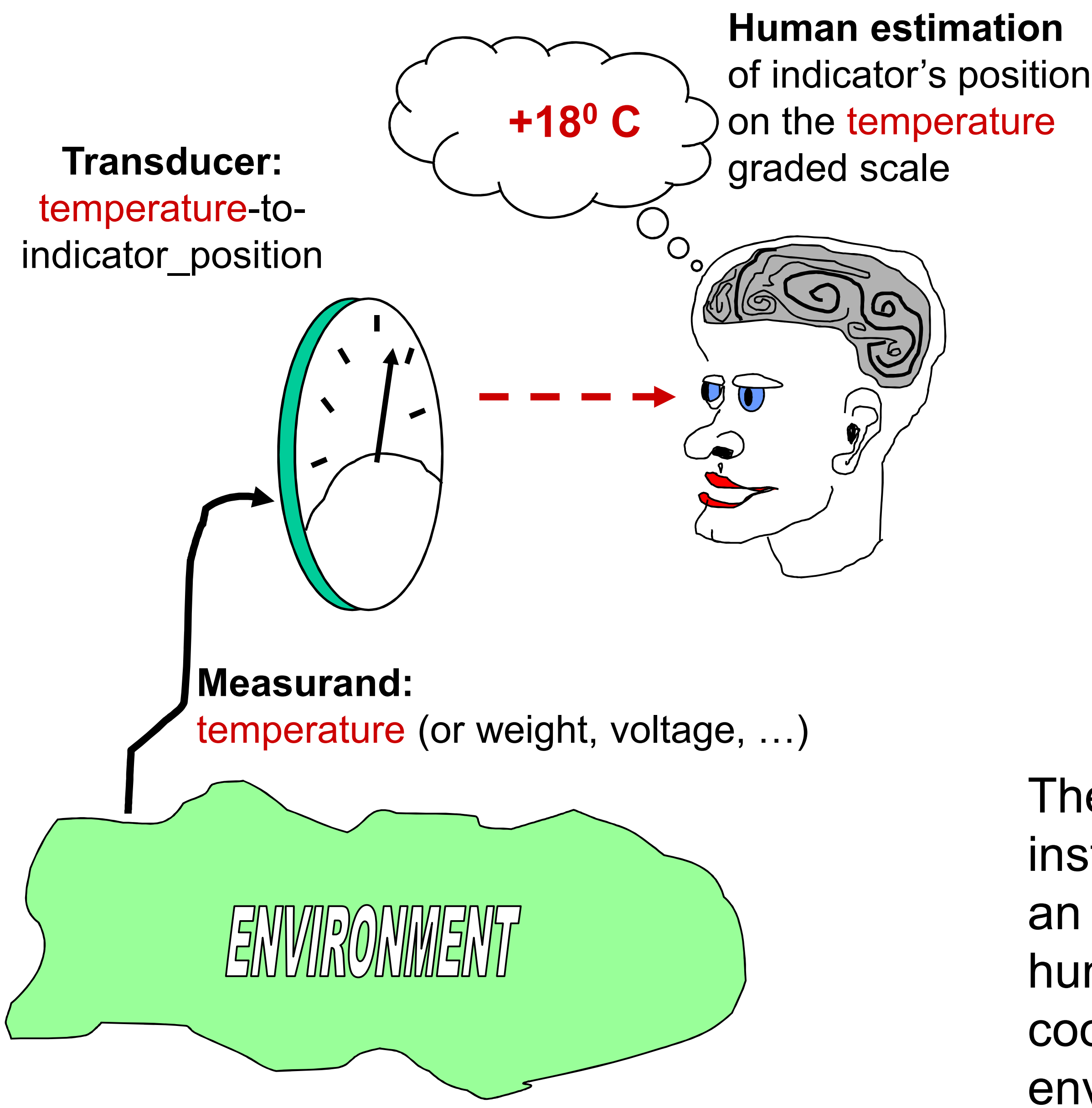
Human-Instrument Symbiotic Partnership for Multimodal Environment Perception

Emil M. Petriu[1], Thomas E. Whalen[2],
Imre J. Rudas[3], Dorina C. Petriu[4] and Marius D. Cordea[1]

[1]University of Ottawa, ON, Canada, [2]CRC, Ottawa, ON, Canada,
[3]Budapest Tech Polytechnic Institution, Hungary,
[4]Carleton University, Ottawa, ON, Canada

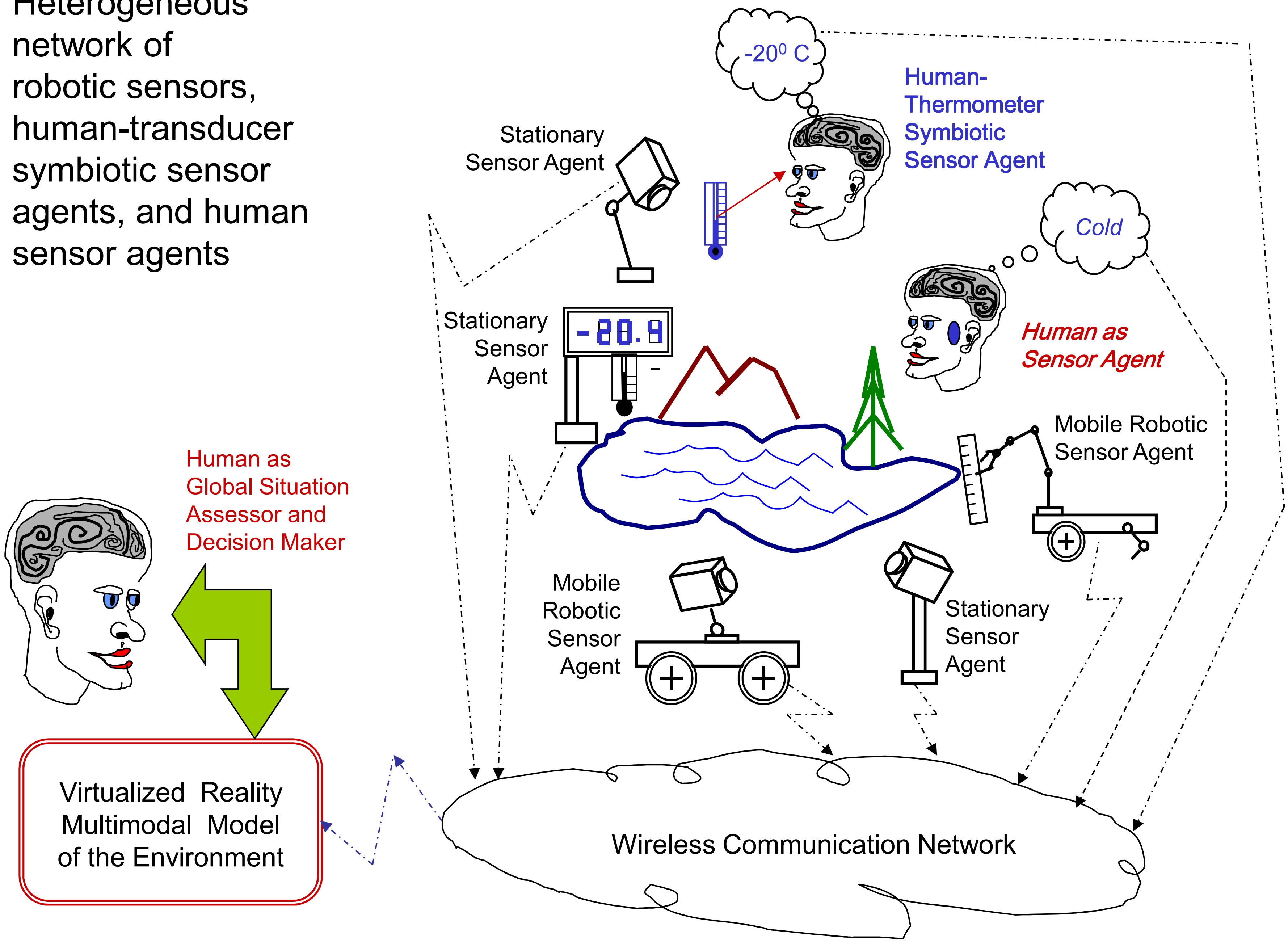
Discussing the aims of the human-computer symbiosis, Licklider writes in his seminal paper “Man-Computer Symbiosis, ”*IRE Trans. on Human Factors in Electronics*, Vol. HFE-1, pp. 4-11, March 1960.“*It seems likely that the contributions of human operators and equipment will blend together so completely in many operations that it will be difficult to separate them neatly in analysis. That would be the case if, in gathering data on which to base a decision, for example,both the man and the computer came up with relevant precedents from experience and if the computer then suggested a course of action that agreed with the man's intuitive judgment*.”

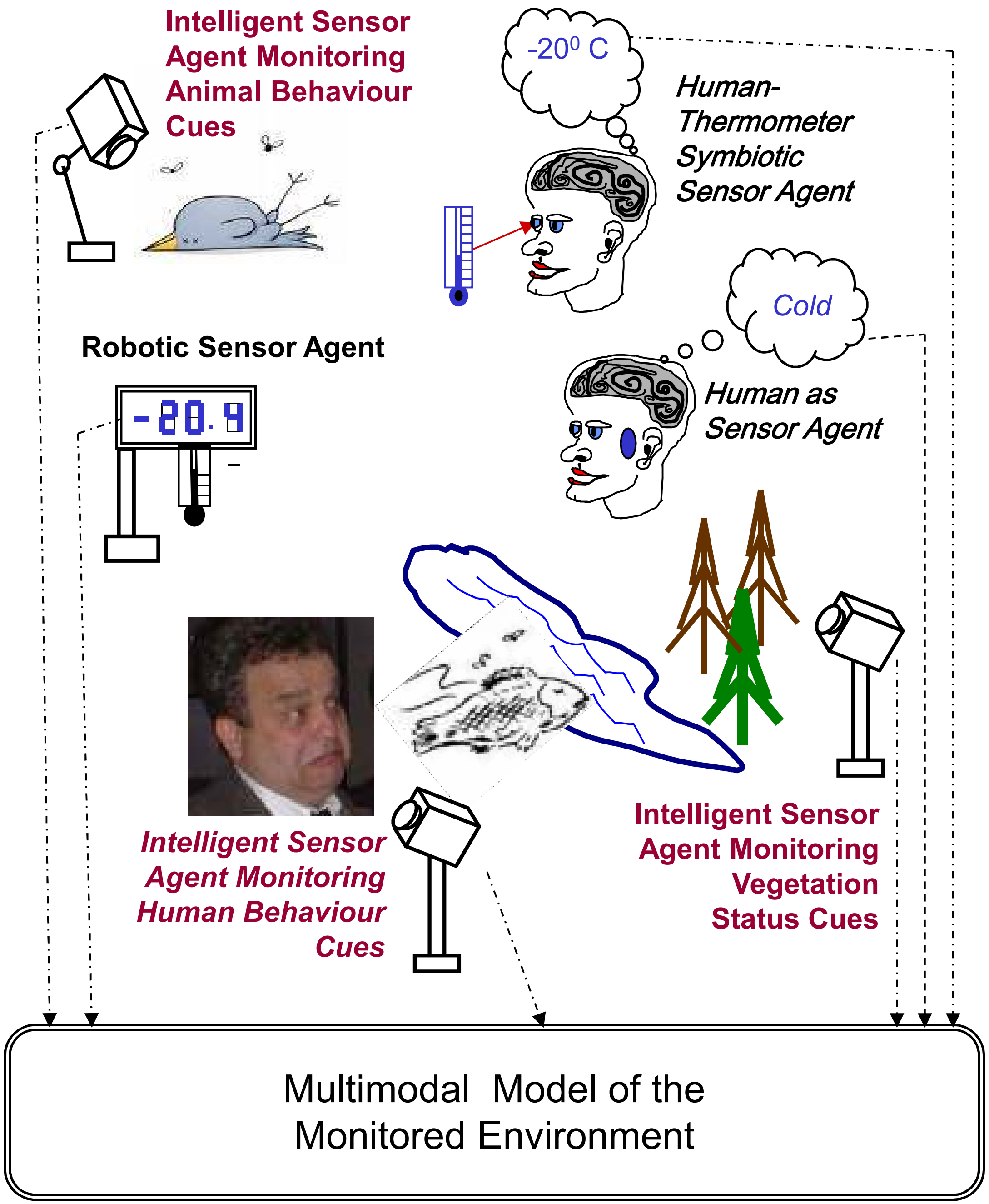
Pursuing Liklider’s idea this paper proposes a methodological framework for a symbiotic human-instrument partnership for multimodal environment and situation assessment.



The classic analog instrument paradigm: an early example of human-transducer cooperation for environment sensing.

Heterogeneous network of robotic sensors, human-transducer symbiotic sensor agents, and human sensor agents





Heterogeneous network of robotic sensors, human-transducer symbiotic sensor agents, human sensor agents, and *intelligent sensor agents capable to comprehend human and animal behaviour, and vegetation status.*

Symbionts combine intrinsic machine-sensing reactive behavior with higher-order human-oriented world-model representations of the immersive virtual reality.

Humans are valuable in a symbiotic partnership to the degree that their capabilities complement those of the computers/machines.

Humans are very high-bandwidth creatures:

- their visual system is capable of perceiving more than a hundred megabits of information per second, and
- their largest sense organ, the skin is capable of perceiving nearly that much as well.
- human speech conveys information in the form of intonation and inflection as well as the actual words uttered.

People communicate through "body language“ which includes facial expressions and eye movements.

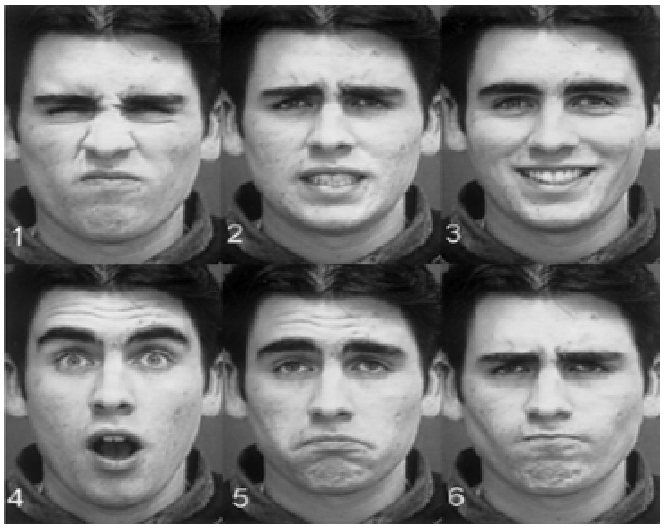
Human sensor information is “fuzzy quantizied” while the machine sensor information, both the symbiotic analog_ transducer & human, and the fully automated digital one, is “sharp & concatenated quantized”

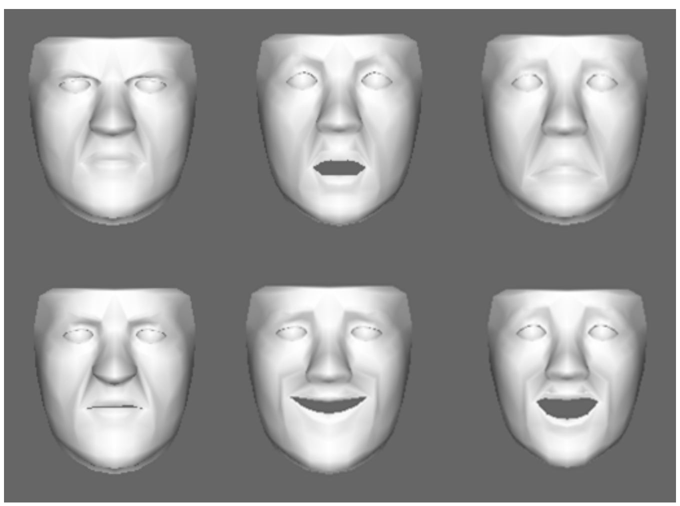
It is possible to reduce the uncertainty of the measurements involving humans as sensors part of multisensor systems, by using Fuzzy Cognitive Maps, NNs, and Associative Memories.

Dempster-Shafer theory of evidence approach is used to incorporate human-like uncertainty management and inference mechanisms in our context-aware multi-sensor data fusion system. This approach allows us to incorporate time-variable weights representative of sensor precision which will improve the sensor fusion accuracy in dynamic environments.

Linguistic pattern recognition techniques and semantic model representations are used to develop a semantic level situation assessment system that will allow understanding of the dynamics of a complex scene based on multimodal sensor data streams.

Facial Expression Recognition using a 3D Anthropometric Muscle-Based Active Appearance Model





- Facial Action Coding System (FACS)
  - 7 pairs of muscles + "Jaw Drop" = Expression Space
- Muscle "contractions" control mesh deformation in "Anthropometric-Expression (AE)" space
- Texture intensities are warped into the geometry of the shape
  - Shape: apply PCA in AE space
  - Appearance: apply PCA in texture space
- Model defined by rigid (rotation, translation) and non-rigid motion (AE)
- Model instances synthesized from AE space,

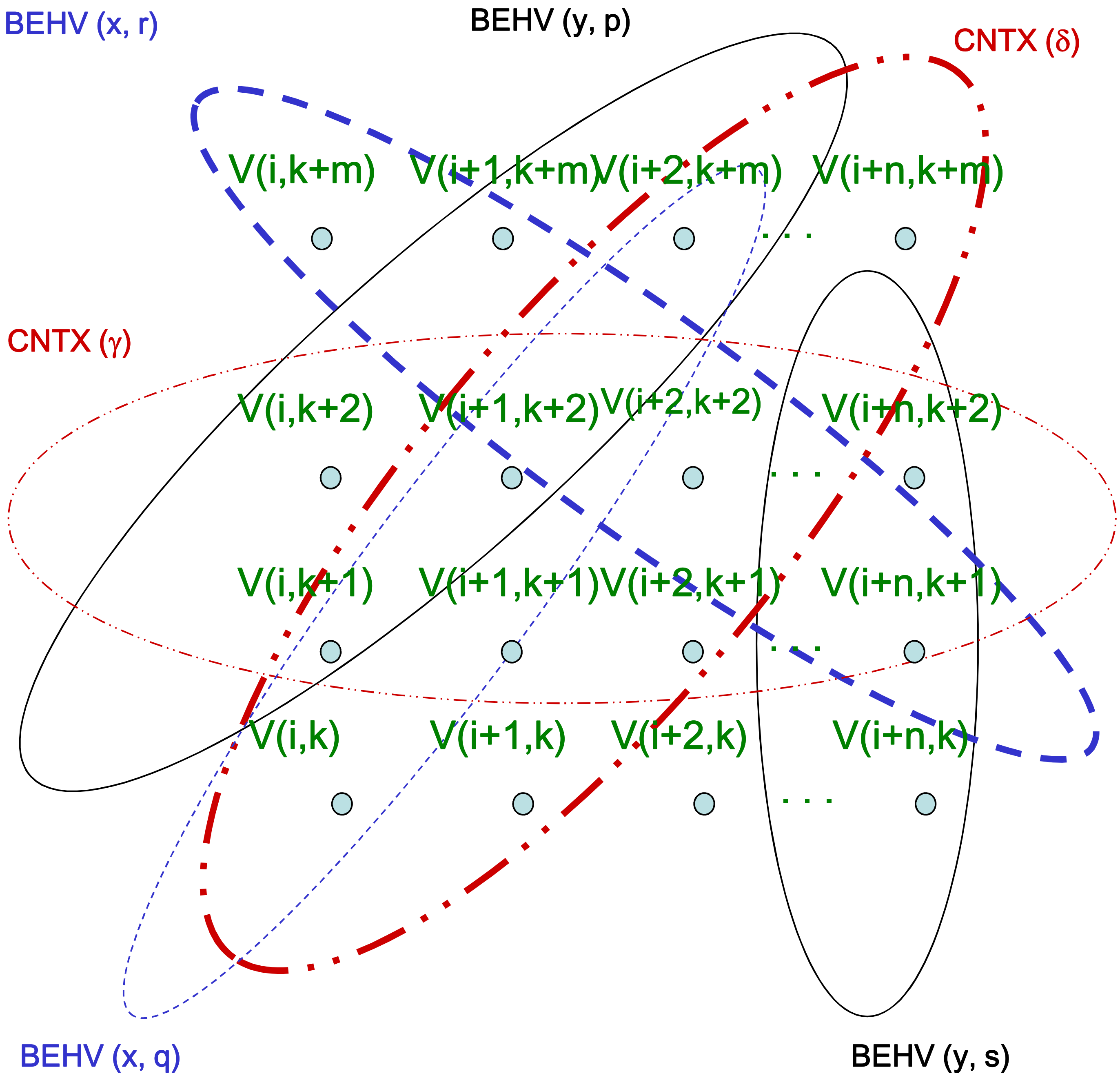
Facial Expression Recognition

- Person Dependent

AU	Signification	No.	Correct	False	Missed	Confused	Recognition Rate
0	Neutral	20	19	1	0	0	95%
1	Inner Brow Raiser	24	21	0	3	0	87.5%
2	Outer Brow Raiser	52	41	1	9	1	78.8%
4	Brow Lowerer	42	41	0	0	1	97.6%
12	Lip Corner Puller	51	48	3	0	0	94.1%
15	Lip Corner Depressor	18	17	0	1	0	94.4%
26	Jaw Drop	74	55	0	19	0	74.3%
Total		281	242	5	32	2	86.1%
False Alarm: 1.7%, Missed: 11.3%							

- Person Independent

AU	Signification	No.	Correct	False	Missed	Confused	Recognition Rate
0	Neutral	20	17	3	0	0	85.0%
1	Inner Brow Raiser	10	8	0	2	0	80.0%
2	Outer Brow Raiser	27	20	1	5	1	74.0%
4	Brow Lowerer	24	21	1	1	1	87.5%
12	Lip Corner Puller	17	13	0	4	0	76.4%
15	Lip Corner Depressor	13	10	1	2	0	76.9%
26	Jaw Drop	24	14	0	10	0	58.3%
Total		135	103	6	24	2	76.2%
False Alarm: 4.4%, Missed: 17.7%							



<u>Context-based plausible meaning of the specific behaviour of a human agent</u>: Estimating the value **V** of an environmental parameter of interest based on the specific behaviour ***BEHV*** of a human agent, which is function of the respective parameter and the context ***CNTX.***

In the previous figure: the human agent "**x**" exhibits the *behaviour* **BEHV (x, r)** , which may occur for any of the following *environmental parameter values* **{V(i, k+m), V(i+1, k+m), V(i+2, k+2), V(i+n, k)}**, *in the context* **CNTX (d)** defined by the following values of the environmental parameter of interest **{V(i+2, k+m), V(i+n, k+m), V(i+1, k+2), V(i+2, k+2), V(i, k+1, V(i+2, k+1), V(i, k), V(i+1, k)}.**
It can be concluded that this **specific behaviour in the given context** occurred because of the **specific value V(i+2, k+2) of the environmental parameter** of interest, which is the value that is shared by the definition domains of the behaviour **BEHV (x, r),** and the context **CNTX (d).**

We adopted a ***two-tier context definition:*** *1st tier* includes four basic object characteristics: *location, identity, time, and activity*; all other possible contextual characteristics belong to the 2nd tier and are considered as *attributes of the primary context properties*.