A Rigorous View of Mode Confusion

Jan Bredereke

University of Bremen

Mode Confusion

- a kind of automation surprise
- in shared-control systems
- humans use a mental model of the technical system
  - can get out of sync
- examples:
  - Airbus A320 in Mulhouse airshow, 1988
  - Airbus A320 near Strasbourg, 1992

What Exactly is a Mode Confusion?

- literature??
- our work:
  1. a definition
  2. causes
  3. what to do

Basic Idea of a Definition

- "The user must not be surprised" (with respect to safety)
- mental model: specification
  reality: implementation
- formally: a refinement relationship in an abstracted description

Automonous Wheelchair "Rolland"

- joystick-to-motor line wiretapped
- ring of sonar sensors
- safety module
- driving assistant
  - turning on the spot skill
  - obstacle avoidance skill
  - . . .

Where are the Modes?

- modes of wheelchair: find in its software?
- black-box view!
  - user can see wheelchair behaviour only
  - relevant events are environment events
    - ▷ "joystick pushed forward"
    - ▷ "motor starts to move"
- requirements level
  - formalism: CSP, . . .

Relating Mental Model and Reality

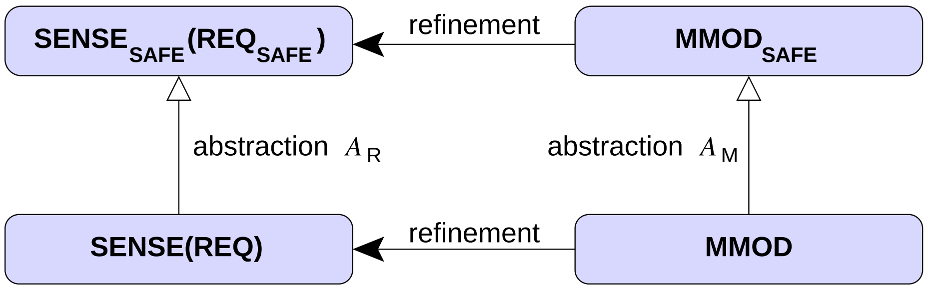
- user must perceive reality through his senses
  - environment events $\neq$ mental events !
  - "wall gets close" $\neq$ "see that wall gets close"
- formally: function over behaviours:
  SENSE: environment events $\rightarrow$ mental events
- rigorous specification/implementation relation:

$$\mathrm{MMOD} \ \sqsubseteq_F \ \mathrm{SENSE}(\mathrm{REQ})$$

  - CSP, failure refinement. See paper.

Two Refinement Relations

- abstraction to safety-relevant part



Rigorous Definitions

Definition 1 (Potential future behaviour)
A potential future behaviour is a set of “failures”.

(failure = trace + set of refusals)

Definition 2 (mode)
A mode of the perceived reality $\mathrm{SENSE}_{\mathrm{SAFE}}(\mathrm{REQ}_{\mathrm{SAFE}})$ is a potential future behaviour.
A mode of the mental model $\mathrm{MMOD}_{\mathrm{SAFE}}$ is a potential future behaviour.

Rigorous Definitions [2]

Definition 3 (mode confusion)
A mode confusion between the perceived reality $\mathrm{SENSE}_{\mathrm{SAFE}}(\mathrm{REQ}_{\mathrm{SAFE}})$ and the mental model $\mathrm{MMOD}_{\mathrm{SAFE}}$ occurs if and only if

the perceived reality is not a failure refinement of the mental model, i.e., iff

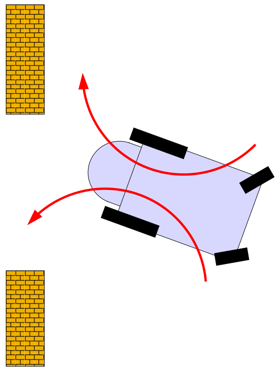
$$\mathrm{MMOD}_{\mathrm{SAFE}} \quad \not\sqsubseteq_F \quad \mathrm{SENSE}_{\mathrm{SAFE}}(\mathrm{REQ}_{\mathrm{SAFE}})$$

Application to Autonomous Wheelchair “Rolland”

- extracted mental model by user interview
- got requirements by reverse engineering C++ code
- both specifications written in CSP
  - 1200 lines of CSP
- model-checking refinement
  - commercial tool FDR

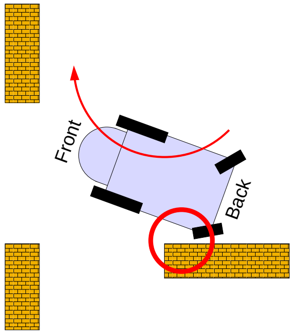
Wheelchair: Obstacle Avoidance Skill

- re-inforces user command to either
  - pass left (through doorway)
  - pass right (turn away from door)
- steers back after avoidance complete
- implicit mode transitions



Mode Confusion Found while Modelling Wheelchair

- unexpectedly hidden obstacle
  - danger in forward curve:
    - ▷ back of wheelchair swerves out
    - ▷ may hit obstacle behind user's head
  - automation prevents accident
    - ▷ changes direction/speed
  - user doesn't notice event
    $\rightarrow$ wheelchair and mental model behave differently



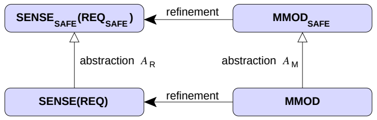
Mode Confusions Found by Model-Checking Wheelchair

- user's senses work at different speeds
  - vision, tactile, motion-detection
  - perceive reaction before cause
  - is general problem
- wrong mental model of "halt" routine
  - speed command = 0 cm/s $\rightarrow$ steering angle = "straight"
  - is relevant: "can you do this narrow curve?"

Mode Confusions Found by Model-Checking Wheelchair [2]

- wrong abstraction in user's mental model of old joystick position
  - wheelchair steers back when obstacle passed
  - except if joystick moved
  - did not work in abstracted mental model
- (found above known problem, too)
- (proved that no further mode confusion exists)

Classification of Mode Confusion Problems



- derived from rigorous definitions:
- classification:
  - 1. incorrect observation by the user
  - 2. incorrect knowledge of the user
  - 3. incorrect abstraction by the user
- classification is by cause
- leads to recommendations for avoiding mode confusions
  - $\rightarrow$ details in paper

Summary

- rigorous definitions of "mode" and "mode confusion"
  - mental model/reality like specification/implementation
  - rigorous modelling approach
    - ▷ black-box view
    - ▷ precise interfaces
- new classification by cause
- solutions:
  - recommendations for design
  - foundation for detection by model checking
    - ▷ successful practical application

Thank you.

Recommendations for Avoiding Mode Confusions [1]

- correct observation by the user:
  - check: can user physically observe all relevant events?
  - check: are user's senses sufficiently precise?
  - solution: add feedback event
  - check: do all relevant events become concious?
    - ▷ psychology!

Recommendations for Avoiding Mode Confusions [2]

- correct knowledge of the user:
  - document requirements rigorously
    - ▷ training material complete
    - ▷ also learnable
  - avoid non-determinism in requirements (complexity!)
  - check: do imprecise sensors introduce non-determinism?
  - solution for non-determinism: add feedback event

Recommendations for Avoiding Mode Confusions [3]

- correct abstraction by the user:
  - psychology!
  - document explicitly what is safety-relevant

Future Work

- try out recommendations
  - $\rightarrow$ psychology experts for non-technical ones
- more application domains beyond aviation and robotics

Failure

Definition 4 (Failure of a process P)
is a pair (s, X) of a trace s ($s \in \mathit{traces}(P)$) and a "refusal" set X of events.
The events in X may be blocked by P after the execution of s.

Failure Refinement

Definition 5 (Failure Refinement)
Process P refines process S in the failures model, written $S \sqsubseteq_F P$, iff
$\mathit{traces}(P) \subseteq \mathit{traces}(S)$ and also
$\mathit{failures}(P) \subseteq \mathit{failures}(S)$.

Description Without Internal State

- refer to history of events only
- example:
  - wheelchair has approached a wall
  - wheelchair has not moved back yet
  - $\rightarrow$ wheelchair must not move forward
- formalism: CSP, ...

Getting an Explicit Mental Model

according to Rushby [1]:

- from training material
- from user interviews
- by user observation