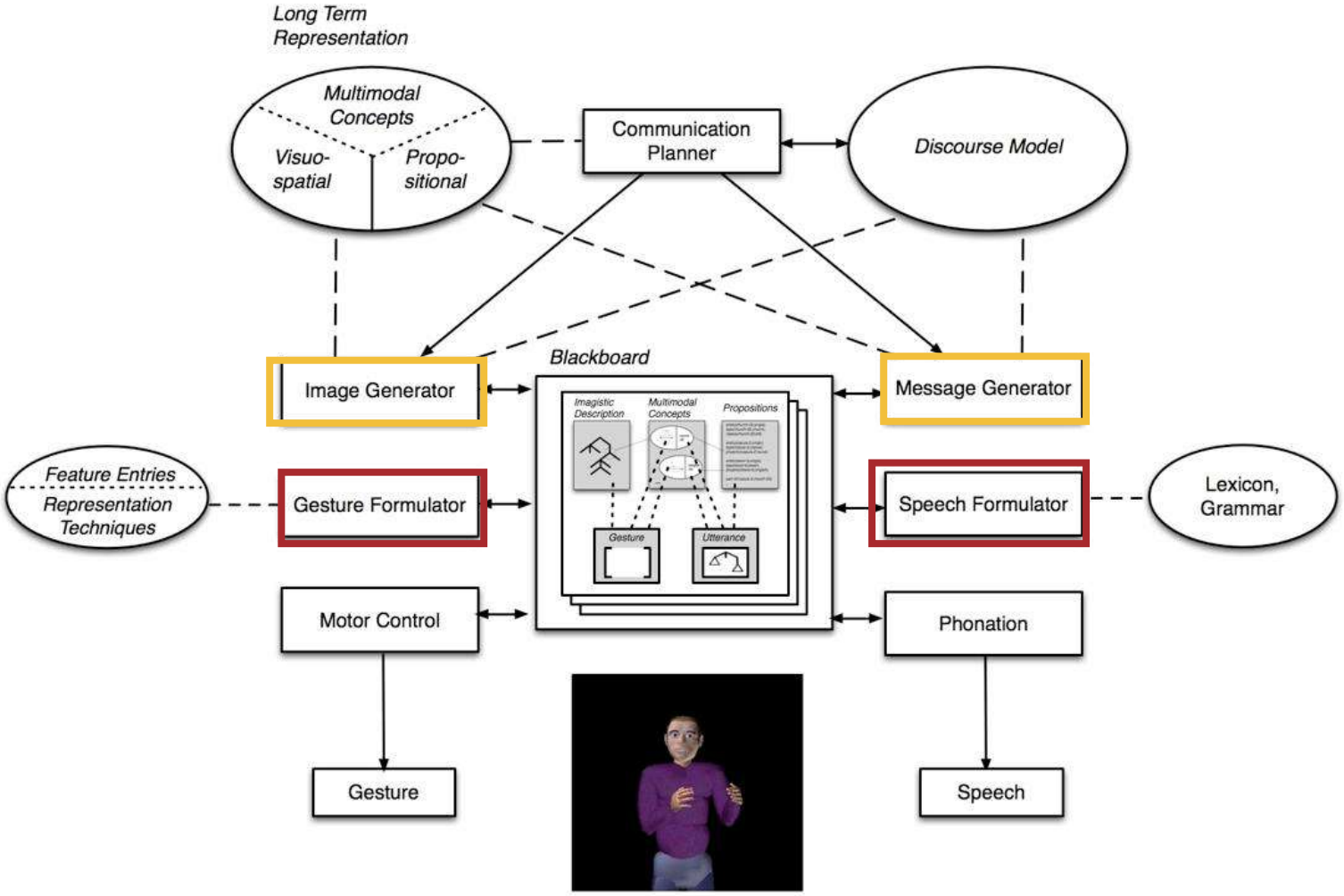
Automatic generation of speech and gesture Bergmann - Kopp

Task direction domain

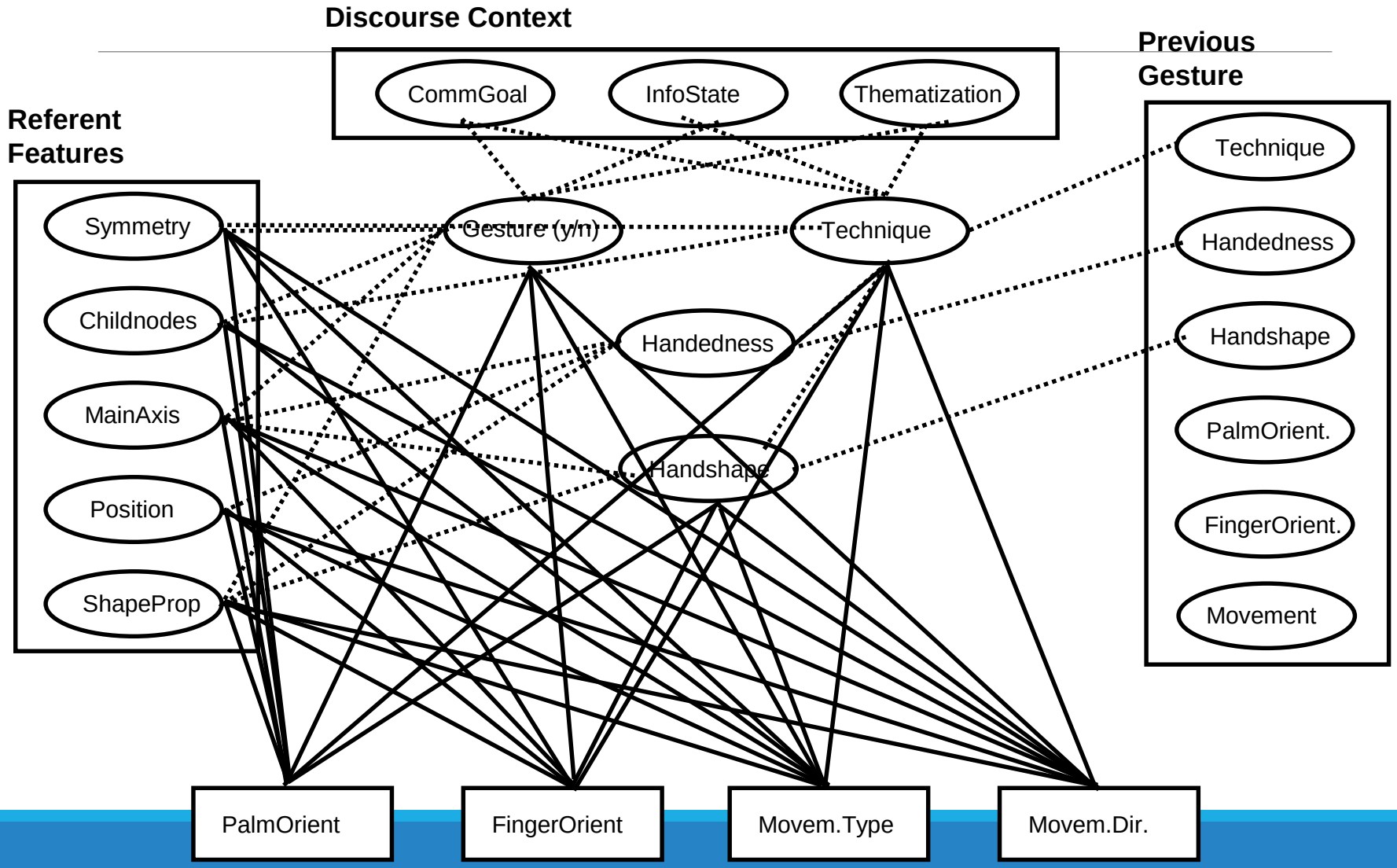
Interweaving of multimodal content planning and micro-planning of speech and gesture

- 2 main tasks
  - select the content to say/display
  - derive the form of coordinated language and iconic gestures

Overall architecture Bergmann - Kopp

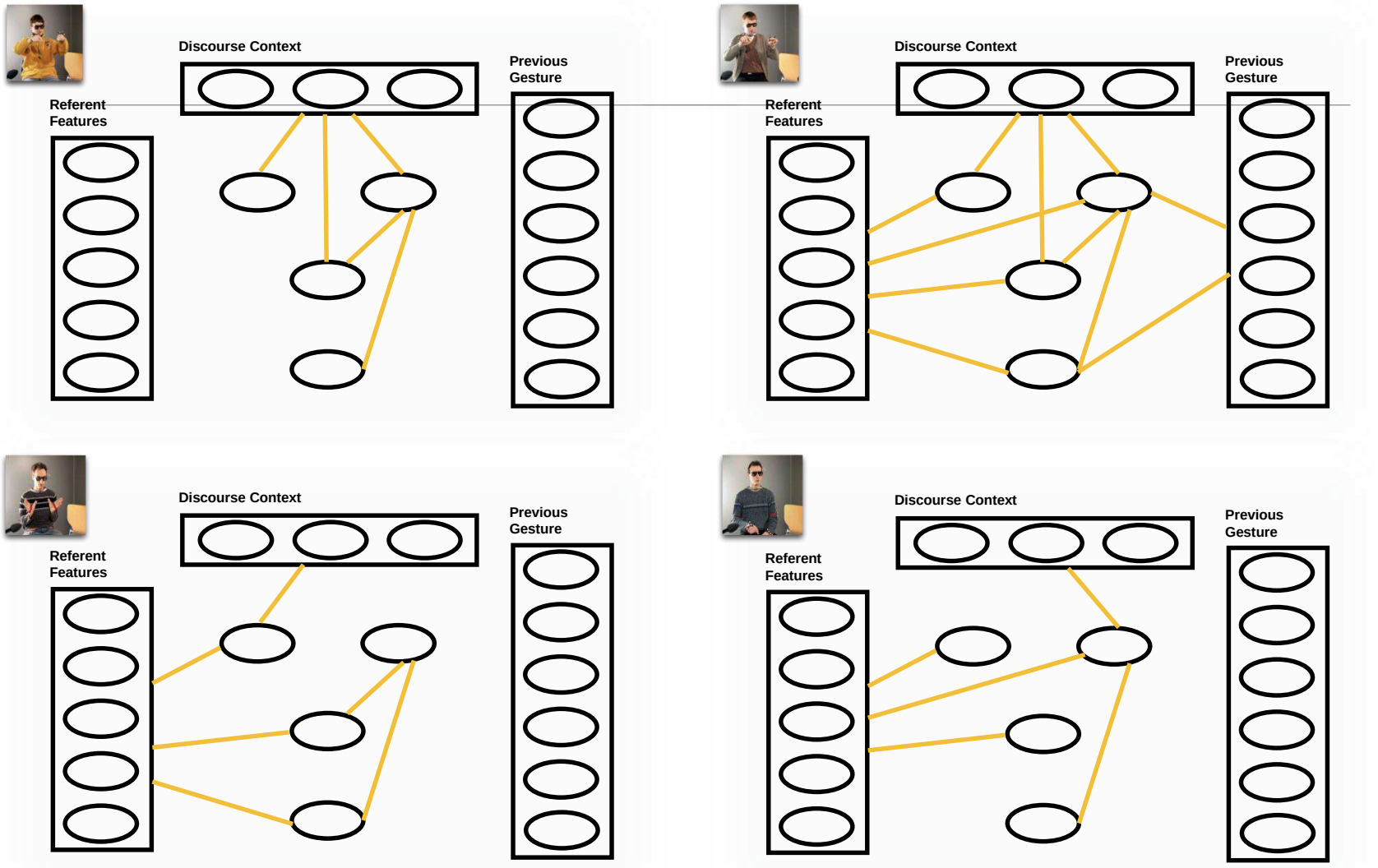


Building Networks Bergmann - Kopp



Individual Speaker Networks -

Bergmann - Kopp



Idiosyncrasy in the production process of iconic gestures

Automatic generation of speech and gesture Bergmann - Kopp

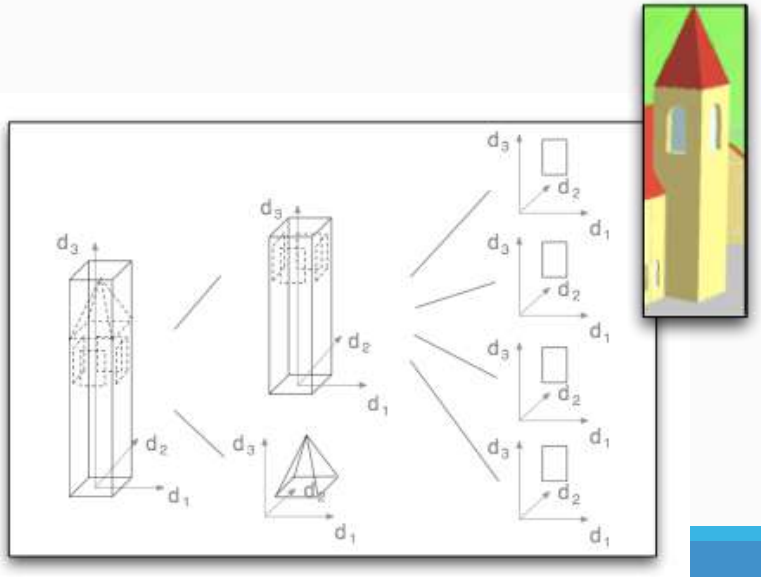
4 production modules:

- Image Generator: derivation of imaginistic description for objects
- Preverbal Message Generator: selection of facts from propositional memory
- Speech Formulator: sentence planning (use of SPUD)
- Gesture Formulator: compute gesture morphology.

Content Representation Bergmann - Kopp

Computational Imagery

- Object shape represented as Imagistic Description Trees (Sowa & Wachsmuth, 2005)
  - Hierarchical structure
  - Extents in different dimensions
  - Possible underspecification



Linguistic spatial representation

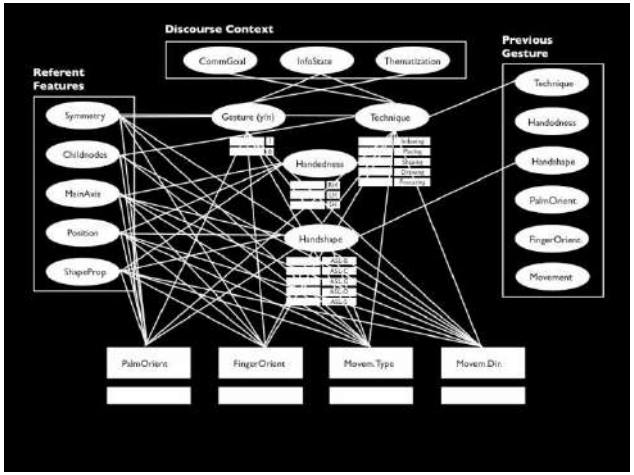
- Knowledge drawn upon by speech formulation
- Logical formulae based on a formal ontology of domain knowledge
- Partly reflects imagistic knowledge

```
shared (entity (church-5, single)).
shared (instance_of (church-5, church)).
private (entity (window-3, single)).
private (instance_of (window-3, window)).
private (property (window-3, round)).
private (relpos (window-3, middle)).
private (entity (churchtower-1, single)).
private (instance_of (churchtower-1, tower)).
private (relpos (churchtower-1, left)).
private (part_of (church-5, churchtower-1)).
private (entity (churchtower-2, single)).
private (instance_of (churchtower-2, tower)).
private (relpos (churchtower-2, right)).
private (part_of (church-5, churchtower-2)).
...
```

Generation Example Bergmann - Kopp

Gesture Formulator

- Bayesian Decision Network to fill gesture feature matrix

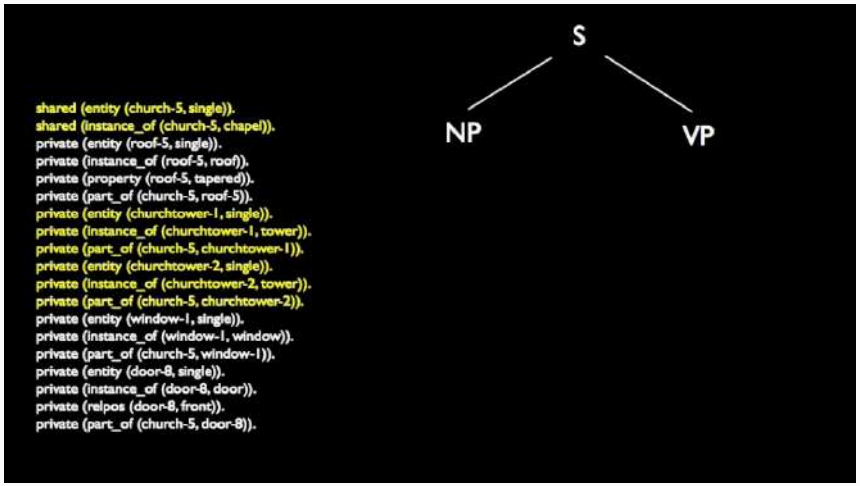


Gesture matrix

LOC: Periphery Left
TRAJ: Linear
MOVEMENT DIR: Down
HANDSHAPE: ASL-bent-5
PALM DIR: Down

Speech Formulator

- Microplanning with SPUD (Stone et al., 2003)
  - Grammar-based microplanner using a Lexicalized Tree Adjoining Grammar (LTAG)



Surface Realization Bergmann - Kopp

Gesture matrix

LOC: Periphery Left
TRAJ: Linear
MOVEMENT DIR: Down
HANDSHAPE: ASL-bent-5
PALM DIR: Down

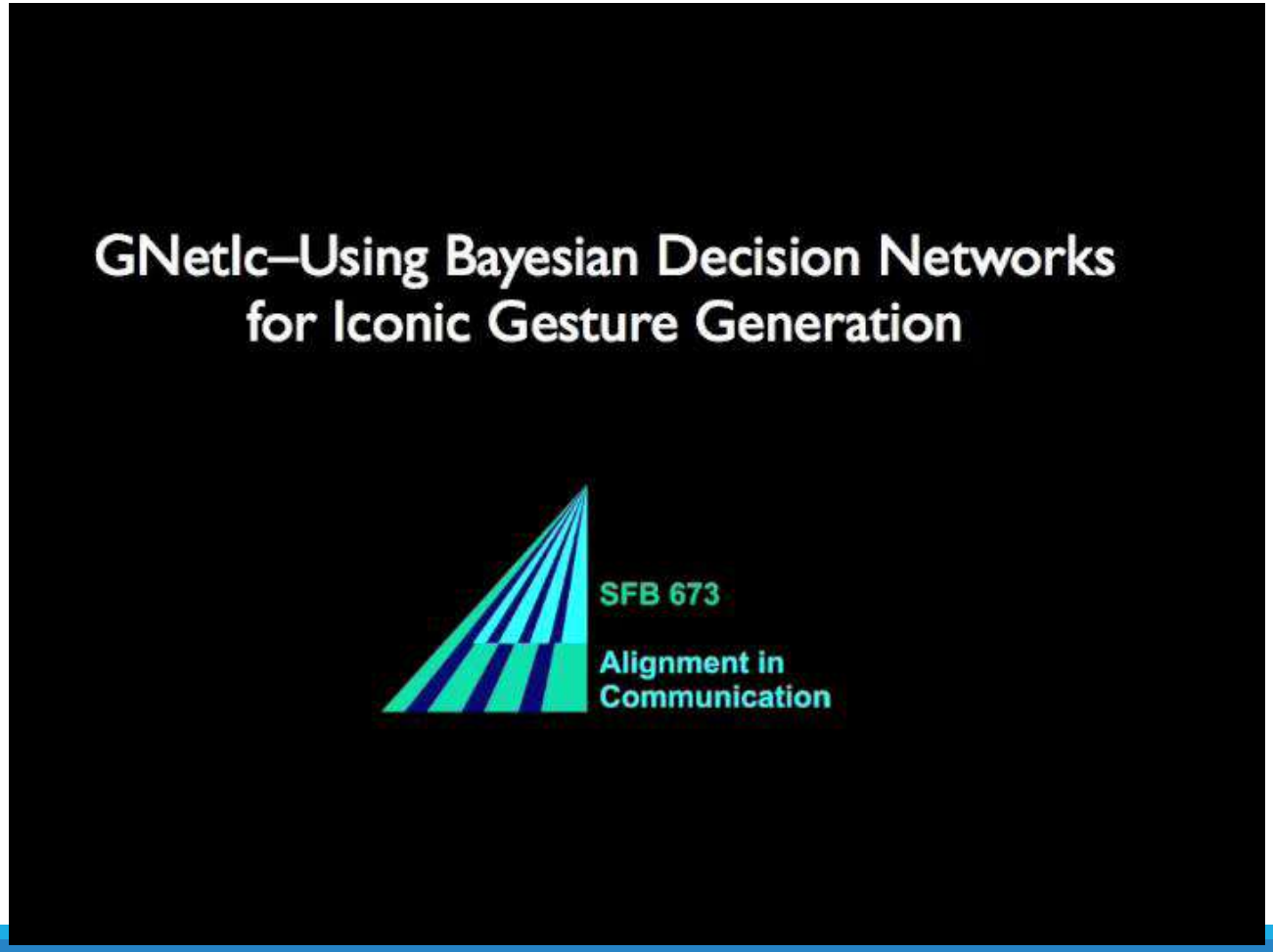
```
<constraints>
  <parallel>
    <static slot="HandShape" value="BSifinger"/>
    <repeat_alt times="3">
      <dynamic slot="HandLocation">
                    <dynamicElement type="curve">
                      <value type="start"  name="LocShlder LocCentLeft
LocFar"/>
                      <value type="end"    name="LocShlder LocCentLeft
LocNorm"/>
                      <value type="normal" name="DirU"/>
                      <value type="shape"  name="LeftC"/>
                      <value type="extension" name="0.6"/>
                    </dynamicElement>
      </dynamic>
    </repeat_alt>
  </parallel>
</constraints>
```

MAX/ACE (Kopp & Wachsmuth, 2004)

- On-the-fly speech synthesis and movement planning
- Scheduling and co-articulation of speech and gestures, incremental chunks (intonation phrase + gesture phrase)

"the church has two towers"

Modeling Results Bergmann - Kopp



Speech driven animation

Speech driven approach

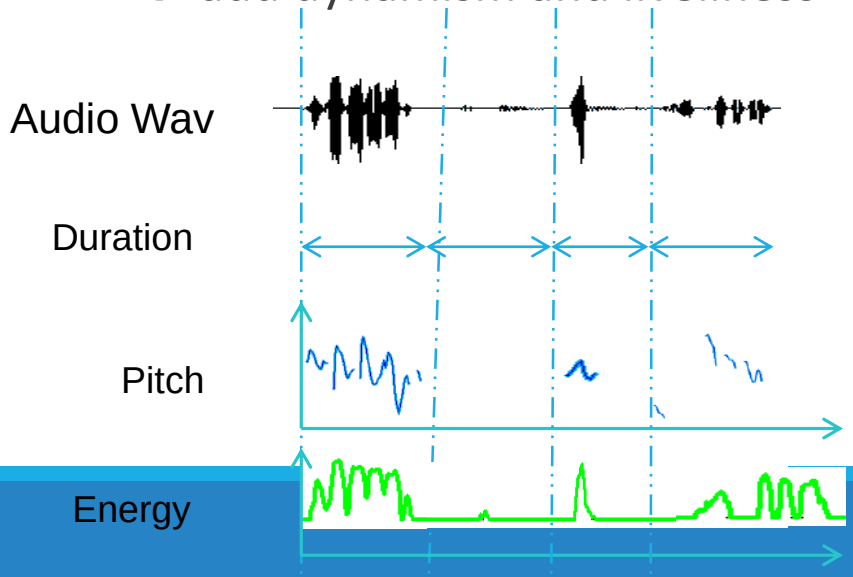
Input: speech features

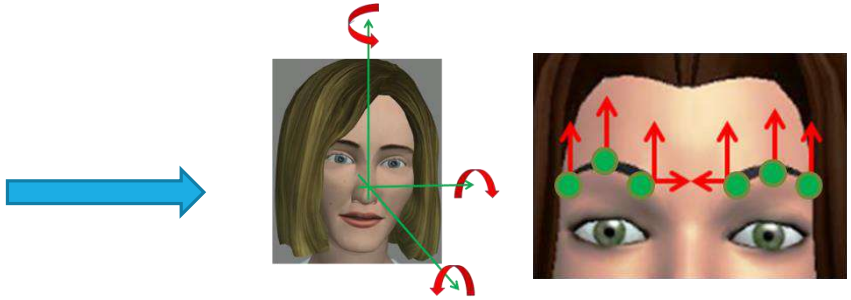
Output: 3D animation parameters

→ infer lip movement, facial expression, head movements, gesture from audio cues

→ capture link betweeb features

→ add dynamism and liveliness





Speech driven approach

Learn a model from real data

Input: speech features

Output: 3D animation parameters

→ infer lip movement, facial expression, head movements, gesture from audio cues

→ capture link betweeb features

→ add dynamism and liveliness

Diffulties:

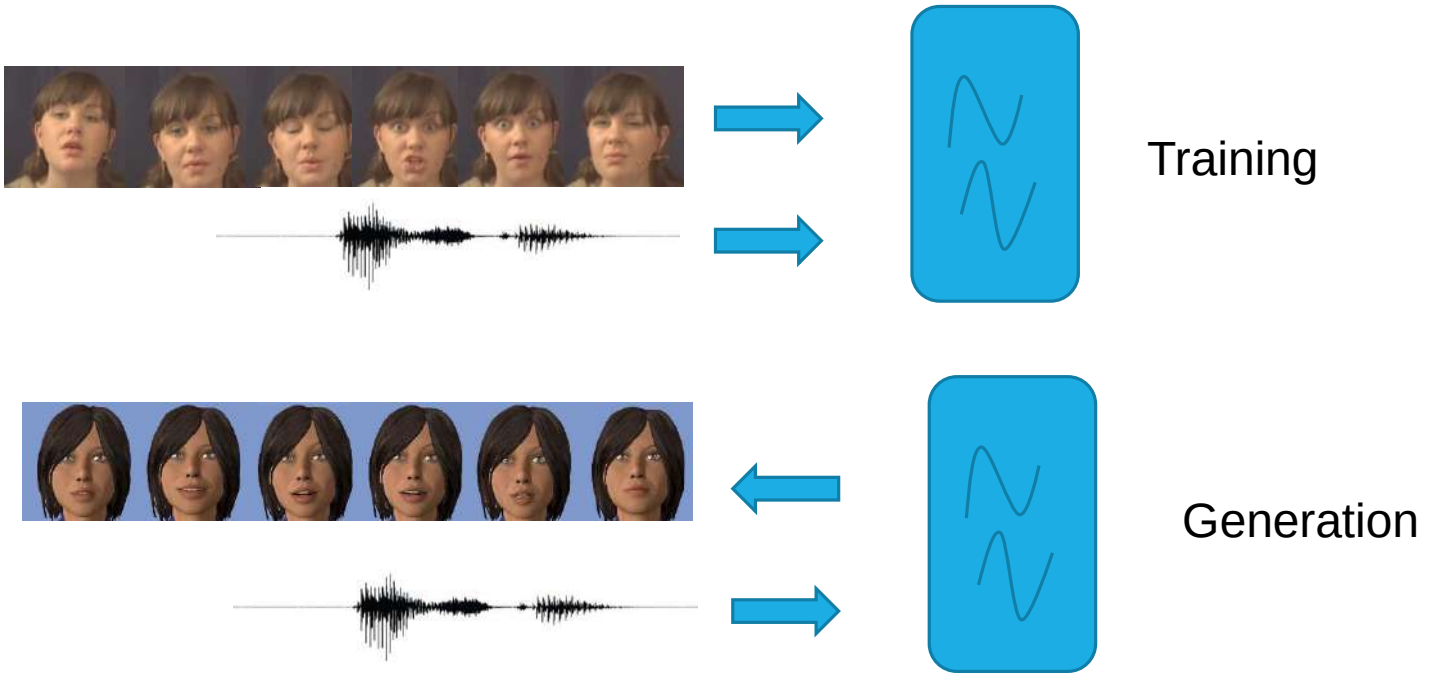
→ no explicit linguistic or semantic information

→ many-to-many mapping between speech data and motion

Statistical Models

HMM: capture temporal relationship between speech data and motion

→ use HMM to compute lip shapes, facial expressions, head movements…



Joint models – Carlos Busso

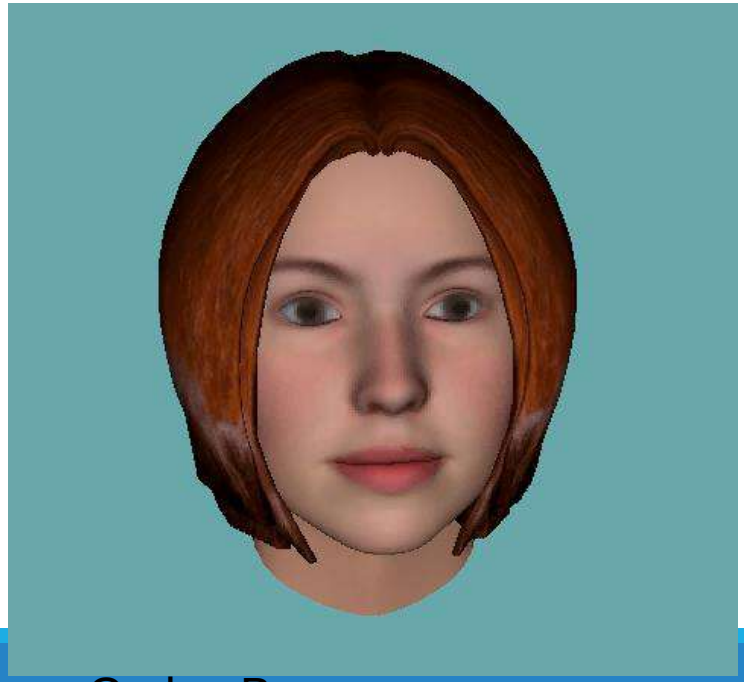
Separate model to capture links between speech and:

- Head motions
- Eyebrows movements

→ do not capture the coherence of multimodal signals

Prosody → model → head

Prosody → model → eyebrow



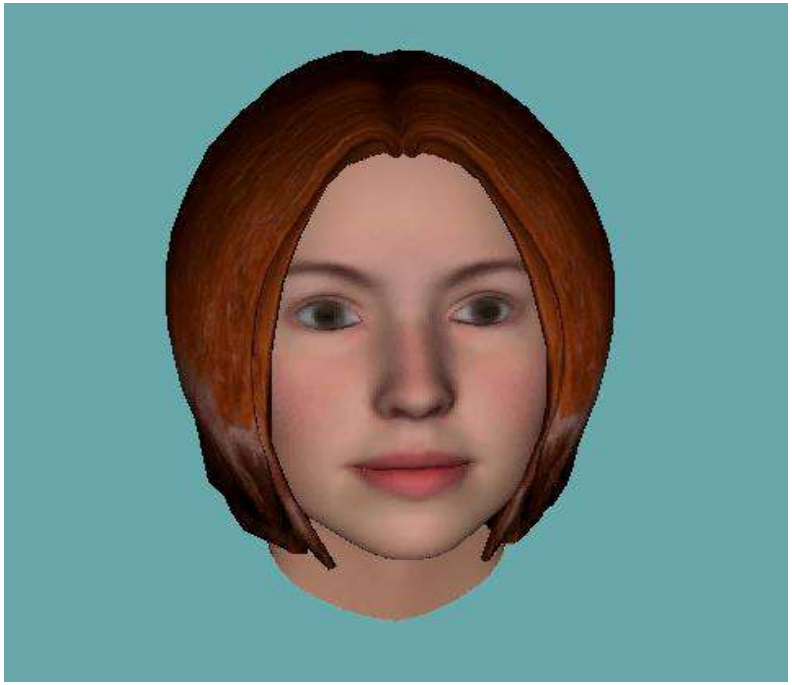
Joint models – Carlos Busso

Prosody → head + eyebrows

Evaluations:

Perceptive study:

- rating of agent's naturalness



Joint statistical Model – Yu Ding

Learned from real data

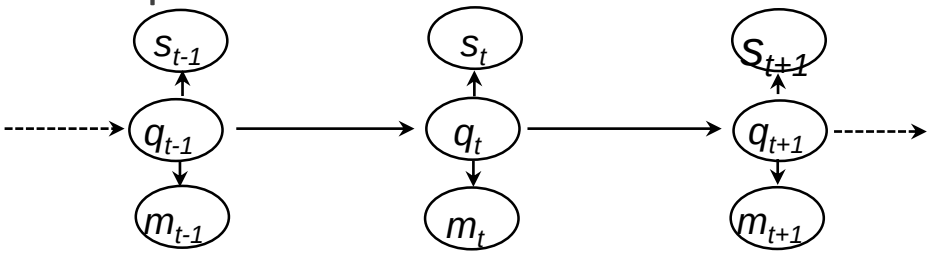
Head and eyebrows animation driven from prosody

Capture the direct influence of previous data and speech on current data

Joint Fully parameterized HMM (FPHMM)

- Audiovisual corpus of actors expressing emotions (Gall *et al.*, 2010)

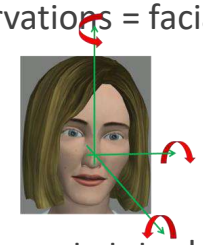
In 'classic' HMM: non influence of previous context or observational variables

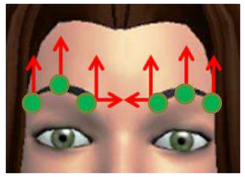


Joint statistical Model – Yu Ding

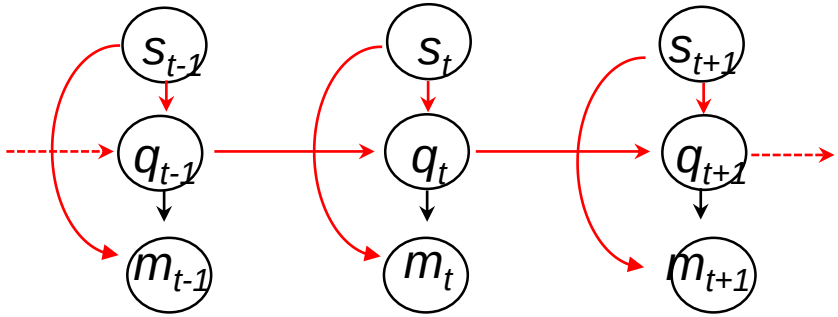
Joint Fully parameterized HMM (FPHMM)

- Based on contextual HMM
  - contextual variables = speech features (S_t)
  - Observations = facial and head parameters

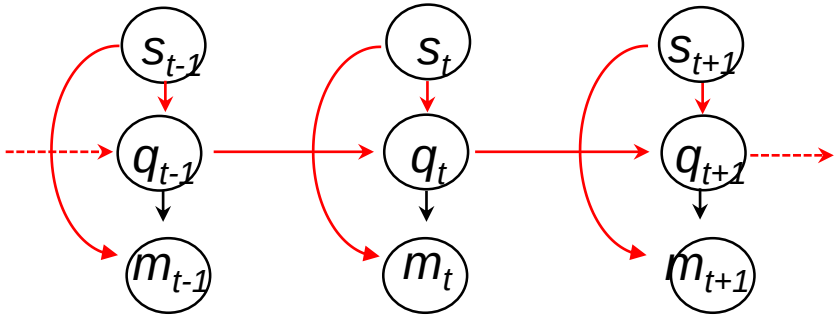




  - A t, current state depends from previous state and contextual variables (speech parameters)
    - q(t) = f(q(t-1), S(t-1))



Joint statistical Model –
Yu Ding



- Validation with users' perceptive studies and objective measures (reconstruction error)

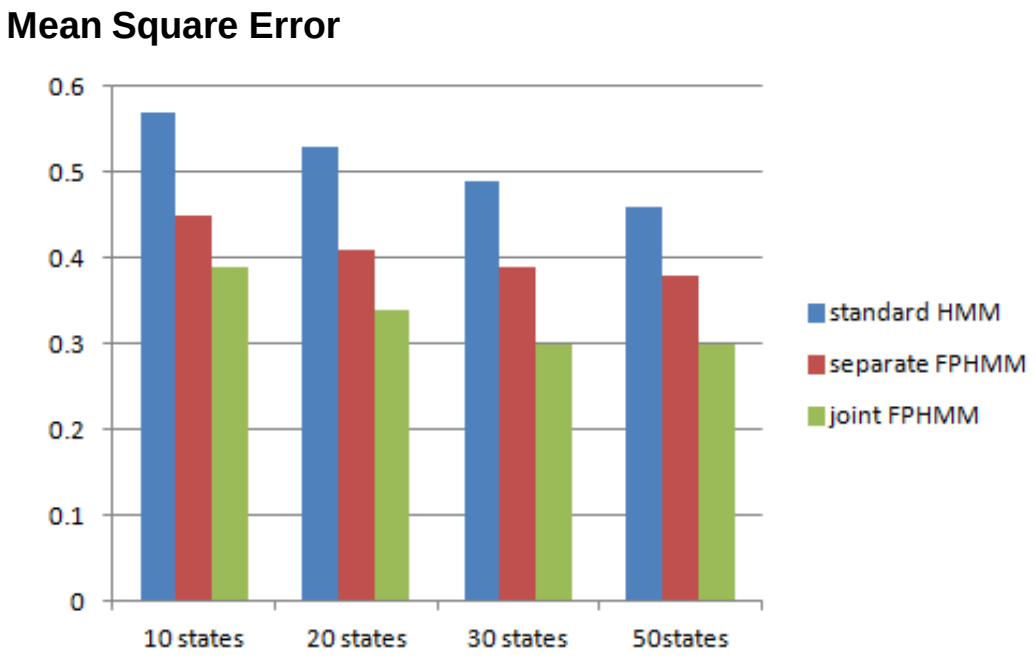
Joint statistical Model – Yu Ding

Compare ***reconstruction error*** between the synthesized and the real motion signals.

Results:

1) PFHMM > standard HMM

2) Joint FPHMM > Separate FPHMM

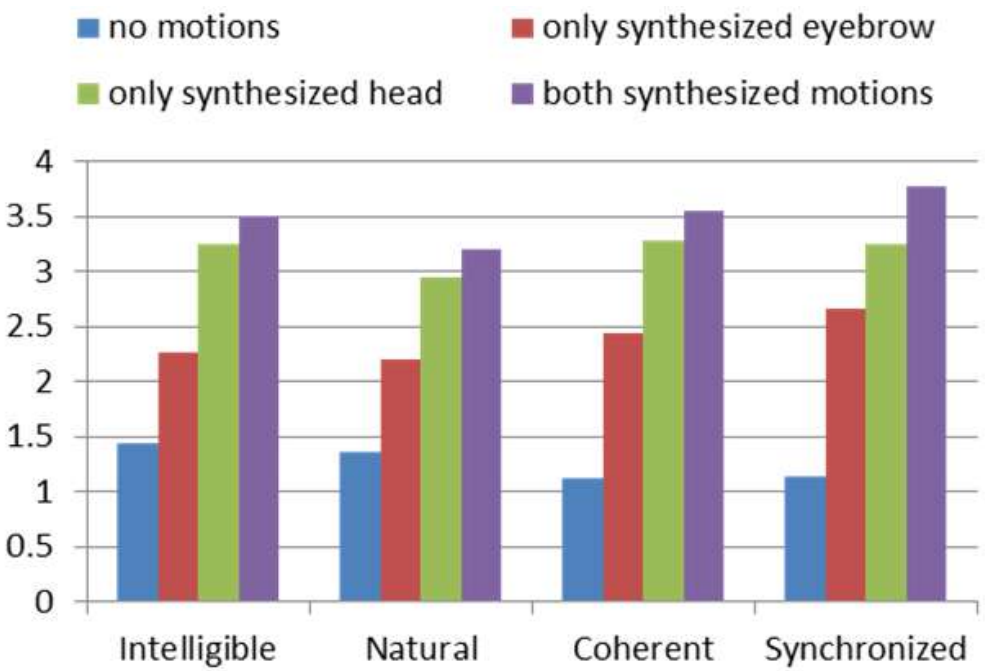


Joint statistical Model – Yu Ding

SUBJECTIVE EVALUATION

Model configuration

Joint FPHMM with 30 states

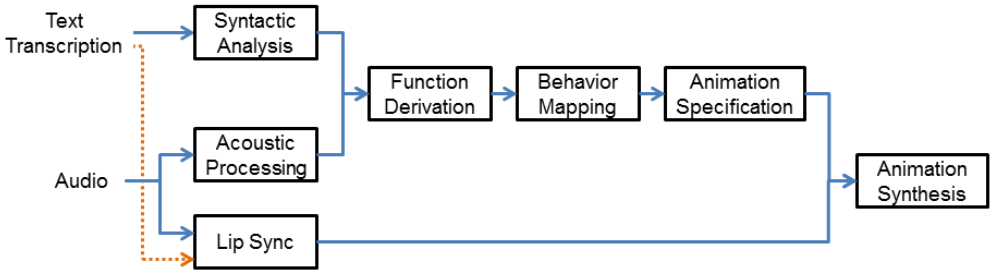


Both modal motions **>** only head motions **>** only eyebrow motions **>** no motions

Gesture models

Performance from Speech
Marsella et al

Automatic generation of multimodal behavior animation from audio signal by inferring the acoustic and semantic properties of the utterance.



Acoustic Processing: pitch extraction, stress, emotion recognition
Syntactic analysis: parsing of the sentence to find its structure

Performance from Speech
Marsella et al

Function Class	Description	Example word	Example behavior
Affirmation	Agree, accept	Okay, yes	Big nod, tilt nod
Interrogative	Direct or indirect questions	What, where, when	Gesture question, eyebrow raise
Spatial relation	Actual or metaphor	Beyond, further	Gesture away

3. Function Derivation: extract communicative functions tied to NVB; 91 rules along with a dictionary of 170 words and phrases employed specifically by the lexical analysis rules.
4. Behavior Mapping: 97 function-behavior mapping rules; many-to-many mapping; probabilistic mapping
5. Animation Specification: scheduling of the behaviors
6. Animation Synthesis: compute animation; coarticulation between gestures

Cerebella Marsella et al

Smart Body & Cerebella

Hanson Robotics + Embody

Cerebella: start-up Embody Digital

Hanson: human-like robot ‘Sophia’

- computation on the fly of:
  - Lip shape
  - Facial expression
  - Head movement
  - Gesture

- From text and sentiment analysis

Video:

https://www.veracode.com/blog/security-news/live-rsa-sophia-social-humanoid-robot

Multi-objective adversarial gesture generation

Ylva Ferstl
Trinity College Dublin

Michael Neff
UC Davis

Rachel McDonnell
Trinity College Dublin

Gesture model

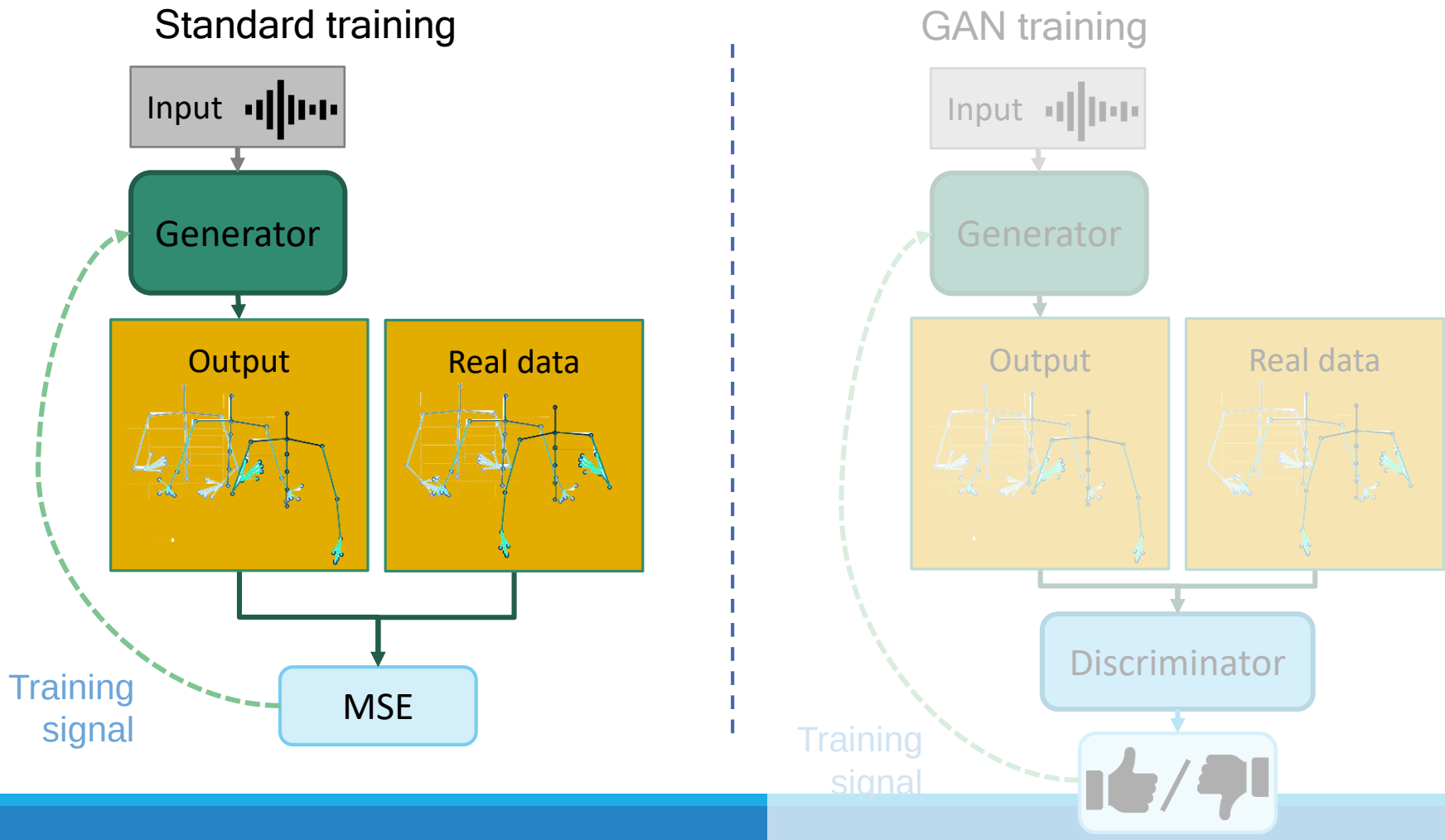
Non-deterministic gestures → large error loss

Even if animation is credible

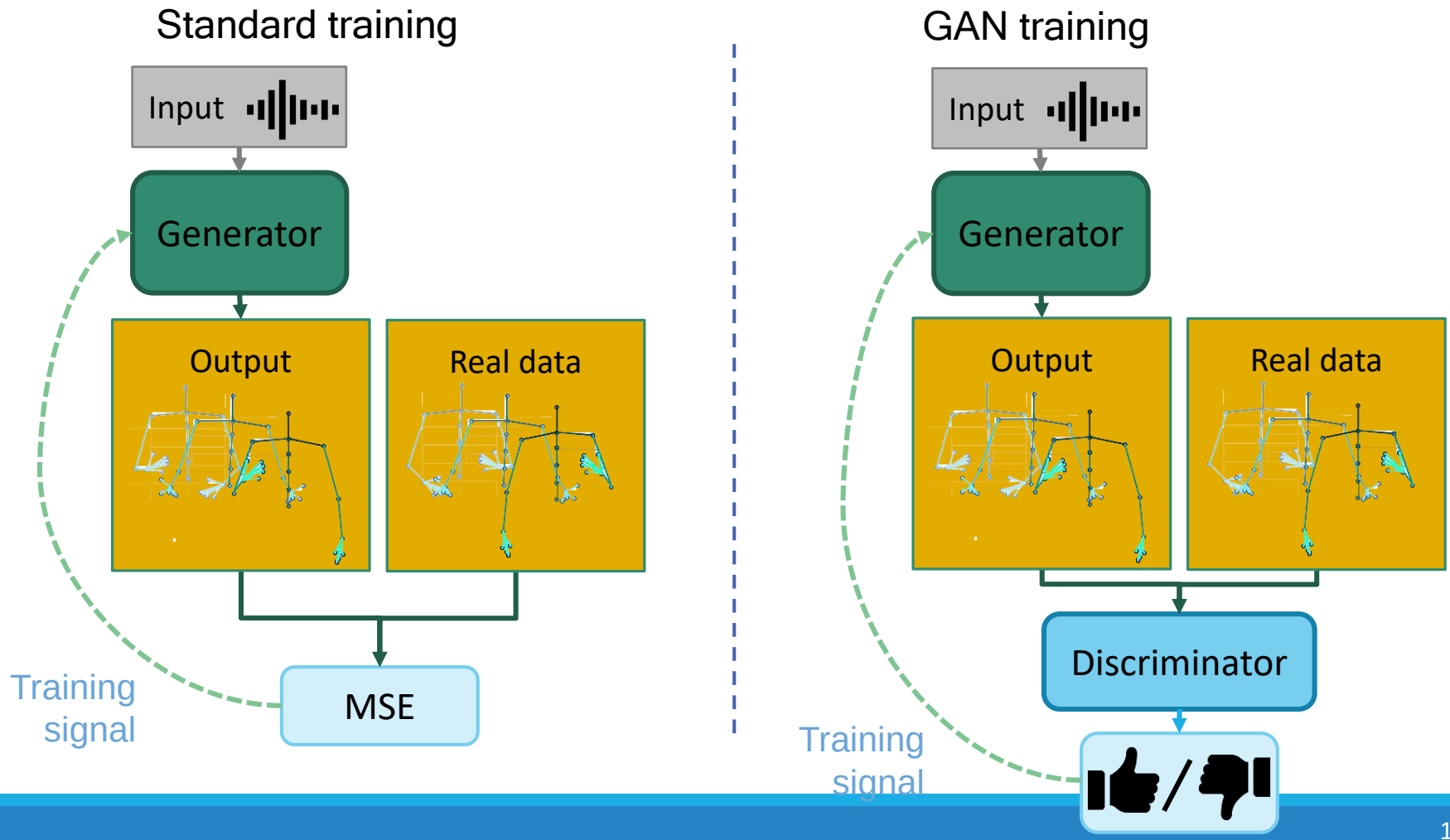
→ use of recurrent neural network

1. train a speech-input-motion-output RNN with a generative adversarial paradigm
2. Capture gesture phases (in particular stroke) by training discriminator network using only gesture phases

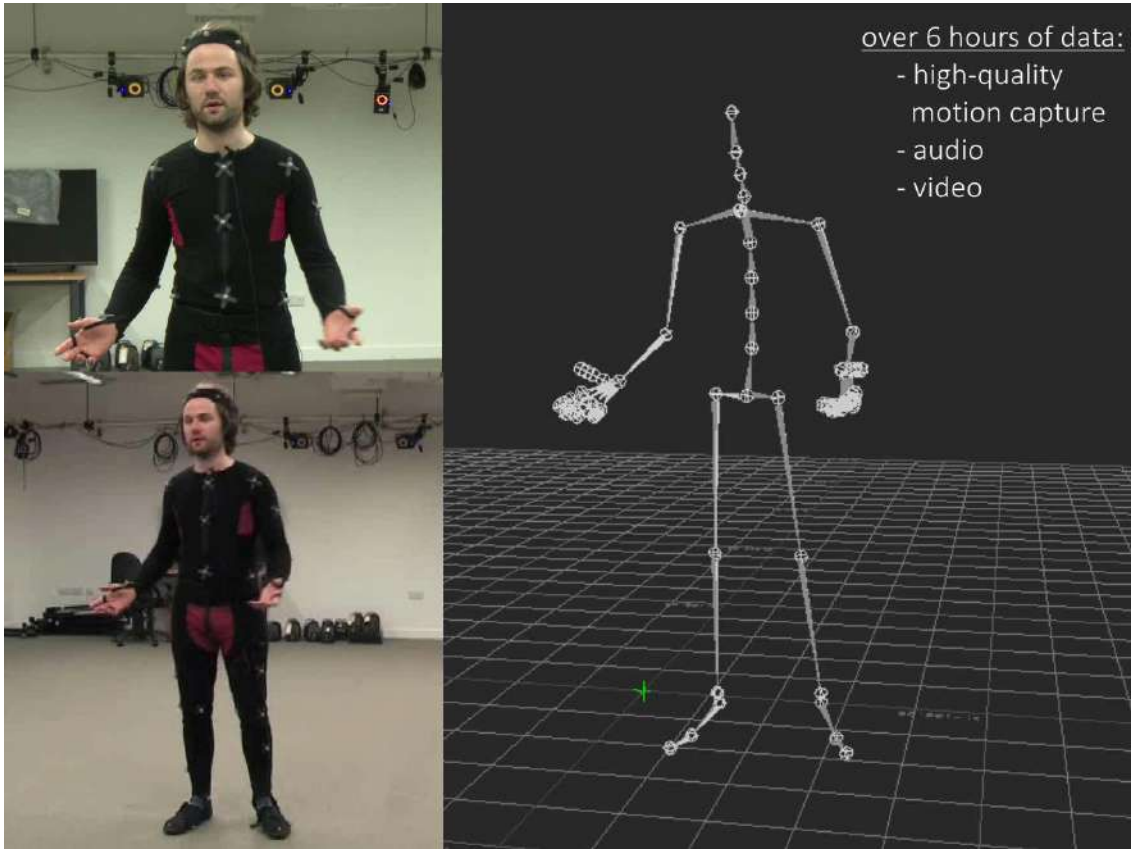
Generative Adversarial Networks (GANs)



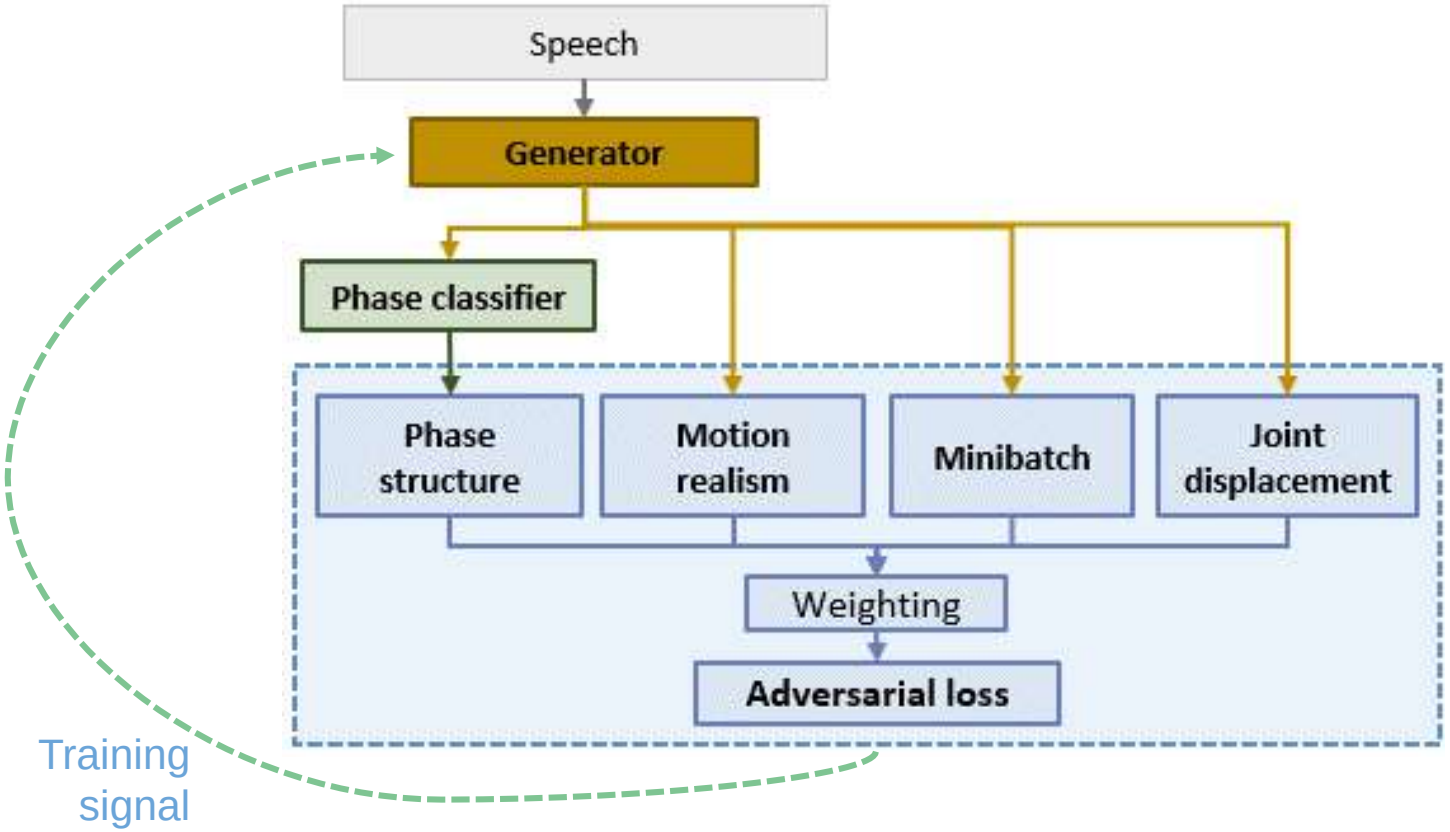
Generative Adversarial Networks (GANs)



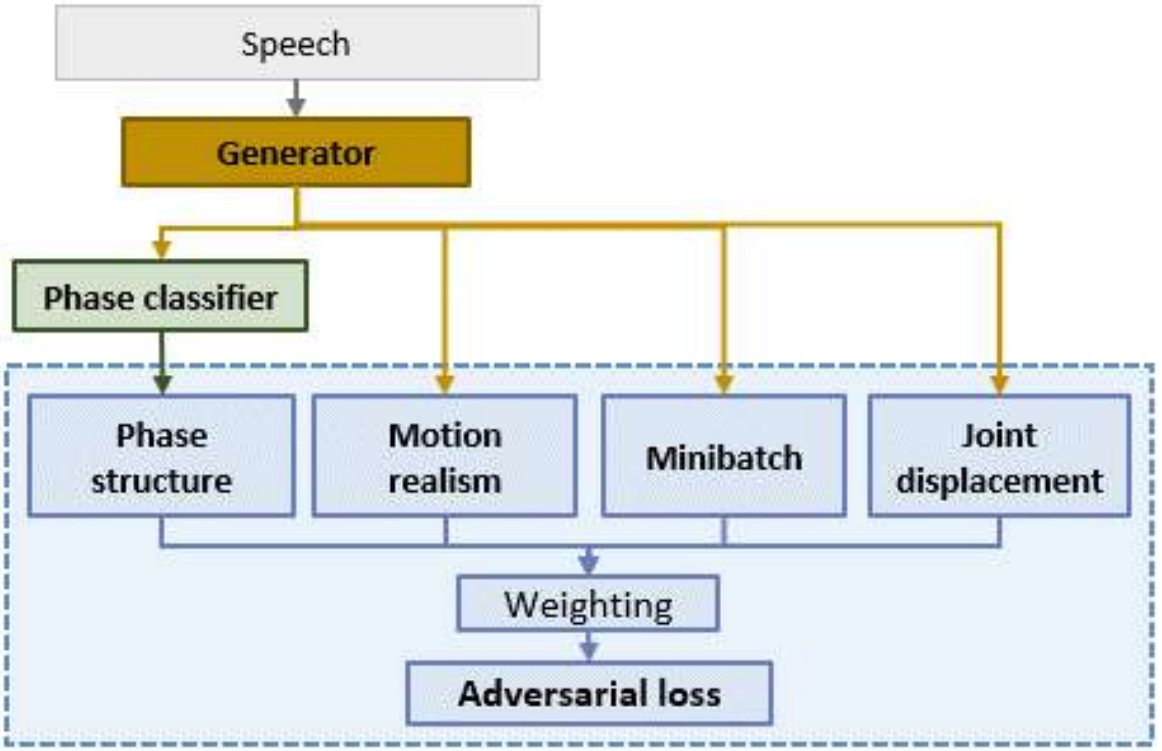
Dataset


over 6 hours of data:
- high-quality
motion capture
- audio
- video

System overview



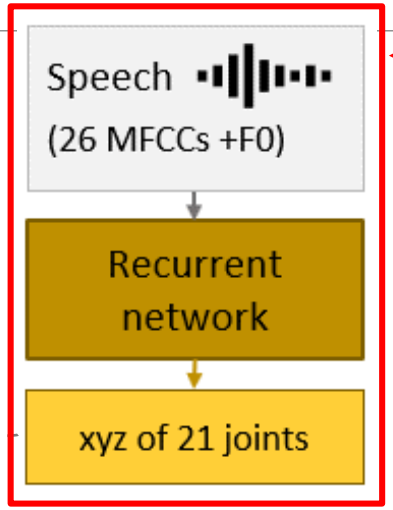
Gesture generator

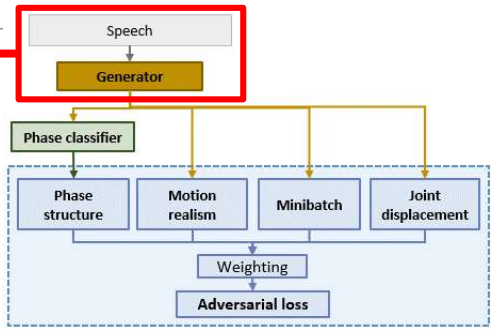


Gesture generator

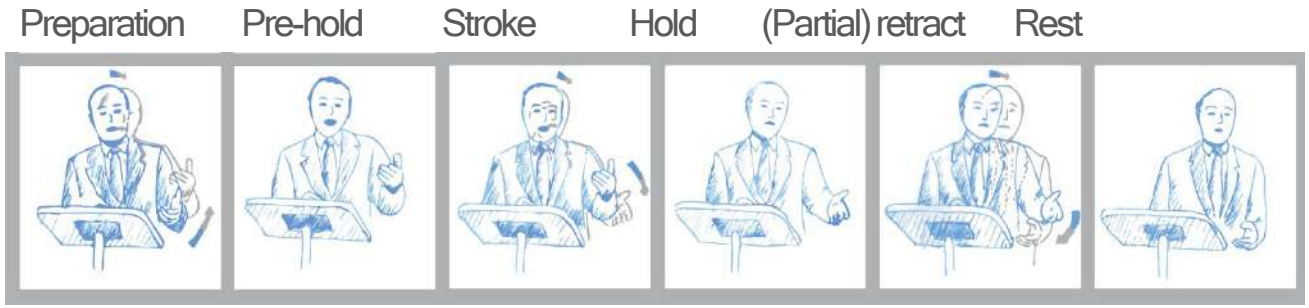
Input: 27 prosodic speech features

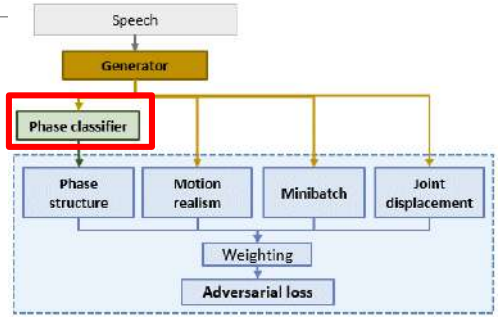
Output: 3D positions of 21 joint motions





Phase classifier

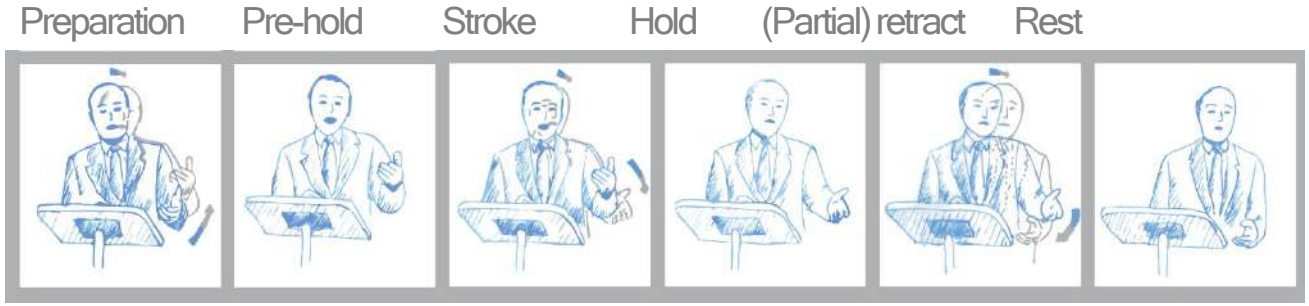


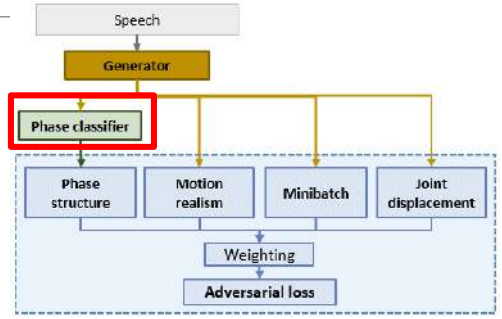


Ren, A. (n.d.). http://web.mit.edu/pelire/www/gesture-research/index.html

Input: 21 joints
Output: gesture phase labels

Phase classifier





Ren, A. (n.d.). http://web.mit.edu/pelire/www/gesture-research/index.html

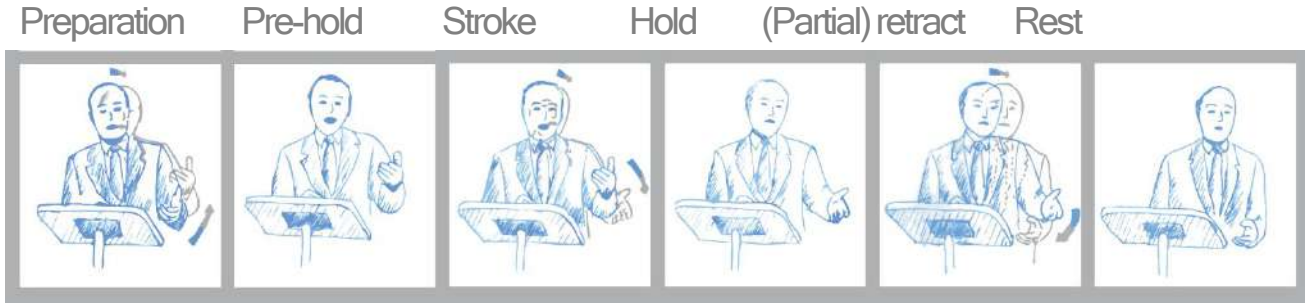
Input: 21 joints
Output: gesture phase labels

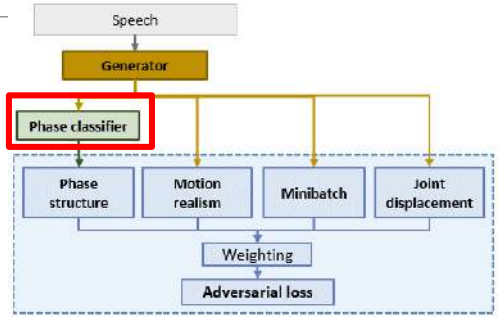
Modelling (realistic) gesture dynamics is hard!

⇒ Implicit modelling: Big data

⇒ Explicit modelling: Labels

Phase classifier

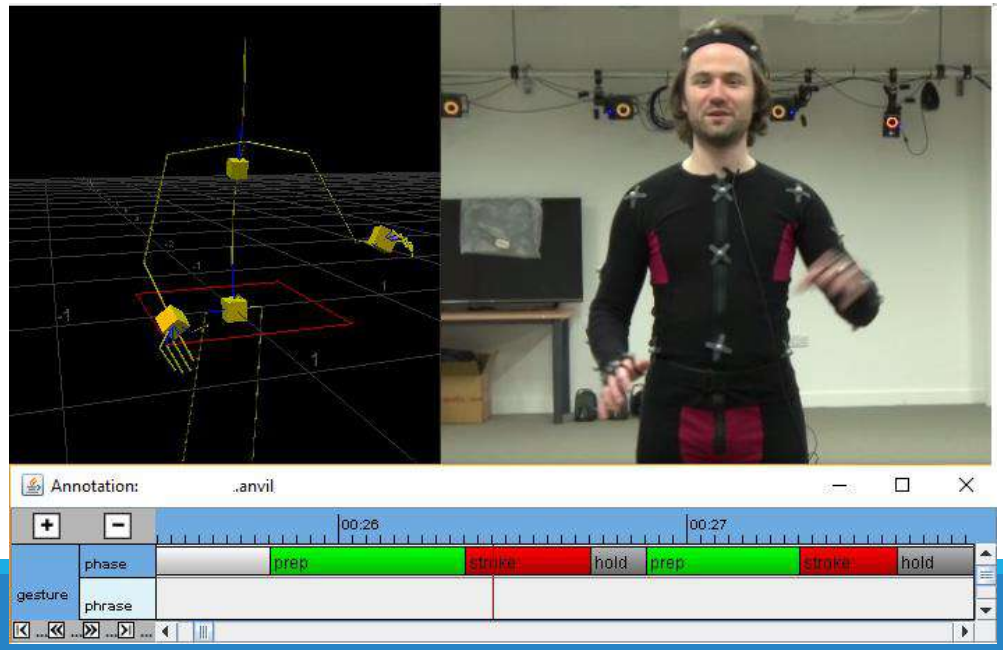




Ren, A. (n.d.). http://web.mit.edu/pelire/www/gesture-research/index.html

Phase data collection:

- Hand-annotation of 226 minutes of dataset

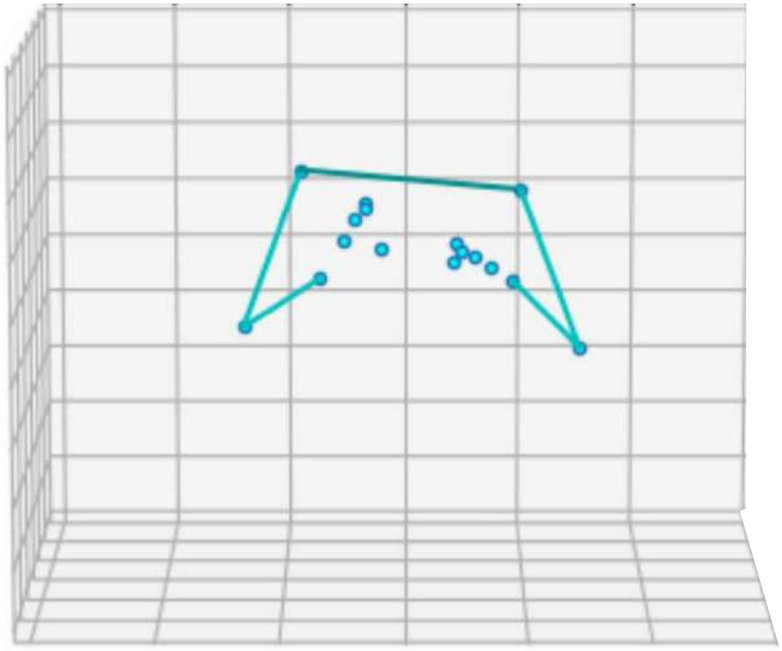


ANVIL annotation tool
[Kipp 2014]

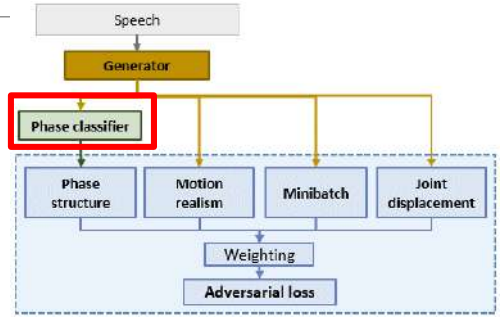
Phase classifier results

Gesture phase	F-score
	4 classes
Preparation	0.64
Stroke	0.79
Hold	0.83
Partial retract	-
Retract	-
'None'	-
'Other'	0.64
Overall	0.76

Automatic labelling of unseen data *(shown at 50% speed):*

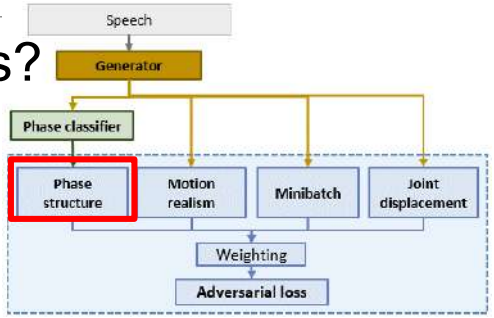


preparation



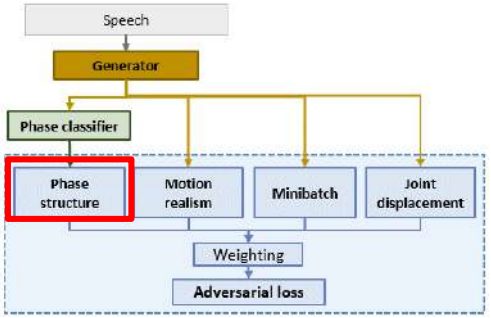
Adversarial training objectives - Phase structure

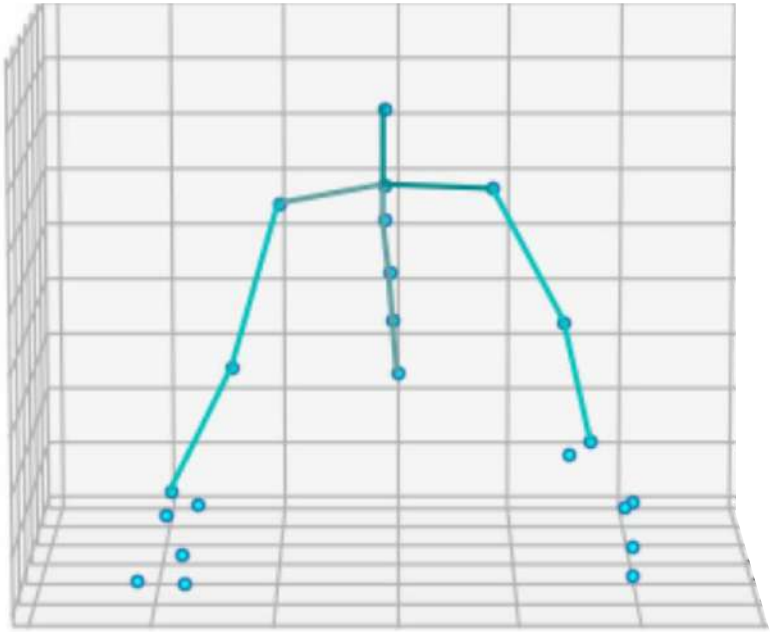
- Task: Does the motion have realistic gesture phases?
- Input: Sequence of phase labels + F0
- Output: 1/0 (real/fake)



Adversarial training objectives - Phase structure

- Task: Does the motion have realistic gesture phases
- Input: Sequence of phase labels + F0
- Output: 1/0 (real/fake)





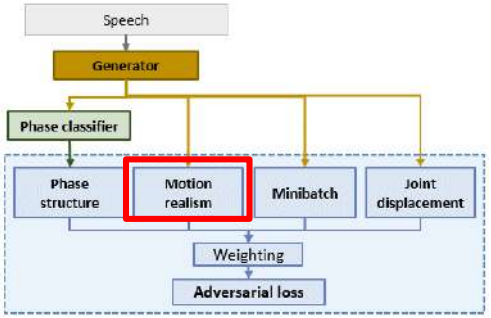
No phase discriminator

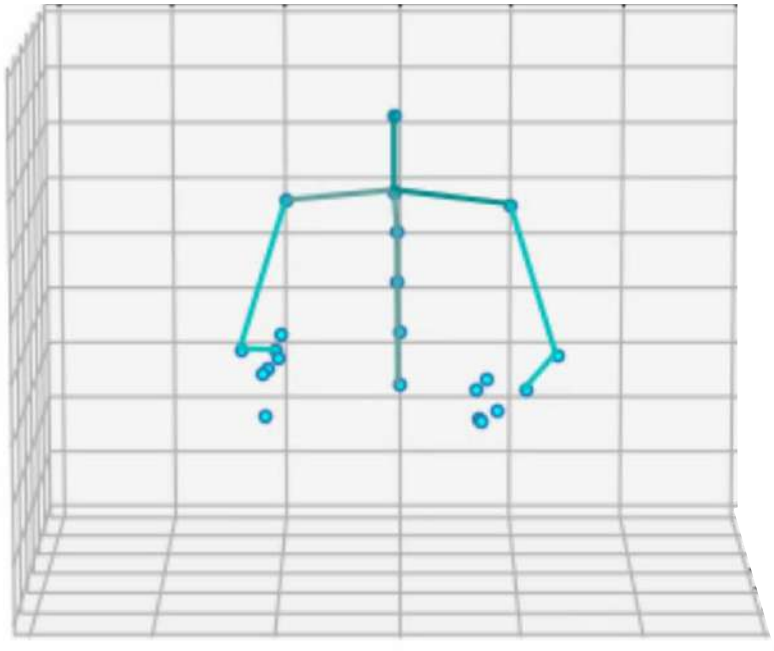
Adversarial training objectives - Motion realism

Task: Does this look like (humanoid) motion corresponding to the speech prosody?

Input: Joint positions + speech features

Output: 1/0 (real/fake)

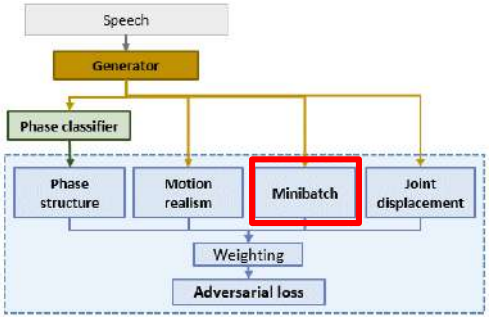


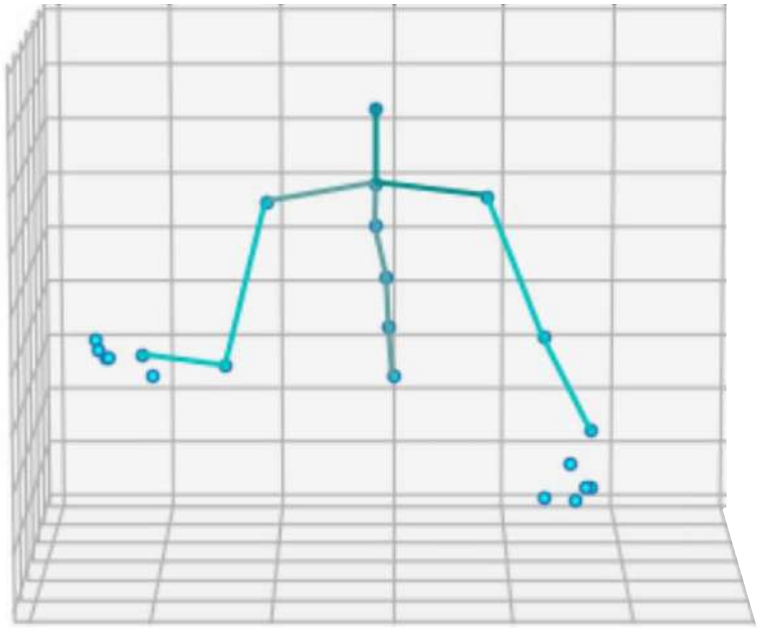


No motion realism discriminator

Adversarial training objectives - Minibatch discrimination

- Task: Detect repetitive motion
- Input: Joint positions
- Output: 1/0 (real/fake)

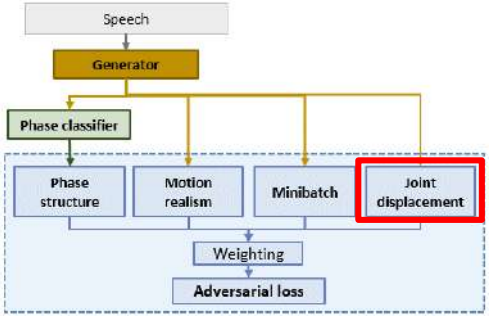


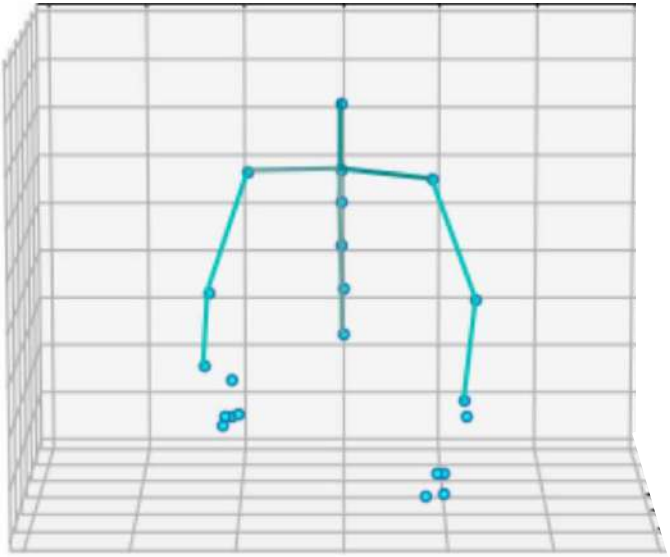


No minibatch discriminator

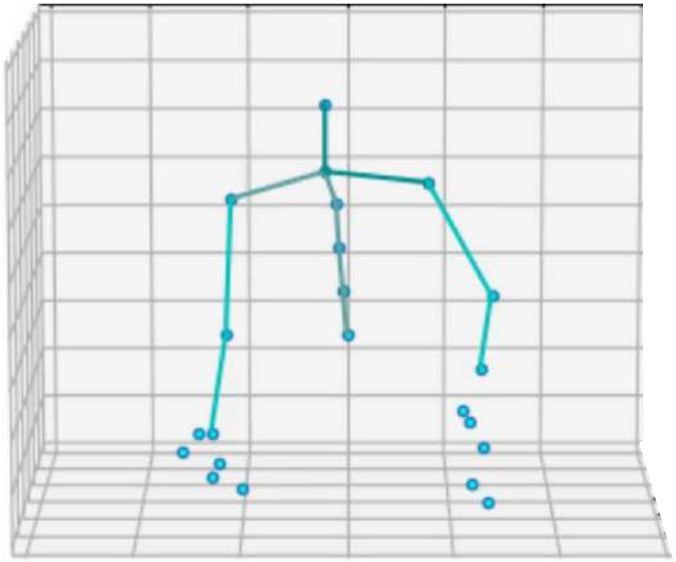
Adversarial training objectives - Joint displacement

- Task: Judge motion speed (+reduce jitter)
- Input: Per-frame joint offset
- Output: 1/0 (real/fake)





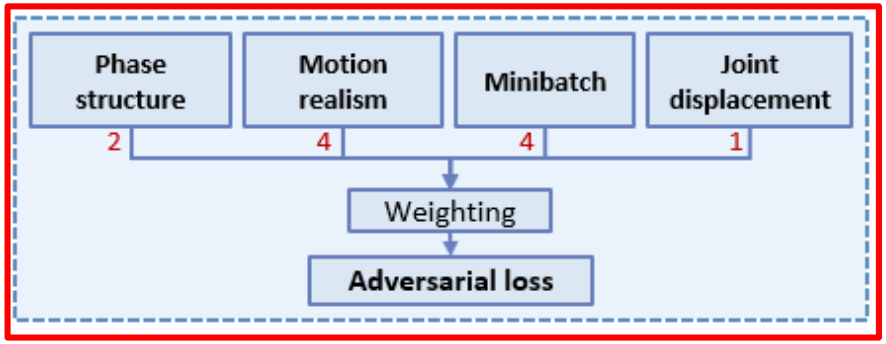
No joint displacement discriminator

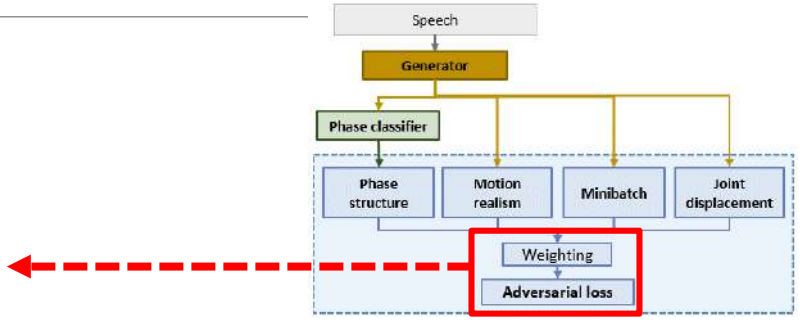


No joint displacement discriminator
+ no phase discriminator

Training loss

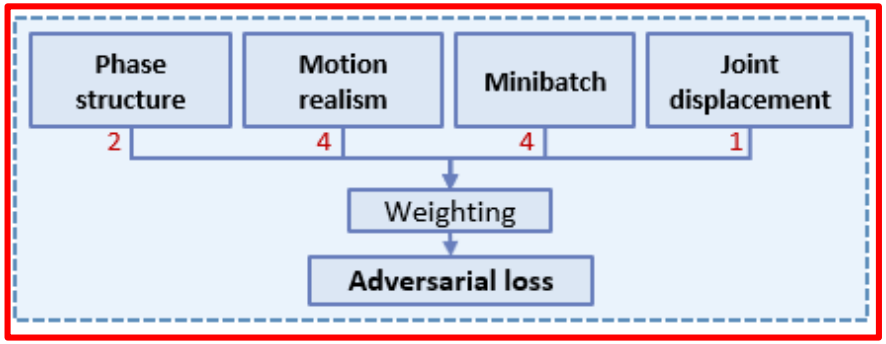
Adversarial error weighting

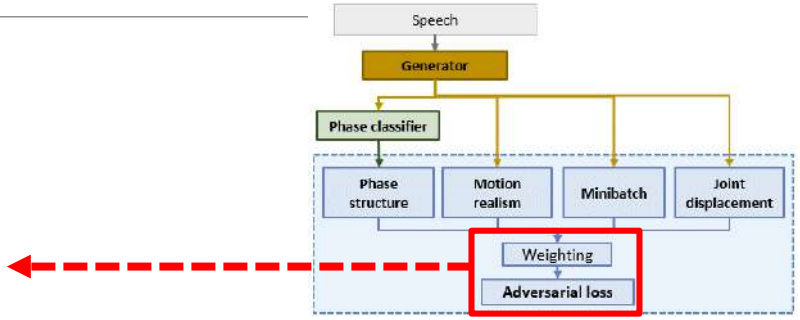




Training loss

Adversarial error weighting

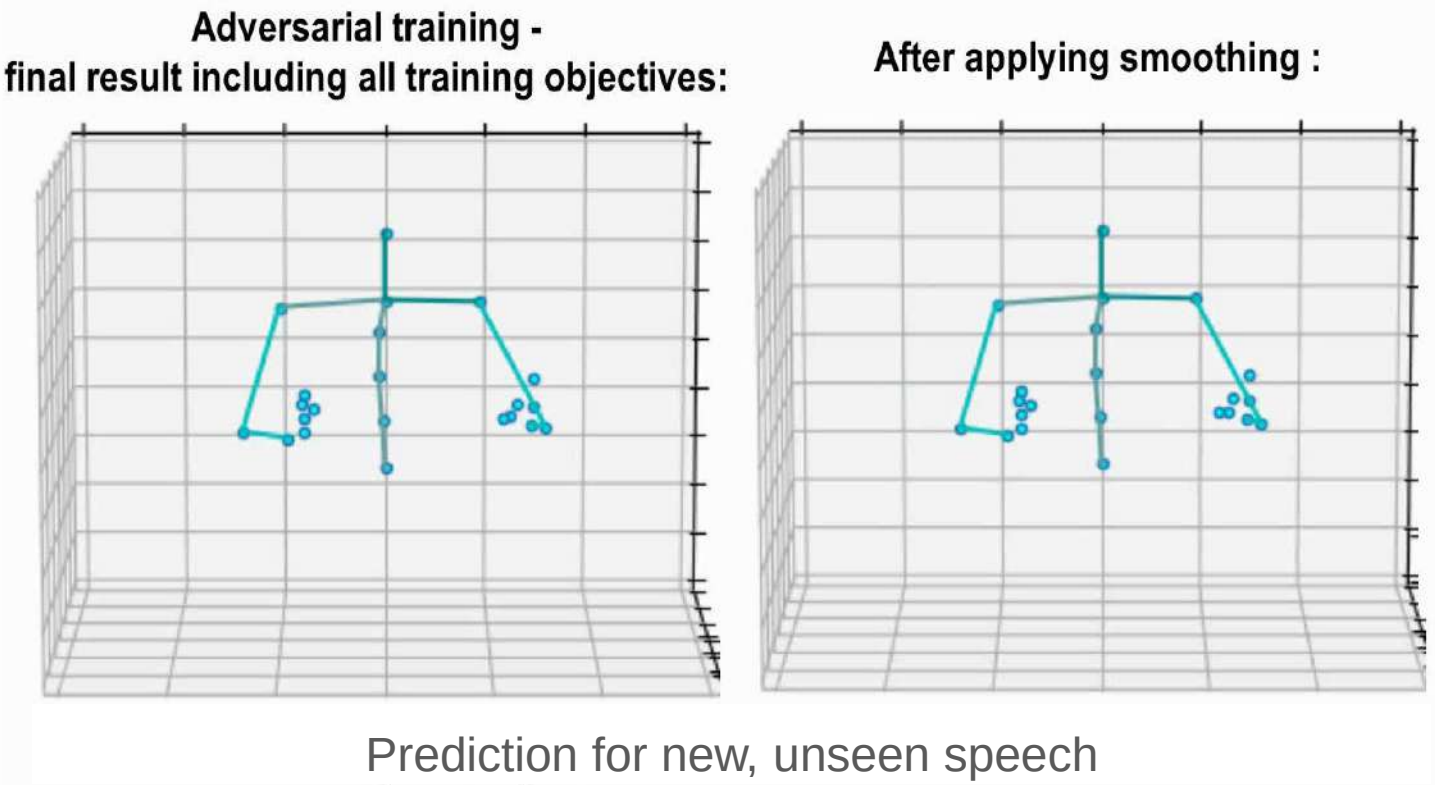




Objective loss penalties:

- Continuation loss: highly reduced with discontinuity penalty
- Finger distance: avoid unrealistic finger positions

Results: Final output: Smoothing



Prediction for new, unseen speech

Metaphoric gestures

GESTURE GENERATION FROM IMAGE SCHEMAS

BRIAN RAVENET, CHOÉ CLAVEL, CATHERINE PELACHAUD

Metaphoric Gesture

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

Eg:

Idea = object with form and location

Sideways flip of the hand → Disregard an idea

Embodied cognition theories (Kendon, Barsalou)

- same sensory and motor representations
- to perceive and act on the world

And

- To reason and communicate about abstract concepts

Metaphoric Gesture

Communication of message

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

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 hold tightly

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), plan:
  - timing (when to place a gesture)
  - shape (which gesture form and movement)
- Capture mental imagery from text and map it into gesture

Image Schemas

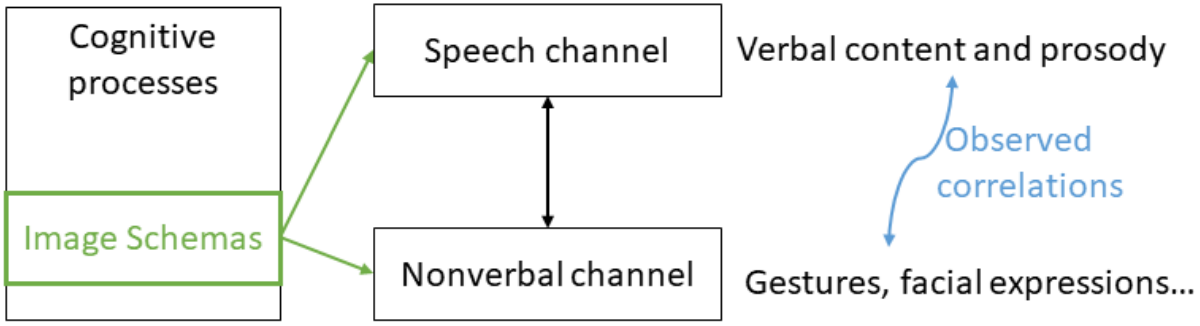
Image Schemas: allows for manipulation of spatial, temporal and compositional concepts (container vs object and whole vs split for instance).

- Image Schemas
  - gestural grammar (Mehler, Lücking, Abrami, 2015): bridge between natural language and gesticulation.
  - 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.

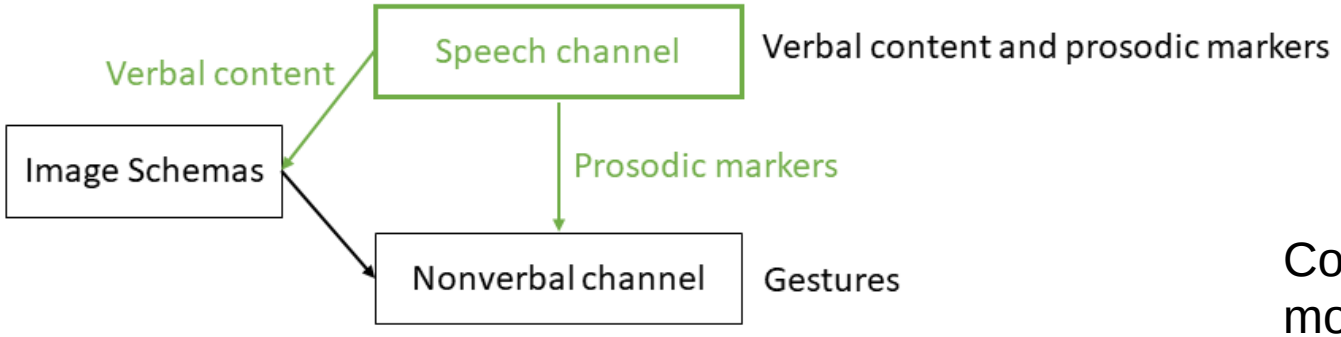
Ideational Unit

- Ideational units by Calbris (2011)
  - Hierarchical structure of discourse
  - Rhythmic-semantic group
- Gesture invariant: invariant properties that are critical for the meaning of the gesture.
- Link between Image Schema and gesture invariant
- Similarity constraint within IU
- Demarcative function within an IU:
  - Ensure following gestures are distinguishable one from another
  - Changes between them are meaningful

Computational Model

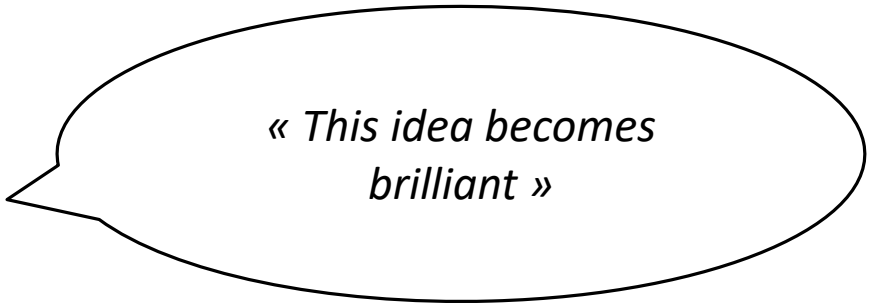


Theoretical model



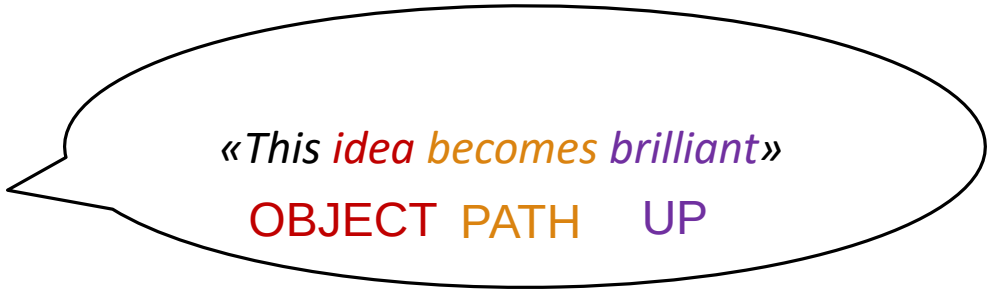
Computational model

Extract Image Schemas from the text



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

Extract Image Schemas from the text



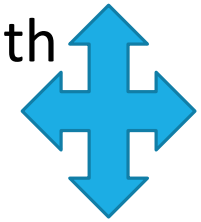
Textual speech of the agent → Extract Image Schemas → Associate Image Schemas to Gesture Primitives → Combine primitives into ideational units → Produce co-verbal metaphoric gestures

Associate Image Schemas to gesture primitives

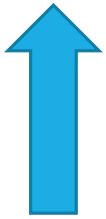
Based on Geneviève Calbris

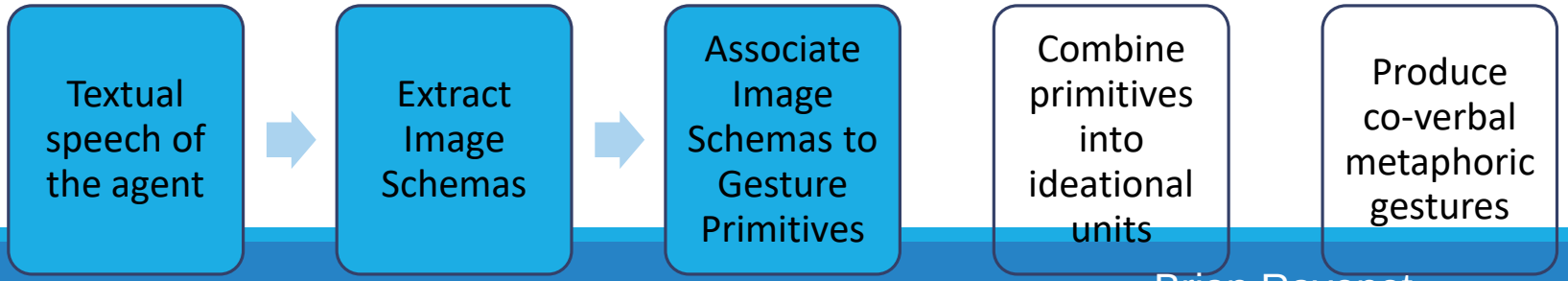
- Object

- Path



- UP



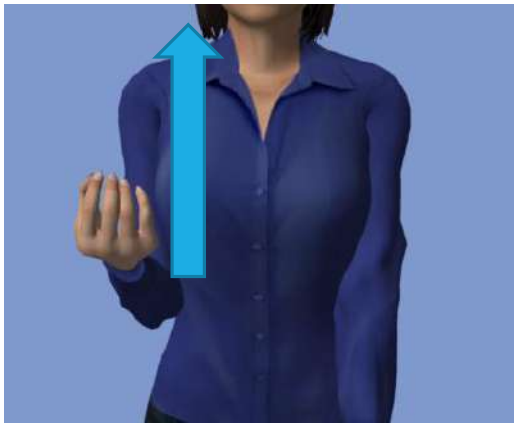


Associate Image Schemas to gesture primitives

« This idea becomes brilliant »

OBJECT

PATH

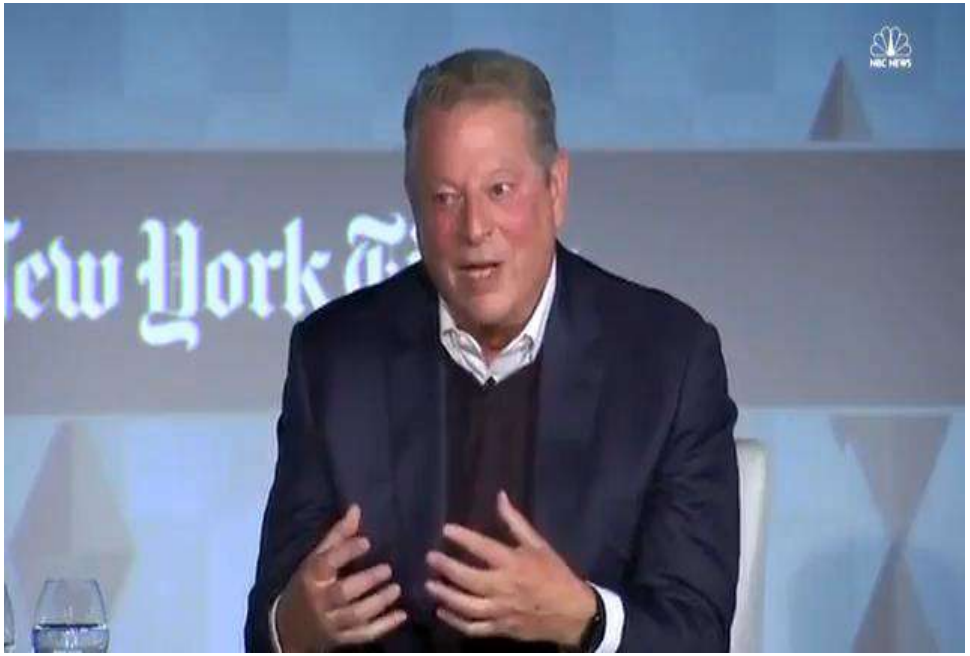


UP

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

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

Real-time interaction

Interaction

Interaction is by no means a one way communication channel between parties.

Interaction is like ‘dancing’, joint action, highly synchronized, embedded in social context

Speaker and listener adapt their behaviors to each other

- Speaker monitors addressees attention and interest in what she has to say
- Listener selects feedback behaviors to show the speaker that he is paying attention, agreeing, understanding, etc

Tight dynamic coupling between both interactants

Turn-Taking

During conversation, interactants speak and listen in turn.

Most of the times, one person speaks at the time.

While there are cultural variants, it is

Exchange of speaking turns

- Often smooth turn exchange
- How possible? Which signals involved?
- How speaker signals she ends her speech?
- How listener perceived these cues? How does he signals he wants to take the speaking turn?

Turn-taking

- Organisation of turns: allocation, construct, exchange of turns
- First study as part of conversation analysis: Sacks, Schlegoff & Jefferson in the 70's
- Mainly look at linguistic cues
- Lately NVB cues are also considered

Turn-Taking Structure

Three main components (Sacks, 1974):

- *Turn-taking component:*
  - Turn-Construction Units TCUs: content of utterances
  - Transition-Relevant-Point TRP: point where a change of turn may happen
    - Point of syntactic, intonational and pragmatic turn completion
- *Turn allocation component*
  - Current speaker selects next speaker
  - Self-selected next speaker
- *Rules:* govern turn construction, minimize gaps and overlaps.
  - Transfer turn to next speaker selected by current speaker
  - Turn grabbing by self-selected next speaker
  - No-one grabs the turn
    - Current speaker continues until next TRP
    - Silence or end of conversation

Turn-Taking: Cognitive Process

Globally, gaps between turns: short, around 200ms (between 100 and 500ms, correspond to 1-3 syllables)

Theory based on language production:

- 600ms latency minimum for language production

→Participants predict when turn is to finish to prepare their response in advance

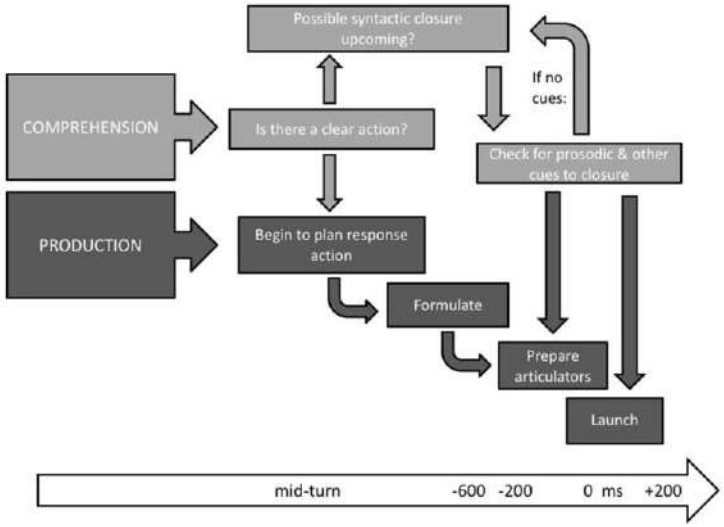
→There is some overlap between production and comprehension despite common processing resources

Listener that wants to take the turn has:

- two processes: comprehension and production, running simultaneously
- Predictions systems: one for self and one for the other

Interleaving comprehension and production in listener wanting to take speaker turn

- Word production: min 600 ms
- Pre-articulation preparation: 200 ms
  - Vocal track
  - Breathing (inhalation prior to speak)
- Comprehension:
  - Check for cues of end of turn:
    - Syntax, prosody, nonverbal behaviors



Levinson & Robeira, 2015

Turn-Taking Cues

Linguistic cues

- Syntax: clause end
- Lexical

Nonverbal behaviors

- Gaze: different patterns
  - Speaker gazes away when starting the turn and gazes at when ending it
  - Listener gazes at speaker
- Gesture: arms go to rest at the end of speaking turns
- Breathing: inhalation before starting to speak
- Posture: body gets ready to speak
- Prosody
  - Final syllable duration
  - Final drop in pitch or loudness

Backchannel signals

They serve as signalling of feedback to speaker

They are not an attempt to take the turn

They occur after a TRP (transition relevant point) or silence

They provide information about the communicative function: **[Allwood'93, Poggi'05]**

- contact
- perception
- understanding
- attitudinal reactions

Backchannel signals

Verbal and non verbal signals

Emitted in a non-intrusive way during the speaker's turn

Emitted with different levels of intentionality

- indicative, displayed and signalled

They depend strongly on culture

- shape
- frequency

Backchannel signals: Lexicon

agree and **accept**: *nods*

like: *smile*

understand: *smile* and *raise eyevrows*

interested: *nod* and *raise eyebrows*

disagree and **refuse**: *shake*

dislike: *frown* and *tension of the lips*

don't understand: both *tilt and frown*; *frown*

not interested *eyes roll up; tilt and gaze away*

disbelieve: *tilt and frown* and *raise of the left eyebrows*

Mimicry

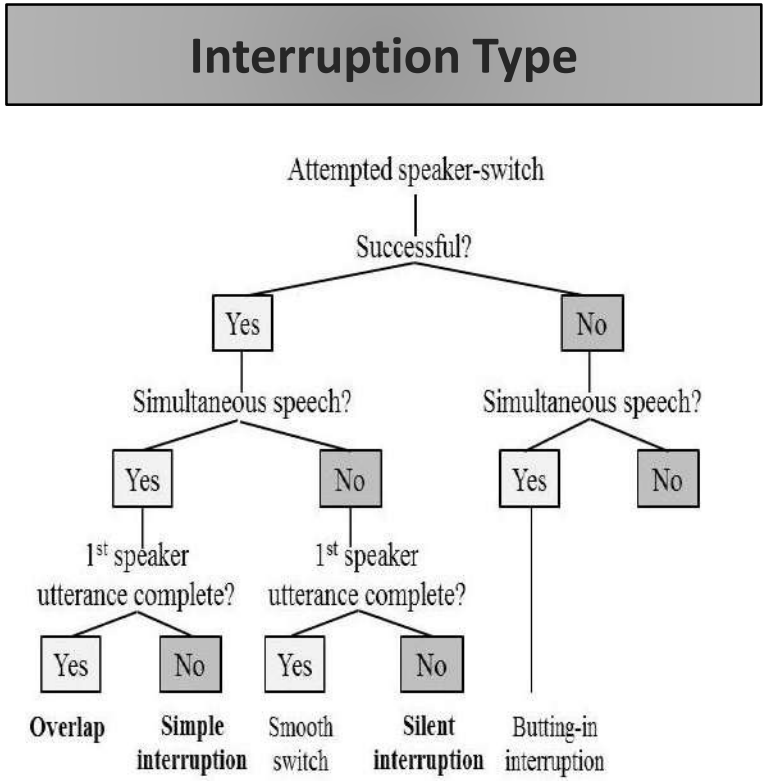
Imitation of the other's behaviour

Synchronization of behaviours

In human interactions mimicry can happen:

- acoustic [Giles'75, Webb'72, Cappella'81...]
  - accent, speed, pause duration...
- visual [Chartrand'75, Warner'96...]
  - type of movement
  - quality of movement
- As a sign of empathy[Neumann'00, Hatfield'94...]
  - emotional state

Interruptions



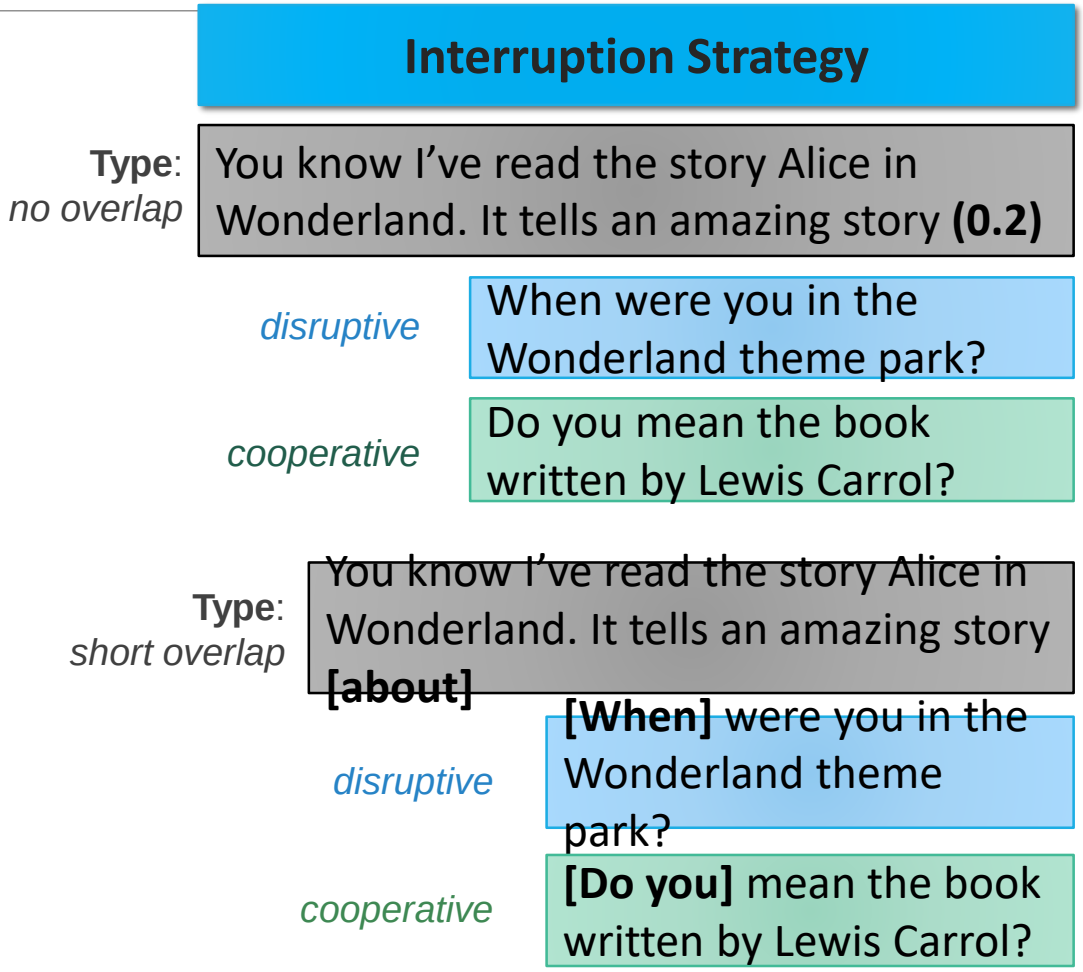
Interruption Types: Categorisation based on timing and overlap of turns (Beattie 61)

- Silent interruption: First utterance incomplete & No overlap
- Simple Interruption: First utterance incomplete & Overlap (shorter)
- Overlap: First utterance complete & Maximum overlap

Interruptions

Interruption Strategies: Based on content of interruption (Murata 94)

- Cooperative interruptions: intended to help the speaker by coordinating the process/and or content of the ongoing conversation."
  - E.g. Agreement, assistance, clarification, …
- Disruptive interruptions: pose threats to the current speaker's territory by disrupting the process and/or content of the ongoing conversation."
  - E.g. Disagreement, floor taking, topic change, …



Glas&Cafaro)

Turn-Taking systems

Previous models:

- Talkie-walkie type:
  - one person at a time
  - Interpretation of speaker's speech at the end of turn
  - Generation of next speaker's speech after end of turn detection

Prediction of end of turn based on turn-taking cues

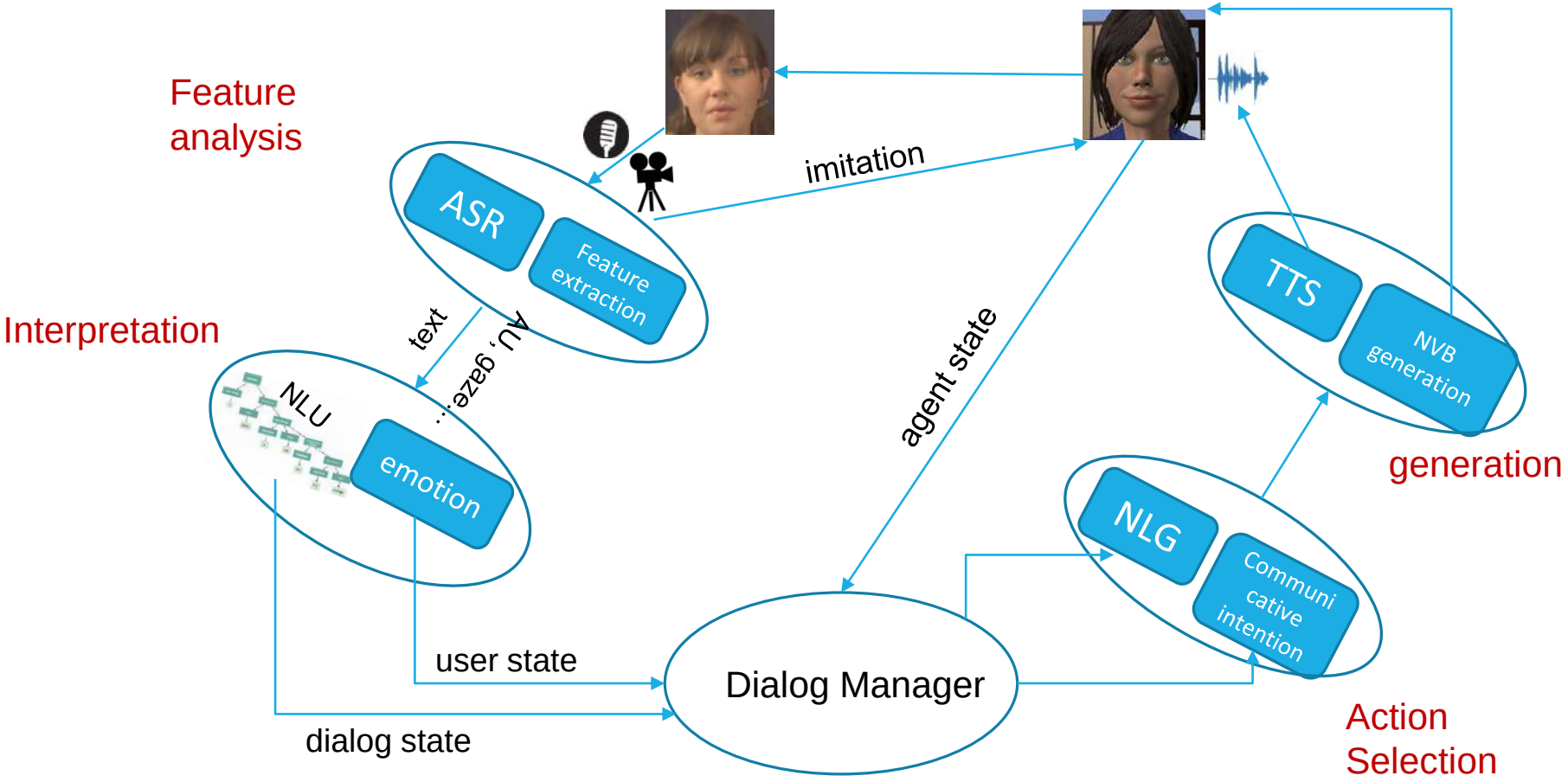
Incremental language processing model

Incremental dialog system

Human-Agent System

- Integrate components in a real-time architecture
  - audio and video analysis
    - low-level feature extractors
    - classifiers for epistemic-affective states
    - ASR
  - action planning
    - turn-taking
    - generate natural language utterances
    - verbal and non-verbal backchannels
    - continuous NVB expressivity – may involve mimicry
  - system behaviour
    - Nonverbal behavior, TTS

Human-Agent System



SEMAINE Project

SEMAINE
THE SENSITIVE AGENT PROJECT

"Sustained Emotionally coloured Machine-human Interaction using Nonverbal Expression"

Sensitive Artificial Listener:

- Subjects dialog with characters
- 4 Characters with 4 distinct emotional traits
- Goal of the application is to induce emotions in subjects

Our goal: model ECAs with different emotional traits that can:

- produce appropriate backchannels
- sustain an emotionally colored communication with a user

Concept of SAL

- Four characters with an emotional agenda
  - Obadiah tries to make users sad
  - Spike tries to make them angry
  - Poppy tries to make them happy
  - Prudence tries to make them reasonable

Building SAL characters

- Consider trait models of personality
  - Influence behaviours
  - Stable, fundamental properties of an individual
  - Link to emotional traits
- Use of Eysenck's model: Extraversion and Neuroticism
- Individual differences in predisposition to experience certain emotions
  - Neuroticism - negative emotionality
  - Extraversion - propensity towards positive emotion
- From personality to behaviour
  - Eg. Extraversion:
    - linked with positive affect, associated with general level of activation & behavioural approach

Representing distinctive characters

Baseline – agent's global behaviour tendency

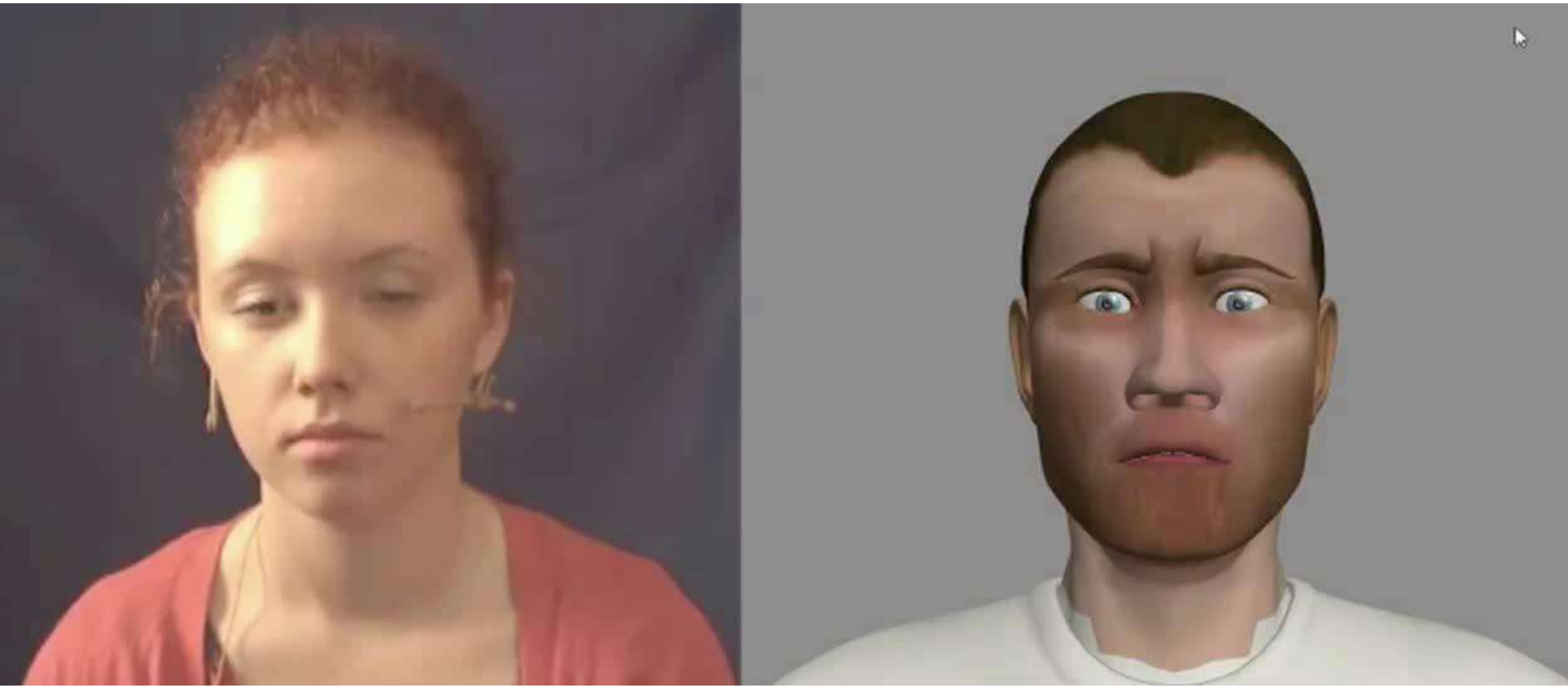
- modality preference…
  - e.g. face, head, gaze, gesture, torso
- …and behaviour expressivity
  - parameters influencing quality of movement - frequency, speed, spatial volume, energy, fluidity, & repetitivity
- expressivity parameters defined for each modality

Agent's mental state

- Tendency to display given back-channels (eg positive vs negative)

Obadiah

EXAMPLE - SPIKE



Joint action – Clark, 96

Signaling: A acts so B gets a meaning.

- coordinating actions between interactants
- include linguistic and nonverbal signals
- Speaker-A acts an intention
- Addressee-B recognizes it

4 levels of shared action: action ladder, from level 1 to level 4

Speaker A's part	Addressee B's part
4 - A is proposing a joint project *w* to B	B is considering A's proposal of *w*
3 - A is signaling that *p* for B	B is recognizing that *p* from A
2 - A is presenting signal *s* to B	B is identifying signal *s* from A
1- A is executing behavior *t* for B	B is attending to behavior *t* from A

Joint Action - example

Examples: A goes to B; B says “I’ll be there”

Level 1: A and B have engaged on a joint action where B says something and knows that A will listen

Level 2: A and B are similarly engaged in a joint action where B utters the words "I'll", "be" and "there" knowing that A will identify them

Level 3: B knows that A is engaged in recognizing this signal as a proposition

Level 4: B's part in this joint proposal is to wait, A's part is to finish what he or she is doing.

Joint Action

3 methods of signaling:

- Describing-as
- Indicating
- Demonstrating

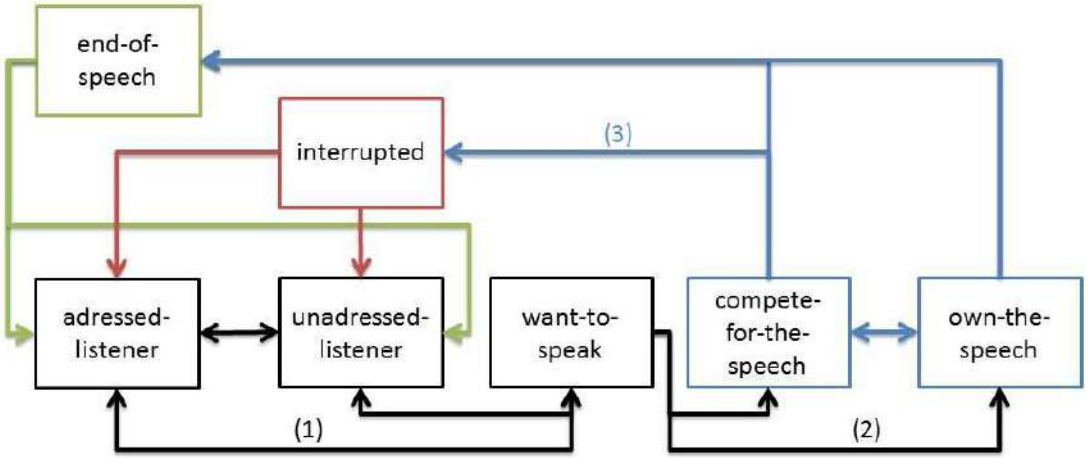
Instrument	Describing-as	Indicating	Demonstrating
Voice	Words, sentences	“I”, “Here”	Tone of voice
Hands, arms	Emblems	Pointing	Iconic
Face	Facial emblems	Pointing	Smiles
Eyes	Winking	Eye gaze	Wide eyes
Body	Junctions	Pointing	iconic

Example: A sees B

- *Describe-as*: A says ‘Hello’
- *Indicate*: A speaks ‘hello’ and gazes at B
- *Demonstrate*: A smiles, raises eyebrow; B interprets ‘A greets B with enthusiasm’

Turn Taking Model

Ravenet



A state-machine allowing the agent to change between speaking and listening state

Based on Clark's joint-action theory

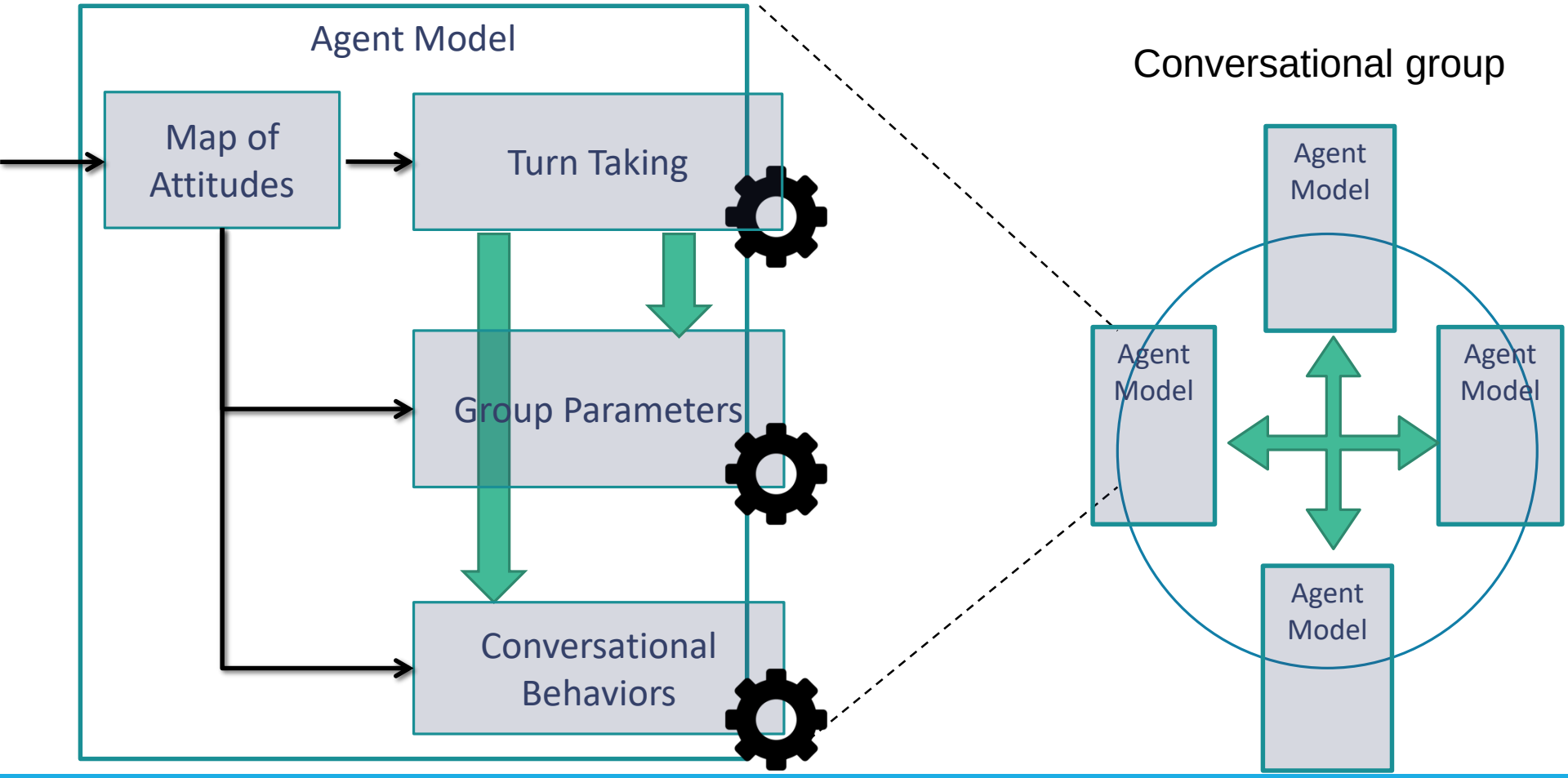
- A turn is the result of the success of the actions Speaking and Listening

The model uses the activities of others as well as the expressed attitudes to trigger the transitions between the different states following rules extracted from the literature.

The model allows overlapping as well as interruption

Group Model

Ravenet



Group Model: group formation and cohesion parameters

Ravenet

Three parameters identified for managing the group cohesion and formation.

- The interpersonal distance between each agent:
  - A difference in status or an hostile attitude leads to a higher distance.
- The gaze behavior
  - A friendly or dominant person receives more gaze.
- The body orientation
  - A dominant person receives more direct orientation.

Implementation

Ravenet

Greta

Unity3D

Impulsion (Pedica, Vilhjalmsson)

Experimental Design

Ravenet, Cafaro, Biancardi

Friendly

dominant

Adaptation mechanisms

BEATRICE BIANCARDI

SOUMIA DERMOUCHE

Interaction

Human-human interaction and human-agent interaction involve:

- Exchange of social signals
- Imitation
- Synchronization
- Adaptation
- Rapport building
- Impression management
- Engagement
- Interpersonal relationship
- ….

Adaptation

Function:

- Favor engagement
- Enhance user's experience
- Signal interpersonal relationship
- Increase rapport
- …

Can be signalled through:

- Imitation, backchannel, verbal alignment, synchronization

Different levels:

- strategy
- behavior
- cue

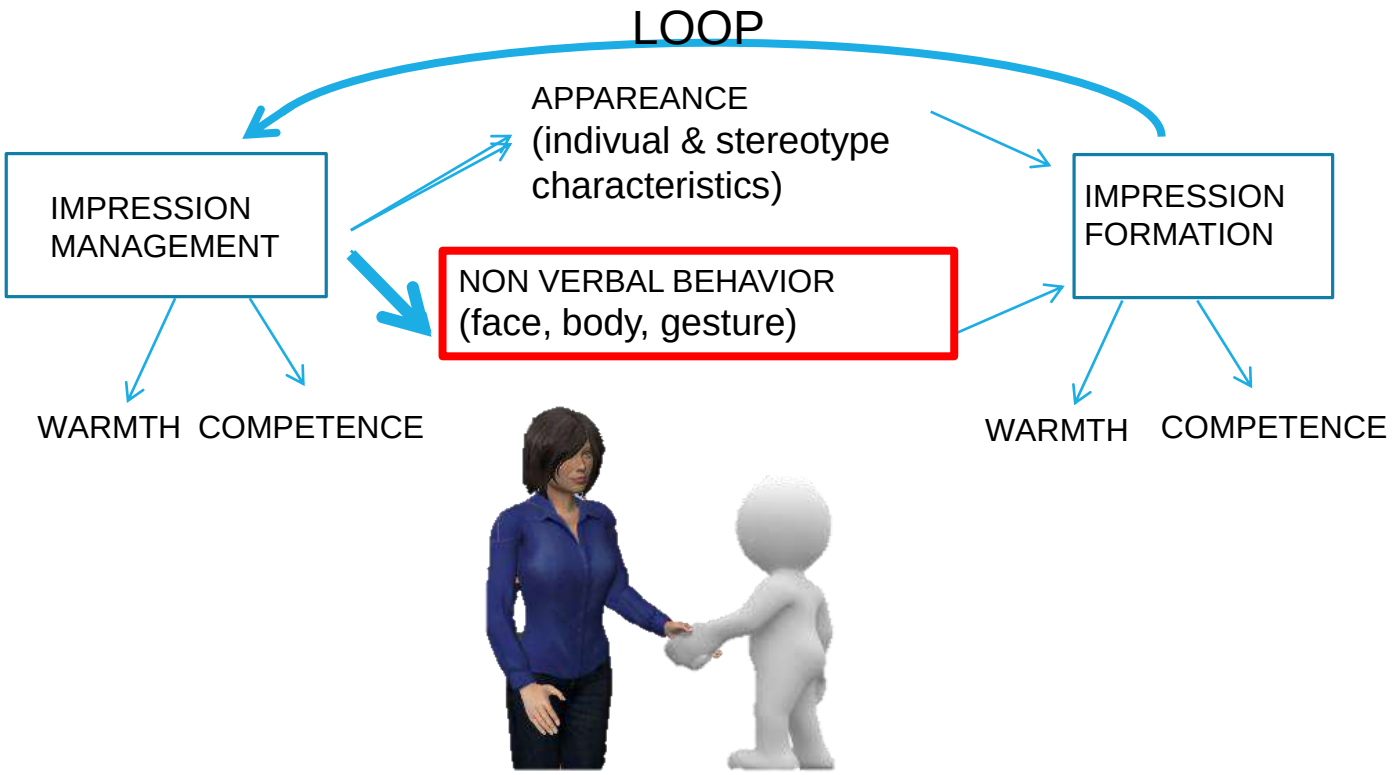
Adaptation

Conduct studies where agent adapts to user to enhance engagement

3 studies

- Adaptation at nonverbal behaviors level
- Adaptation at conversational strategies level
- Adaptation at cues level

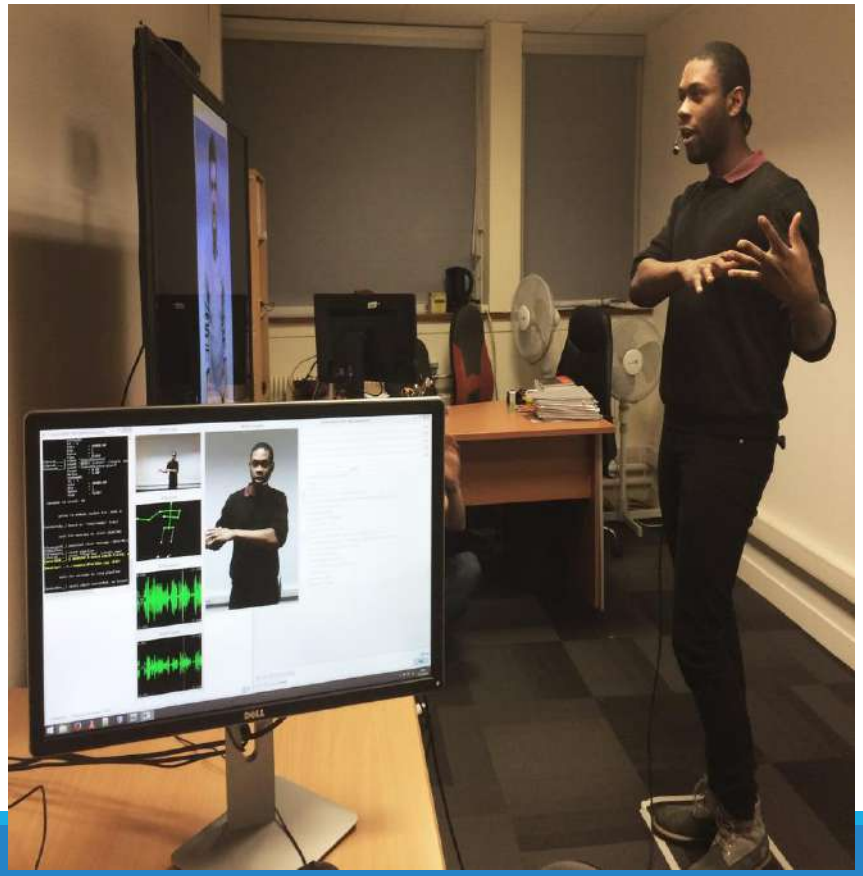
Impression Management
(Biancardi et al)

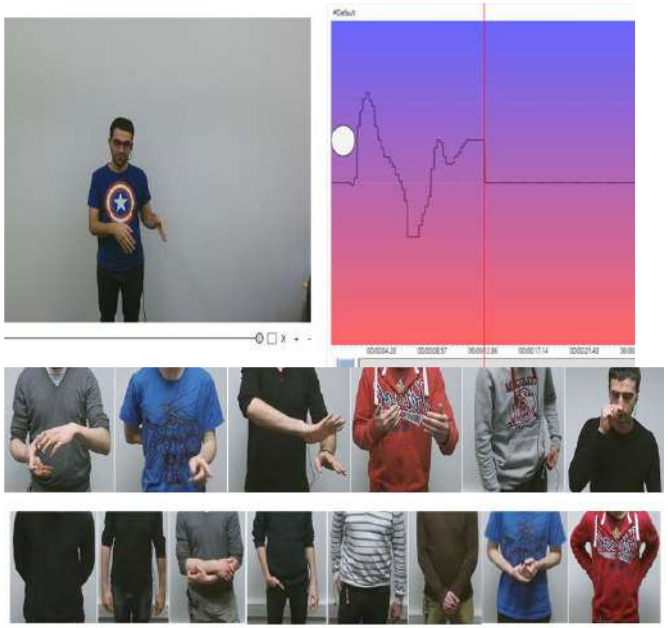


Impression of W&C in Human

Corpus analysis of human-human interaction

- What are the nonverbal cues associated to the impression of warmth and competence





Impression of W&C in Human

ARMS REST POSES	no_rest_poses	arms_down	arms_behind
Warmth	2.2 ****	0.60 **	0.18 ****
Competence	1.6 ****	0.80 n.s.	0.83 n.s.
	arms_crossed	hands_crossed_down	hands_crossed_middle
Warmth	0.08 **	1.36 n.s.	1.00 n.s.
Competence	0.27 ***	1.46 n.s.	1.00 n.s.
	hands_on_hips	hand_on_hip	hand_in_pocket
Warmth	3.6 n.s.	0.60 n.s.	1.23 n.s.
Competence	1.5 n.s.	0.90 n.s.	0.4 **

TYPE OF GESTURES	beat	ideational	adaptor
Warmth	1.4 *	3.09 ***	0.84 n.s.
Competence	1.6 **	1.3 n.s.	0.88 n.s.

SMILING	smile
Warmth	9.67 ****
Competence	0.64 ****

compensation effect

Odds ratios: measure of association

Red: positive association

Blue: negative association

Impression of W&C in Agent

Experimental study

Research questions:

- Is a virtual agent perceived differently in terms of W&C according to its NBV?
- If so, what are the NVB associated to the perception of W&C?
- Do our expectations and a-priori of an ECA influence our impressions?

Independent Variables

Independent Variables: Agent description

Hypothesis: Agent's descriptions (autonomy/human-driven) affect users' judgments about the agent.

Between-subject experiment

Agent

Avatar

Independent Variables: type of gestures, frequency of gestures, rest poses, frequency of smile

hypothesis: similar results as in human study

Within-subject experiment

Independent Variables

Dependent Variables

Warmth
• Kind • Pleasant • Friendly • Warm

Competence
• Competent • Effective • Skilled • Intelligent

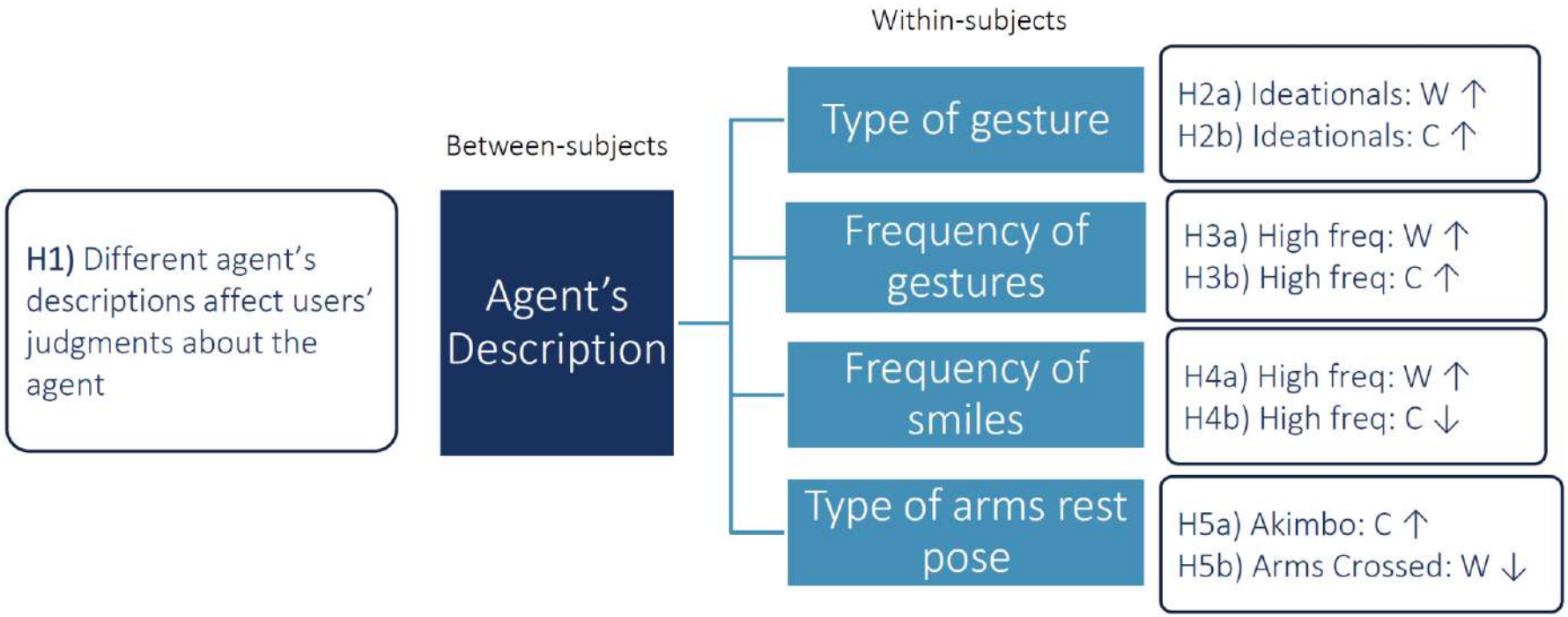
(Aragonés et al., 2015)

Please rate, on a scale from 1 (Not at all) to 7 (Extremely), how much Alice is:

E.g. kind

Not at all						Extremely
1	2	3	4	5	6	7
○	○	○	○	○	○	○

Experimental Design



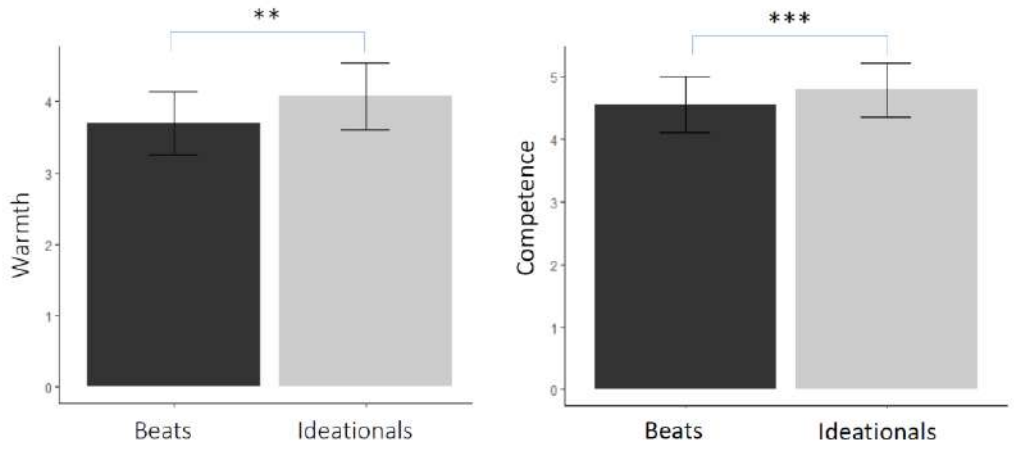
Results

Effect on type of gestures

Within-subjects

Type of gesture

H2a) Ideationals: W ↑
H2b) Ideationals: C ↑



Effect on frequency of gestures

Frequency of gestures

H3a) High freq: W ↑
H3b) High freq: C ↑

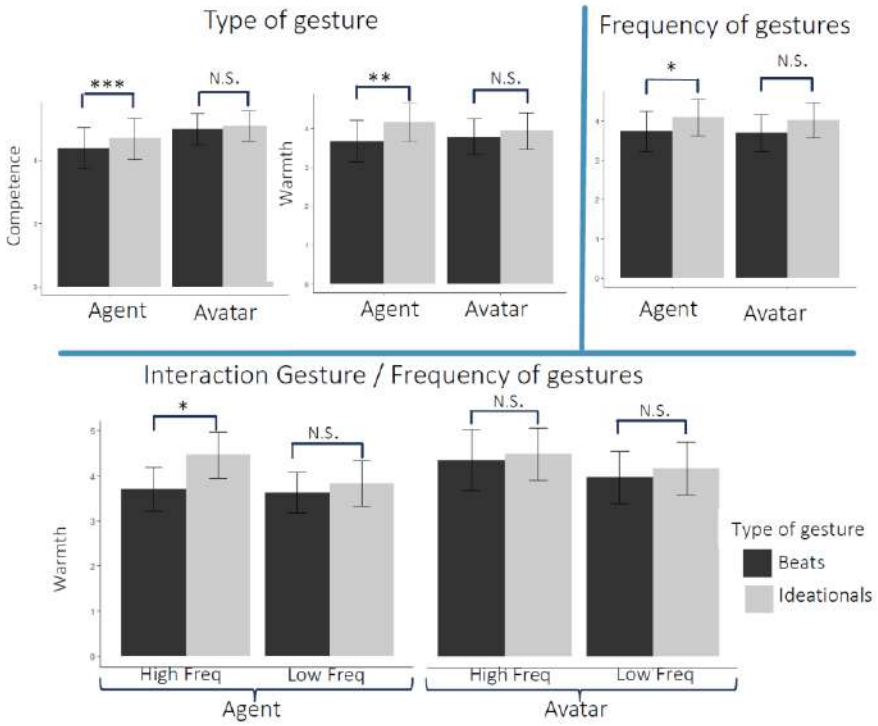
No compensation effect of the smile

Results

H1) Different agent's descriptions affect users' judgments about the agent

Between-subjects

Agent's description



- AGENT: Significant effect of gesture types and rest poses

Adapting agent's behaviour according to user's impressions.

Study: investigate the effect of adapting agent's behaviour to detected user's impressions.

(Q1) *Is it possible to elicit different impressions of W&C by adapting the agent's behavior according to the detected user's impressions?*

(Q2) *Is it possible to influence user's perception of the interaction by maximizing agent's warmth (or competence) during the interaction?*

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

Adapting agent's behaviour according to user's impressions

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

Detection of user's impressions of agent's warmth and competence → from the analysis of their facial expressions.

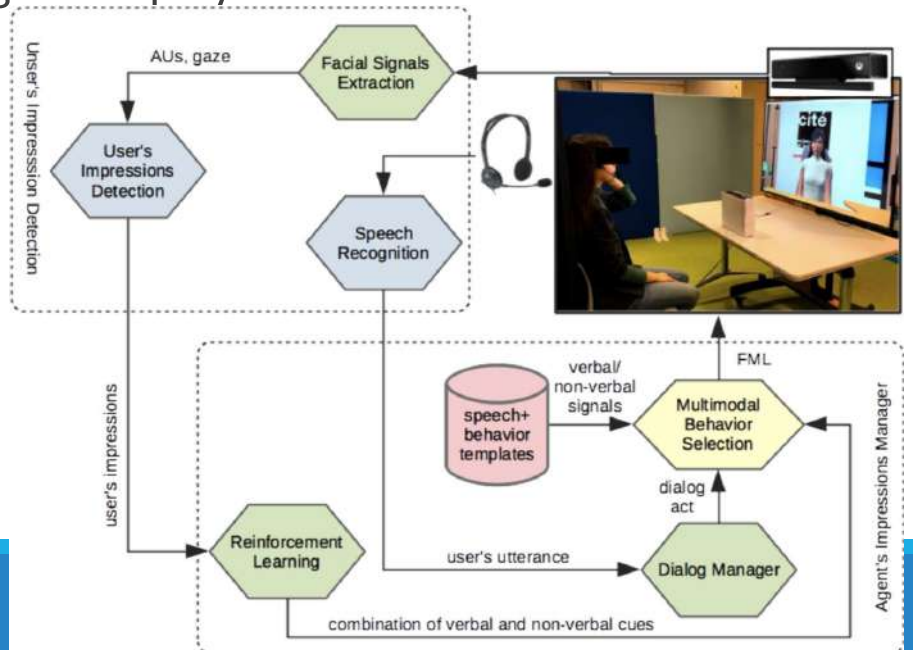
Agent's impression manager: Flipper, dialog manager + Reinforcement learning

→ verbal and non-verbal behaviors agent displays next

→ rewards: detected impressions

Communicative Intention

- Types of gestures
- Arm rest position
- Smile
- Verbal



Adapting agent's behaviour according to user's impressions

Experiment

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

- 71 participants
- 34% women
- 28% in range of 18-25 years old, 18% in range 25-36, 28% in range 36-45, 15% in range of 46-55, and 11% over 55 years old
- 3 conditions: Warmth, Competence, Random

pre-questionnaire: NARS

post-questionnaires:

- Perception of W & C
- Perception of interaction

Pre-questionnaire

NARS

Items
1. I would feel uneasy if virtual characters had emotions.
2. I would feel relaxed talking with virtual characters.
3. I feel comforted being with virtual characters that have emotions.
4. The word "virtual character" means nothing to me.
5. I would hate the idea that virtual characters were making judgements about things.
6. I would feel very nervous just standing in front of a virtual character.
7. I would feel paranoid talking with a virtual character.
8. I am concerned that virtual characters would be a bad influence on children.

Nomura et al. (2006)

Independent Variables

WARMTH: the agent adapts its behaviors according to user's warmth impressions, with the goal to maximize her impression of warmth.

COMPETENCE: the agent adapts its behaviors according to user's competence impressions, with the goal to maximize her impression of competence.

RANDOM: the agent randomly chooses its behaviors without considering user's reactions.

Post-questionnaire

Dependent Variables

Warmth

- Kind
- Pleasant
- Friendly
- Warm

Competence

- Competent
- Effective
- Skilled
- Intelligent

(Aragonés et al., 2015)

Overall perception of the interaction

Measure	Question
satisfaction	*I am satisfied with my interaction with Alice.*
continue	*I would like to talk with Alice again.*
like	*I liked Alice.*
learnfrom	*I have learned something from Alice.*
expo	*Alice made me want to visit the exposition (if you haven't yet)*
rship	*I would describe Alice as a complete stranger vs a close friend.*
likeperson	*I would describe Alice just as a computer vs like a person.*

Bickmore et al. (2011)

Hypotheses

H1: the agent would be perceived *warmer* when it adapts its behaviors according to user's warmth impressions; that is, in the WARMTH condition compared to the RANDOM condition.

H2: the agent would be perceived *more competent* when it adapts its behaviors according to user's competence impressions; that is, in the COMPETENCE condition compared to the RANDOM condition.

H3: when the agent adapts its behaviors, either in WARMTH or COMPETENCE condition, this would improve user's overall experience compared to the RANDOM condition.

Results

Warm condition:

- No significant difference but a trend in the perception of warmth and competence compared to the Random condition
- Significant differences in regard to NARS: apriori effect
  - Positive apriori participants found agent's warmer than negative apriori ones

Competence condition:

- Significant difference: the agent appears more competent than in the Random condition

Adaptive agent: higher impression of W & C; but not always significant differences

Results

Example:

to maximize warmth, the system learned in the previous video:

- Gesture type: No gesture
- Arm rest: Akimbo
- Smile: yes
- Verbal: supplication

Participant's perception from questionnaire:

- warmth: 3.75 (Max 5)

Conclusion

Adaption at behavior level to maximize participant's impression

Capture behaviors that play a role in impression formation

Adaptation at speaker turn

But

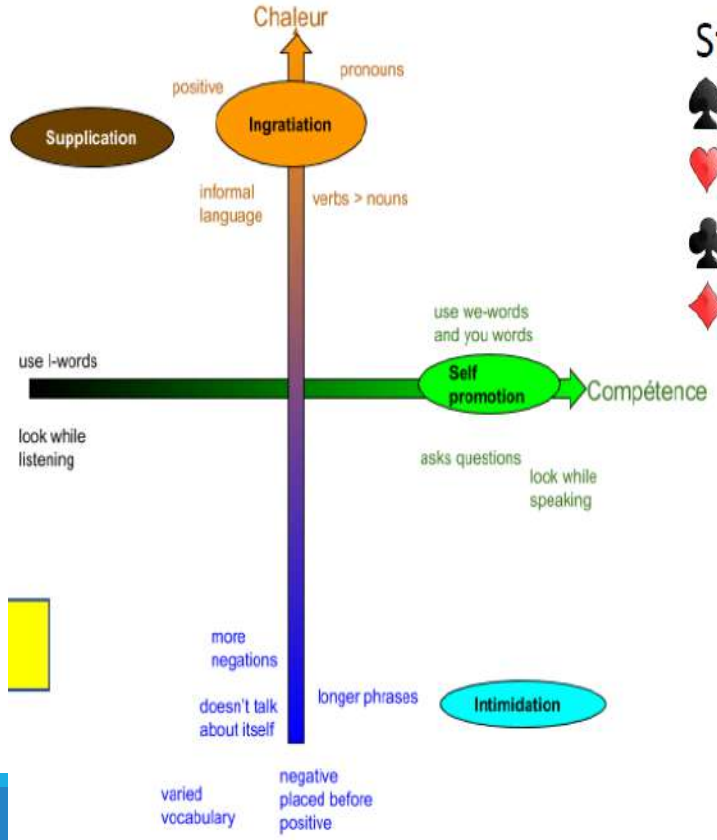
No assured behaviors coherency across turns

No consideration of conversational strategy

Study 2: Adapting agent's strategies according to user's engagement

Use of verbal and nonverbal behaviors strategies to act on other's impression of self

They influence perception of Warmth and Competence



Strategies :

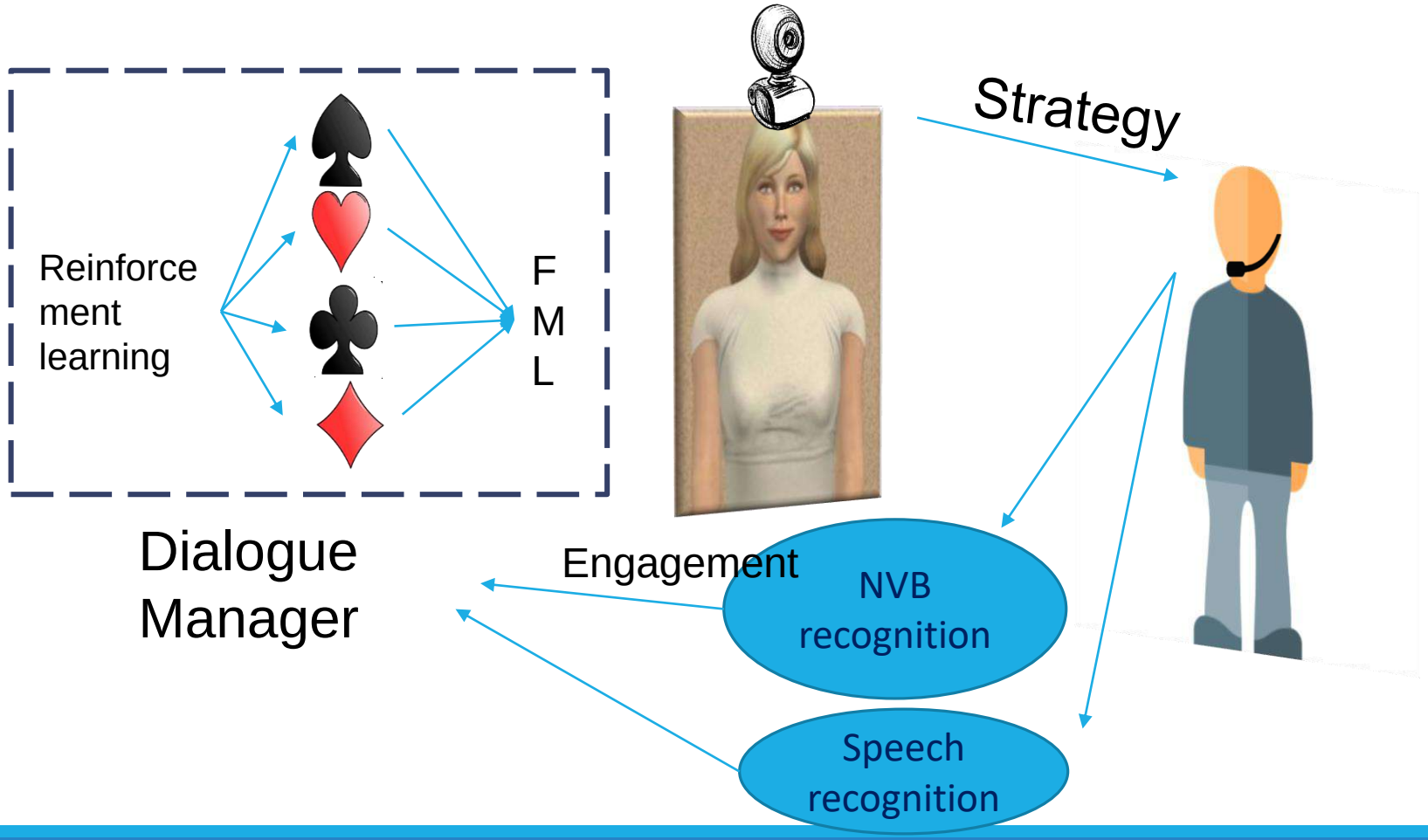
- ♠ Ingratiation — **Wa +; Co N/A**
- ♥ Supplication — **Co ↘ ; Wa ↗**
- ♣ Self-promotion — **Co +; Wa N/A**
- ♦ Intimidation — **Wa ↘ ; Co ↗**

Strategy	Translated sentence
Ingratiation	"You can test some games, if you wanna."
Supplication	"I dunno about the other exhibits of the museum, but here you can test some games, it's cool!"
Self-promotion	"In this exhibit, you can test some videogames."
Intimidation	"In this exhibit, you can try out some games on different platforms."

Jones and Pittman (1982)
Pennebaker (2011)
Callejas et al. (2014)

Adapting agent's strategies according to user's engagement

Agent's goal: maintain human engaged in interaction



Experiment

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

- 75 participants (30 female)
- The majority in the 18-25 or 36-45 age range

Pre-Questionnaire: NARS

Post-Questionaire

Independent Variables

- IMPR: adaptation model
- RAND: no model; random choice
- INGR: no model; ingratiation
- SUPP: no model; supplication
- SELF: no model; self-promotion
- INTIM: no model; intimidation

Hypotheses

H1ingr: the agent in INGR condition would be perceived as *warm* by users

H1supp: the agent in SUPP condition would be perceived as *warm* and *not competent* by users

H1self: the agent in SELF condition would be perceived as *competent* by users

H1intim: the agent in INTIM condition would be perceived as *competent* and *not warm* by users

H2a: the scrores of the overall perception items would be higher in the IMPR condition compared to all other strategies

H2b: the agent in IMPR condition would influence how it was perceived in terms of W&C

Results

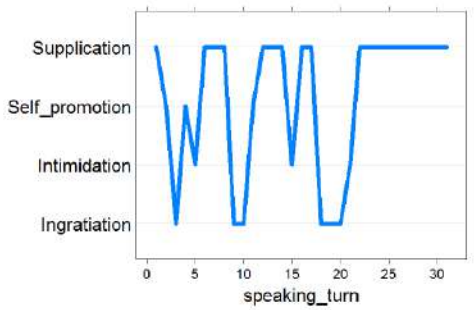
Primacy of warmth dimension:

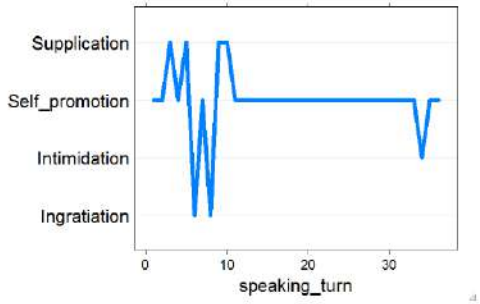
- Supplication, ingratiation: agent appears warmer
- Self-promotion: same level of warmth as supplication and ingratiation → halo effect with competence

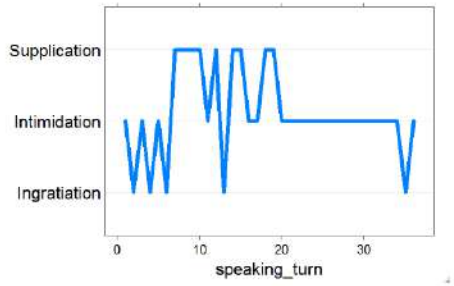
Stronger impact of negative impression over positive one (Peeters and Czapinski, 1990)

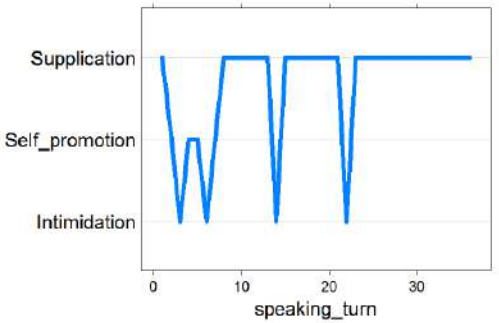
- Intimidation: agent appears colder

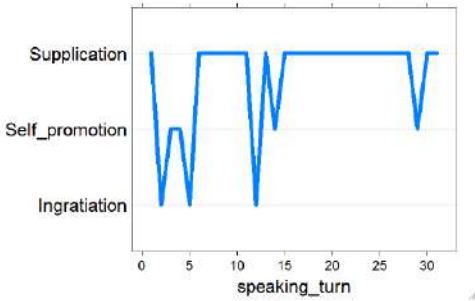
Results

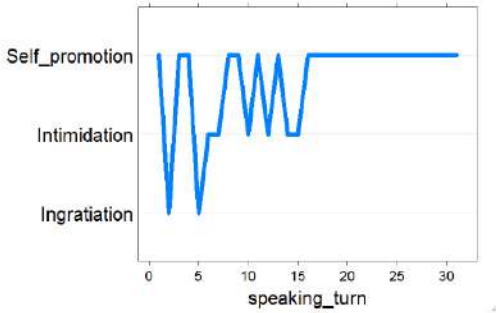












Convergence: 52±22 %

Conclusion

Adaption at strategy level to maximize participant's impression

Adaptation at speaker turn

But

No consideration of rapport building

No model of synchronization nor imitation

No explicit adaptation at participant's signal level

Adapting Agent's behavior to User's cues

Aim: develop a computational model that

- Captures the adaptation of interactants at the signal 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

Corpus

Corpus: NoXi

Automatic annotations:

- Smile activation and intensity
- Head rotation
- Gaze direction

Semi-automatic annotation

- Conversational state: who speaks

Prediction model

IL-LSTM: Interaction Loop LSTM

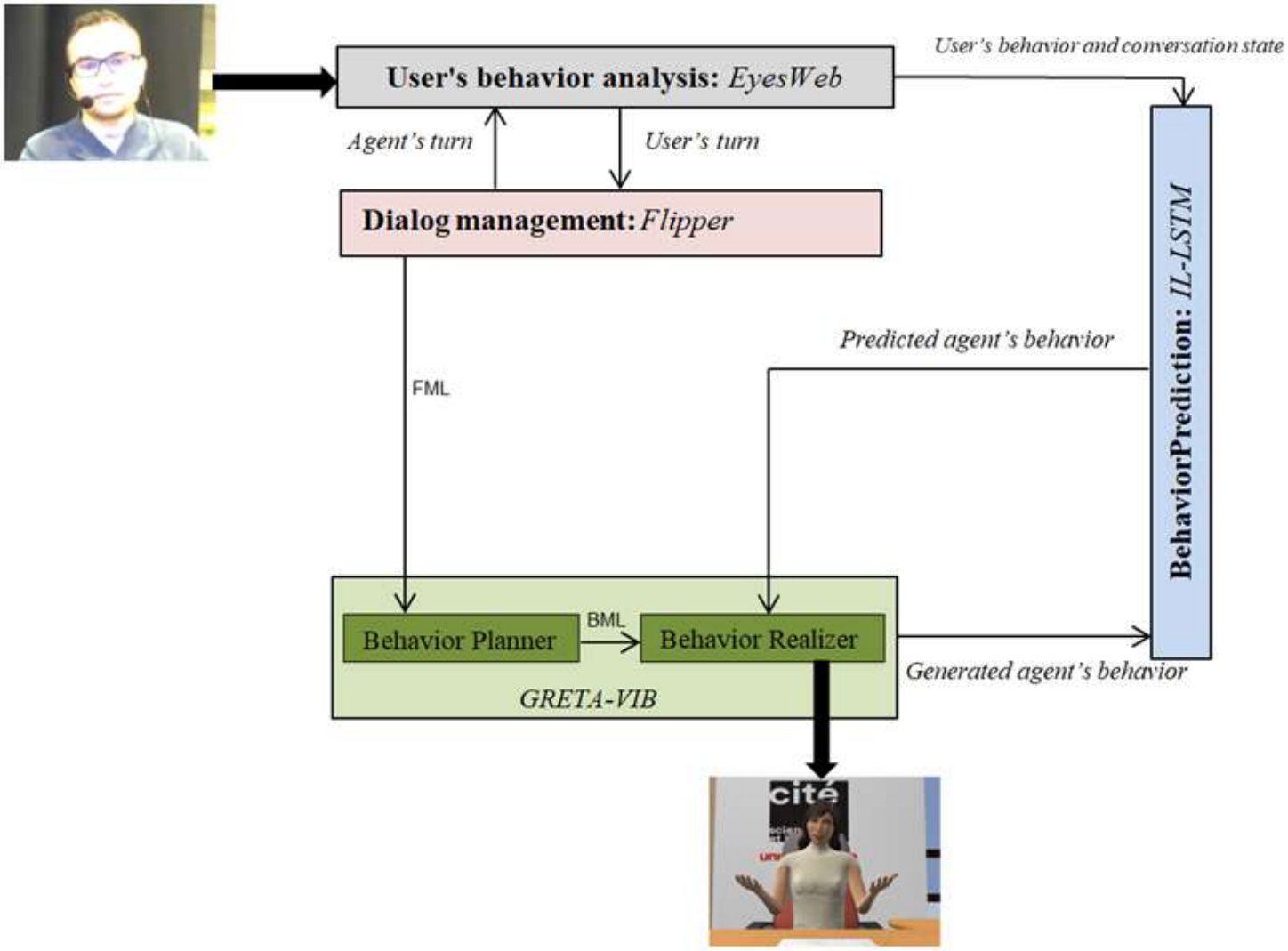
LSTM: models sequentiality and temporality of non-verbal behaviors over time

Input: sequence of features observed during a sliding window of n seconds; Data from both interactants

- Smile activity and intensity
- Head rotation
- Gaze

Output: predicted value of 4 features

Architecture



Agent's Behavior Generation

Input:

- Time window of 20 frames
- 20 frames of user's behaviors from EyesWeb
- 20 frames of agent's behaviors

Output: next agent's frame: Blend of agent's frame computed from

- intent planner
- IL-LSTM

(f0, ..., f20) → f21

(f1, ..., f21) → f22

(f2, ..., f22) → f23

…

Evaluation

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.

Evaluation

NARS: apriori attitude of participants towards the agent

PEFiC model [van Vugt et al., 2006]

- Competence
- Realism
- relevance

IAS questionnaire: perceived friendliness of the agent

Engagement (involvement and distance) and satisfaction of the user

Evaluation

Realism	Alice resembles to a real life person
Competence	Alice is effective
	Alice is qualified
	Alice is competent
	Alice is intelligent
Relevance	Alice motivated me to go and see the exhibit on video game
	Alice taught me something
Friendliness	Alice is kind
	Alice is warm
	Alice is agreeable
	Alice is sympathetic
Involvement	Alice gives me a good feeling
Distance	I dislike Alice
Satisfaction	I am satisfied about my interaction with Alice
	I appreciated Alice
	I would like to talk again with Alice

Evaluation

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

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

Affective Computing

Affective Computing (Picard,1997)

Give computers the ability to :

- recognize emotions
- express emotions (facial expressions, speech…)
- « have » emotions

Why considering emotion in HCI?

Emotion

Introduction

First scientific description of emotions

(C. Darwin, 1872)

Sneering, Defiance: Uncovering the canine tooth on one side.—

- lips are retracted
- grinning teeth exposed
- upper lip being retracted in such a manner that the canine tooth on one side of the face alone is shown; the face itself being generally a little upturned and half averted from the person

Introduction

Damasio, 1994

Primary emotions : innate, fast reaction, limbic system, « feeling »

- Example : dangerous object approaching => run away

Secondary emotions : arise later in individual development, limbic + cognitive system « thinking »

- Example : death of a loved one

Emotion

"Emotion is defined as an *episode* of interrelated, synchronized changes in several components in response to an event of major significance…" (Scherer, 2000)

Components:

- Neurophysiological and autonomous nervous pattern (in central and nervous systems)
- Motor expression (in face, gesture, gaze)
- Feeling (subjective experience)
- Action tendencies (action readiness)
- Cognitive processing (mental processes)

Emotion

Associated with key cognitive functions:

- Focusing mental, sensory resources
- Influencing beliefs
- Informing decision-making
- Preparing action and reaction
- Learning and long-term adaptation

Emotion

Why do we have emotions (Scherer)?

- Evolutionary significance of emotions
- emotion as a social signaling system
- emotion affords behavioral flexibility
- information processing
- regulation and control (we monitor our own physiological changes)

The variables of Emotion (Lazarus, 1991)

What constitutes an emotion ?

Four kind of observable variables :

- Actions
- Physiological Reactions
- What people say
- Environmental events and context

The variables of Emotion (Lazarus, 1991)

Actions

- Attack, avoidance, moving toward or away from a place or person…

- Two types of actions :
  - Volitional (goal) : smile to create a social effect
  - Expressive (unintended) : smile while being satisfied about something

The variables of Emotion (Lazarus, 1991)

Physiological reactions :

- Autonomic nervous system (heart rate, skin conductance), brain activity, hormonal secretions
- Sometimes phenomena of emotion, sometimes independent (physical effort…)
- Some are observable without instruments (face reddens or pales)
- Some are measurable only with instruments (brain activity)

The variables of Emotion (Lazarus, 1991)

What people say about emotions

- I’m angry, I’m proud…
- Subjective, not observable
- Only access to emotion interpretation
- Validity ? We don’t know what is happening behind…

The variables of Emotion (Lazarus, 1991)

Environmental events and contexts

- Social, cultural, physical events
- Cultural interpretation of threats ?
- Reason why an individual is anxious about an event ?

The variables of Emotion (Lazarus, 1991)

Non observable variables relevant to emotion :

- *actions tendencies* : private impulses as muscle tension, may be not recognized.
- *subjective emotional experience* : we only have clues about them
- *person-environment relationship* : person needs in relation with environment constraints: strategy to change person-environment relationship
- *appraisal processes* : cognitive construction of emotion

Sentic Modulation (Picard, 1997)

Apparent to others

- Facial expression
- Voice intonation
- Gestures, movement
- Posture
- Pupilary dilatation

Less apparent to others

- Respiration
- Heart rate, pluse
- Temperature
- Electrodermal response, perspiration
- Muscle action potentials
- Blood pressure

Emotion Representation

Discrete:

- Ekman: 6+ universal prototypes: anger, disgust, fear, joy, sadness, surprise (shame and embarrassment)
- Often used for emotional displays

Dimensional:

- Russell, Plutchik: e.g. arousal and valence
- Often used for recognition tasks

Appraisal:

- Subjective evaluation of situation, object, person
- Often used for emotion triggering

Discrete Emotion

There is an innate neural motor program that leads to specific response pattern

Independence from cognitive process

prototypical facial expression: full-blown facial expression of emotion

Universal emotions found in all cultures for encoding and decoding

Notion of « display rules »: modulation of the expression depending on culture

Paul Ekman

6 facial expressions of emotions universally recognized: anger, disgust, joy, fear, sadness, surprise

- shame, embarrassment

follower of Darwin

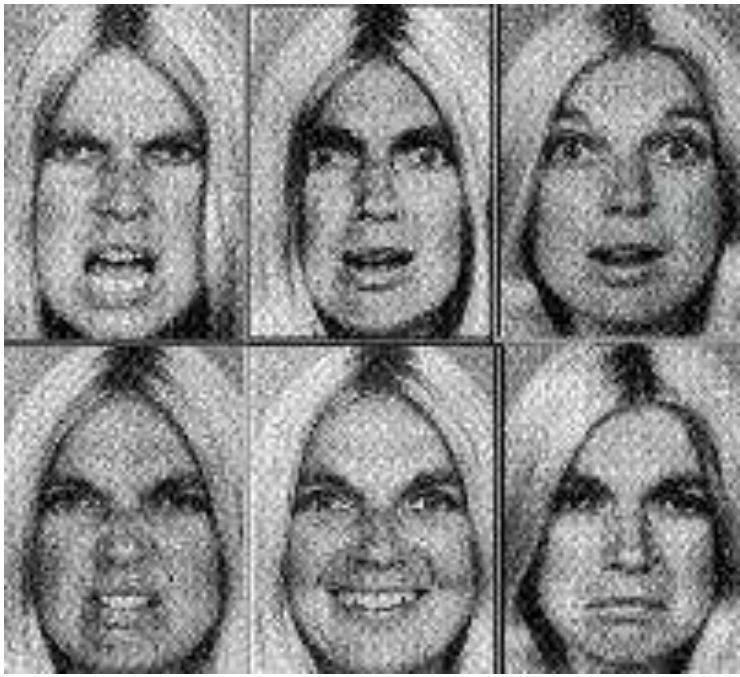
facial expression of

emotion: for communication

FACS: Facial Action

Coding System

Display rules



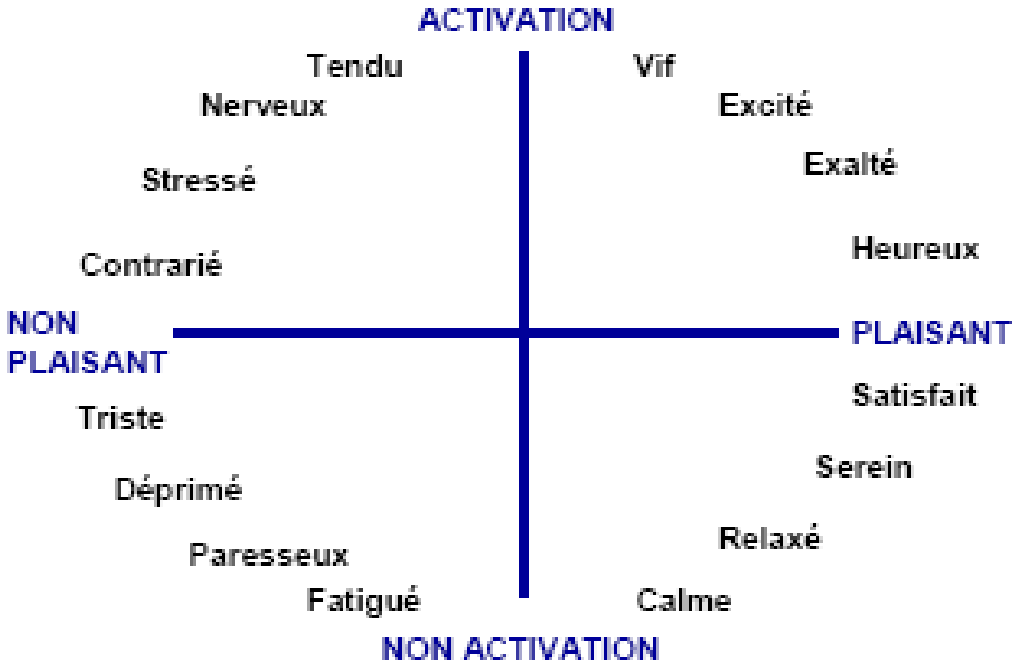
Ekman

Dimensional representation

Plutchik, Russell's wheel

arousal (active-passive) /valence (positive-negative)

Link between facial expressions and points in space.



Dimensional Representation

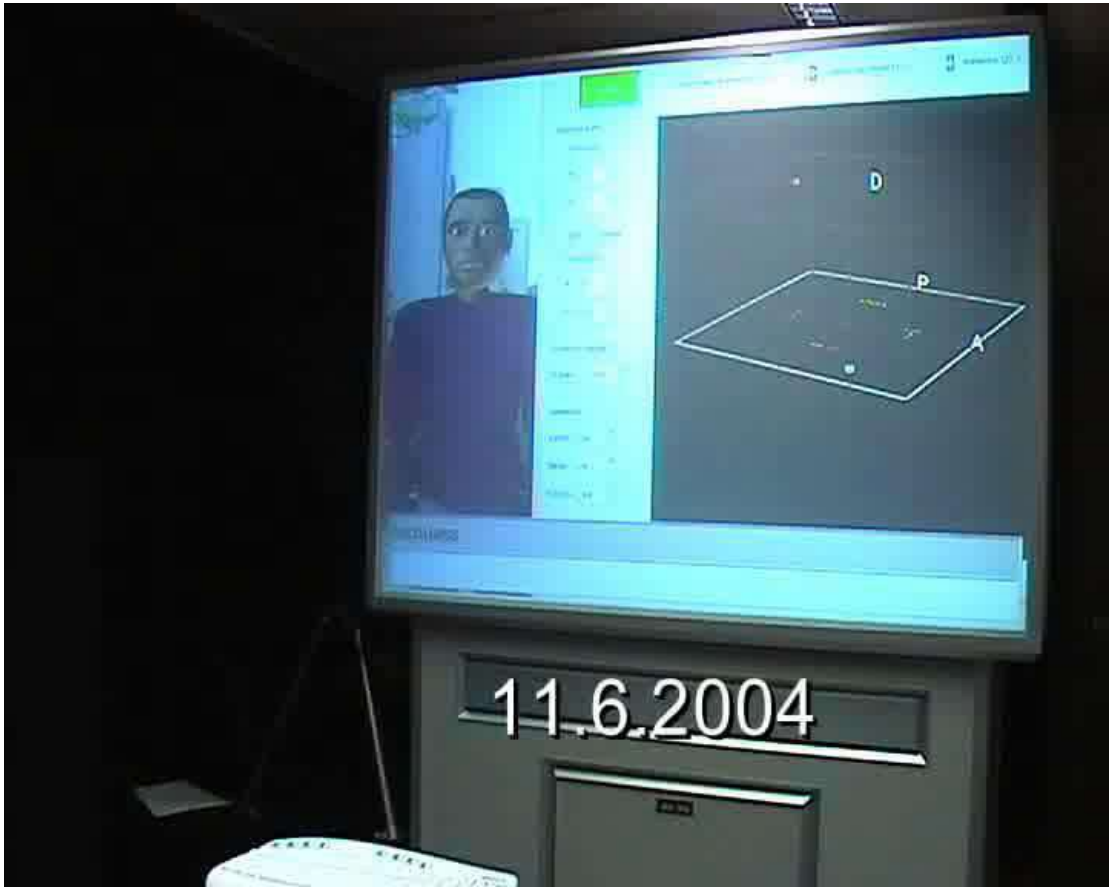
Other representations:

3D: PAD (Pleasure-Activation-Dominance) (from Merhabian's theory)

Studies by (Fontaine, Scherer et al) showed the need of 4 dimensions to encode similarities & differences of emotions

4D: evaluation pleasantness, potency-control, activation-arousal, and unpredictability

DEMO – MAX (Kopp et al)



Appraisal model

Definition of appraisal (Scherer 00): « Evaluation of the significance of an object, event or action to a person, including an evaluation of one's coping activities. It can occur at various levels of the central nervous system and need not be conscious »

Appraisal model

Evaluate continuously environment (Scherer 00):

- *5 Stimulus Evaluation Checks (SECs)*:
- *novelty* (change in the pattern of external/internal stimulation; a novel event has occurred or is expected?)
- *Intrinsic pleasure* (is the situation pleasant or not; approach tendency/avoidance tendency)
- *goal/need significance* (is the situation relevant to one's goal)
- *copying potential:* Evaluating the causation of a stimulus event nd the coping potential available to the organism
- *Norm/self compatibility:* Evaluating whether the event, particularly an action, conforms to social norms, cultural conventions, and whether it is consistent with internalized norms or standards as part of the self-concept or ideal self

OCC Model

Ortony Clore Colins, 1988: computational model of emotions

appraisal theory

consider 3 classes of emotion: emotions triggered by

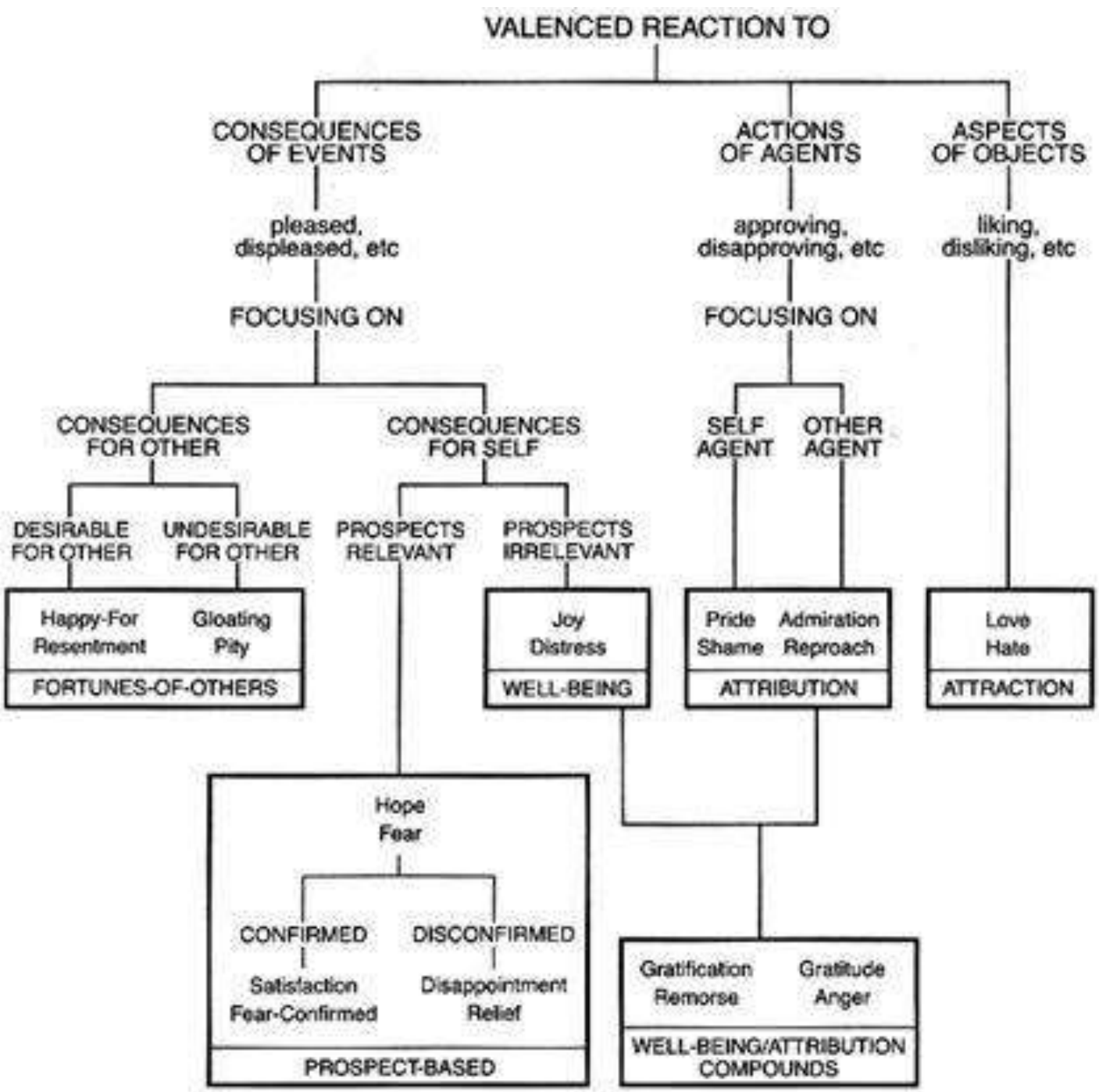
- an event affecting person's goal: eg joy, hope, fear
- an event affecting a principle or a standard: eg shame, proud, reproach
- by animated or inanimated, concrete or abstract objects: love, hate

model 22 emotions

OCC Model

GROUP	DESCRIPTION	EMOTION TYPE AND NAME
Well-being	Appraisal of a situation as an event.	*Joy*: an event is desirable for self. *Distress*: an event is undesirable for self.
Prospect-based	Appraisal of a situation as a prospective event.	*Hope*: a prospective event is desirable. *Fear*: a prospective event is undesirable.
attribution	Appraisal of a situation as an accountable action of some agent.	*Pride*: approving of one's own action. Admiration: approving of another's action. *Shame*: disapproving of one's own action. *Reproach*: disapproving of another's action.
attraction	Appraisal of a situation as containing an attractive or unattractive object.	*Liking*: finding an object appealing. *Disliking*: finding an object unappealing.

OCC Mod


VALENCED REACTION TO
CONSEQUENCES OF EVENTS
ACTIONS OF AGENTS
ASPECTS OF OBJECTS
pleased, displeased, etc
approving, disapproving, etc
liking, disliking, etc
FOCUSING ON
FOCUSING ON
CONSEQUENCES FOR OTHER
CONSEQUENCES FOR SELF
SELF AGENT
OTHER AGENT
DESIRABLE FOR OTHER
UNDESIRABLE FOR OTHER
PROSPECTS RELEVANT
PROSPECTS IRRELEVANT
Happy-For
Resentment
Gloating
Pity
FORTUNES-OF-OTHERS
Joy
Distress
WELL-BEING
Pride
Shame
Admiration
Reproach
ATTRIBUTION
Love
Hate
ATTRACTION
Hope
Fear
CONFIRMED
DISCONFIRMED
Satisfaction
Fear-Confirmed
Disappointment
Relief
PROSPECT-BASED
Gratification
Remorse
Gratitude
Anger
WELL-BEING/ATTRIBUTION COMPOUNDS

OCC Model

GROUP	DESCRIPTION	EMOTION TYPE AND NAME
Well-being	Appraisal of a situation as an event.	*Joy*: an event is desirable for self. *Distress*: an event is undesirable for self.
Prospect-based	Appraisal of a situation as a prospective event.	*Hope*: a prospective event is desirable. *Fear*: a prospective event is undesirable.
attribution	Appraisal of a situation as an accountable action of some agent.	*Pride*: approving of one's own action. Admiration: approving of another's action. *Shame*: disapproving of one's own action. *Reproach*: disapproving of another's action.
attraction	Appraisal of a situation as containing an attractive or unattractive object.	*Liking*: finding an object appealing. *Disliking*: finding an object unappealing.

EMOTION EXPRESSION

Facial Expression of Emotion

Expressions of emotions

- help in understanding the (ambiguous) messages
- make the communication more natural and plausible
- influence attention, user's performance, and satisfaction

Basic emotions

- Anger, fear, sadness, happiness, disgust, surprise
- Universally recognized (Ekman)
- Basic emotion = family of related states (Ekman 75)

Facial Expression of Emotion

Complex expressions in real-life may occur

- superposition of emotions,
- quick succession of different emotions,
- masking of one emotion by another emotion,
- suppression or overacting of an emotion,
- inhibition of an expression of emotion,
- voluntary expressions.

These expressions are used to

- express emotional states,
- express attitudes, intentions,
- communicate interpersonal relations,
- influence the perception of the displayer,
- obtain some goals, to influence the behavior of the others

Facial Expression of Emotion

Facial expressions leak information about lies (Ekman, 1992)

- **Micro expressions**
  - The felt emotion is unconsciously expressed for a second
- **Masks**
  - The felt emotion is masked by a wrong emotion
- **Timing**
  - Longer on- and offset times are good indicators of a wrong surprise
- **Asymmetry**
  - Asymmetric facial activity seem to indicate lies

Facial Expression of Emotion

People recognize

- expressions of two emotions
- masked, superposed, or sequential expressions
- distinction between expressions of felt emotion from fake emotion

COMPUTATIONAL MODELS OF FACIAL EXPRESSION

Facial Expressions

Problem: how to define a multimodal expression for a given communicative function, emotion?

Different approaches following theories of emotion:

- discrete: limited number of fundamental emotions (called "*primary*" or "*basic*") characterized by a specific expression
- Dimension: emotion is a point in space; link between emotion, dimension and facial expression
- Appraisal: emotion arises from a series of sequential evaluation checks of the surrounding stimuli; each step is linked to a facial response.

Discrete Representation

Creation of new expression as the linear combination of existing ones

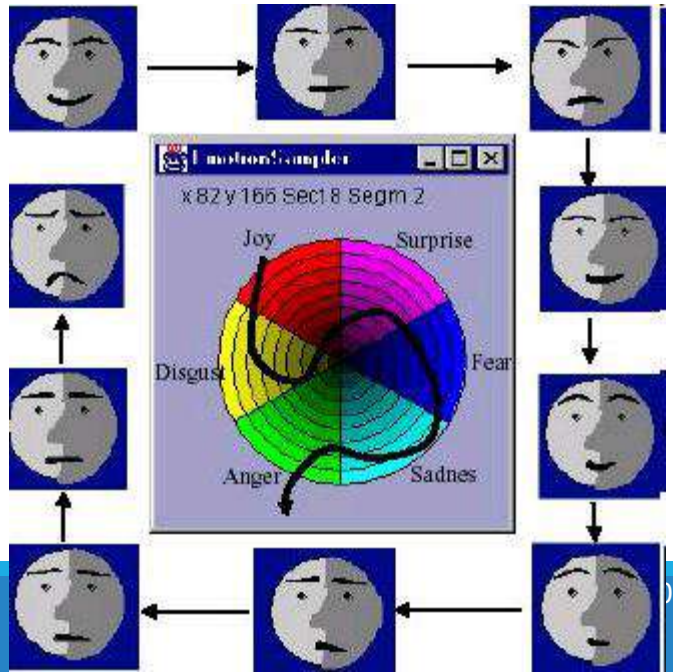
MPEG-4 includes the 6 prototypical expressions and allows their combination

EmotionDisc: bi-linear

interpolation between 2

basic expressions and

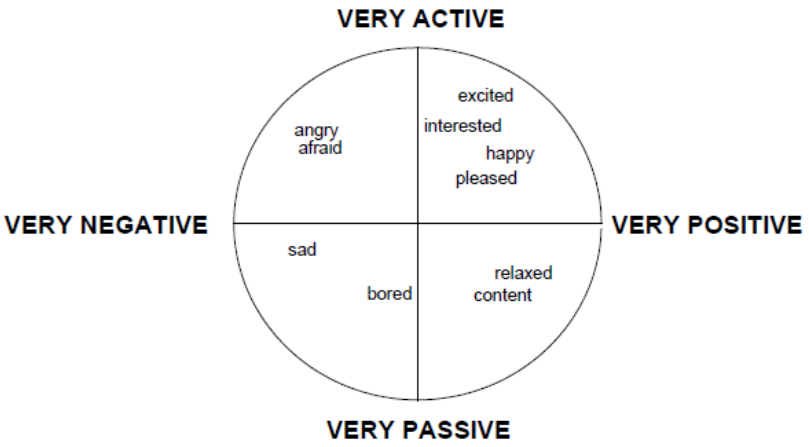
the neutral one.



Dimensional representation

Label of discrete emotions are placed on dimensional space (2D for (Tsapatsoulis et al, 02) & 3D for (Albrecht et al, 05))

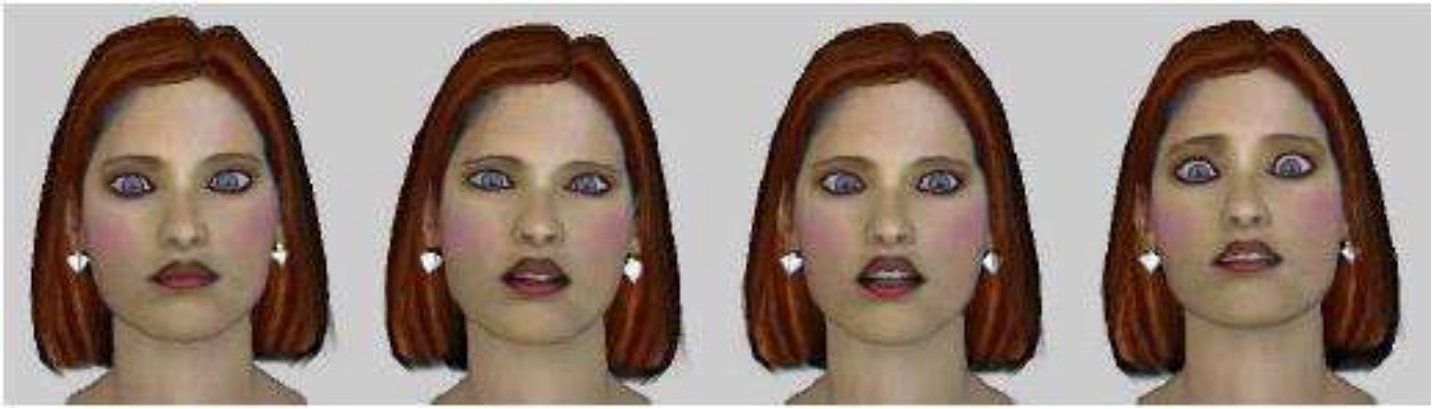
New expression corresponding to a point in the emotion space is computed by interpolating between the expressions of the 2 closest emotions



Appraisal Representation

Animation (Lisetti & Paleari, 06):

1) convert each SEC (Sequential Evaluation Checks (Scherer)) to AUs;
2) convert AUs to face parameters (Haptek)
3) find appropriate intensities for the AUs
4) exploit the temporal and intra-SEC correlation adapting AUs intensities.

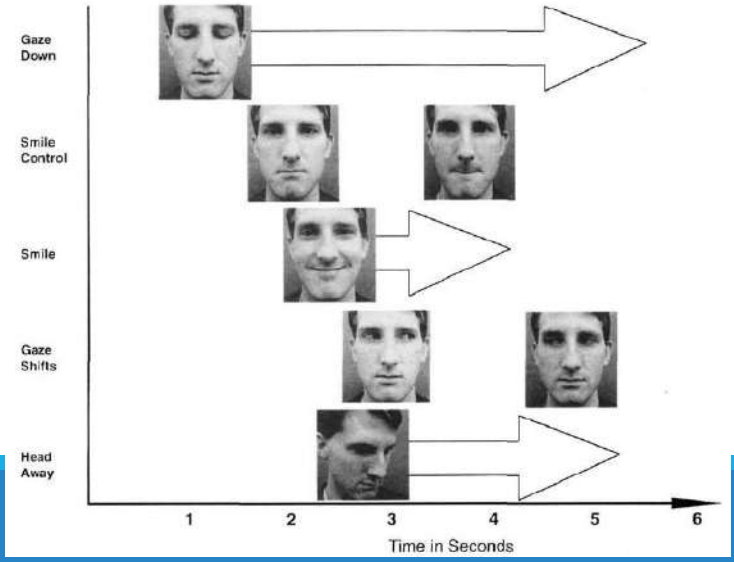


Multimodal Sequential Expressions

Go beyond static mono-modal expressions

Many emotions are expressed by sequences (or combination) of multimodal signals rather than monomodal signals (eg static facial expressions)

Data obtained from theory and literature:
(Keltner (1995); Shiota et al (2003); Harrigan & O'Connell (1996); Rozin & Cohen (2003))



D. Keltner, B. N. Buswell
Embarrassment: Its Distinct Form and Appeasement Functions

Expression Representation

Definition of a representation scheme for emotional expression:

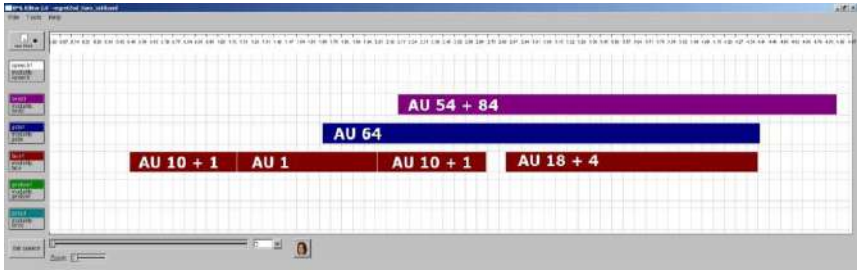
- signals description across modalities
- their partial temporal order
- their constraints

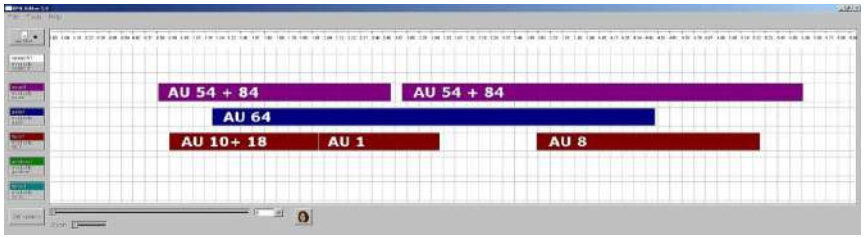
Behavior set: set of signals (frown, nod, torso backward, …)

Constraint set: describes the temporal and spatial constraints among the signals in the behavior set

- Temporal constraint: start-signal s*i* < end-signal s*j*
- Spatial constraint: signals s*i* and s*j* cannot co-occur

Expressions of Emotions





Regret

Niewiadomski

Face Dynamics, Rachel Jack

Facial expressions of emotion are not universal

So far, knowledge constrained by theory

→ Using data-driven psychological methods

Develop a 3D facial model driven by Action Units AU (FACS, Ekman 2002).

Each AU defined by its temporal course

Generate facial animations by combining randomly:

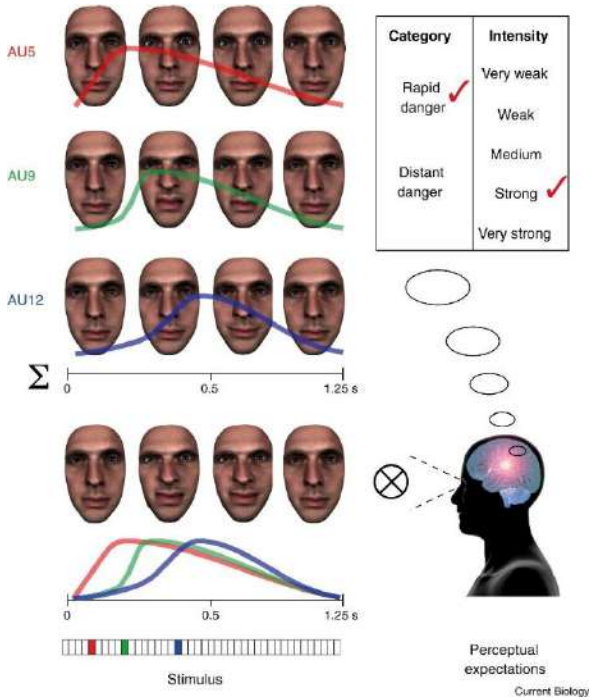
- set of AUs

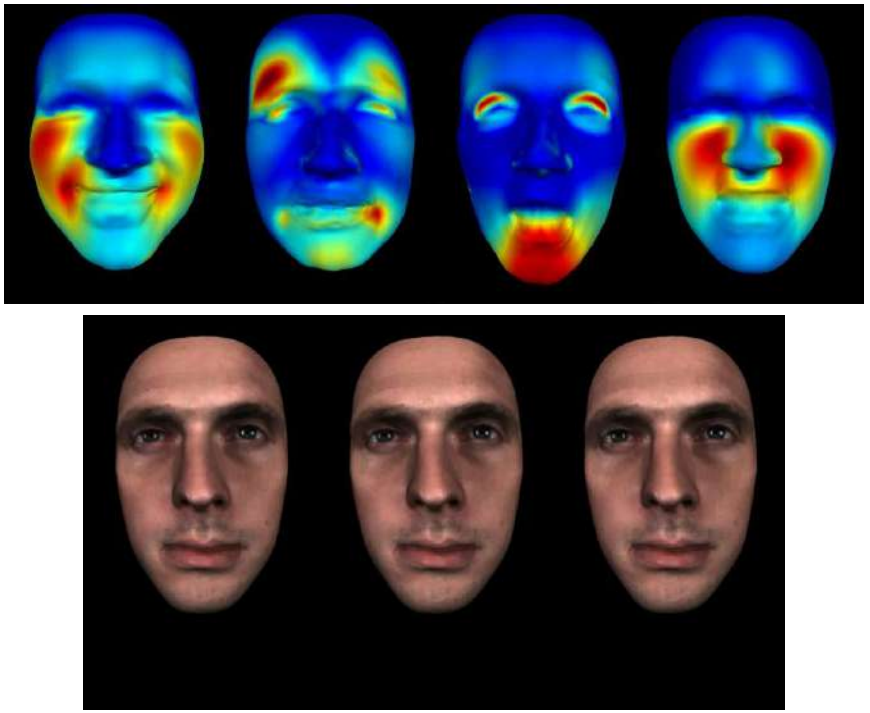
- temporal course for each AU

- acceleration curve for each AU

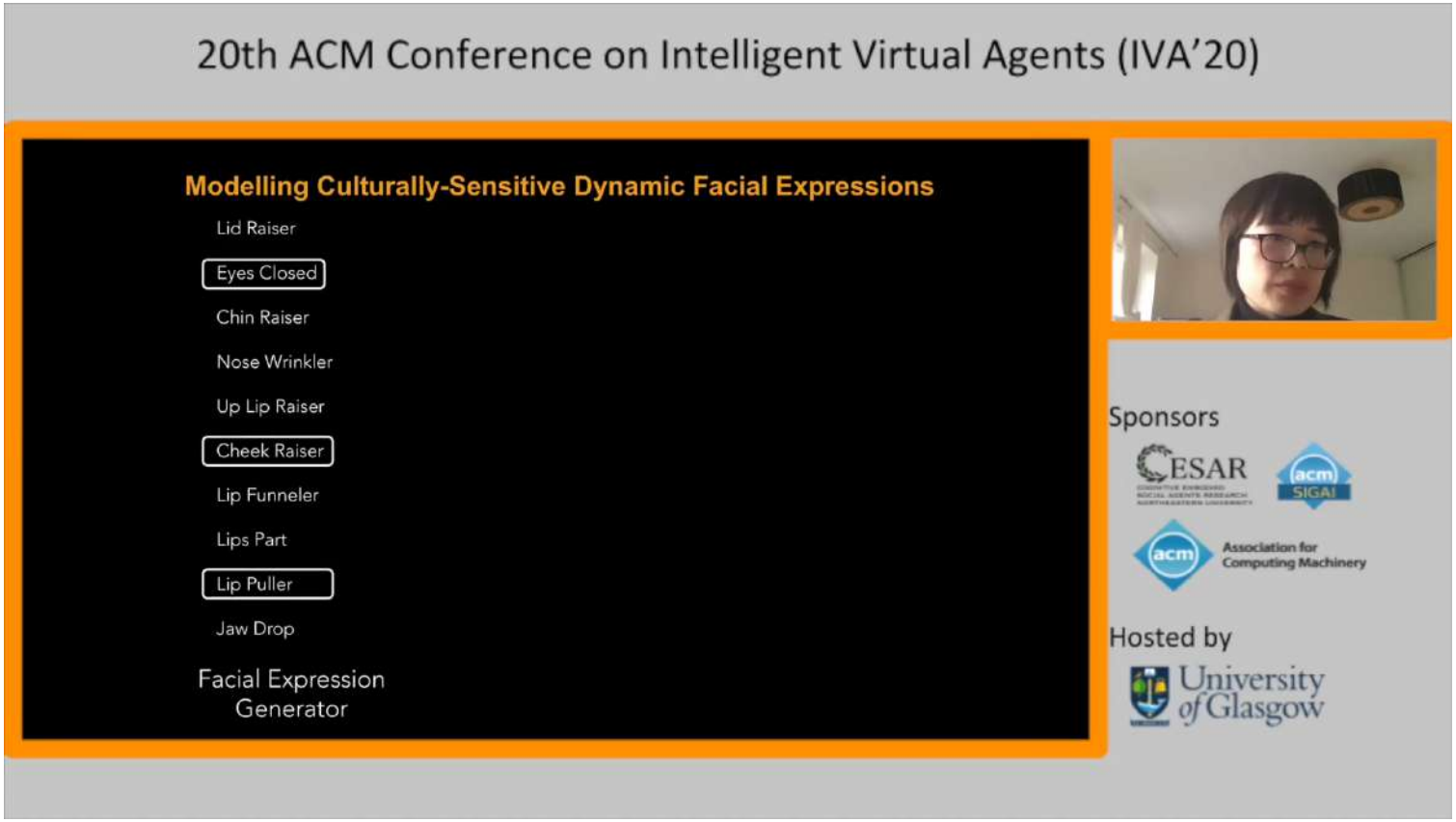
Each facial animation is evaluated through perceptual tests

Stimuli Generation





Stimuli Generation

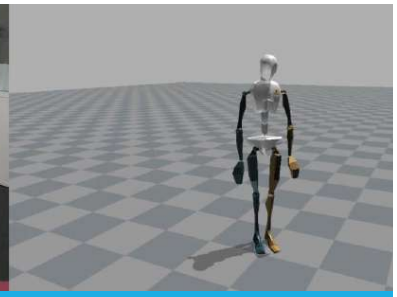


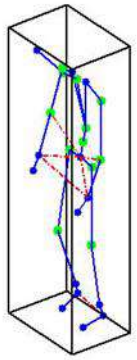
Multimodal expression of emotion

Emotions are displayed through body shape and expressivity (Berthouze)

Characterization of body movement and expressivity for emotion

- Actors: Gemep (Bänziger and Scherer 2007)
- Mocap database:
  - Emilya (Fourati&Pelachaud 2016)
  - IEMOCAP corpus (Busso et al. 2008)
  - Mockey database (Tilmanne et al., 2011)





Multimodal expression of emotion Ennis, McDonnell

Characterize

- Body shape and movement
- Body expressivity and dynamism

Multimodal behaviors

- Temporal coordination of modalities
- Blend or mix-match of information

Evaluation Studies

Conduct Evaluation Study

Theory-driven hypotheses

- Based on ***a priori statements*** formulated in the theory or on preliminary observations of a phenomenon
- Follows **deductive reasoning** (or top down reasoning)

Data-driven hypotheses

- Based on specific **observations** / findings of previous studies
- Follows **inductive reasoning** (or bottom up reasoning)

Hypotheses

Descriptive hypotheses

- Aims to predict about the relationships between variables
- E.g.: What is the multimodal behavior of a given emotion?

→ correlation design: study through correlation measures

→Hypothesis: predictions about the relation between a given emotion and its multimodal expression

Causal hypotheses

- Aims to predict about the causes of a particular behavior
- E.g.: What causes a given multimodal behavior?

→ experimental design: manipulate variables to measure their impact

→ hypothesis: predictions about the causes of a particular behavior

Variables

Independent variable

Variable that is manipulated by the researcher so that the different levels of the variable change between or within groups of subjects in the experiment

-> *the factor which the experimenter thinks will affect behavior*

Dependent variable

Variable of interest that is measured from an individual

-> *the DV operationalizes the concepts of interest related to a particular research question*

Extraneous variable

A factor external to the experimental design that could affect the D.V.

If control is not sufficient, the observed effect of the I.V. on the D.V. could be attributed to the extraneous variable, or the extraneous factor could mask a real effect of the I.V.

Participants

Recruiting participants

- Online survey
- M-Turk
- Laboratory

Sampling participants:

- Age
- Culture
- Gender
- Personality
- …

Design

Design

- ***a variable is manipulated***
- other variables that could affect the results are ***controlled***

→ ***uncover causal relationships*** between variables

Between-subjects design

- Different individuals are assigned to the different levels of the independent variable
- most frequent problem: Individual differences across groups

Within-subjects design

- The same individuals are assigned to the different levels of the independent variable.
- Avoid **order effects**: Random assignment *of the levels* → *E.g. Latin square*

Scales of measurement

Nominal scales: Non-ordered categorical responses (gender, marital status, mood category)

Ordinal scales: The response categories of a measure contain an ordering

Continuum of measurement:

- E.g.: “not anxious”, “a bit anxious”, “very anxious”, “extremely anxious”
- Rank ordering of stimuli
- Response categories are not equally spaced on the continuum

=> Do not involve numerical values

=> Qualitative data

Interval scales:

- Involve numerical categories that are equally spaced on a continuum
- E.g. Likert scales

=> Quantitative data

Perceptive Study

- *context-free level* of evaluation: only the perception of stimuli is evaluated without considering any information about the context;
  - E.G: present images or videos of virtual agents expressing an emotion and ask users to indicate the recognized emotion types and intensity through (forced choice) questionnaire.
  - some elements of the social context may impact users' perception: e.g. gender of the virtual agent/user

- i*n-context level* of evaluation: the perception of stimuli is evaluated in a particular context of interaction.
  - Measure effects of stimuli in context; e.g. measure expression of emotion on user
  - Control condition
  - Objective measures (eg through sensors) and subjective measures (questionnaires)

BELIEVABILITY

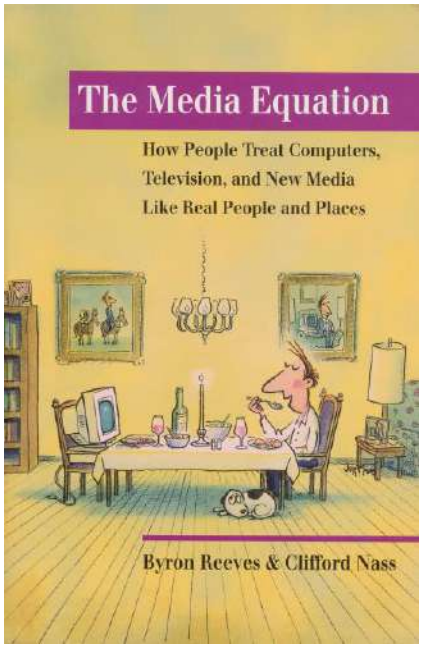
Believability

Properties of human-machine communication :

Machine viewed as social and emotional actor; much more than tools (Reeves and Nass, 96)

Mediation equation

- Take any social psychology finding between 2 humans
- Replace one human with computer and test if effect still hold
- E.g., Person A “catches” emotion from person B → Person A “catches” emotions from computer



Believability

Definition by Loyall 1997: *« a character is considered to be believable if it allows the audience to suspend their disbelief »*

includes:

- physical appearance
- emotions and personality
- social capabilities
- consistency in actions and behaviors

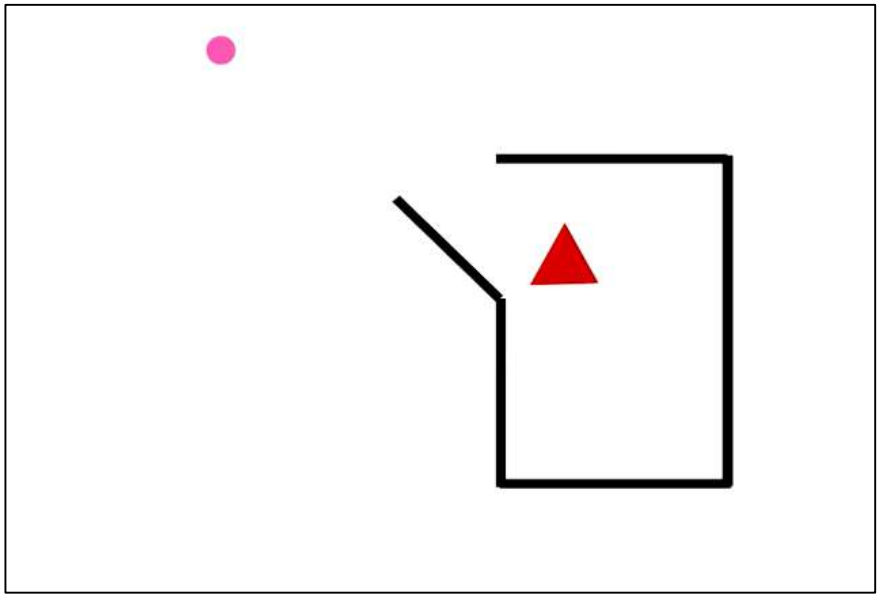
Believability

Believable agent:

- may act directly on subject's life (eg sell him an object)
- affords trust, emotion response, comprehension etc since
  - she is responsive to user's concern
  - she interacts with him
  - she shows interest
  - …

Believability
A little experiment…

What happens in this video ?



Believability
A little experiment…

Heider and Simmel experiment 1944 :

Study of causality attribution

→Result : most observers interpreted the picture in terms of actions of animated beings, chiefly of persons

Social Effects

"S*ocial* effects" are effects that occur when one is faced with another person that don't occur (or are substantially reduced) when one is alone or facing a non-human partner

- In-group favoritism / out-group bias
- Social facilitation/inhibition
- Impression management
- Emotional contagion
- Social proxemics
- Display rules
- Persuasion

Social effects are weak or absent when interacting with machines

Social Effects

How to measure social effects?

1. Endow agent with some characteristics
2. Have participants interact with agent
3. Measure social effects
   1. Subjective measures through questionnaires
   2. Objective measures: behavior analysis, physiological data, …

Mind perception theory emphasizes importance of emotion (Greg, Wegner)

- How we treat other entities depends on extent to which we attribute them “a mind”
- People organize other minds in 2 broad dimensions: Do they think? Do they experience emotion?

Social Effects

What if we tell participants agents are driven by human (intelligence)?

Use of Wizard of Oz setting

Experience:

Participants are either told they are interacting with an autonomous agent or an agent (avatar) controlled by a human (WoZ)

→ agents controlled by a human are perceived more trustworthy, more cooperative…

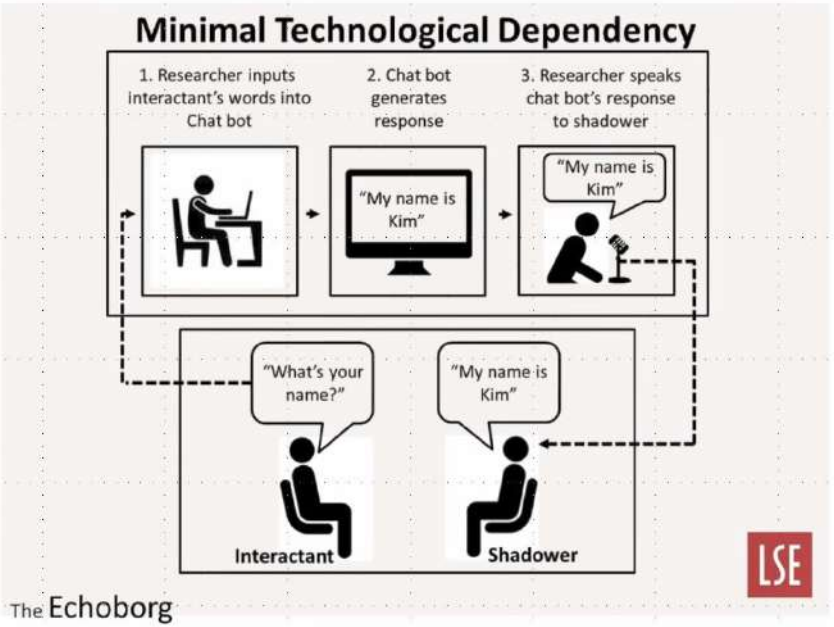
Possibility: Endow agents with human-like capabilities: visual, behavior, emotion, action…

→ Uncanny effects

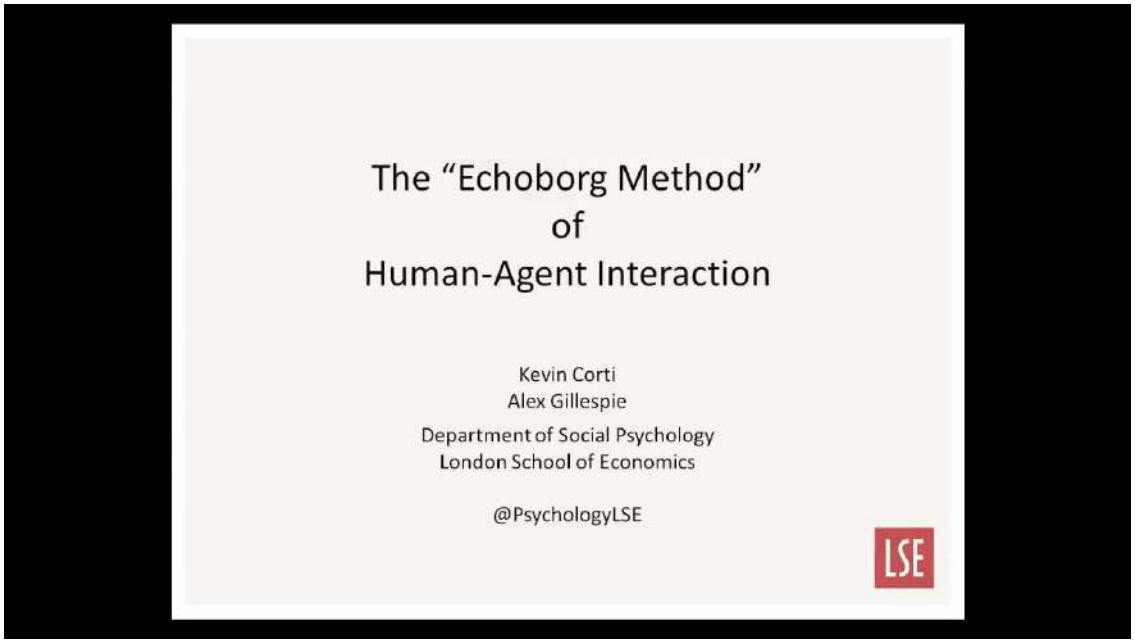
Echoborg Method of Human-Agent Interaction, Corti, Gillespie

Settings:
WoZ controls a human confederate and not an agent

- Participants interact with a confederate
- Confederate is controlled by a WoZ
- WoZ:
  - Reports on a chatbot (Cleverbot) what the participants have said
  - Tells the answer of chatbot to confederate
- Confederate says aloud the answer



Echoborg Method of Human-Agent Interaction, Corti, Gillespie



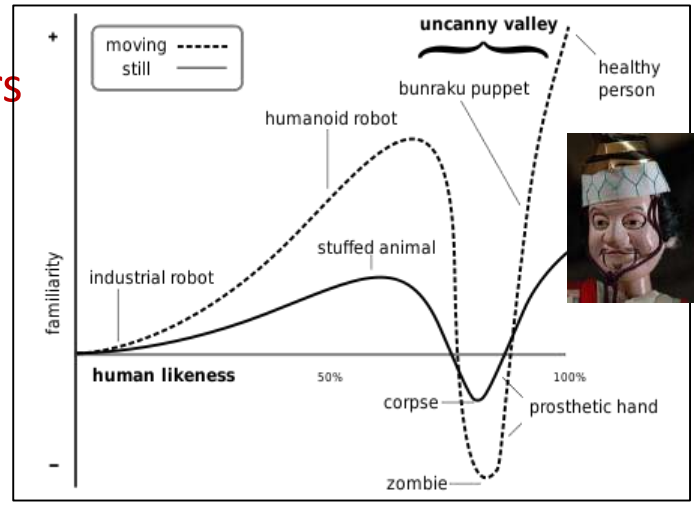
UNCANNY VALLEY

The Uncanny Valley (Mori, 1970)

Concept introduced by Mori, 1970

- Link between human-like qualities (degree of anthropomorphism) and sense of familiarity/affinity (degree of acceptance/rejection)
- 2 curves: static / moving entities
- Principle : human features look and move almost, but not exactly, like natural human beings

→ revulsion from human observers



MORI 1970 (translated by MacDorman and Minato, 2005)

Bunraku puppet

Android Repliee Q2

Uncanny Valley

Uncanny Valley

Different factors:

- Visual (rendering, appearance)
- Movement (behavior, emotion, social signals)

Different studies have investigated the question of uncanny valley along these dimensions using different types of robots and of virtual agents

Uncanny Valley, Ishiguro

Geminoid, Ishiguro et al

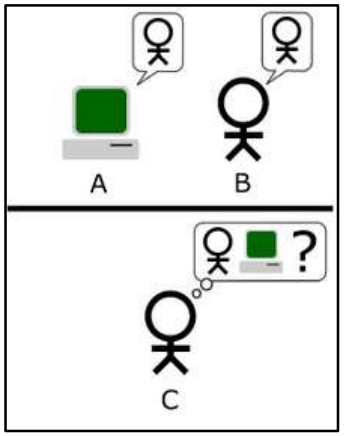
Study:

- 3 conditions: robot, android (Geminoid), human
- Biological appearance or not
- Biological movement or not

The Turing test

Principle: test a machine's ability to exhibit intelligent behavior equivalent to, or indistinguishable from, that of a human.

- Appearance: static view
  - Pass the Turing test



Uncanny Valley

- Neurobiological correlates (Saygin et al, 2011)
- Movement
  - 3 stimuli:
    - Robot: nonbiological appearance and motion → congruent
    - Android: biological appearance, nonbiological motion → non congruent
    - Human: biological appearance and motion → congruent
- Participants view video of stimuli doing simple movement
- Record of fMRI activities
  - → Different brain activities when there is discrepancy between appearance and movement due to expectancy errors.

Saygin et al, The thing that should not be: predictive coding and the uncanny valley in perceiving human and humanoid robot actions, Oxford University Press, 2011

Photo-realism

High development in computer graphics

Real-time processing for video-games, virtual reality

Can capture:

- 3D models: scanner
- Rendering: realistic reflections, physically based materials and photometric lighting
- Animation: motion capture

Impact:

- Uncanny valley hypothesis
- In VR, sense of social presence
- Emotional response

Photo-Realism: Study 1, McDonnell

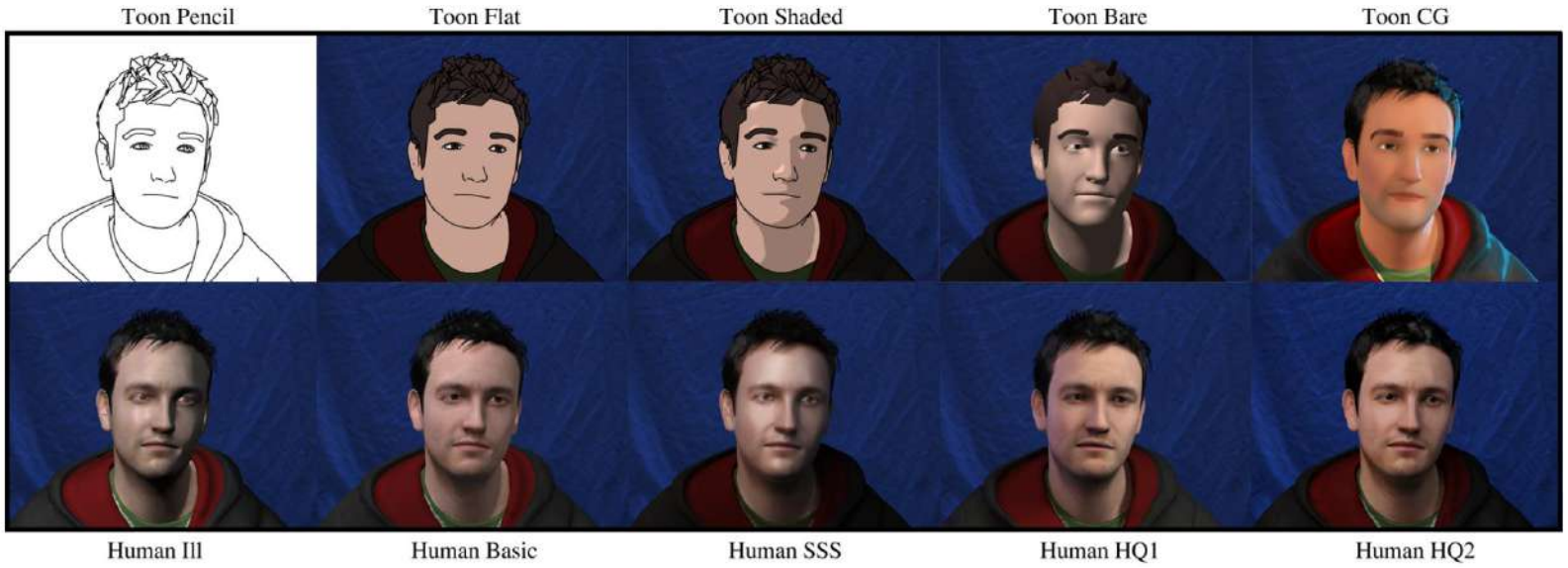
2 sets of stimuli: static images and movements (with and without artifacts)

Hypothesis: « participants would give abstract characters more positive ratings than characters that they considered to have almost photo-realistic appearance”

Perceptive study:

- Extremely abstract – Extremely realistic
- Extremely unappealing – Extremely appealing
- Extremely unfamiliar – Extremely familiar
- Extremely eerie – Extremely re-assuring
- Extremely unfriendly – Extremely friendly
- Extremely untrustworthy – Extremely trustworthy

Photo-Realism: Study 1



McDonnell, R., Breidt, M., Bülthoff, H. 2012. Render me Real? Investigating the Effect of Render Style on the Perception of Animated Virtual Humans ACM Trans. Graph. 31 4, Article 91 (July 2012)

Photo-Realism: Study 1

Results:

- Cartoon characters :
  - highly appealing,
  - more pleasant than characters with human appearance when large motion artifacts were present
  - more friendly than realistic styles
- Movement changes only how familiar we find the characters, and also how appealing or pleasant they are considered.
  - Highly unappealing characters are considered more so when movement is applied,
  - Motion anomalies are considered more unpleasant on human than on cartoon render styles

Non interactive stimuli

Photo-Realism: Study 2, McDonnell

Virtual Reality setting

Virtual characters and environment: 3 rendering styles:
- lowest, mid-range and highest

Movement: motion capture of a human actress

Measure:
- Sense of social presence: through questionnaires and behavioral responses
- Place illusion (sense of « being there »): degree of engagement with virtual environment and proxemy with virtual character
- Emotional response: show empathy to virtual character

Results:
- Photorealism
  - increases place illusion
  - Is more visually appealing
- Stylized agent:
  - Animation more realistic
  - Increase of congruent emotional responses

Zibrek, Martin, McDonnell, Is Photorealism Important for Perception of Expressive Virtual Humans in Virtual Reality?, ACM Transactions on Applied Perception, 1(1), 2019

Photo- Realism, Soul Machine

Soul Machines: https://www.soulmachines.com/

Creator: Mark Sagar (worked on Avatar)

AVA: Autodesk Virtual Agent

Extremely realistic virtual agent: visual, audio and emotion

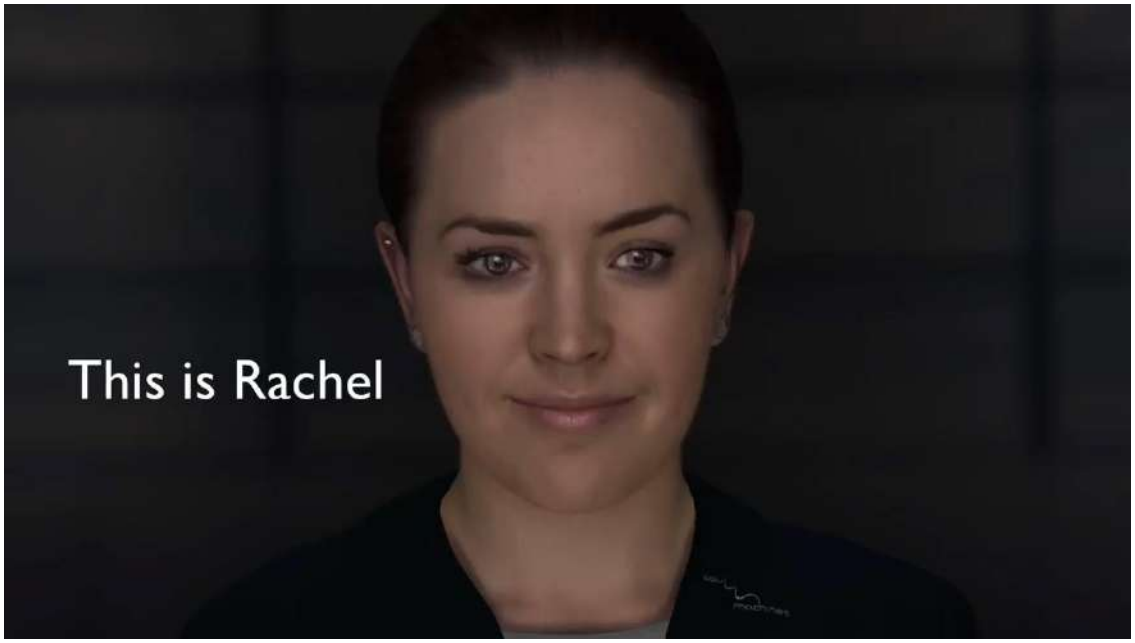
Includes all visual effect from cinema

Aims: develop hyper-real Digital Humans with emotional intelligence

Go beyond:

- Robotic look to hyper realism graphics and animation
- Generic character to create character with own personality
- Scripted dialogs to create context aware interaction

Soul Machines



NEON SANSUMG

TV Anchor - Corean

Mind Perception and Uncanny Valley, Greg, Wegner

Hypothesis: the sense of eeriness from human-like robots may arise from the attribution of a mind in robots

Two dimensions of mind perception:

- Emotion experience: feel and sense emotion
- Agency: capacity to act and do

Create uncanny scales:

- Uncanny: Participants rated extent they felt ‘‘uneasy,’’ ‘‘unnerved,’’ and ‘‘creeped out,’’ on 5-point scales from ‘‘not at all’’ (1) to ‘‘extremely’’(5)

Perceived experience: rate perceived experience

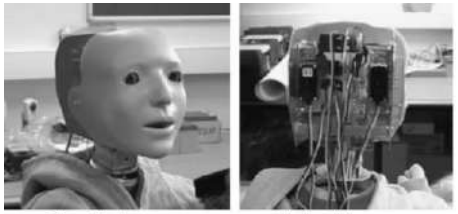
- This robot has the capacity to feel
- pain’’ and ‘‘This robot has the capacity to feel fear,’’—and
- agency—‘‘This robot has the capacity to plan actions’’ and
- This robot has the capacity to exercise self-control.

Gray, Wegner, Feeling robots and human zombies: Mind perception and the uncanny valley, Cognition 125 (2012)

Mind Perception and Uncanny Valley

Study 1: human-like appearance of machine → attribution of experience and feelings of unease

Stimuli: video of mechanical robot (back of Kaspar's head) and of human-like robots (Kaspar)



Results:

- similar agency to the humanlike and mechanical robots
- greater experience to the humanlike than to the mechanical robot

Mind Perception and Uncanny Valley

Study 2:

- Appearance not varied
- Capacity varies
- Stimuli:
  - Super computer: powerful computer
  - Experience: computer able to feel some emotions
  - Agency: can execute actions independently

Results:

- the experience condition significantly more unnerving than either the control or agency conditions

Mind Perception and Uncanny Valley

Study 3:

Stimuli: Description of human with different capacities

- normal human
- unable to ‘‘plan or make goals,’’ or ‘‘do things a normal person can do.’’ In the experience-less condition (agency less)
- unable to ‘‘feel pain, pleasure or fear or otherwise experience what a normal person can experience.’’

Results:

- a person without experience makes people uneasy but not a person without agency

<u>Overall conclusion:</u>

the uncanny valley seems to be driven by a number of factors, including general perceptions of categories, specific facial features

Use of emotion an important design dimension for intelligent machines and should be used with care

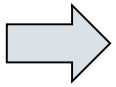
Believability vs Realism

Not only a question of appearance…

- non adapted emotional displays influence the user's evaluation of the agent negatively

(Walker et al., 1994, Becker et al., 2005, Prendinger and Ishizuka, 2001)

- appropriate emotional displays increase the agent's believability (Lim and Aylett, 2007)



socially adapted behaviors required !

Uncanny Valley: some conclusions

Uncanny valley often described on a single realism dimension

Important to realize there are many dimensions of realism

Some reason to believe uncanny valley is disconnect between visual and behavioral realism

As degree of realism increases → need to control it accurately

Easier to match visual fidelity than behavioral fidelity

Any questions?