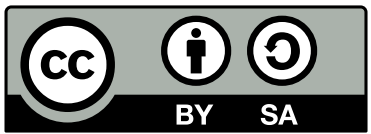
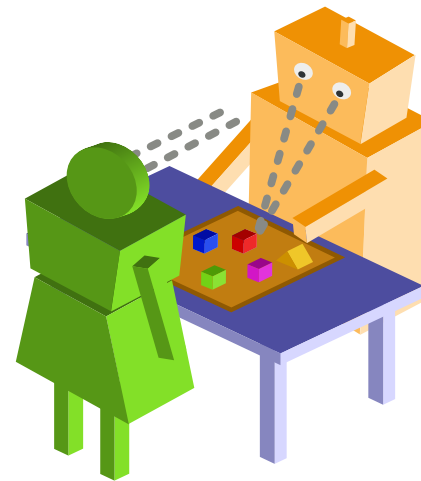


You can download the sources of this presentation here:
github.com/severin-lemaignan/presentation-ros4hri

PAL ROBOTICS

ROS for Human-Robot Interaction
Towards REP-155

ROSCon | Oct 2022



Séverin Lemaignan

PAL Robotics Senior Scientist AI & Social Interactions

situation assessment

symbolic grounding

symbolic reasoning

SYMBOLIC SOCIAL COGNITION FOR ROBOTS

ontologies

perspective taking

cognitive architectures

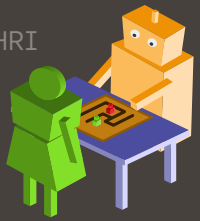
social situation assessment

joint action

ROS4HRI

natural language processing

REAL-WORLD SOCIAL AUTONOMY



DATA-DRIVEN HRI

learning of social policies

large datasets

theory of mind

group dynamics

human-in-the-loop ML

robotics for learning

experimental robotics

CHILD-ROBOT INTERACTION

trust

HUMAN FACTORS

engagement

responsible AI

anthropomorphism

social robotics

participatory design

persuasion

WHY ROS4HRI?

- dealing with humans is actually hard: they keep on disappearing/reappearing; hard to predict where/when; ‘shape’ known at run-time only, etc.
- widely different requirements depending on application: from ‘2D points’ to full online kinematic model.
- no ROS standard for HRI (nothing, nada, rien!)

[Mohamed, Lemaignan, **ROS for Human-Robot Interaction**, IROS 2021]

DESIGN REQUIREMENTS

- representations **application-agnostic**: from point-like crowd simulation, to kineastetic teaching, to social interaction

[Mohamed, Lemaignan, **ROS for Human-Robot Interaction**, IROS 2021]

DESIGN REQUIREMENTS

- representations **application-agnostic**: from point-like crowd simulation, to kineastetic teaching, to social interaction
- does not enforce any specific algorithm or perception pipeline

[Mohamed, Lemaignan, **ROS for Human-Robot Interaction**, IROS 2021]

DESIGN REQUIREMENTS

- representations **application-agnostic**: from point-like crowd simulation, to kineastetic teaching, to social interaction
- does not enforce any specific algorithm or perception pipeline
- however, takes into account what current algorithms can or can not do (eg: kinematic model of human)

[Mohamed, Lemaignan, **ROS for Human-Robot Interaction**, IROS 2021]

DESIGN REQUIREMENTS

- representations **application-agnostic**: from point-like crowd simulation, to kineastetic teaching, to social interaction
- does not enforce any specific algorithm or perception pipeline
- however, takes into account what current algorithms can or can not do (eg: kinematic model of human)
- integrated as much as possible with existing ROS conventions (eg: `robot_state_publisher` for human forward kinematics)

[Mohamed, Lemaignan, **ROS for Human-Robot Interaction**, IROS 2021]

SCOPE

- set of new messages (`hri_msgs`)

SCOPE

- set of new messages (`hri_msgs`)
- topics naming convention

[Mohamed, Lemaignan, **ROS for Human-Robot Interaction**, IROS 2021]

SCOPE

- set of new messages (`hri_msgs`)
- topics naming convention
- `tf` frames naming convention

SCOPE

- set of new messages (`hri_msgs`)
- topics naming convention
- `tf` frames naming convention
- (parametric) kinematic model of humans

[Mohamed, Lemaignan, **ROS for Human-Robot Interaction**, IROS 2021]

SCOPE

- set of new messages (`hri_msgs`)
- topics naming convention
- `tf` frames naming convention
- (parametric) kinematic model of humans
- (a few) global ROS parameters

[Mohamed, Lemaignan, **ROS for Human-Robot Interaction**, IROS 2021]

SCOPE

- set of new messages (`hri_msgs`)
- topics naming convention
- `tf` frames naming convention
- (parametric) kinematic model of humans
- (a few) global ROS parameters

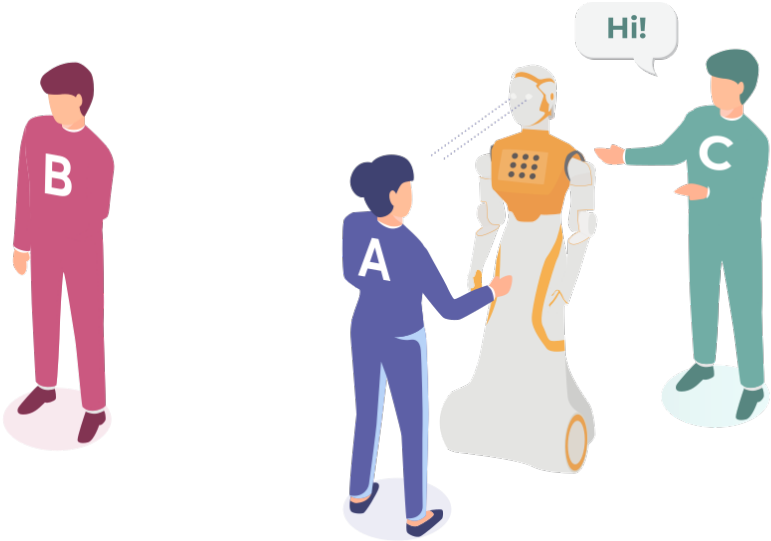
- for now, focus on *perception* only

SCOPE

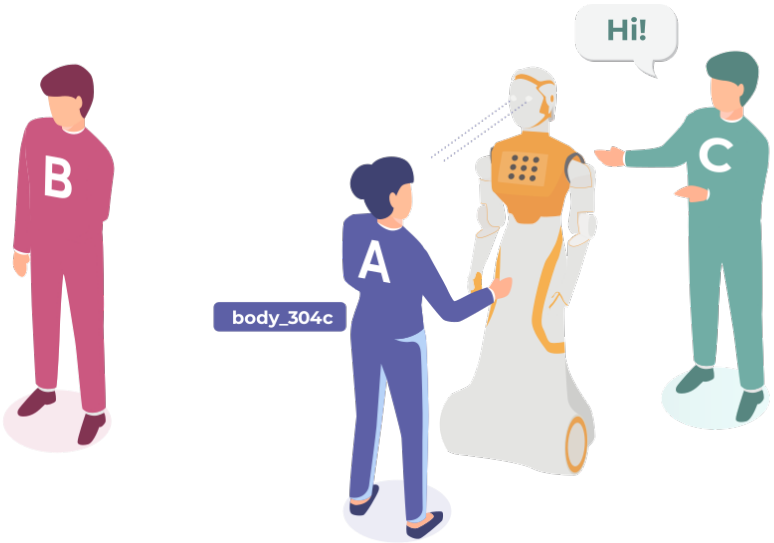
- set of new messages (`hri_msgs`)
- topics naming convention
- `tf` frames naming convention
- (parametric) kinematic model of humans
- (a few) global ROS parameters

- for now, focus on *perception* only
- initially, ROS1

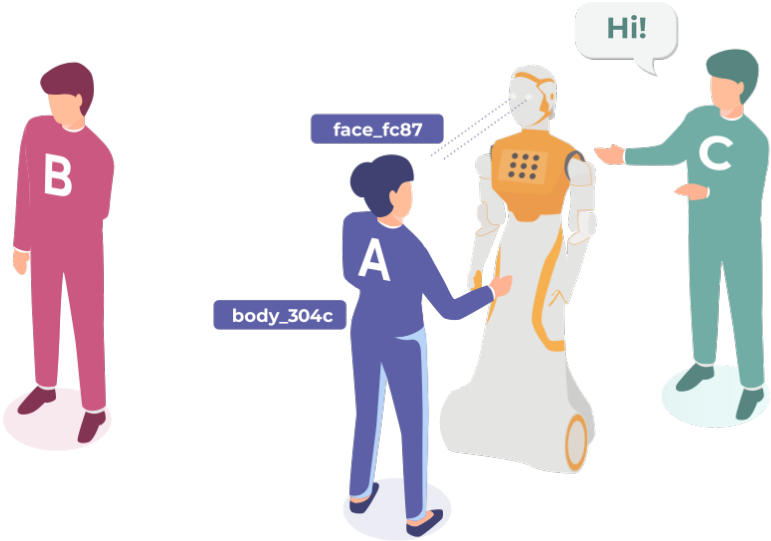
HUMAN REPRESENTATION: PERMANENT VS TRANSIENT IDS



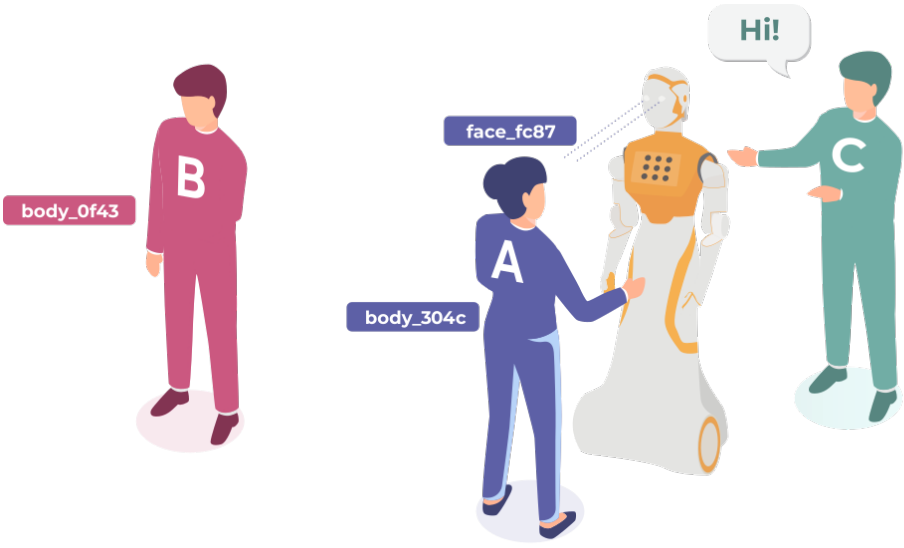
HUMAN REPRESENTATION: PERMANENT VS TRANSIENT IDS



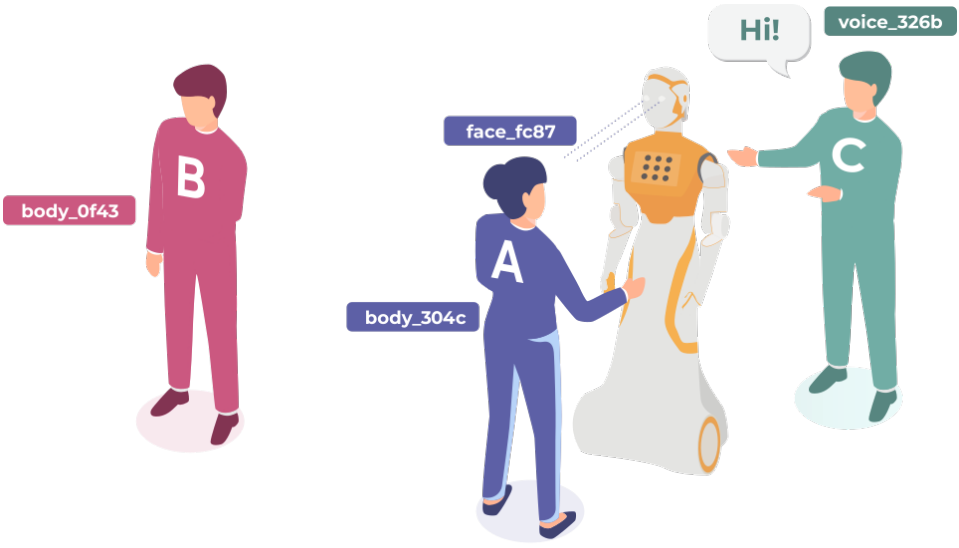
HUMAN REPRESENTATION: PERMANENT VS TRANSIENT IDS



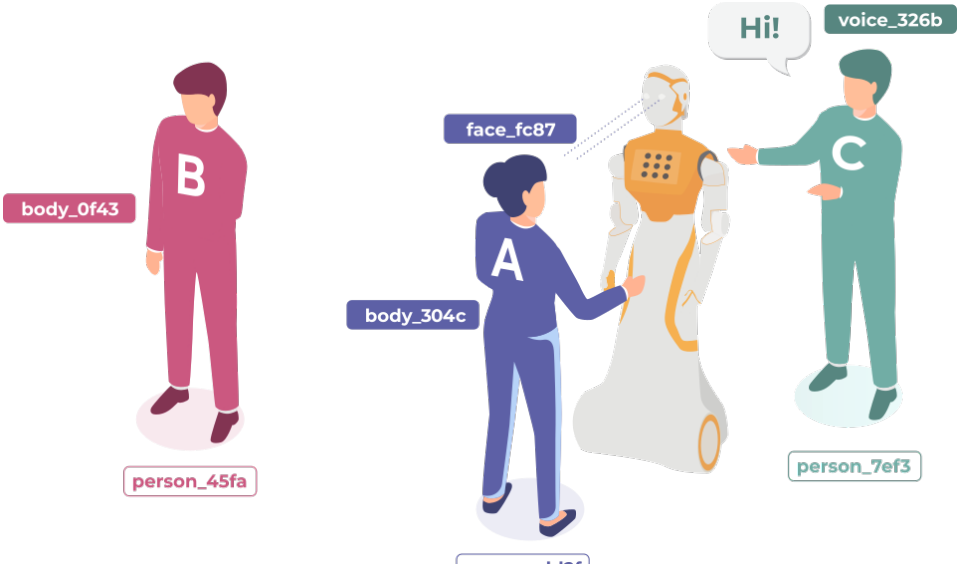
HUMAN REPRESENTATION: PERMANENT VS TRANSIENT IDS



HUMAN REPRESENTATION: PERMANENT VS TRANSIENT IDS



HUMAN REPRESENTATION: PERMANENT VS TRANSIENT IDS



TOPICS STRUCTURE: FACES

Under **/humans/faces/<faceID>/** (eg /humans/faces/bf3d):

Name	Message type	Description
/roi	hri_msgs/NormalizedRoI2D	Region of the face in the source image
/cropped	sensor_msgs/Image	Cropped face
/frontalized	sensor_msgs/Image	Frontalised face
/landmarks	hri_msgs/FacialLandmarks	The 2D facial landmarks extracted from the face
/facs	hri_msgs/FacialActionUnits	The presence and intensity of facial action units found in the face
/expression	hri_msgs/Expression	The expression recognised from the face
/softbiometrics	hri_msgs/SoftBiometrics	Soft biometrics like age and gender of the face

TOPICS STRUCTURE: BODIES

Under **`/humans/bodies/<bodyID>/`** (eg `/humans/bodies/5e4d`):

Name	Message type	Description
`/roi`	`hri_msgs/NormalizedRoI2D`	Region of the whole body in the source image
`/cropped`	`sensor_msgs/Image`	Cropped image of the body
`/joint_states`	`sensor_msgs/JointState`	The joint state of the human body
`/skeleton2d`	`hri_msgs/Skeleton2D`	The 2D points of the detected skeleton
`/posture`	`hri_msgs/BodyPosture`	Recognised body posture (sitting, standing)
`/gesture`	`hri_msgs/Gesture`	Recognised symbolic gesture

3D pose? `tf` frames from joint state + human URDF! I'll come to it in a minute.

TOPICS STRUCTURE: VOICES

Under **/humans/voices/<voiceID>/** (eg /humans/voices/dde2):

Name	Message type	Description
/audio	audio_common_msgs/AudioData	Separated audio stream for this voice
/features	hri_msgs/AudioFeatures	INTERSPEECH'09 Emotion challenge low-level audio features
/is_speaking	std_msgs/Bool	Whether or not speech is recognised from this voice
/speech	hri_msgs/LiveSpeech	The live stream of speech recognized via an ASR engine

TOPICS STRUCTURE: PERSONS

Under **/humans/persons/<personID>/** (eg /humans/persons/45ff):

Name	Message type	Description
/face_id	std_msgs/String (latched)	Face matched to that person (if any)
/body_id	std_msgs/String (latched)	Body matched to that person (if any)
/voice_id	std_msgs/String (latched)	Voice matched to that person (if any)
/alias	std_msgs/String (latched)	ID of other person, if alias
/anonymous	std_msgs/Bool (latched)	if true, anonymous person (not permanent ID)
/engagement_status	hri_msgs/EngagementLevel	engagement status of the person *with the robot*
/location_confidence	std_msgs/Float32	Location confidence; 1 means 'person currently seen', 0 means 'person location unknown'
/name	std_msgs/String	Name, if known
/native_language	std_msgs/String	IETF language codes like EN_gb, if known

TOPICS STRUCTURE: GROUPS

Under **`/humans/groups/<groupID>/`** (eg `/humans/groups/56ef2`):

Name	Message type	Description
`/members`	`hri_msgs/IdLists`	Person ID of the members of the group

Attention: not yet in the REP-155!

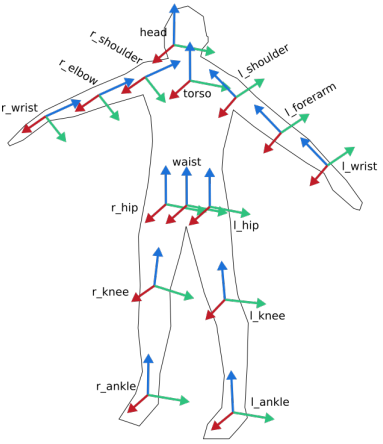
TOPICS STRUCTURE: INTERACTIONS

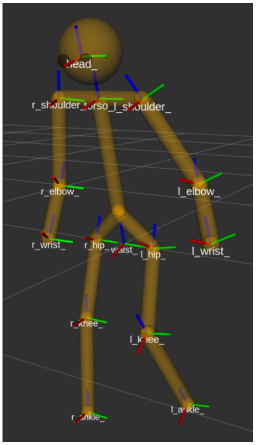
Under **`/humans/interactions/`**:

Name	Message type	Description
`/gaze`	`hri_msgs/Gaze`	estimated gazing behaviours

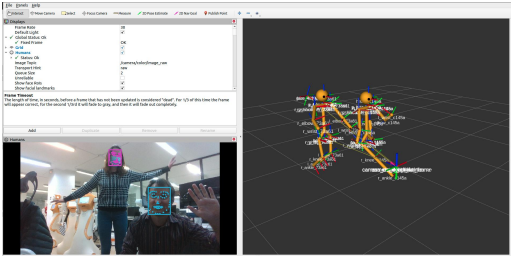
HUMAN PHYSICAL REPRESENTATION

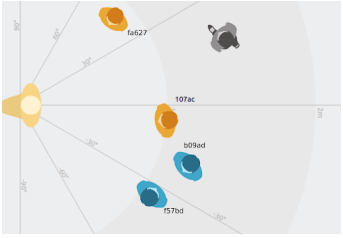
- standard ROS pipeline: joint state (eg OpenPose, mediapipe) -> `robot_state_publisher` + URDF
- URDF generated on the fly, based on person's height (xacro params)
- Follows REP-120 as much as possible.

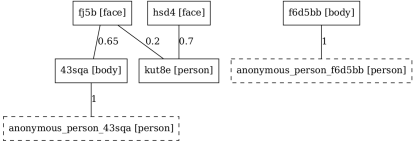




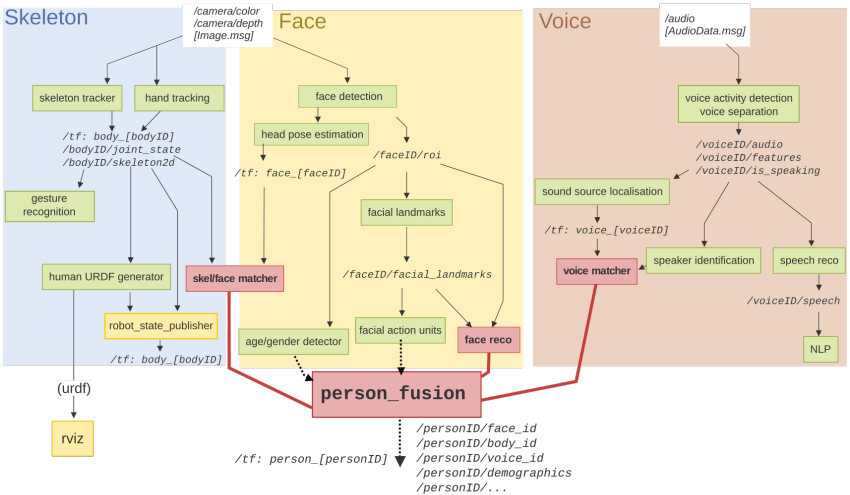
TOOLING

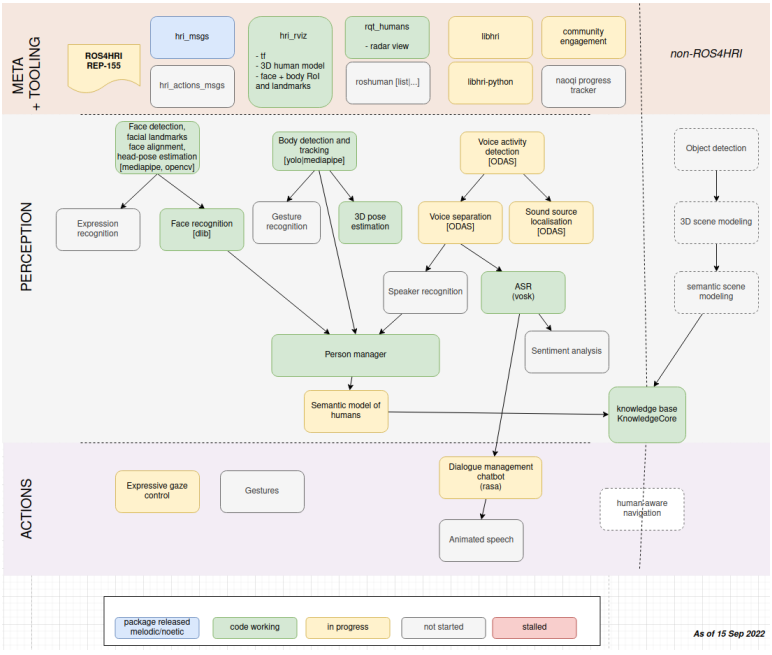



fa627
107ac
b09ad
f57bd


fj5b [face]
hsd4 [face]
f6d5bb [body]
0.65
0.2
0.7
1
43sqa [body]
kut8e [person]
anonymous_person_f6d5bb [person]
1
anonymous_person_43sqa [person]

ONE POSSIBLE PIPELINE (BUT OTHER ARE POSSIBLE!)




META + TOOLING
ROS4HRI REP-155
hri_msgs
hri_actions_msgs
hri_rviz
- tf
- 3D human model
- face + body RoI and landmarks
rqt_humans
- radar view
roshuman [list|...]
libhri
libhri-python
community engagement
naoqi progress tracker
non-ROS4HRI
PERCEPTION
Face detection, facial landmarks face alignment, head-pose estimation [mediapipe, opencv]
Body detection and tracking [yolo|mediapipe]
Voice activity detection [ODAS]
Object detection
Expression recognition
Face recognition [dlib]
Gesture recognition
3D pose estimation
Voice separation [ODAS]
Sound source localisation [ODAS]
3D scene modeling
Speaker recognition
ASR (vosk)
semantic scene modeling
Person manager
Sentiment analysis
Semantic model of humans
knowledge base KnowledgeCore
ACTIONS
Expressive gaze control
Gestures
Dialogue management chatbot (rasa)
human-aware navigation
Animated speech
package released melodic/noetic
code working
in progress
not started
stalled
As of 15 Sep 2022

Thank you!

Slides:
github.com/severin-lemaignan/presentation-ros4hri

(btw, we are always looking for great people to join us: drop me line if you want to know more!)

PAL ROBOTICS