Embodied conversational agents - Challenges for IP

Zsófia Ruttkay
Associate Professor
Human Media Interaction, Dept. of CS, University of Twente, NL
Szent-Györgyi Fellow at PPKE ITK, Bp
http://hmi.ewi.utwente.nl/~zsofi/

Overview of the talk

Intro about ECAs

FA Analysis of facial expressions of humans and cartoons

Non-photorealistic facial animation - CharToon

Emotion Disc

Evaluation of personality cues

Full body

Style in gesture - GESTYLE

Challenges for IP

About ECAs I

- More or less human-like synthetic characters to interact with users
- In roles as:
  - information provider
  - tutor
  - sales assistant
  - ‘translator’ (sign language)
  - avatar in telepresence and VR
  - entertainment, games
  - also medium for experiments
- Research since 10 years, recently own events:
  IVA, Gesture WS, AAMAS workshop series
- Book by J. Cassell et al. (2000, MIT Press)

Examples

Rea from MIT,
Cassell et al.

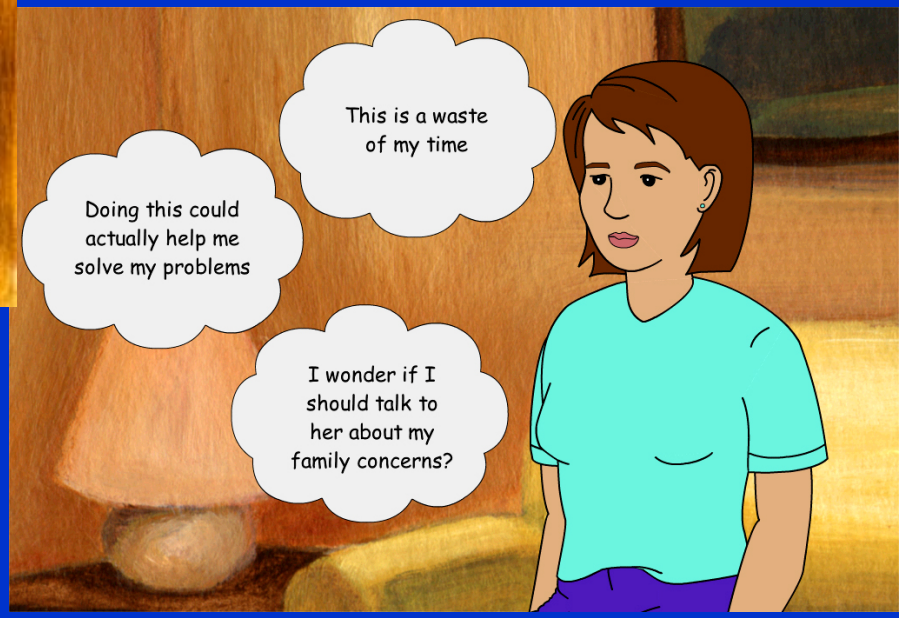
Steve from USC
J. Rickel, L. Johnson et al.

Examples

Examples

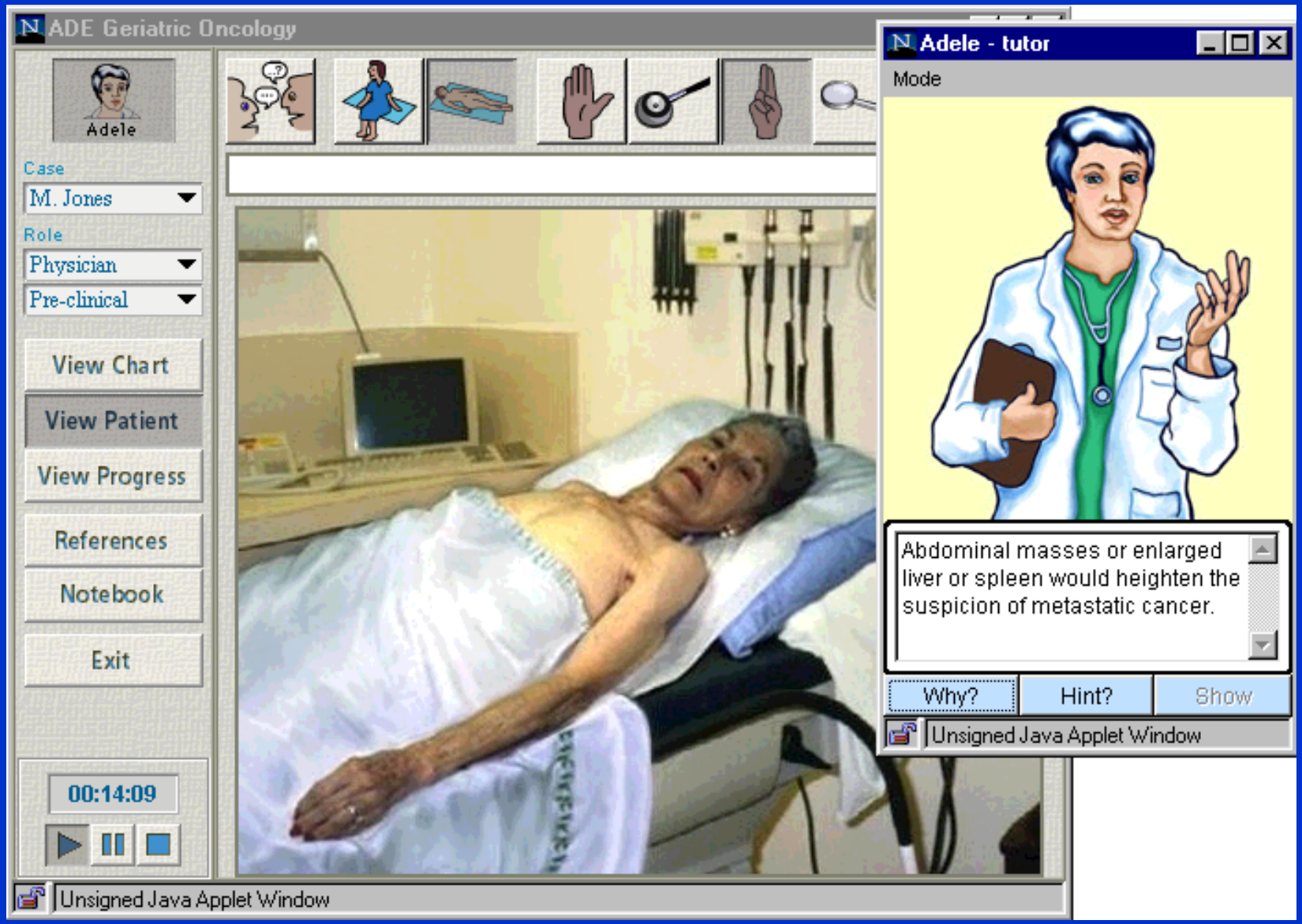
Mission Exercise, USC ISI + ICT

Examples



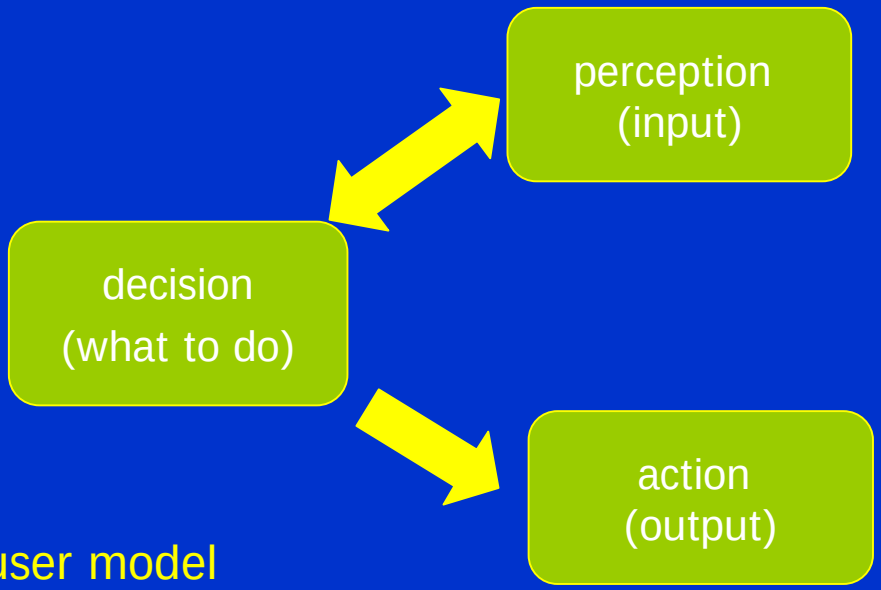
Carmen's Bright Ideas from USC ICT, Marsella et al.

Examples



USC – ISI, Adele, Shaw et al.

About ECAs - general architecture



- nl processing
- (meta) speech understanding
- face interpretation
- gesture recognition
- environment monitoring

(object, resources, situation)

- multimodal output
- related to environment

- user model
- affect model
- discourse model
- domain knowledge
- goals ...

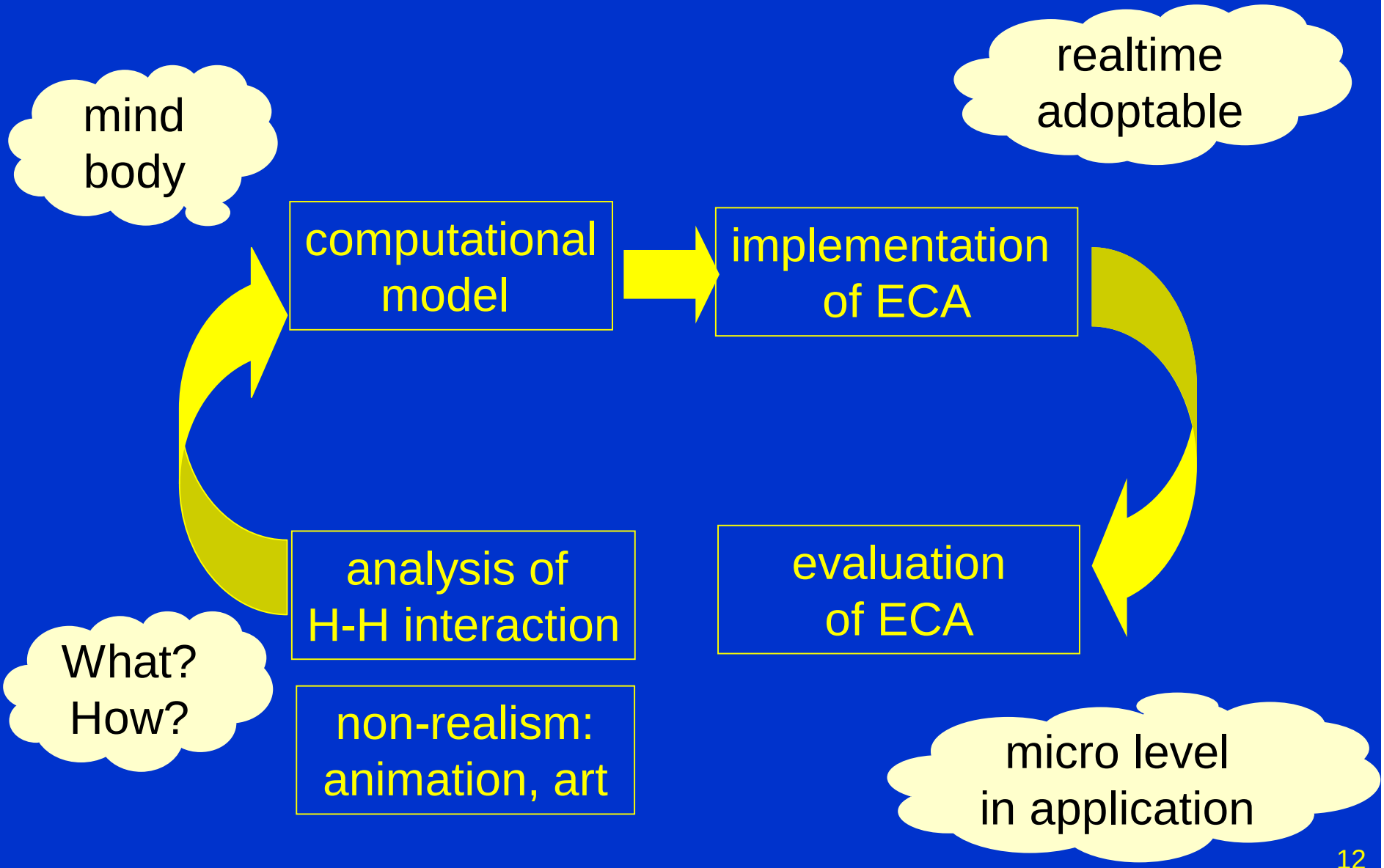
- real-time
- suspension of disbelief
- ‘good’ for ...

About ECAs – Design parameters

- Embodiment: head – torzo - full
- ‘Look’: gender, age, profession, personality
- Design: realistic/not, 2d/3d
- Communication channels & directions
  speech, facial expressions, gestures, body lang.
- What is to be expressed: emotional, cognitive, phys. state
  references to changing surrounding (incl. user)
- Role: monologue – dialogue,
  relation to user

About ECAs – Hierarchy of behavior

About ECAs - development cycle



About ECAs – Some issues

- Mind and body issue
- Why to use ECAs?
  - ease of use: broader user group, less load
  - more effect
  - more fun
- Moral and legal issues: clones, manipulation
- To mimic humans, or less is more?
- Interdisciplinary research

About ECAs – Do people take them?

- Illusion – do people believe, react to as if they were real?
- YES
  Clifford Nass: CASA paradigm
  Computers Are Social Actors
- Series of experiments on effect of micro and macro design
- User's characteristics matter

Overview of the talk

Intro about ECAs

FA ⇨ Analysis of facial expressions of humans and cartoons

Non-photorealistic facial animation - CharToon

Emotion Disc

Evaluation of personality cues

Full body

Style in gesture - GESTYLE

Challenges for IP

Motivations for FA research

•Need for: expressive, animated synthetic faces

• Functions of facial (and other) noverbal signals: - emotions: motivation, feedback
- cognitive state

- speech punctuation
- dialog control
- additional/redundant info (size, location, ...)

- personality and culture of speaker

Questions concerning FA

- Which expressions to consider?

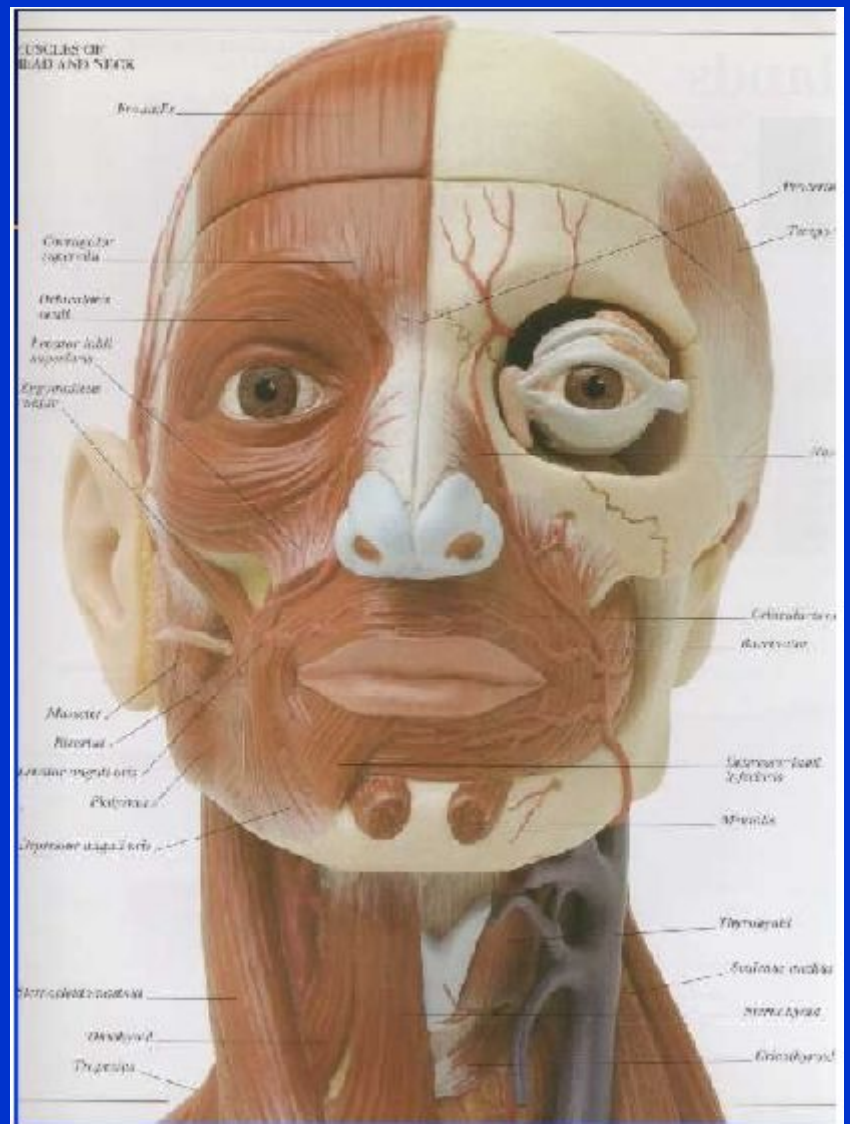
- What are those expressions like?

  - static and dynamical characteristics
  - generic and individual characteristics

- What facial model to use?
  - (non-)photorealism
  - beyond realism/simplification of motions?

- Tool to make synthetic face models and animate them

Facial anatomy



- about 44 pairs of muscles
  - bundle
  - sheet
  - circular
- physiological data missing
- coarticulation
- many parameters
- nonlinear effects

Psychology of facial expressions

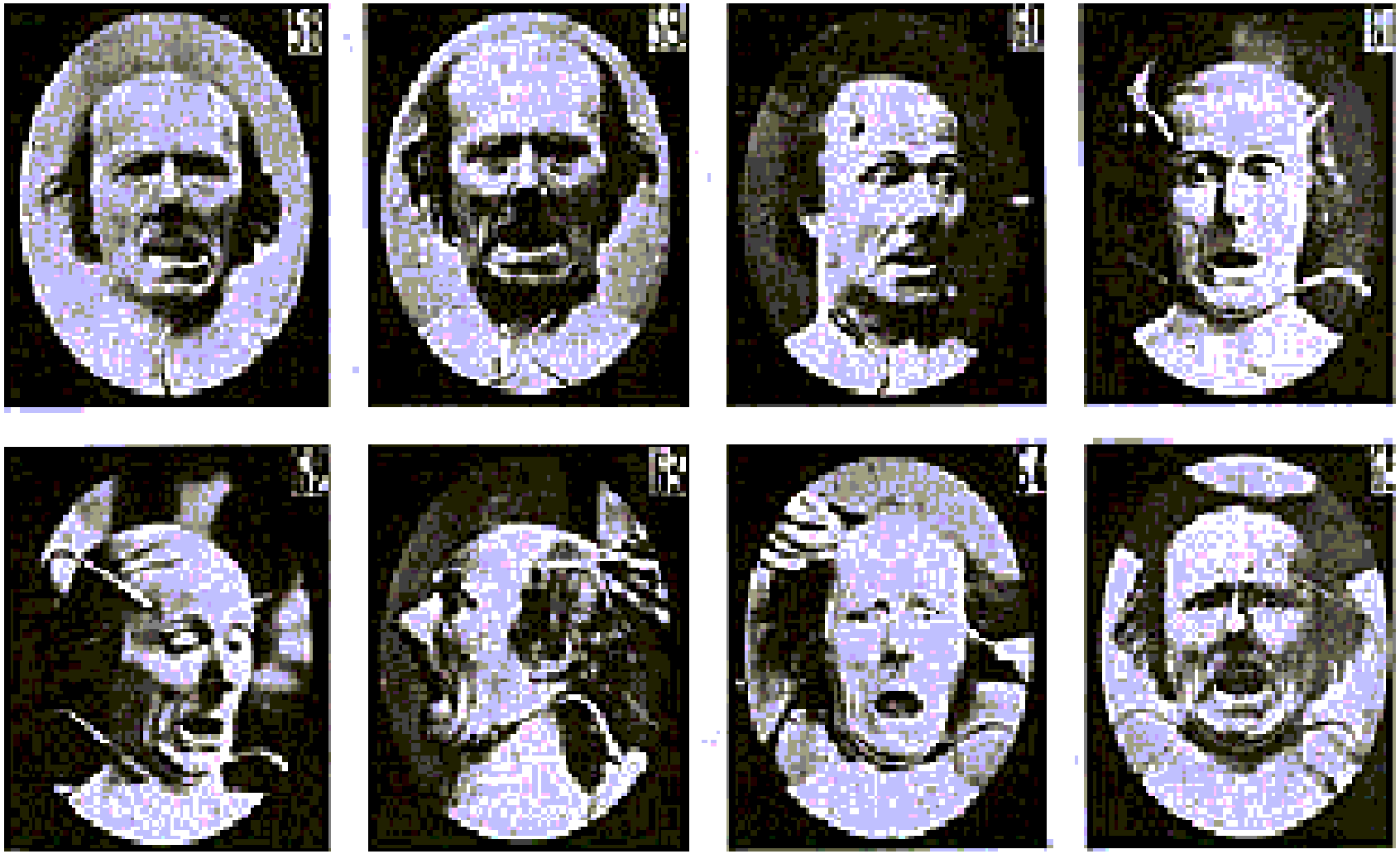
- “meaningful” – physically possible - impossible
- biological need/content/emotional state/discourse regulator/...
- muscle activity – emotion sensing in brain
- difference between real and ‘on demand’ expressions

data collection must be done with care:
- spontaneous/stimulated expression
- natural setting (e.g. in news)
- real people vs. actor?

History ....

- 1882: Darwin: The Expression of the Emotions in Man and Animals

- 1862: Duchenne: photos of expressions induced by electric signals



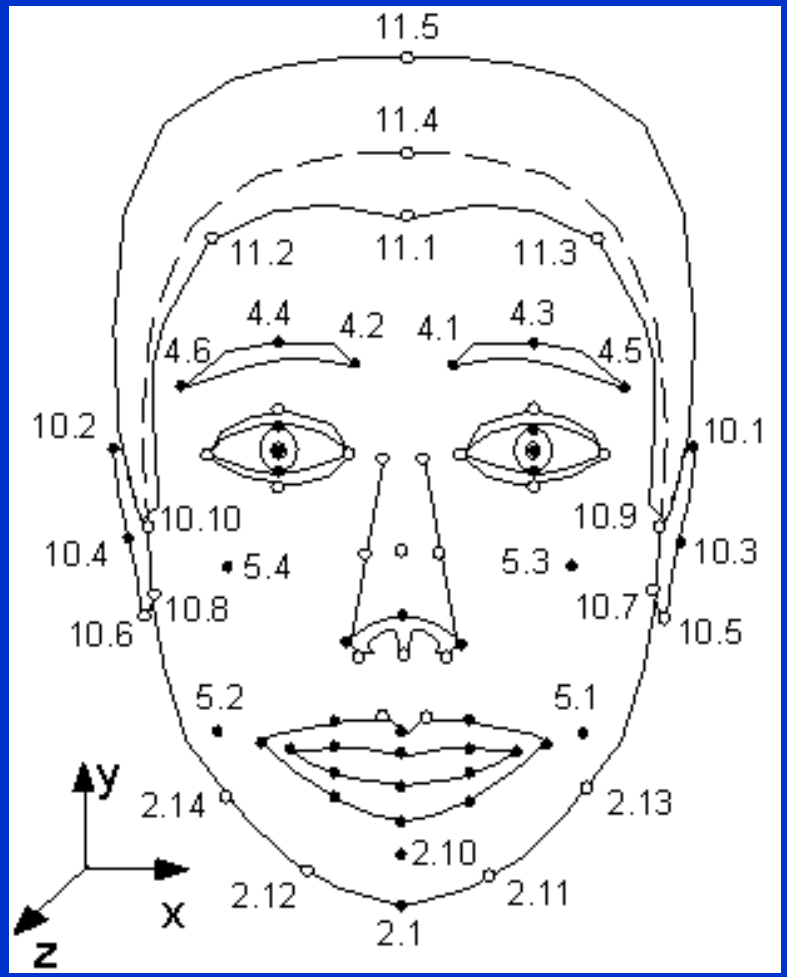
History: coding of facial expressions

- 1970 - ...: Paul Ekman (U. of California)

- 1989: 6 general emotional expressions :
  happyness, surprise,
  sadness, fear, disgust, anger

- FACS [1978] coding system – still preferred in US
- discrete
- visual effectof (groups of) muscles

Coding of facial expressions: MPEG4



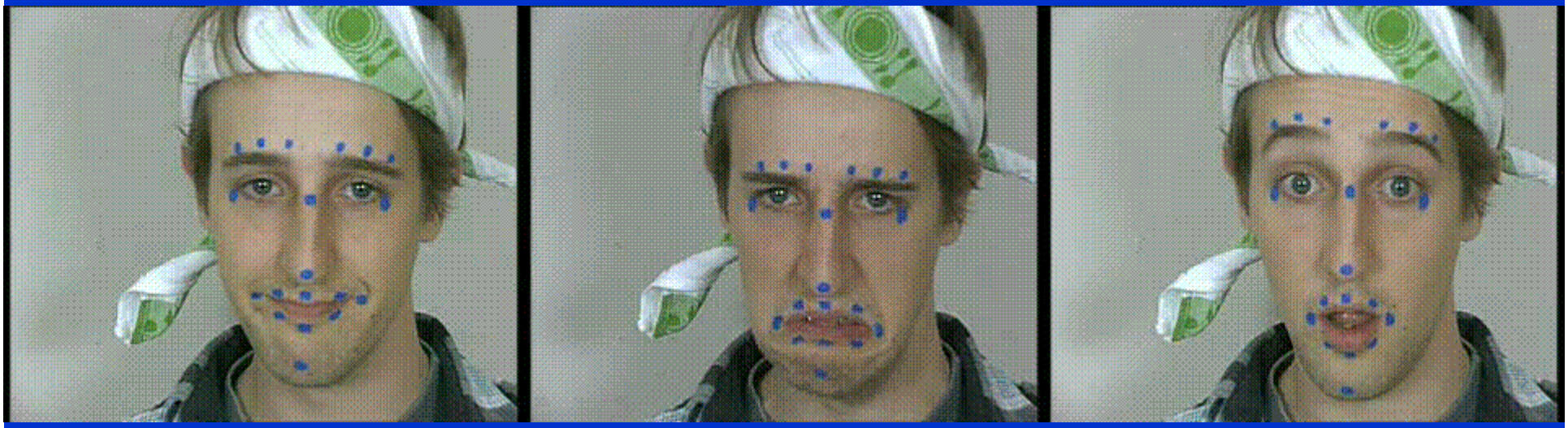
- 1999: ISO standard
- widely used
- 68 normalised parameters (FAPs)
- continous
- black points x, y (z) displ
- min, max, neutral value

Emberi arckifejezések leirása: MPEG4

#	FAP name	FAP description	Unit
31	raise_l_i_eyebrow	Vertical displacement of left inner eyebrow	ENS
32	raise_r_i_eyebrow	Vertical displacement of right inner eyebrow	ENS
33	raise_l_m_eyebrow	Vertical displacement of left middle eyebrow	ENS
34	raise_r_m_eyebrow	Vertical displacement of right middle eyebrow	ENS
35	raise_l_o_eyebrow	Vertical displacement of left outer eyebrow	ENS
36	raise_r_o_eyebrow	Vertical displacement of right outer eyebrow	ENS
37	squeeze_l_eyebrow	Horizontal displacement of left eyebrow	ES

Analysis of facial expressions – own investigation 2002

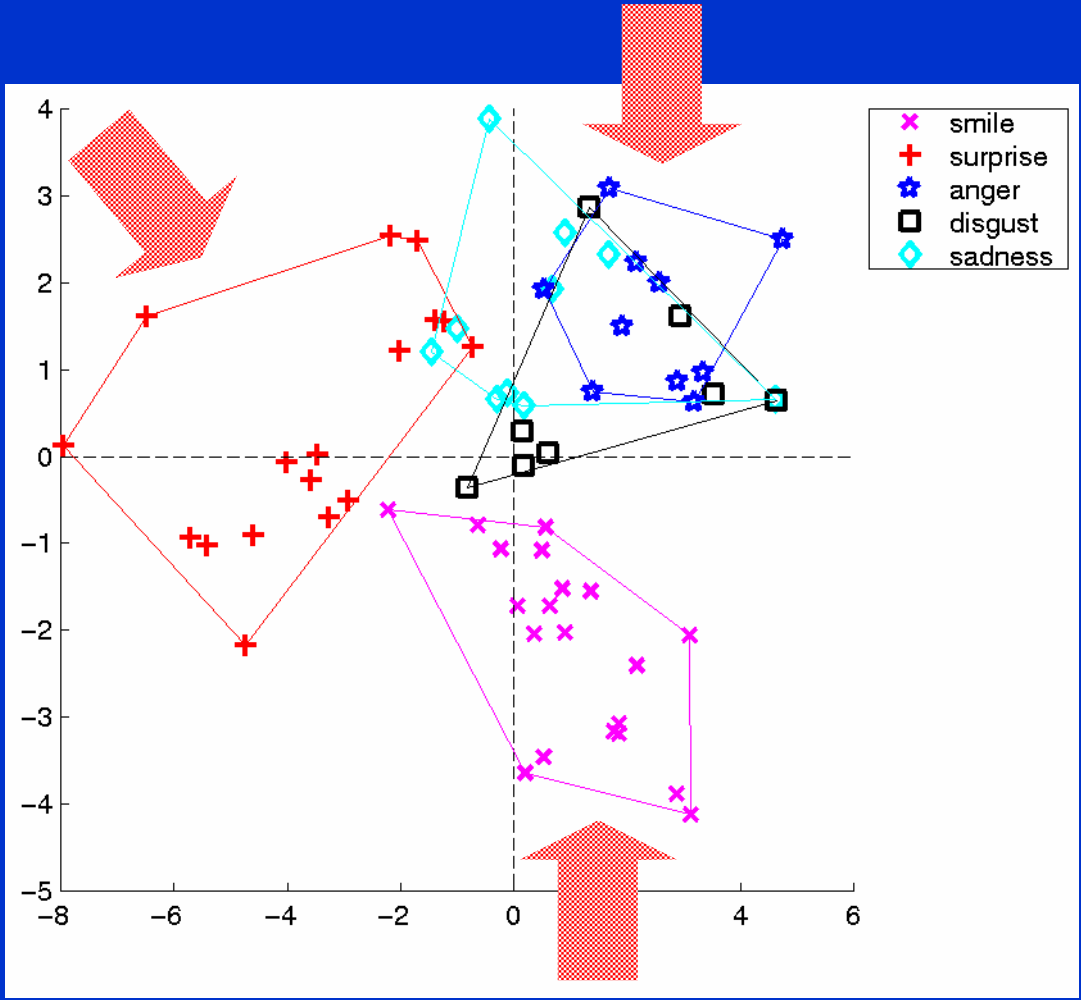
- On demand expressions of untrained subjects
- Track 15 FAPs marked on face



The E-Space: PCA

FAP variables	component 1 47.6%	component 2 26.3%	component 3 9.41%
3 open mouth	-0.17	0.22	0.63
4 lower middle upper lip	-0.20	0.25	-0.31
5 raise middle lower lip	0.17	-0.26	-0.59
6 raise left corner point mouth	0.12	-0.34	-0.04
7 raise right corner point mouth	0.14	-0.39	0.13
12 stretch left corner point mouth	0.15	-0.41	0.19
13 stretch right corner point mouth	0.15	-0.41	0.20
31 raise left inner eyebrow	-0.34	-0.12	-0.04
32 raise right inner eyebrow	-0.34	-0.09	-0.02
33 raise left middle eyebrow	-0.35	-0.14	-0.05
34 raise right middle eyebrow	-0.35	-0.14	-0.02
35 raise left outer eyebrow	-0.35	-0.07	-0.13
36 raise right outer eyebrow	-0.35	-0.10	-0.06
37 squeeze left eyebrow	0.22	0.30	-0.19
38 squeeze right eyebrow	0.23	0.23	0.02

The E-Space: 2 components' graph
(tracked)



Hor: eyebrows

Vert: mouth corners

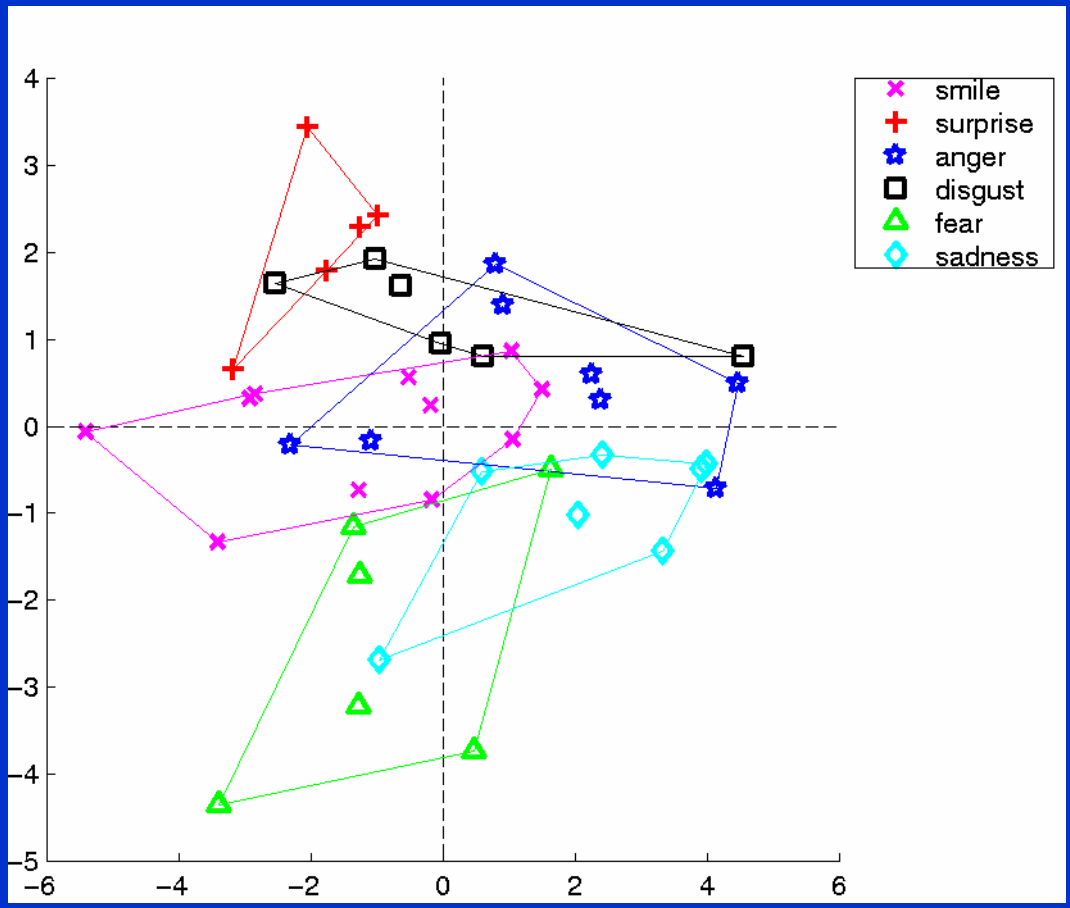
The E-Space: CharToon data

The E-Space: CharToon data

- Experienced animator

- 59 expression stills:
  - Anger: Be_Careful_What_You_Say
  - Smile: Absolute_Joy, Adoration, Smile
  - Other: Doubt_suspicion

The E-Space: Graph (cartoon)



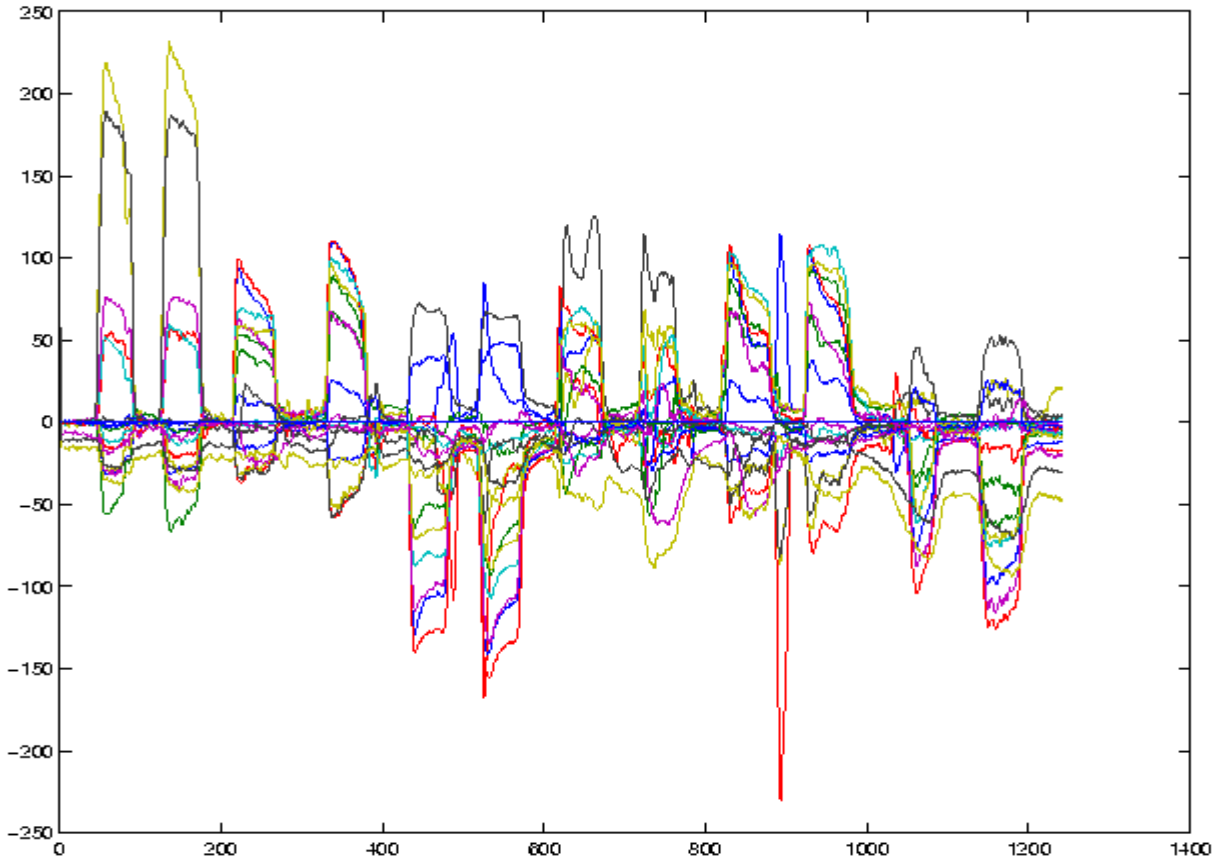
- Hor: eyebrows
- Vert: stretch mouth
- 3rd: smile/squeeze

(all a-symmetric)

The E-Space: Conclusion (stills)

- Tracker FAPs not always sufficient to distinguish (negative) emotions
- More symmetry in tracked expressions
- Animator uses more of the expression space
- smiling shape of the mouth and raising of the eyebrows are most significant

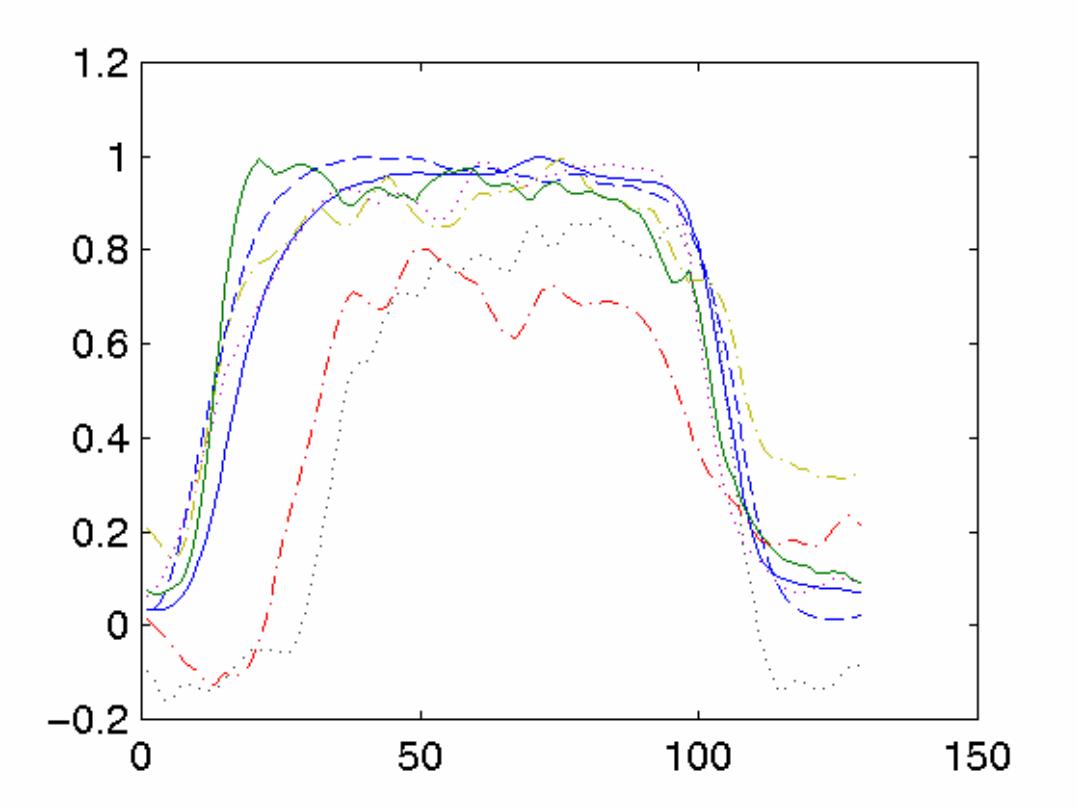
Time curves of FAPs



Actuation of muscles during smile and anger (I. Essa, PhD at MIT, 1995)

Figure 7-10: *Comparison between peak muscle actuations for Smile and Anger Expressions. The dot product of these two vectors is 0.94.*

The E-Space: Time curves



Co-actuation of muscles during smile (I. Essa, PhD at MIT, 1995)

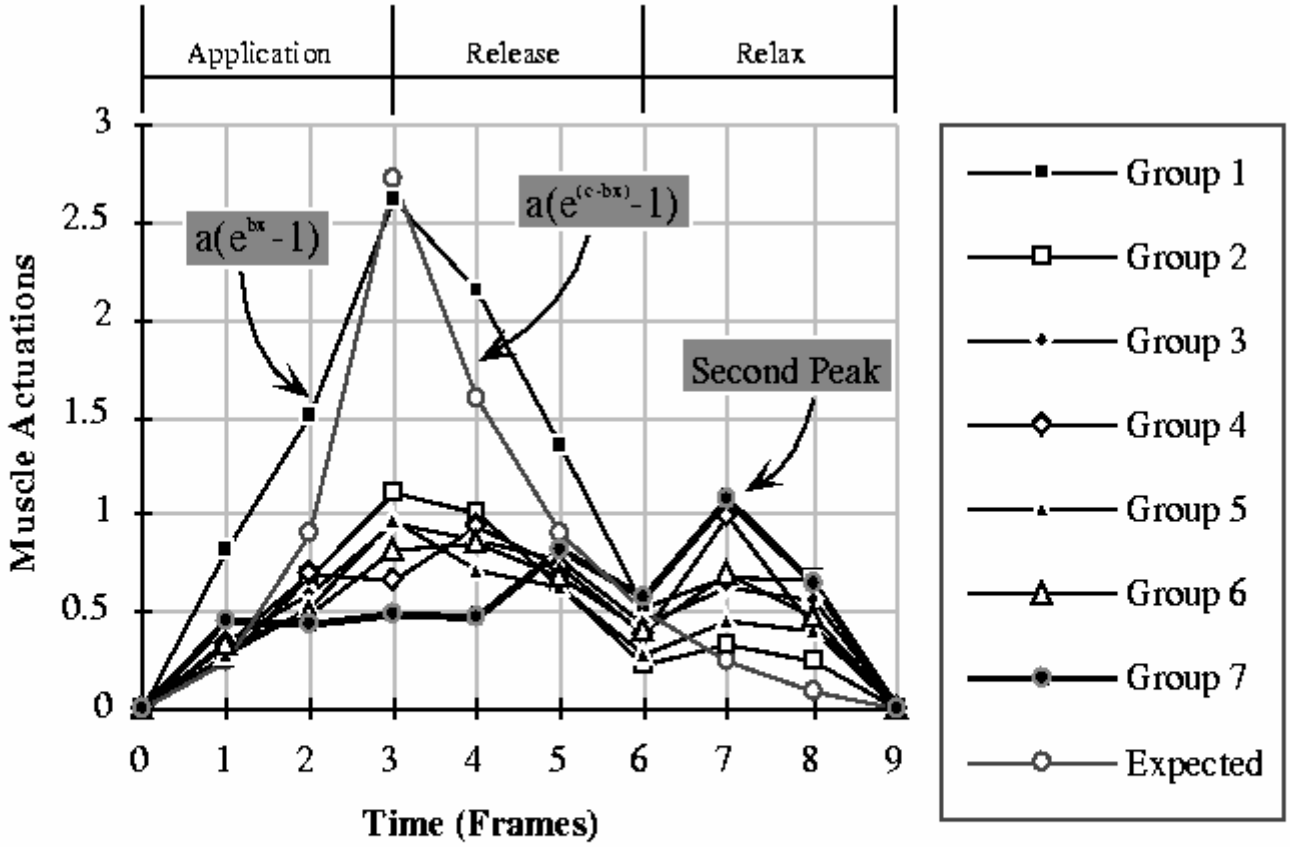
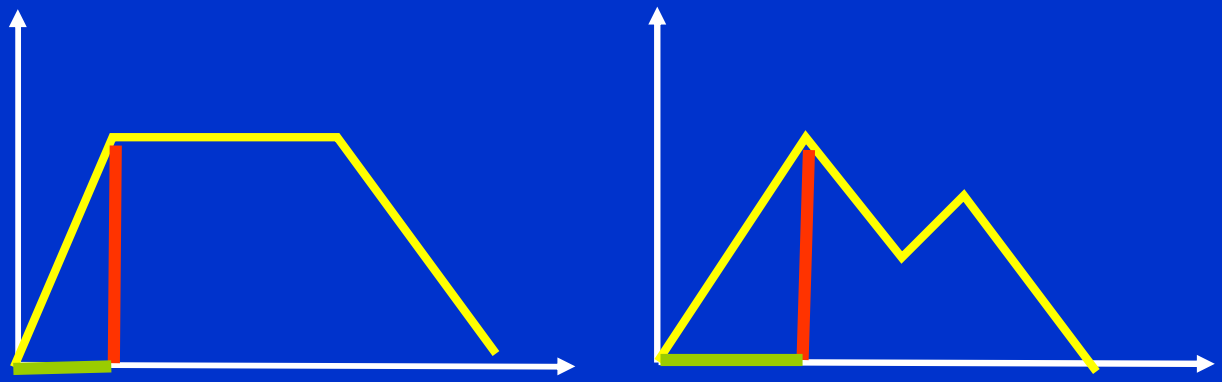


Figure 6-7: *Actuations over time of the seven main muscle groups for the expressions of smiling – lip motion. The plots shows actuations over time for the seven muscle groups and the expected profile of application, release and relax phases of muscle activation.*

Facial expression analysis - dynamism

- Time IS characteristic and important
- Trapezoid activation function, variants (camel, ...)
- Still open questions: shapes, constraints, ...
- No data(base) published (MS either)



Overview of the talk

Intro about ECAs

FA Analysis of facial expressions of humans and cartoons

➡ Non-photorealistic facial animation - CharToon

Emotion Disc

Evaluation of personality cues

Full body

Style in gesture - GESTYLE

Challenges for IP

Facial animation - image morphing

- To make in-betweens for animations (warp)
- To generate a multitude of models (morph)
- To generate caricatures/exaggerated faces

Example tool: FantaMorph

Facial animation – model based

Basic principle:

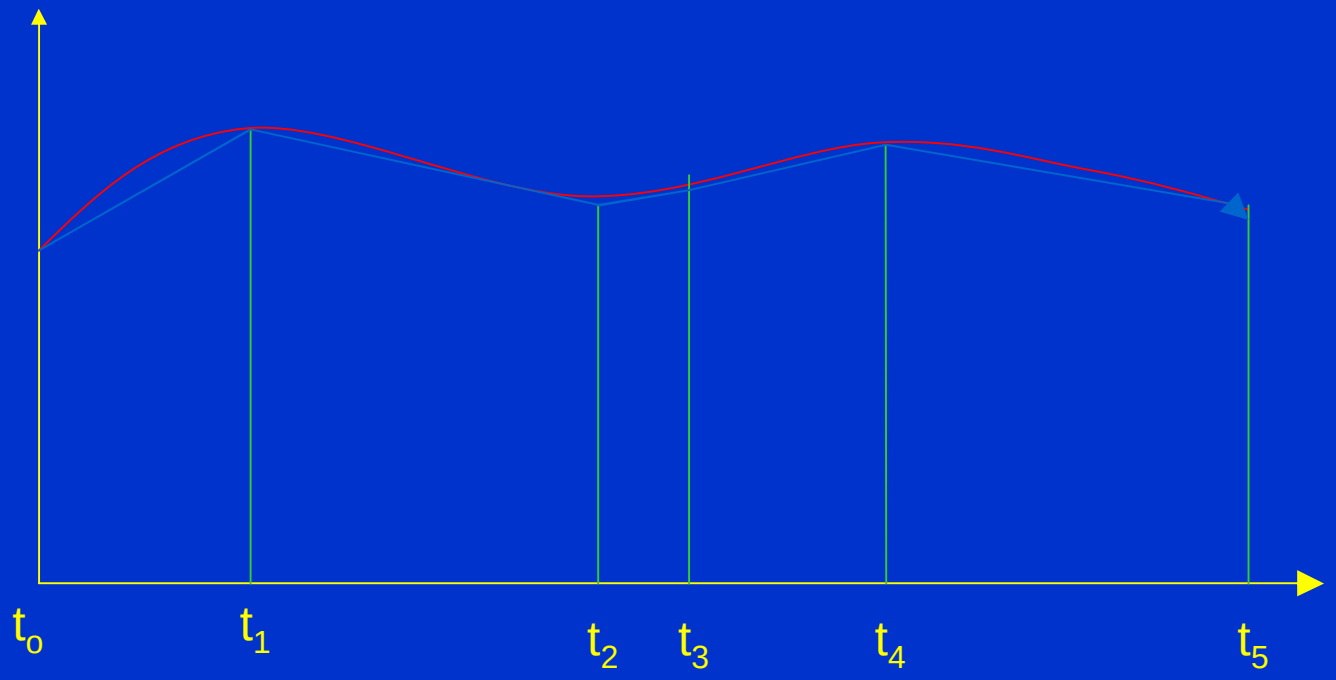
- Deformation principles – 2d skeleton, 3d muscle, ...
- Key positions: values of controllers
- Inbetweens by interpolation

Issues:

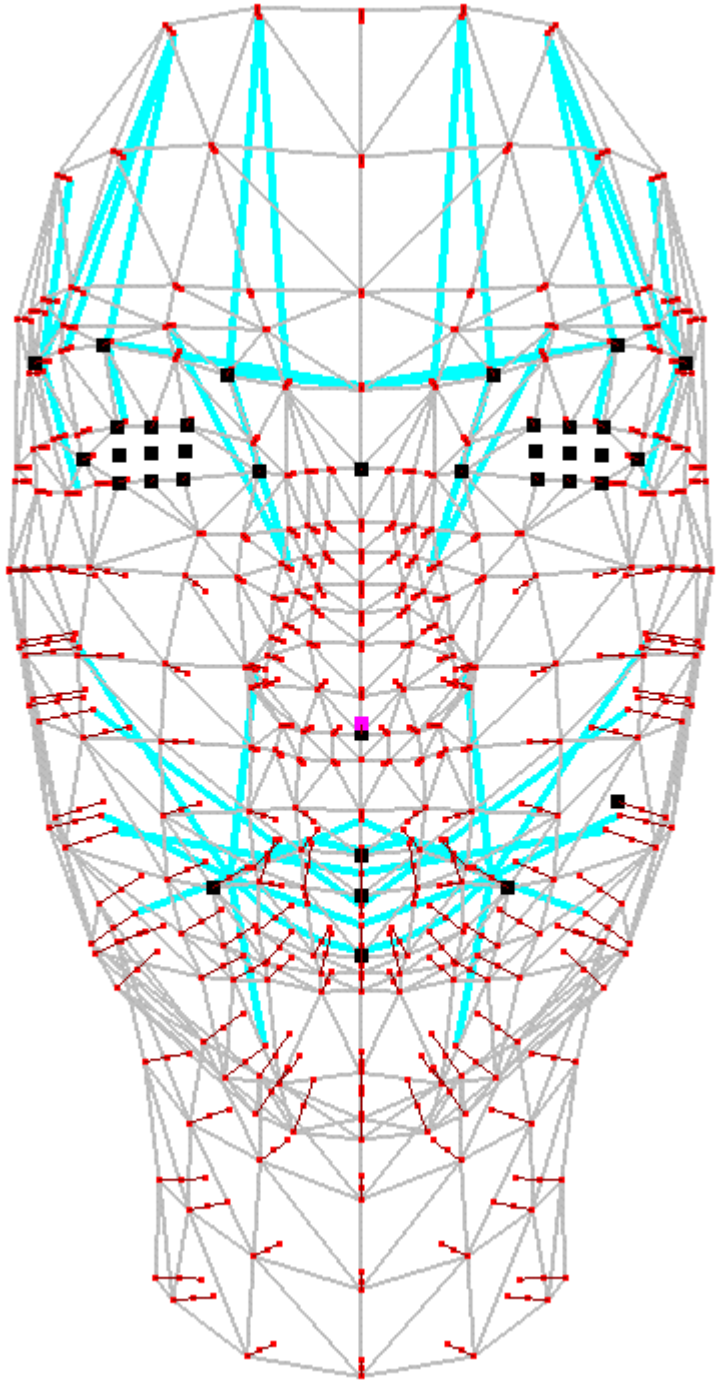
- Reusable, parametrised repertoire
- Blending
- Superposition (talking while smiling)
- Visual speech - coarticulation

Facial animation

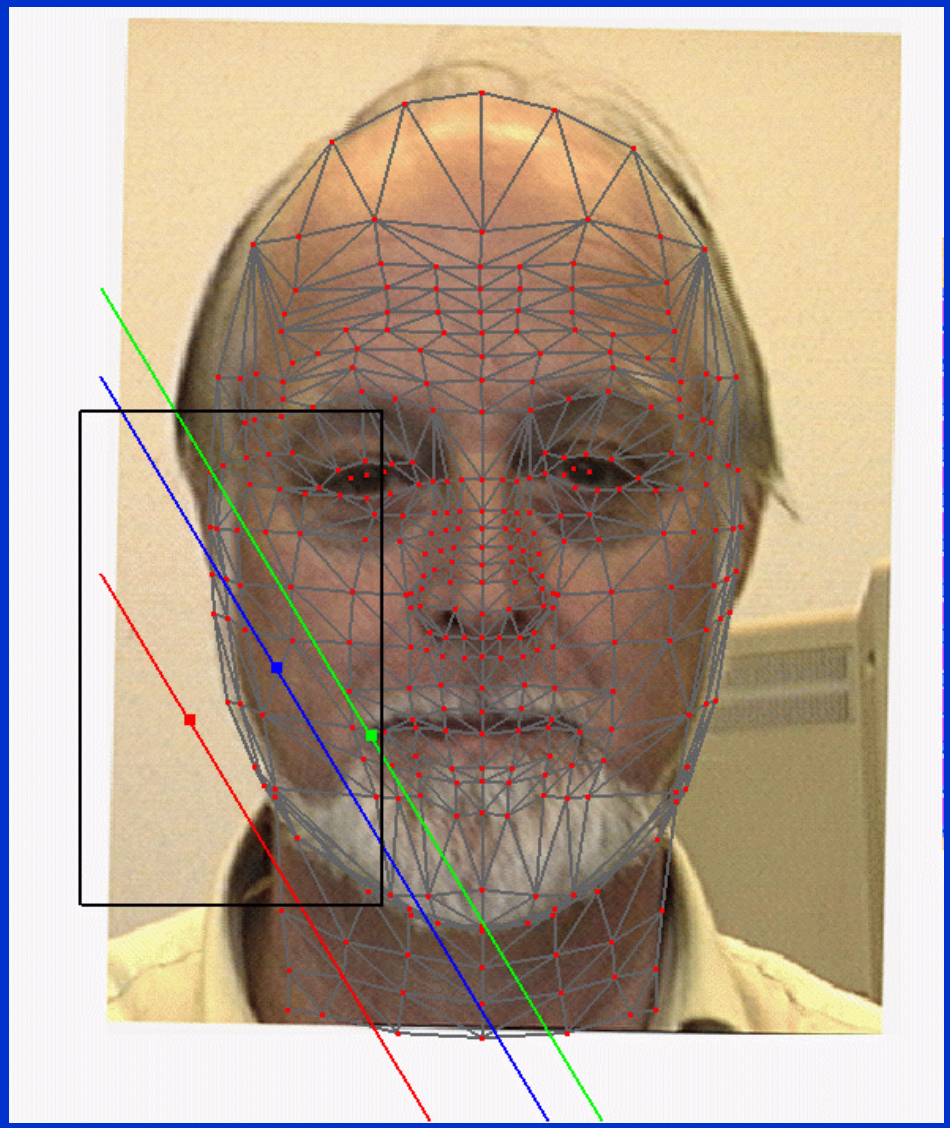
- how to interpolate: linear, C^1, C^2
- how to choose t_i (times for keyframes)

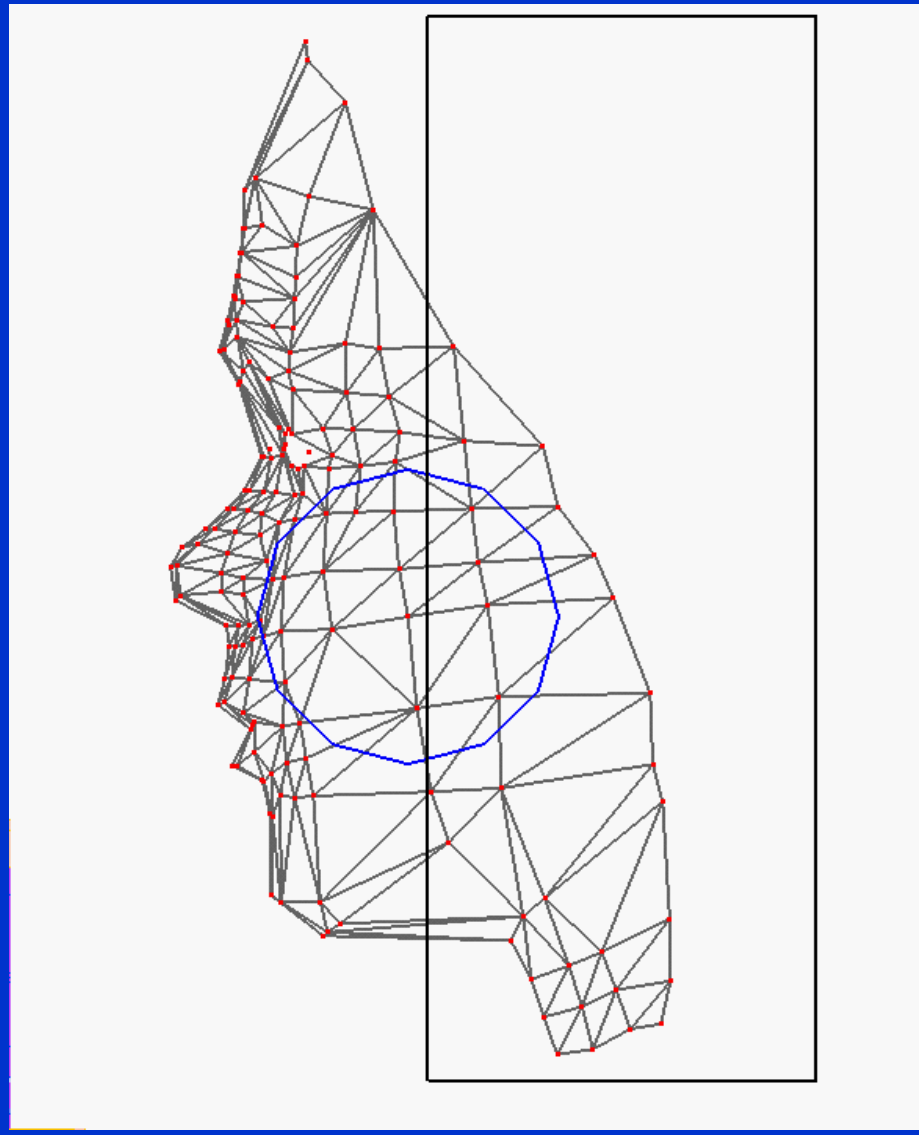


3D model

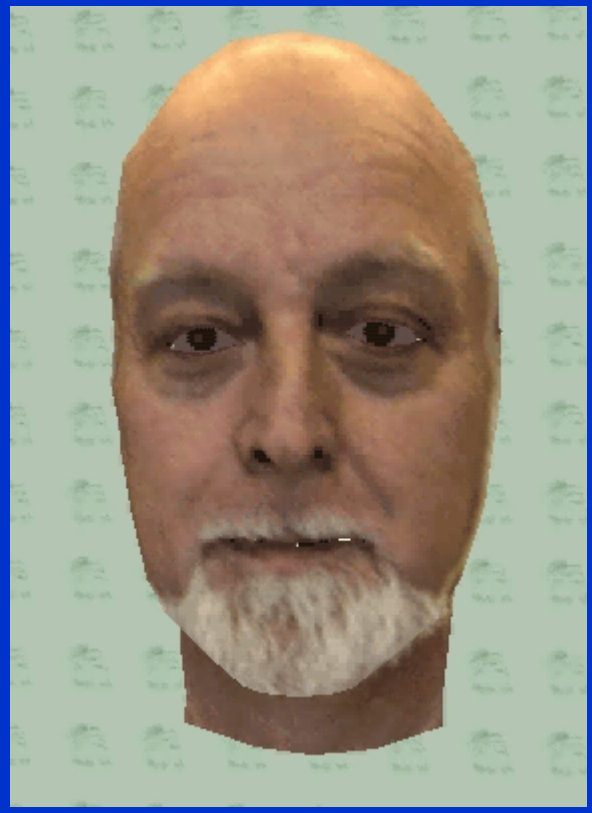


3d facial modelling (CWI)

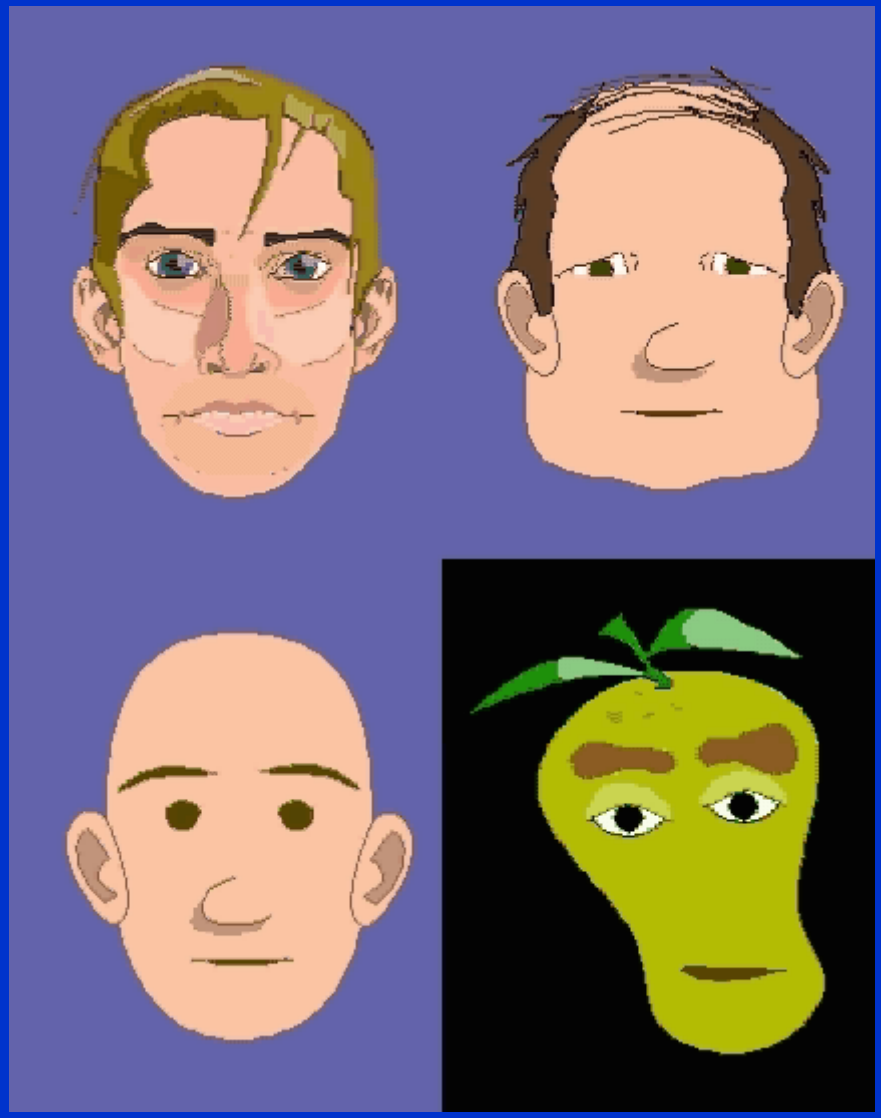


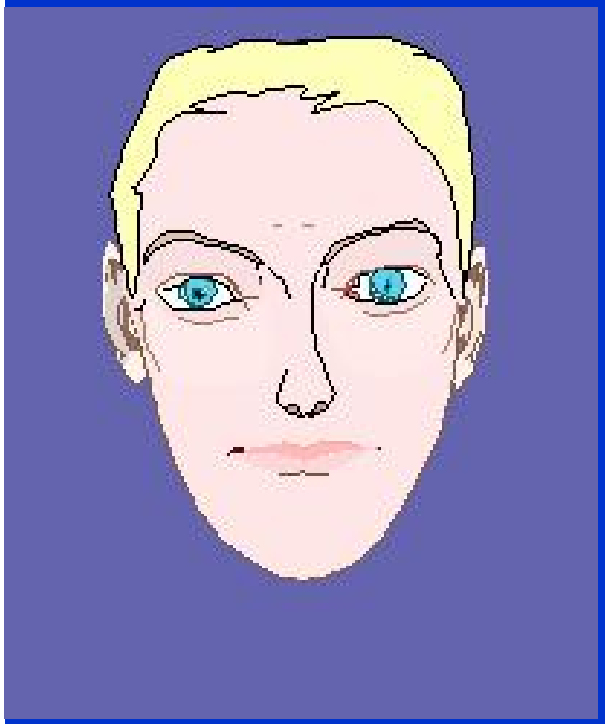


3d facial animation (CWI)

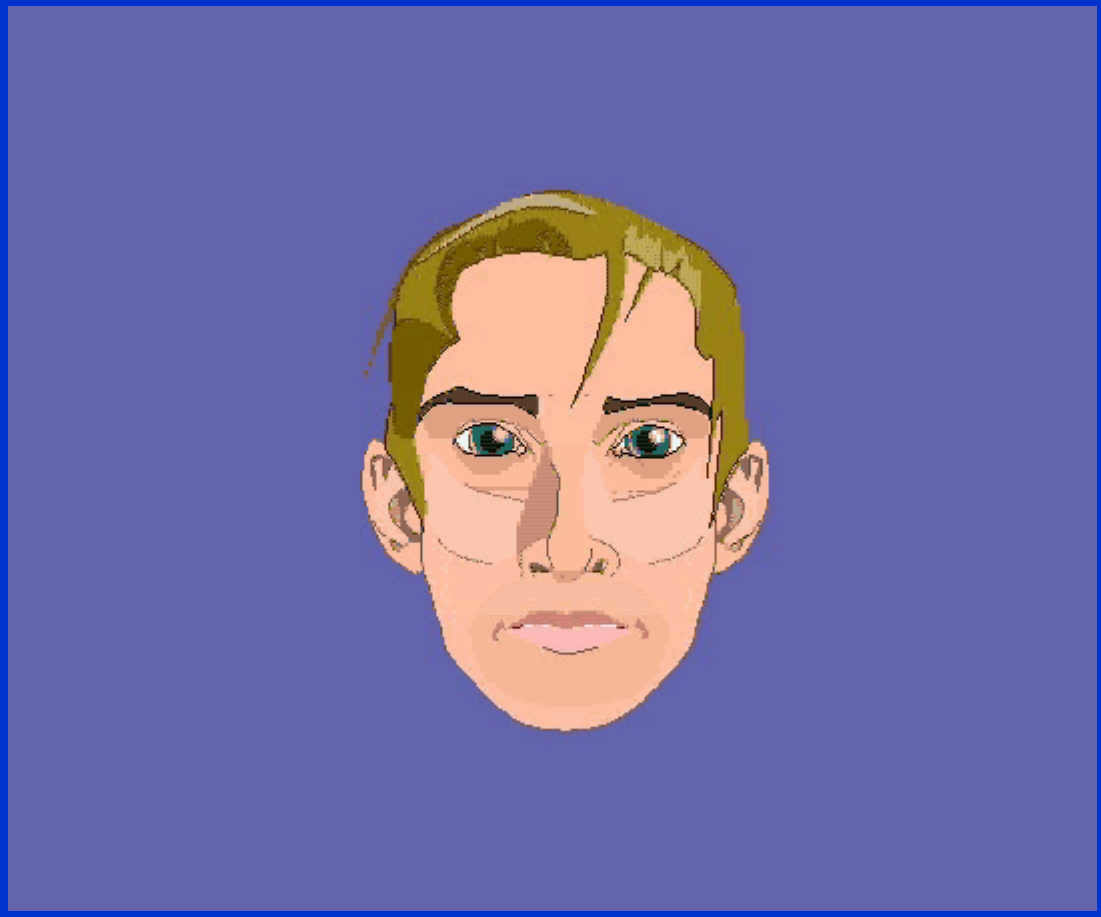


Same animation for different faces

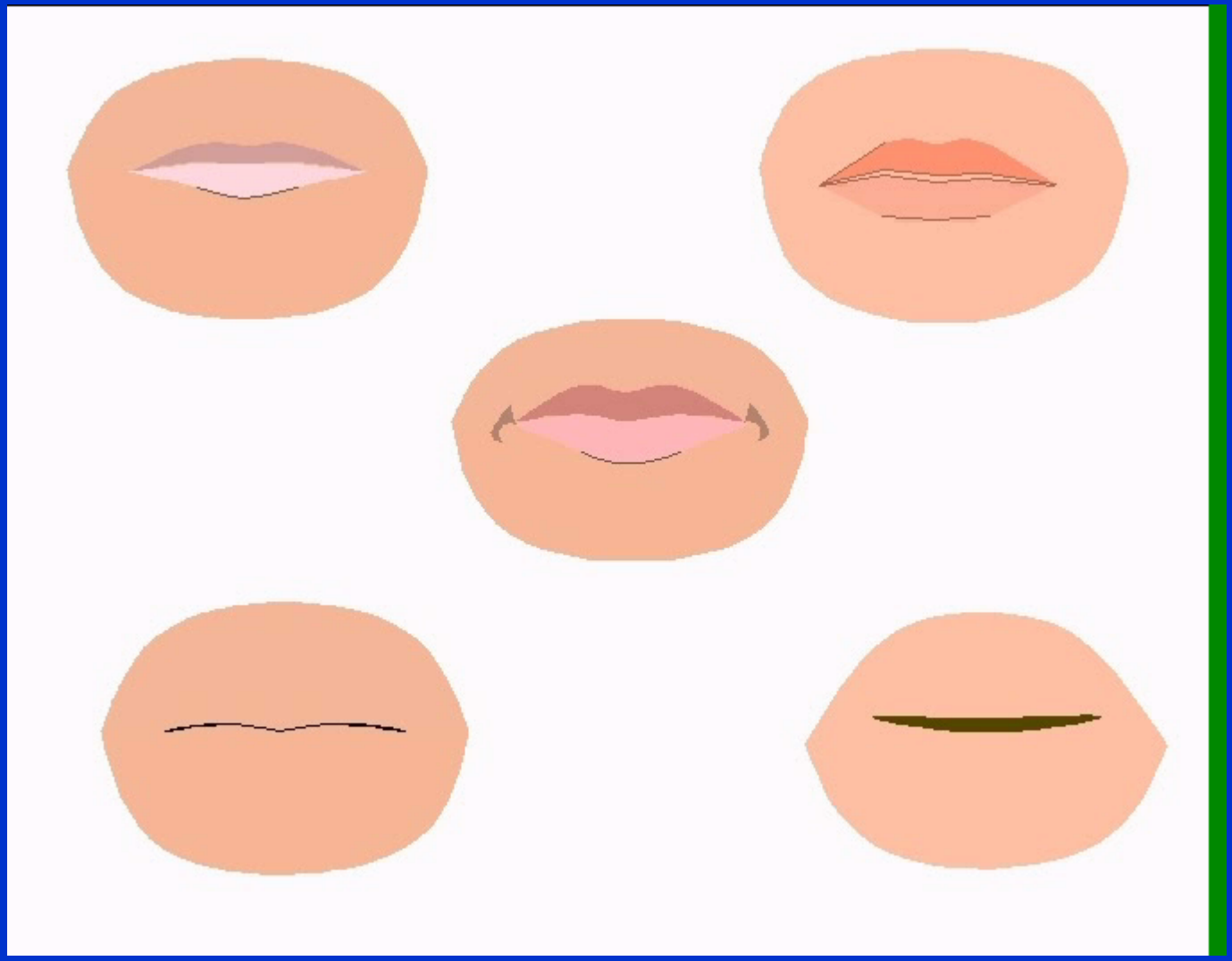




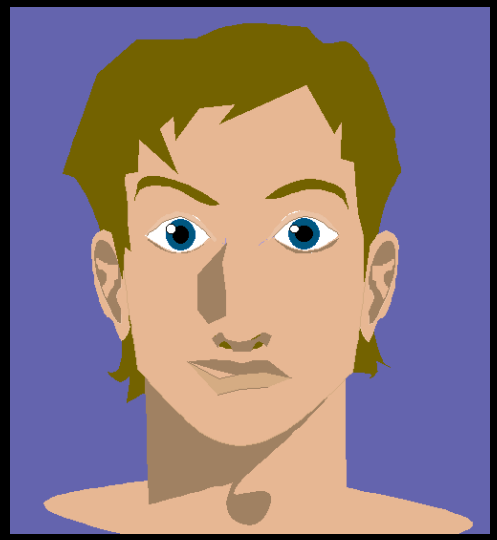
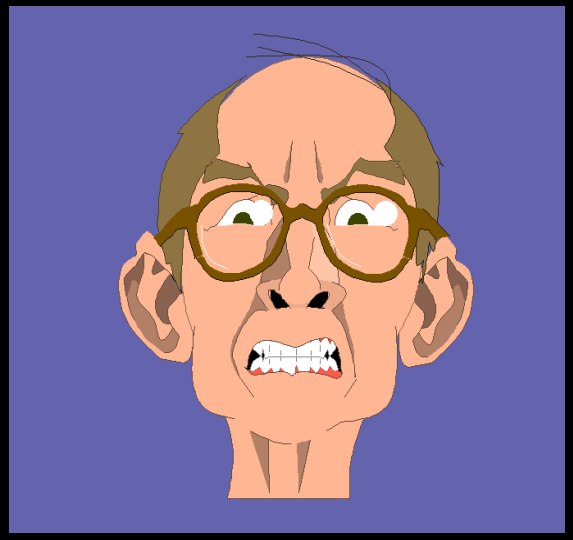
Visual speech



Lip-sync



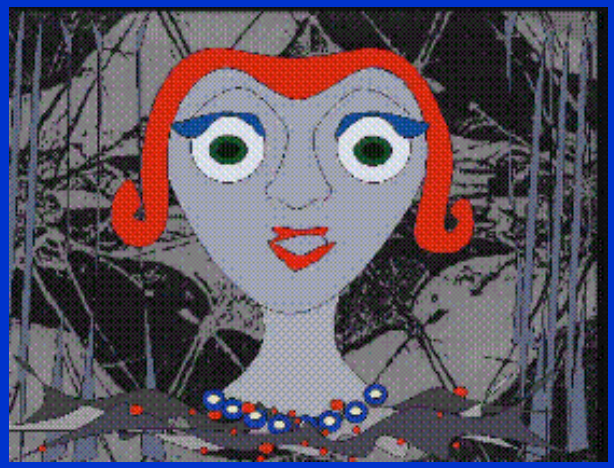
Some 2d head models

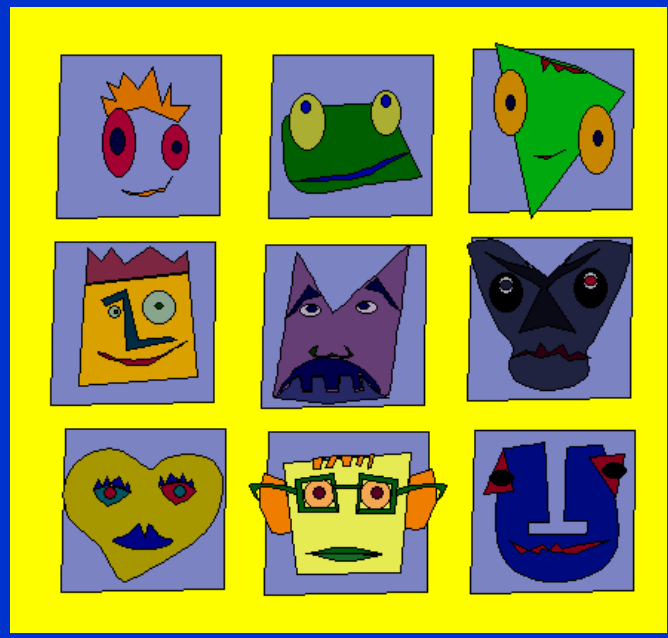


Some movies made with CharToon

by A. Lelievre and Zs. Paal

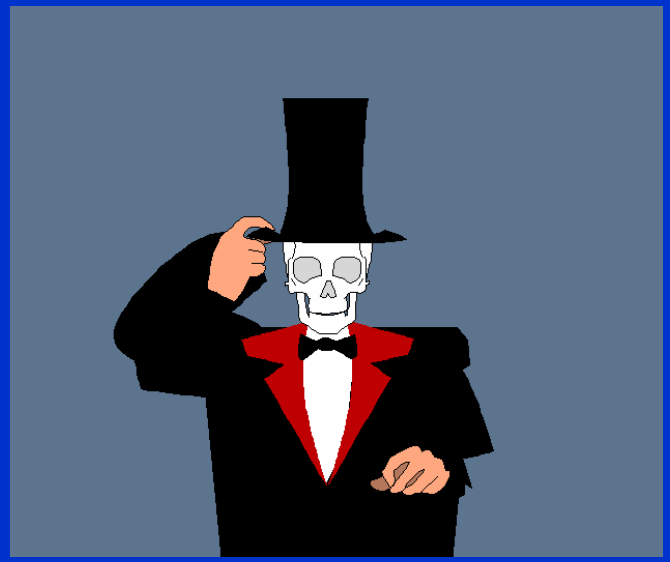


Lily



Nine faces

Frans



Magician

3D realistic vs. 2D cartoon faces

	2D cartoon	3D realistic
model making	easy	very difficult
animation	fast	slow
aesthetics	appealing	disturbing
expressive aspects	presence emotions speech	anatomy details
expressions	beyond realism!	within realism

Component repertoire

Unlimited freedom to design faces from CharToon components

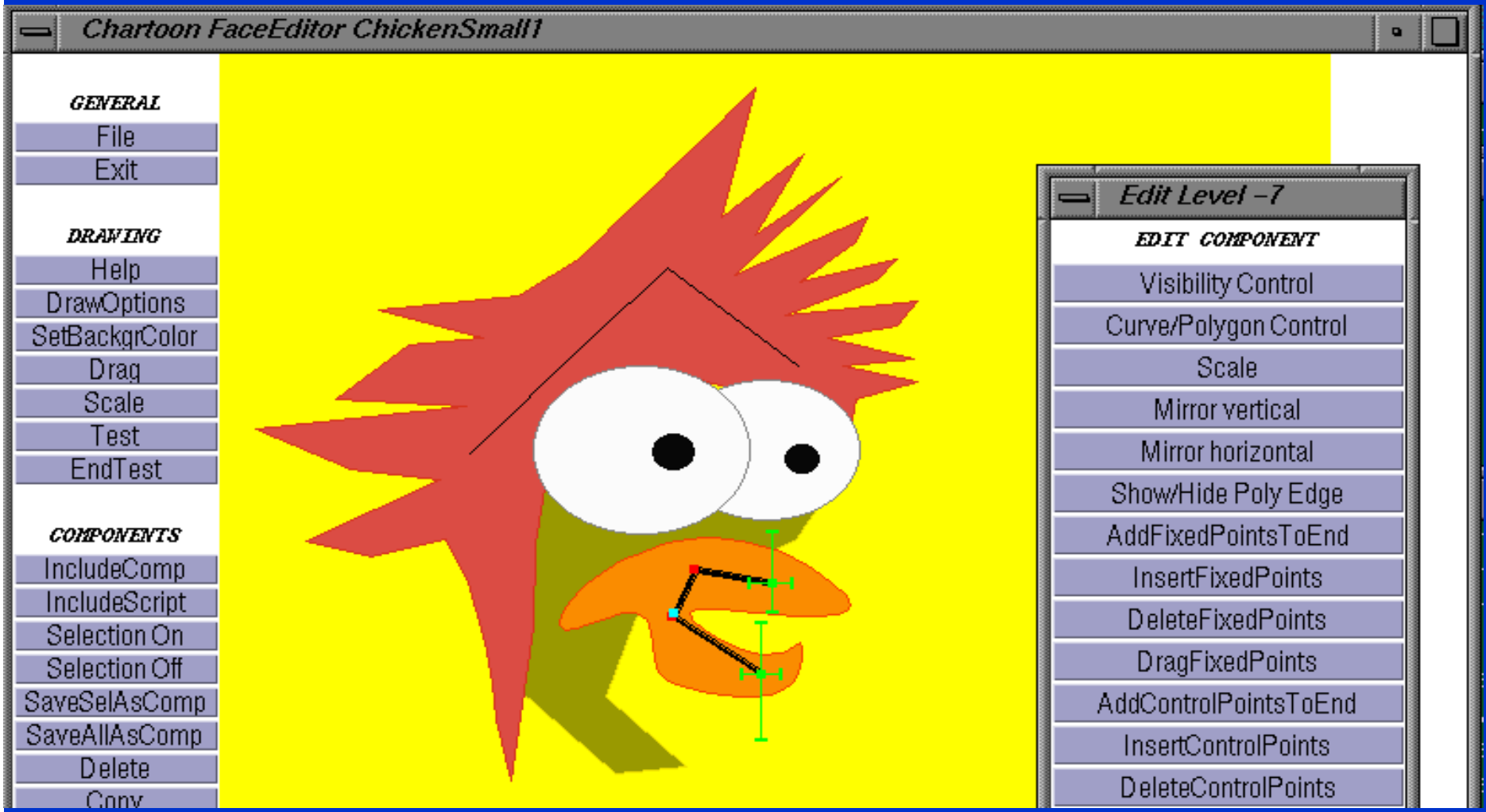
Repertoire of:

- facial feature components
- (still) expressions and animations

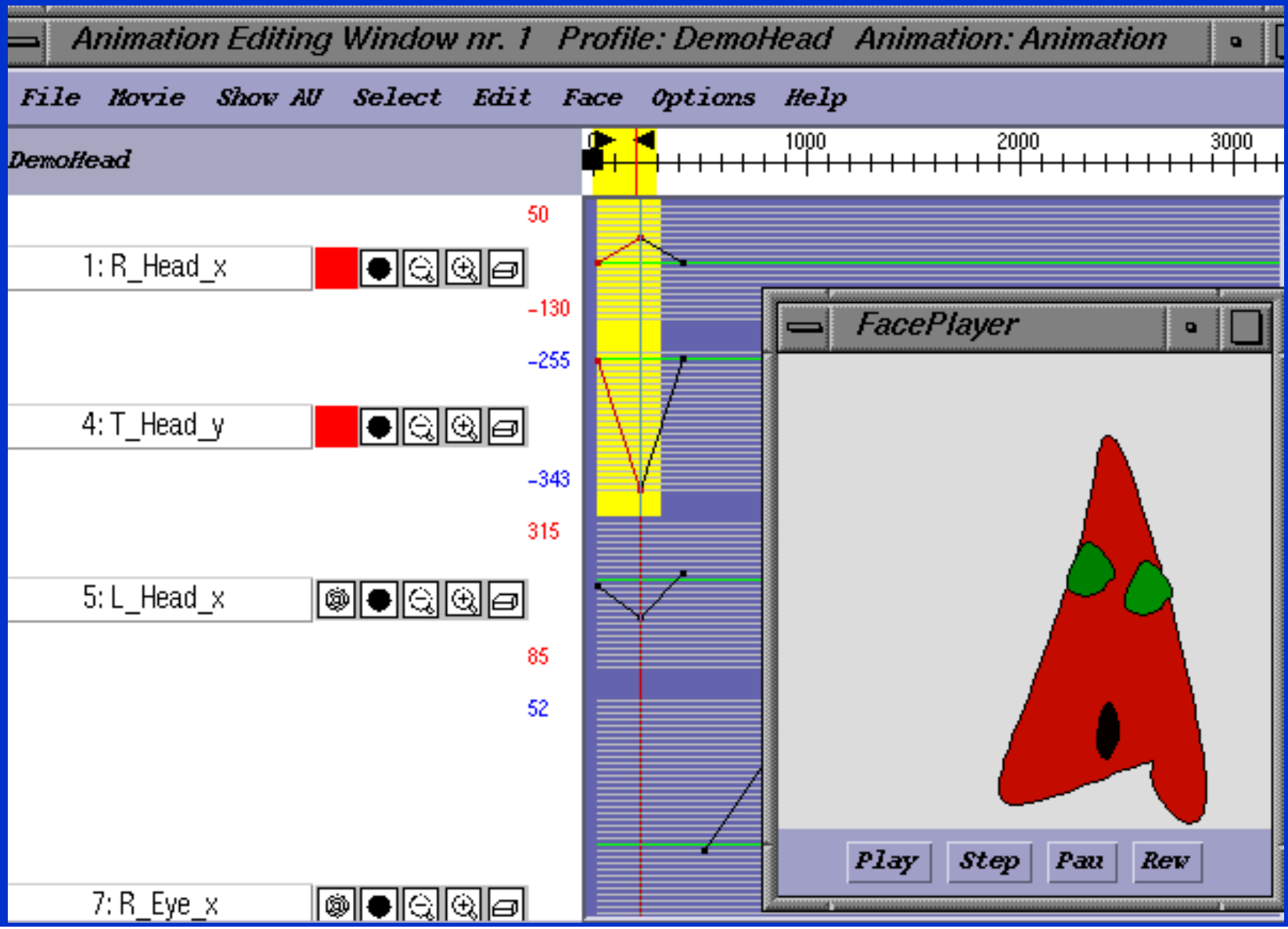
Support to reuse the components:

- technical level (design, UI)
- ‘mix & match’ recipes with respect to expected functionality

Animating faces in CharToon I

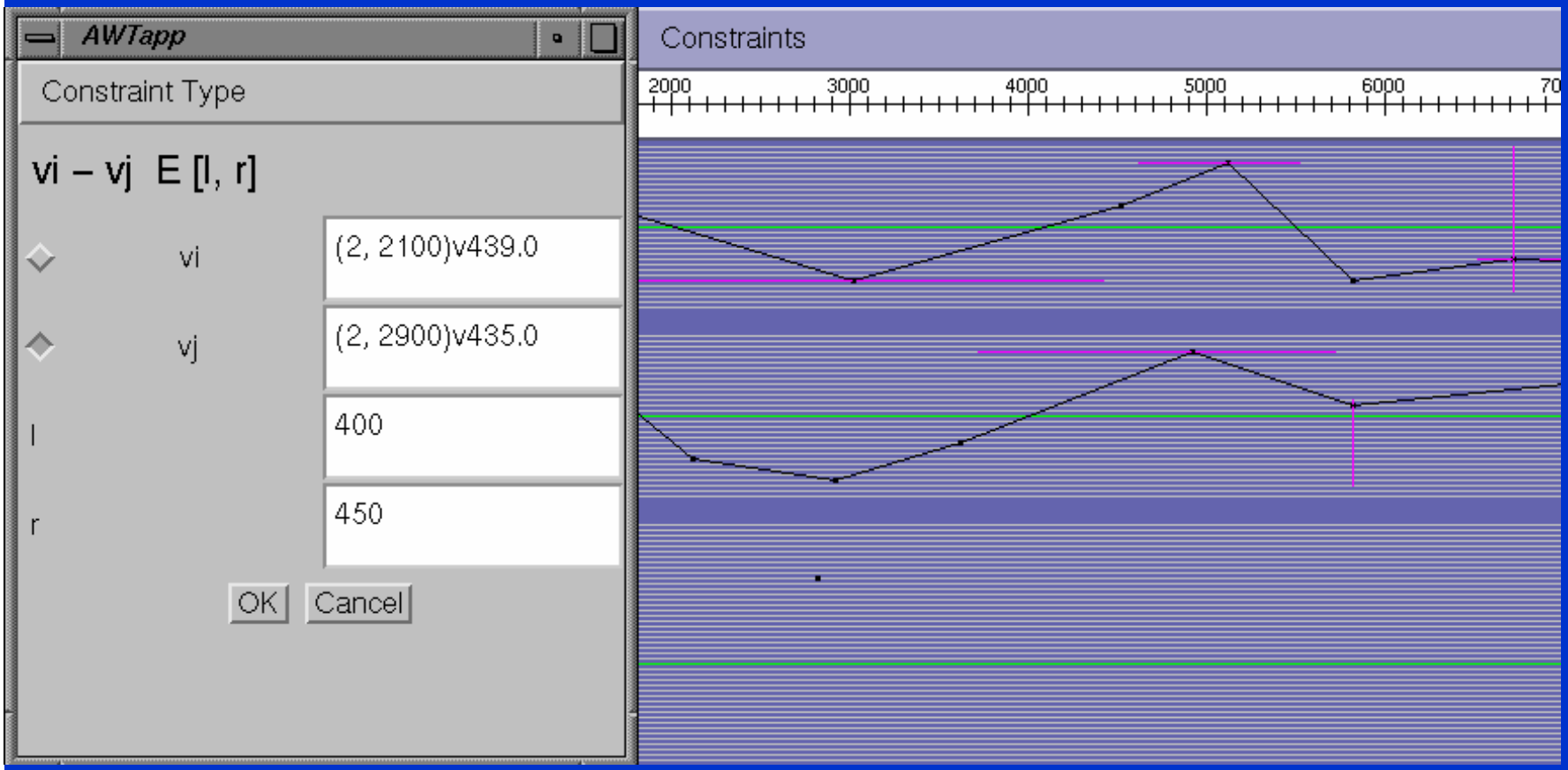


Animating faces in CharToon II

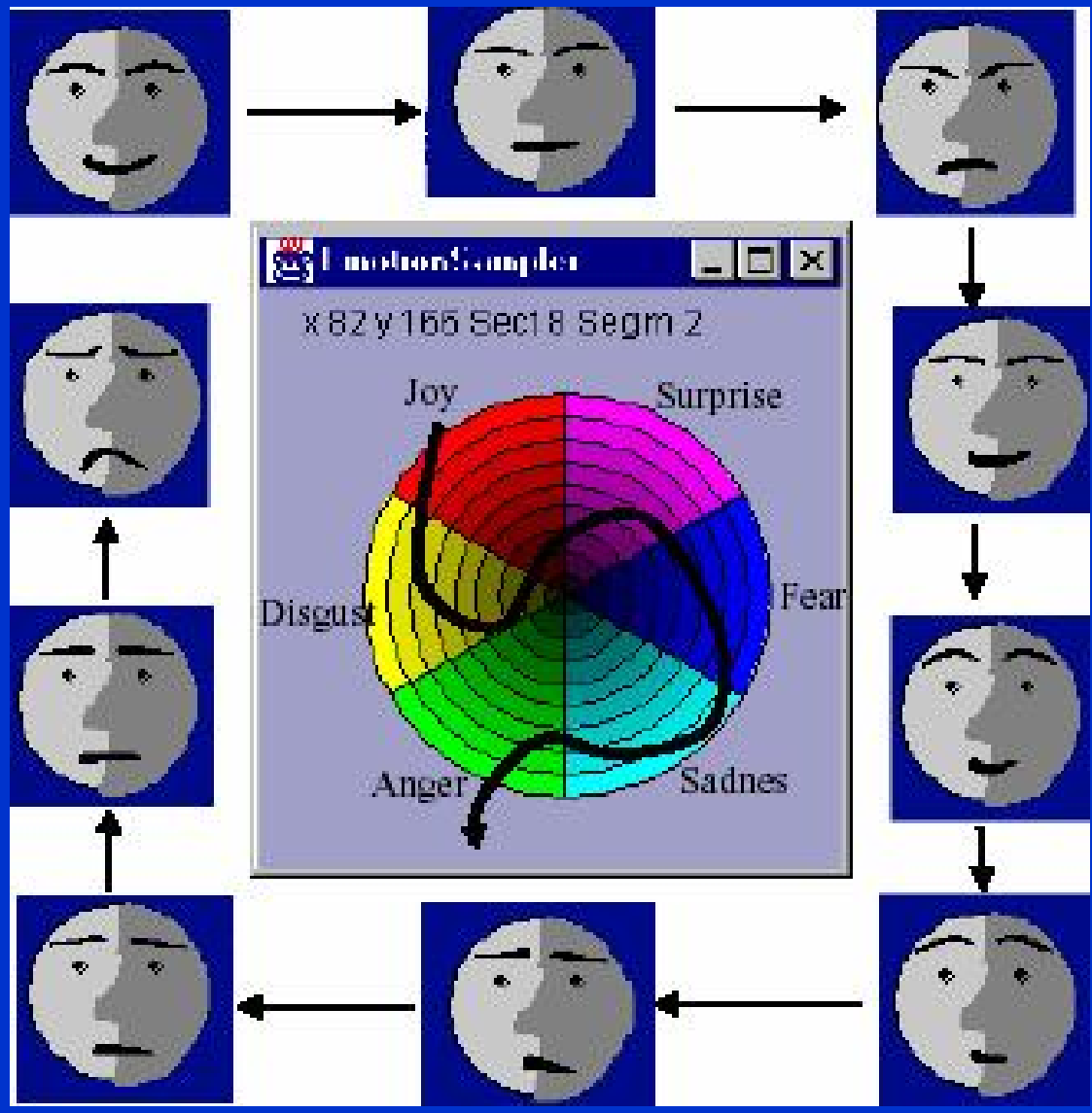


Animating faces in CharToon III

+ constraints technology



Emotion Disc



Emotion Disc

Live Demo

Emotion Disc

- Cyclic neighborhood of 6 basic emotions
- Tool to:

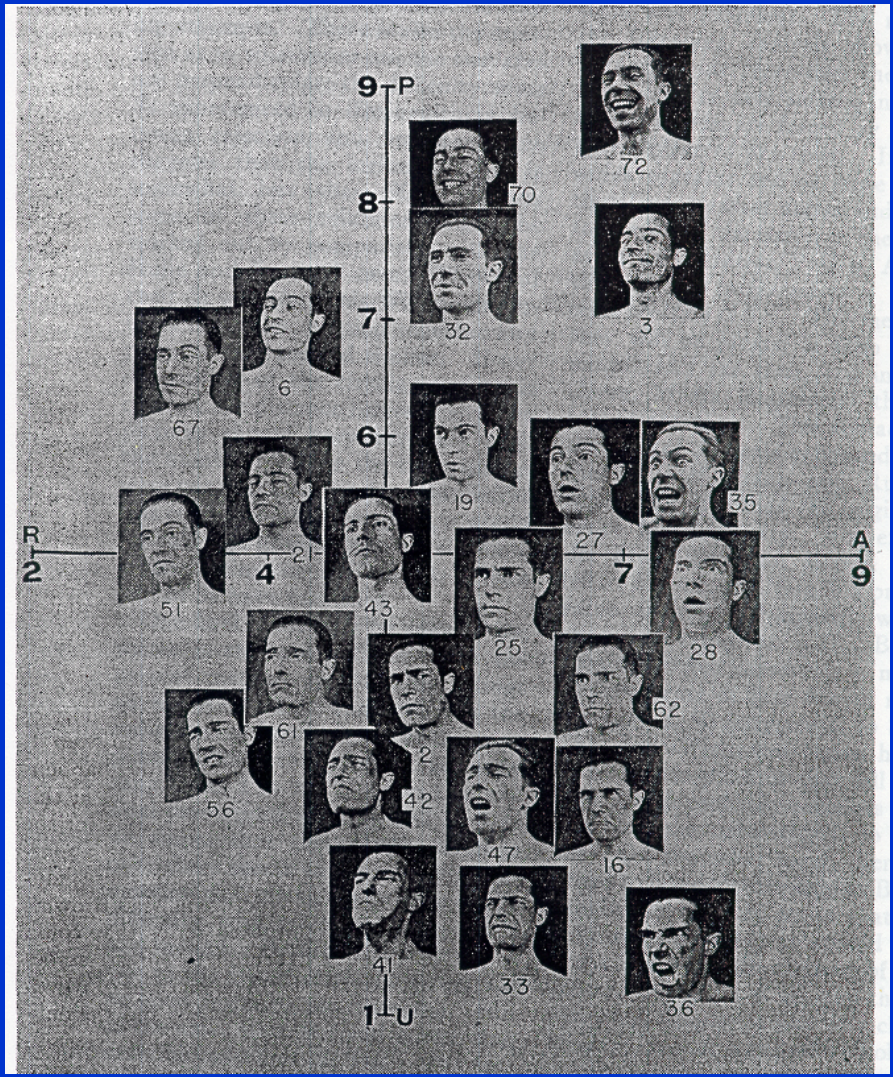
  - explore the expression space

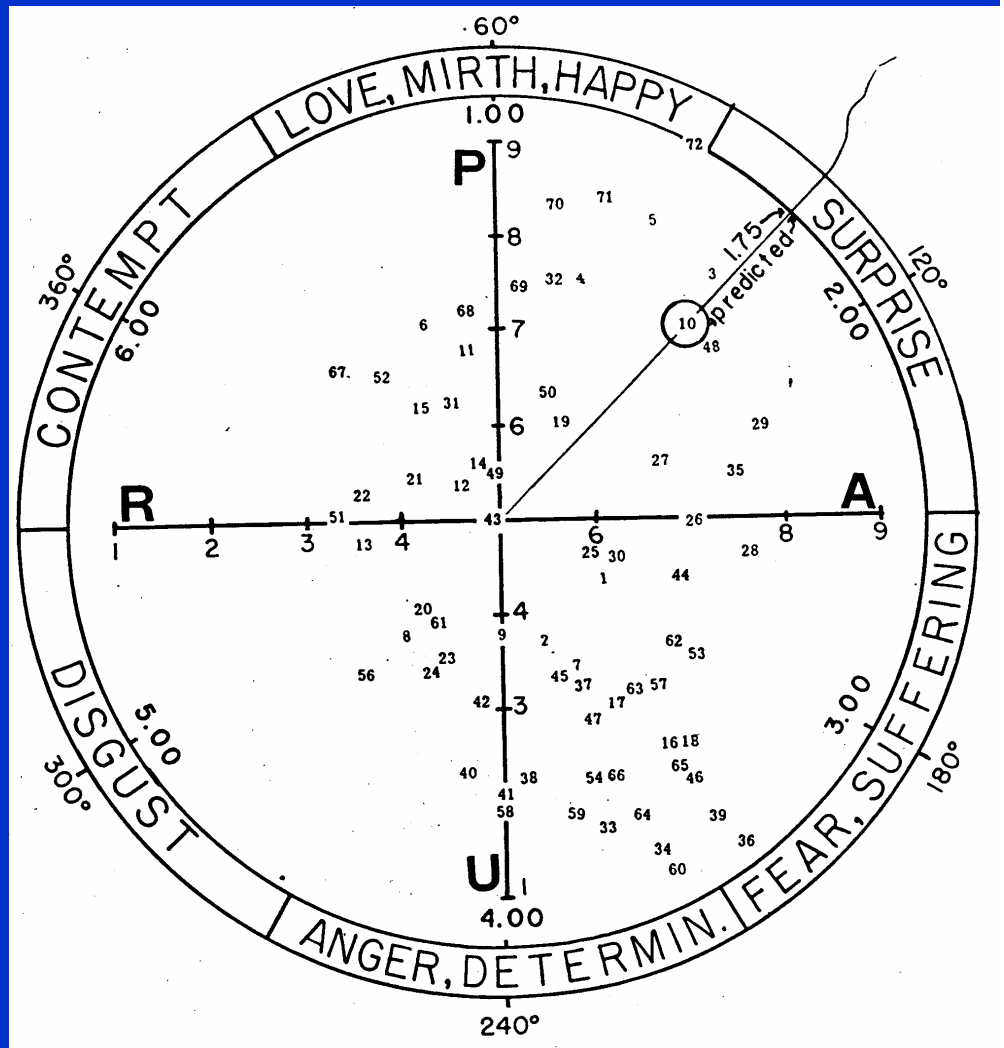
  - control the expressions

- Experiments with simple but still expressive 2D faces

  - how expressive they are?

  - contradicting expressions (2-3 discs)





From: Harold Schlosberg: The Description of Facial Expressions in Terms of Two Dimensions, Journal of Experimental Psychology, Vol. 44. No. 4. Oct. 1952, pp 229-237.

Overview of the talk

Intro about ECAs

FA Analysis of facial expressions of humans and cartoons

Non-photorealistic facial animation - CharToon

Emotion Disc

➡ Evaluation of personality cues

Full body

Style in gesture - GESTYLE

Challenges for IP

Evaluation of talking heads

- Work with E. Krahmer at als.
- Effect of facial signals to express:

  - focus --- eyebrow & pitch
  - personality --- eyebrow, gaze and speech

- See publications, experimental mat. on web page

The evaluation research questions

- Do facial cues contribute to the perception of extraversion?
- How do gaze + brows + speech cues add up?
  - strength
  - inconsistency

Personality for ECAs

- Personality is important:
  - to improve comprehension (Laurel 1993)
  - to like and trust and ... the ECA better
    (Nass and Lee 2000, Isbister and Nass 2000, ...)
  - people attribute personality to agents
    CASA paradigm (Reeves and Nass 1996)
  - fun, variety

Personality for ECAs

- Trait theory (e.g. Allport 1937)
- "Big 5" (e.g. Wiggins 1997):
  Extroversion (dominance),
  Agreeableness, Conscientiousness,
  Emotional stability, Culture
- Cues studied so far:
  - speech (pitch and duration)
  - gaze
  - posture (Isbister and Nass 2000)
  - language

The experiment

	Gaze	Speech	Brows
Extrovert	fixed on user 2 blink	wide pitch range more variation higher tempo	2x accent 3x100 ms stages each
Introvert	“move S-W look aside move center” 2x	low pitch range little variation 20% slower	no movement

2 x 2 x 2 = 8 personality profiles

The experiment

- Examples of stimuli from http://www.cwi.nl/~zsofi/stim.html

Evaluation of ECAs

- Hot topic, evaluation is a MUST
- Zs. Ruttkay, C. Pelachaud: From Brows to Trust (Kluwer, 2004)

Overview of the talk

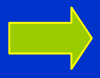
Intro about ECAs

FA Analysis of facial expressions of humans and cartoons

Non-photorealistic facial animation - CharToon

Emotion Disc

Evaluation of personality cues



Full body

Style in gesture - GESTYLE

Challenges for IP

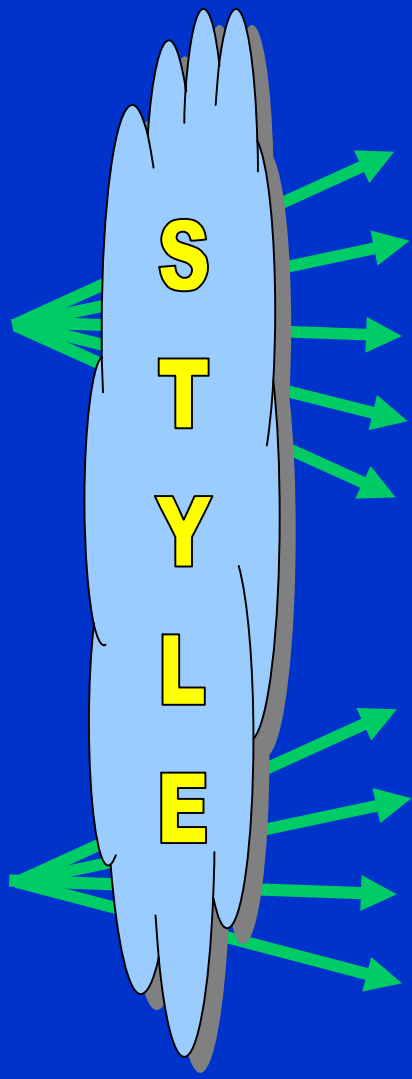
Research questions on gestures

- What gestures do people use?
  - references to environment & user (location, size, time, motion, shape,...)
  - to indicate actions and concepts
  - emotions, cognitive and physical state
  - dialog status
- What gestures should an ECA use?
  - function - application domain
  - redundancy - expressivity
  - realization – non-repetitive, tuned
- Modelling and sw engineering issues:

  motion characteristics,

  performance, reusability, ...

Communicative functions



emphasize

take turn/
start talking

Multi-modal signals

a) raise voice
b) look at listener
c) raise eyebrow
d) beat with hand
e) 11 (?) combinations

a) look at partner
b) raise hand
c) voice signal (hm-hm)
d) ...

Why to bother about style?

- Style is source of information on:
  - gender, age, physical state
  - personality, emotional state
  - ethnicity, profession
  - relationship between speaker and listener
  - situation (public/private, ...)

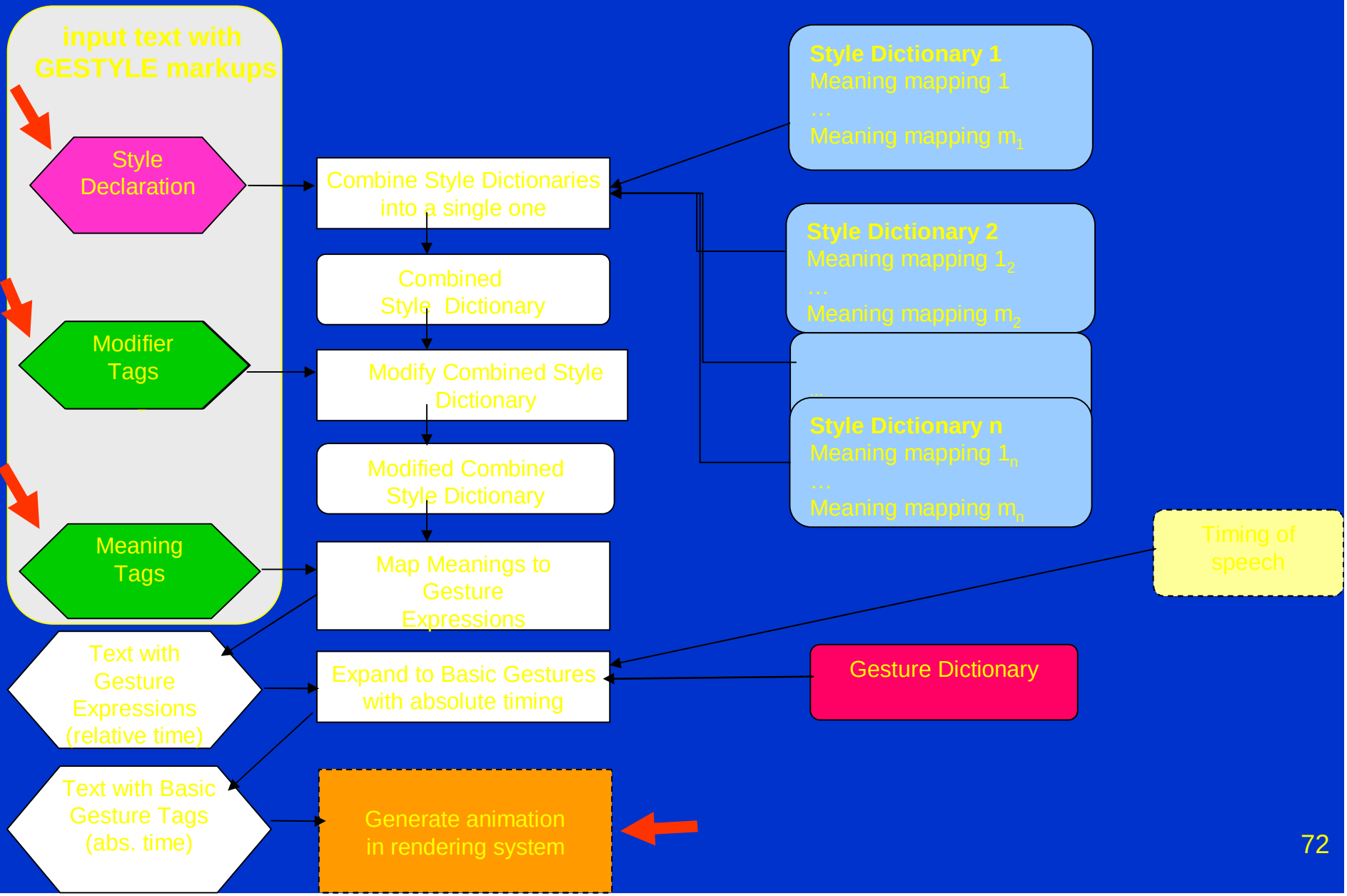
    ...
- variety and joy
- does matter in HCI (Nass et al. 2000, ...)

Our objectives

- ECA with style in
  nonverbal communication (and speech)
- Declare style of ECA
- Generate styled conversation by ECAs
- Framework (can be tailored/extended)
- Easy to use mark-up language
- Experiments!

GESTYLE

GESTYLE: the big picture



GESTYLE: Style dictionary

```
<StyleDictionary Name = "extrovert">

<Meaning Name = "emphasize" CombinationMode =
"DOMINANT">

 <GestureSpec>

        <MannerDefinition intensity="intense"/>

        <UseGest Name="NodAndBeat"/><PAR>

        <UseGest Name="LookAtPerson"/>

        <Probability P="0.7"/>

 </GestureSpec>

 <GestureSpec>

        <MannerDefinition motion_manner="sudden_on"/>

        <UseGest Name="Beat"/>
```

Definition and generation of gestures

- structural characteristics:

  composed of basic gestures by
  par, seq and repeat operators

- performance (manner) characteristics:

  intensity, duration, start_time, precision,
  jerkyness, symmetry, ... parameters

GESTYLE: Combining styles

- Mechanism to resolve conflicting prescriptions on:
  - gestures
  - modalities
  - manner
- Preferences to be given in Style Declaration
- Dynamical modifications of:
  - modality usage
  - manner
  - style dominance

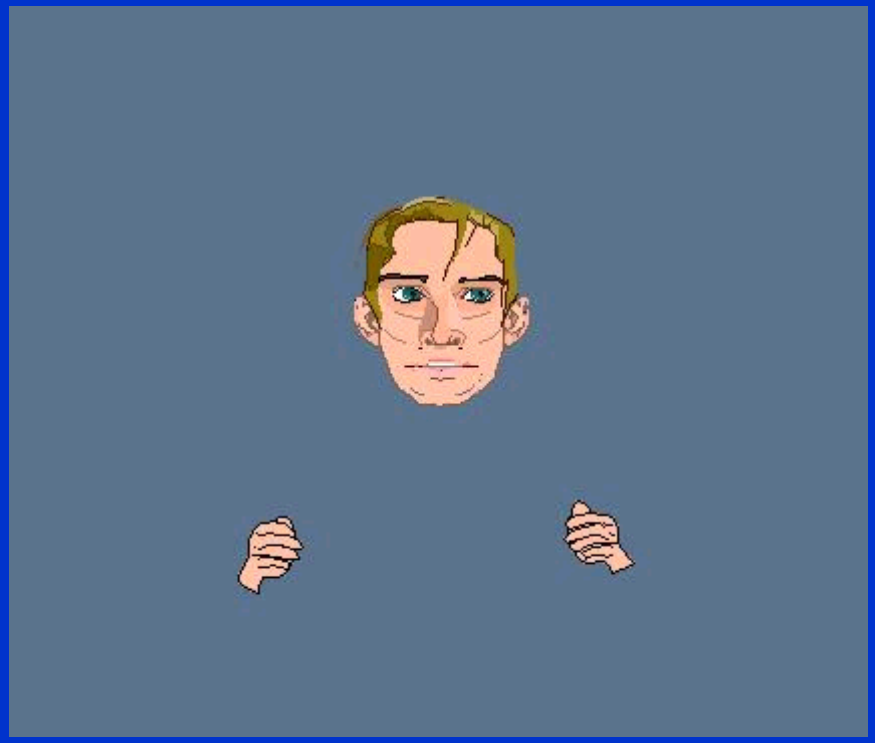
GESTYLE: Gesture repertoire

```
<GestureRepertoire>

<DefGest Name="Nod" GestureLength = "200"/>

<DefGest Name="Beat" GestureLength = "500"
   Symmetry=”right”/>

...

<DefGest Name="NodAndBeat" gesture_length = "500">

   <UseGest Name="Beat" sub_start_time = "0" /><PAR/>

   <UseGest Name="Nod" sub_start_time=”100”/></DefGest>

</GestureRepertoire>
```

Two examles

- Hamlet: CharToon 2D
  - face, eyes, hands
  - .avi
- Yt: H-anim humanoid for the Web
  - hand gestures only
  - real-time, on the fly generation of animation
- Speech by MindMaker's FlexVoice TTS, styled too for Hamlet!

Hamlet



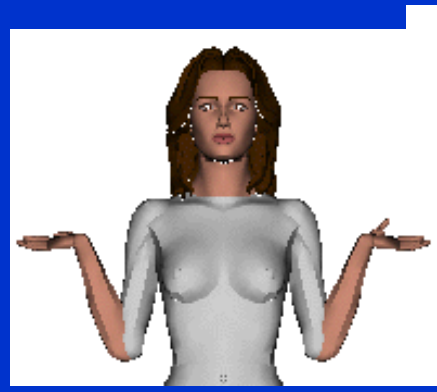
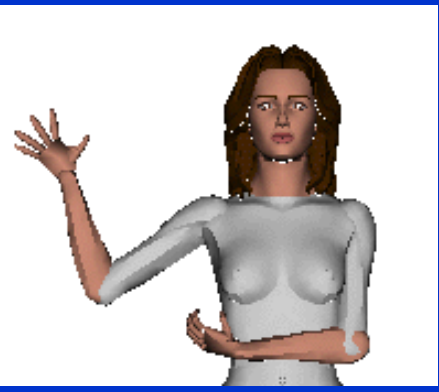
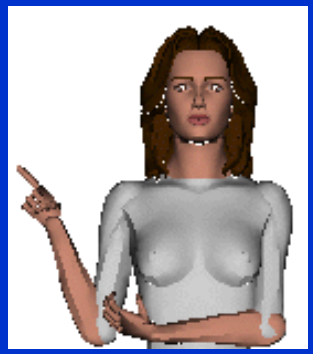
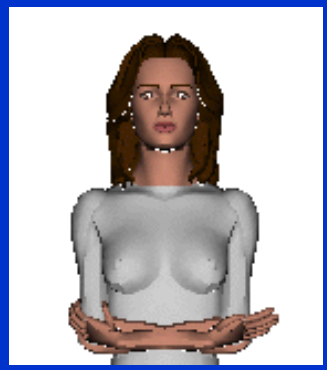
HamletMarkUp1.xml	HamletMarkUp2.xml

StyleDictionaries.xml

Yt

- hierarchical articulated body, 54 DOF of hand joints
- modular gesture definition written in STEP (VU)



emphasis	enumerate	sorry_for	greeting

Further work on hand gestures

- Constraints of the human body
  - –joints, collision, motion characteristics
- In coordination with FA and body motion
- Applications, e.g:
  - physiotherapy on the Web
  - multimodal translation (culture)
  - presenter generator

Overview of the talk

Intro about ECAs

FA Analysis of facial expressions of humans and cartoons

Non-photorealistic facial animation - CharToon

Emotion Disc

Evaluation of personality cues

Full body

Style in gesture - GESTYLE

➡ Challenges for IP

Challenges for IP

Image morphing

- automatic identification of feature points
- interpolation
- per region (e.g. lip sync)
- db/analysis of faces
- combination of morphing and 3d transformations

Face and gesture recognition

- elicitation of characteristics for modeling (off-line)
- real-time recognition of conveyed meaning
- user modeling for interaction control, on-line
- user analysis to learn about how they react to ECAs
- multimodal recognition

Challenges for IP

Texturing - LOD

Image analysis and synthesis

-NPAR (e.g. line drawing) from photos

- exaggeration

- ageing, tiredness and other effects

Questions

Figure 2. Texture mapped face with (a) neutral expression and (b) stern expression.

From: J. Walker, L. Sproull, R. Subramani: Using a Human Face in an Interface, CHI'94, pp 85-91.