Emotion recognition for empathy-driven HRI:
An adaptive approach

Seminar *Intelligent Robotics*
WiSe 2018/19
Presentation by Angelie Kraft

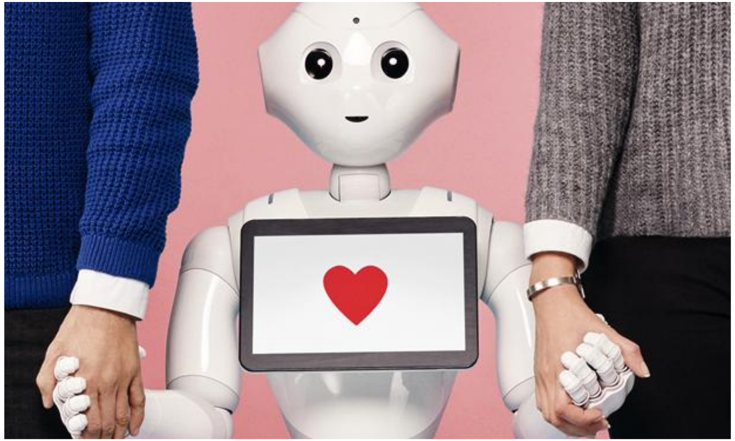
Universität Hamburg
DER FORSCHUNG | DER LEHRE | DER BILDUNG

TAMS

Overview

I. Introduction
II. An Empathy-Driven Approach by Churamani et al. (2018)
III. Evaluation
IV. Conclusion
V. References

I. Introduction



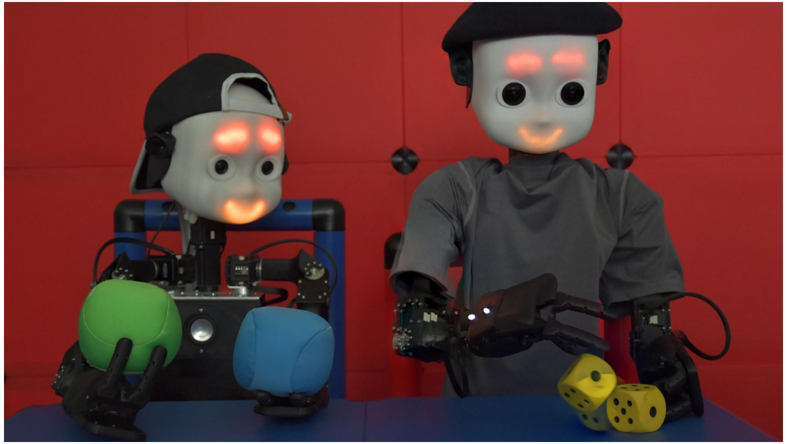
"Pepper" by Softbank Robotics

Future Life with **Pepper** (2016)
[https://www.youtube.com/watch?v=-A3ZLLGuvQY]

https://spectrum.ieee.org/image/MjU4NjkzNA.jpeg

Why do we need robot companions?

- Understanding humans for better **service**
  - Emotion conveys intentions and needs
- Positive **psychological effects:**
  - Autism, dementia, education
- How does Pepper do it?
  - Multi-modal emotion recognition!

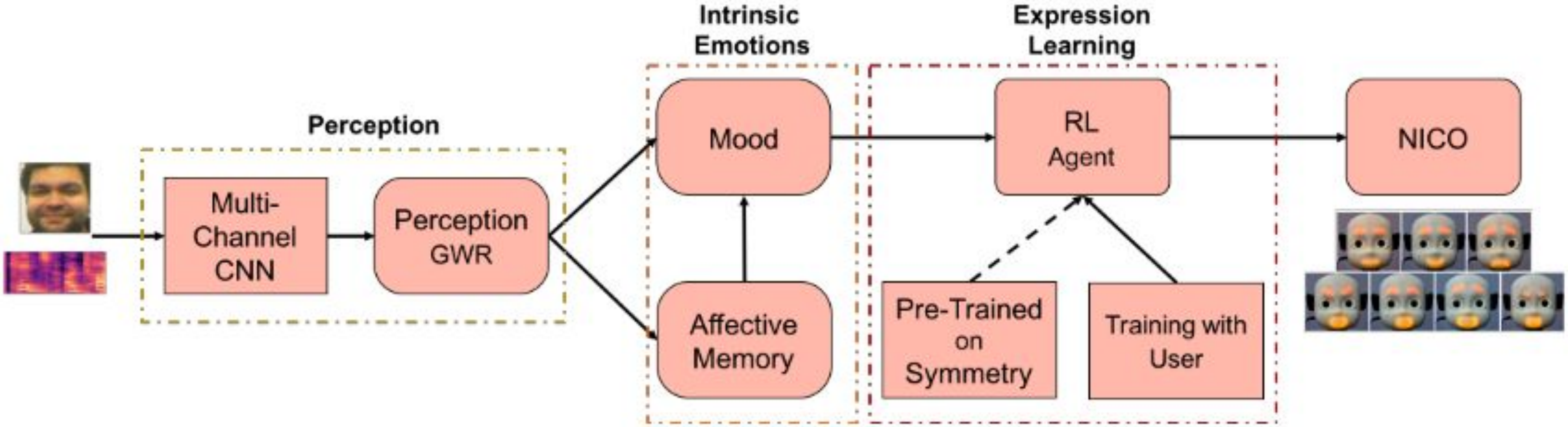


NICO (Neuro-Inspired COmpanion)
by Kerzel et al. (2017)

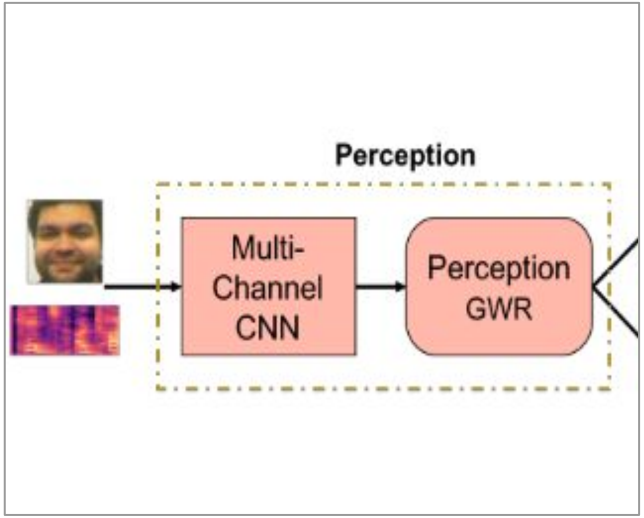
https://www.inf.uni-hamburg.de/en/inst/ab/wtm/research/neurobotics/nico.html

II. An approach to empathy-driven HRI

By Churamani, Barros, Strahl, & Wermter (2018)



Emotion perception module

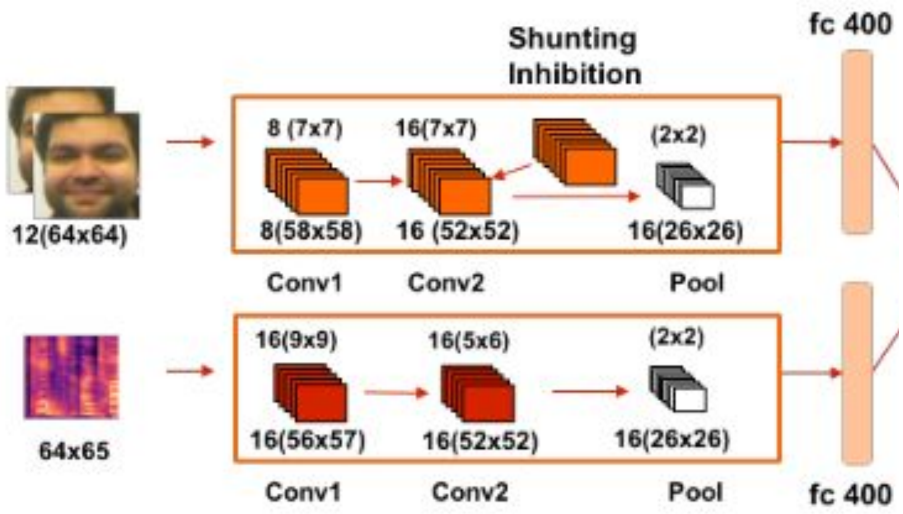


1. **Multi-Channel Convolutional NN (MCCNN):**
   - 1. Channel: Visual information
   - 2. Channel: Auditory information
   - → **Learning**

2. **Growing-When-Required (GWR) network:**
   - Account for variance in stimuli
   - → **Adapting**

Churamani et al. (2018)

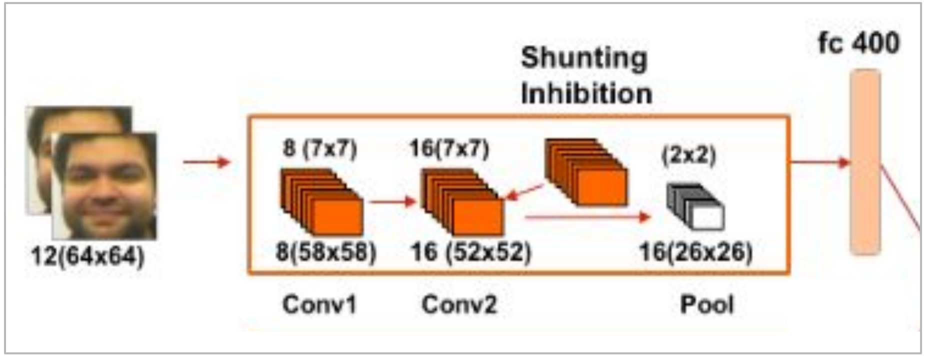
Learning with a Multi-Channel CNN



- Both layers trained equivalently
- Sound transformed into image data:
  - Power spectrum intro "mel scale" frequency

Churamani et al. (2018)

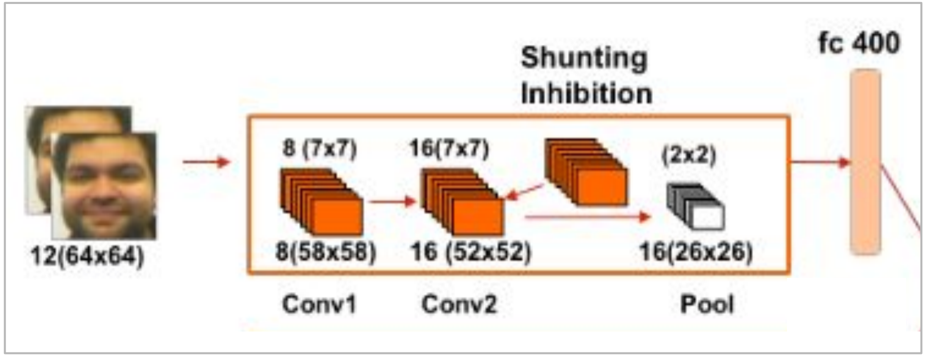
Multi-Channel CNN: Visual channel



- Two convolutional layers:
  - Each filter learns different features
  - First layer: low-level features (e.g. edges with different orientations)
  - Second layer: abstract features (e.g. eyes, mouth)

Churamani et al. (2018)

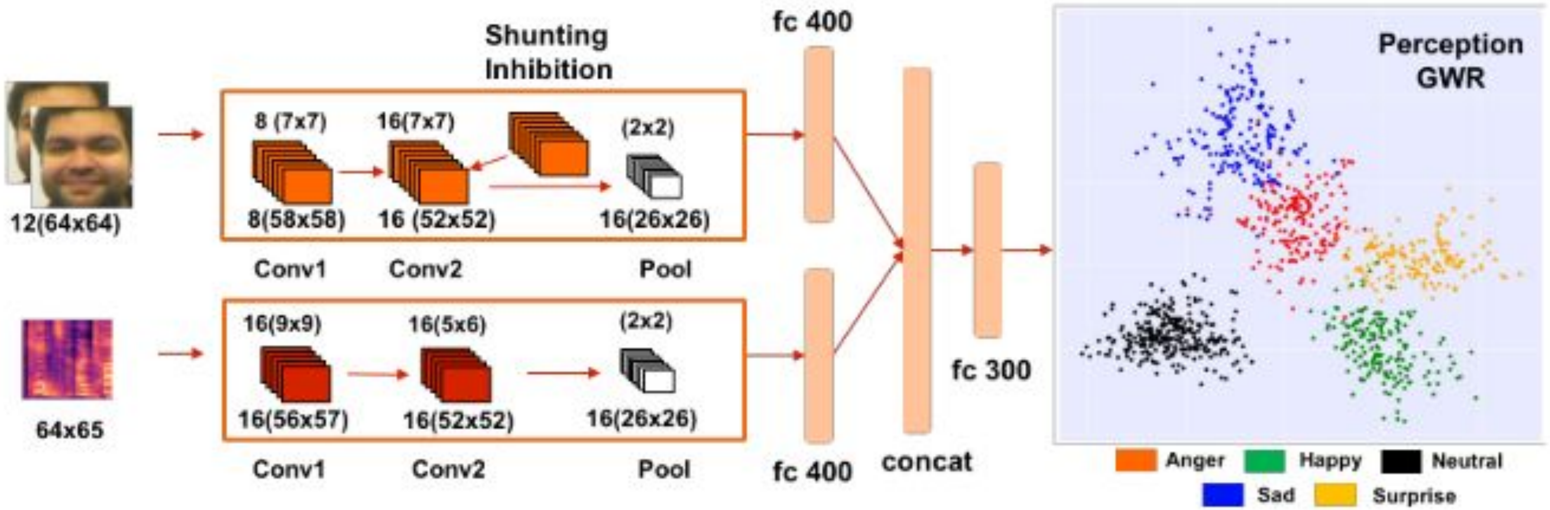
Multi-Channel CNN: Visual channel



- Shunting inhibition for robustness
- Max pooling for down-sampling
- Fully connected layer represents facial features for emotion classification

Churamani et al. (2018)

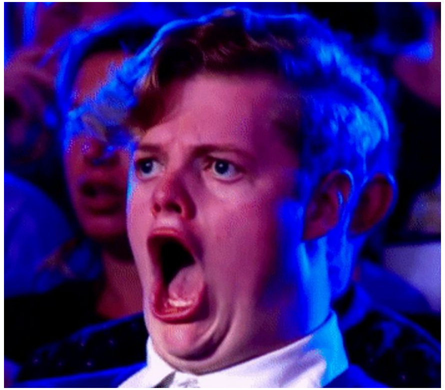
Combining both channels



Churamani et al. (2018)

https://veganuary.com/wp-content/uploads/2016/09/face
-shocked-1511388.jpg

=



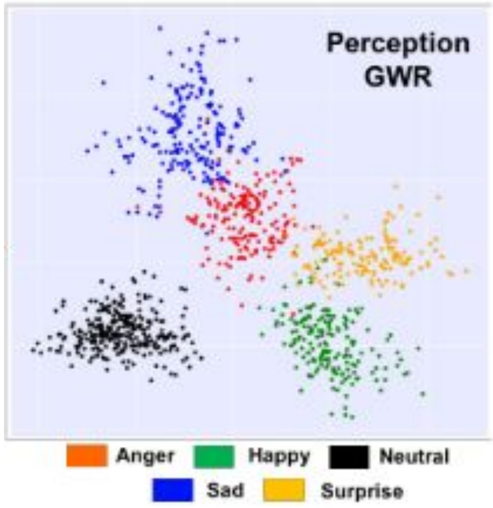
http://hahasforhoohas.com/sites/hahasforhoohas.com
/files/uploadimages/images/shocked-face-gif.png

?

Growing-When-Required

- Is activity of the best-matching neuron high enough?
  - Yes: Keep
  - No: Create new node
- Delete “outdated” edges & nodes

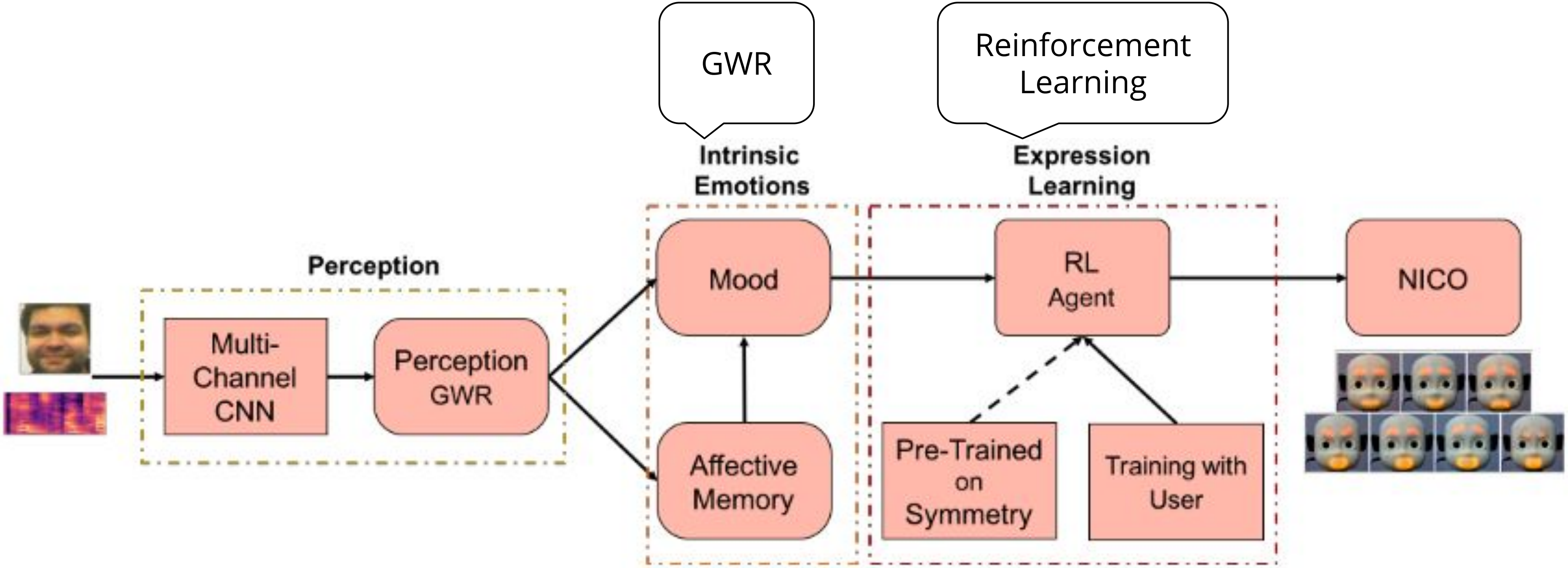
→ Represents emotions in clusters



Churamani et al. (2018)

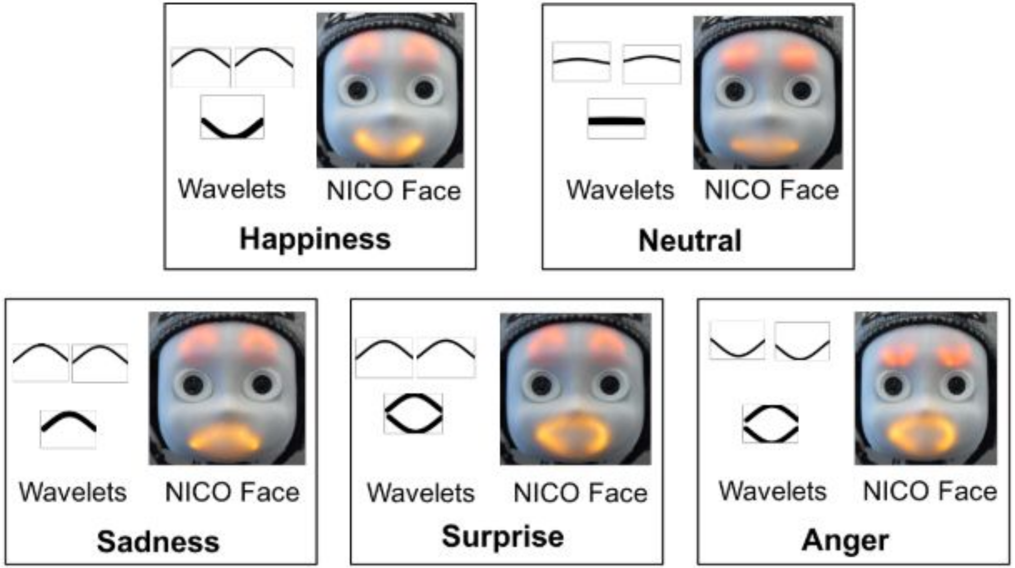
Marsland, Shapiro, & Nehmzow (2002)

Then what?



Churamani et al. (2018)

Emotion expression module



Churamani et al. (2018)

III. Evaluation - Accuracy: SAVEE

- *Surrey Audio-Visual Expressed Emotions*
- Standardized lab-recordings

F: Face channel
A: Speech & Music (Auditory Channel)
AV: Face & Auditory Combined

Class	F	A	AV
Anger	95.4	92.6	100
Disgust	95.6	88.0	100
Fear	89.7	85.5	100
Happiness	100	86.1	95.0
Neutral	100	91.3	100
Sadness	90.0	87.4	96.5
Surprise	86.7	80.5	96.7
Mean	93.9	87.3	98.3

Accuracy in %

Barros & Wermter (2016)

Accuracy: EmotiW

- *Emotion recognition "in the wild"*
- More natural settings

V: Face & Movement (Visual Channel)
A: Speech & Music (Auditory Channel)
AV: Visual & Auditory Combined

Class	V	A	AV
Anger	77.8	70.1	80.3
Disgust	18.7	15.2	23.4
Fear	20.2	7.2	30.8
Happiness	77.8	72.0	81.2
Neutral	70.9	25.4	68.7
Sadness	23.2	16.2	24.5
Surprise	12.1	4.1	14.0
Mean	42.9	30.0	46.1

Accuracy in %

Barros & Wermter (2016)

Comparison with other successful approaches

EmotiW

Methodology	Video	Audio	Both
Liu et al. (2014)	45.28	30.73	48.53
Kahou et al. (2013)	38.1	29.3	41.1
Dhall et al. (2014)	33.15	26.10	28.19
CCCNN	42.9	30.0	46.1

Mean accuracy (%)
on validation split

Barros & Wermter (2016)

GWR vs. no GWR

EmotiW

Accuracy (%)
on validation split

Class	CCCNN	CCCNN+GWR
Anger	80.3	86.4
Disgust	23.4	32.6
Fear	30.8	35.4
Happiness	81.2	85.2
Neutral	68.7	67.1
Sadness	24.5	33.8
Surprise	14.0	17.5
Mean	46.1	51.1

Barros & Wermter (2017)

IV. Conclusion

- Empathy-driven HRI need should account for ...
- **Multi-modality**: e.g. Multi-Channel CNN
- **Interindividual variability**: e.g. Growing-When-Required
- **Context**: e.g. Affective Memory
- Shunting Inhibition for **efficiency, robustness**

- More channels for more multi-modality?
- What if user affect changes instantly?

V. References

Barros, P., Weber, C., & Wermter, S. (2015). Emotional Expression Recognition with a Cross-Channel Convolutional Neural Network for Human-Robot Interaction. In *IEEE-RAS 15th International Conference on Humanoid Robots (Humanoids)* (pp. 582–587). Seoul, Korea: IEEE.

Barros, P., & Wermter, S. (2016). Developing Crossmodal Expression Recognition Based on a Deep Neural mModel. *Adaptive Behavior*, *24*(5), 373–396. https://doi.org/10.1177/1059712316664017

Barros, P., & Wermter, S. (2017). A Self-Organizing Model for Affective Memory. In *International Joint Conference on Neural Networks (IJCNN)* (pp. 31–38). IEEE.

Churamani, N., Barros, P., Strahl, E., & Wermter, S. (2018). Learning Empathy-Driven Emotion Expressions using Affective Modulations. In *Proceedings of International Joint Conference on Neural Networks (IJCNN)*. IEEE. https://doi.org/10.1109/IJCNN.2018.8489158

Marsland, S., Shapiro, J., & Nehmzow, U. (2002). A Self-Organising Network that Grows When Required. *Neural Networks*, *15*(8-9), 1041-1058.

Matthias Kerzel, Erik Strahl, Sven Magg, Nicolás Navarro-Guerrero, Stefan Heinrich, Stefan Wermter. NICO – Neuro-Inspired COmpanion: A Developmental Humanoid Robot Platform for Multimodal Interaction. *Proceedings of the IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)* (pp. 113 - 120). Lisbon, Portugal. 2017.

Mordoch, E., Osterreicher, A., Guse, L., Roger, K., & Thompson, G. (2013). Use of Social Commitment Robots in the Care of Elderly People with Dementia: A Literature Review. *Maturitas*, *74*(1), 14-20.

V. References

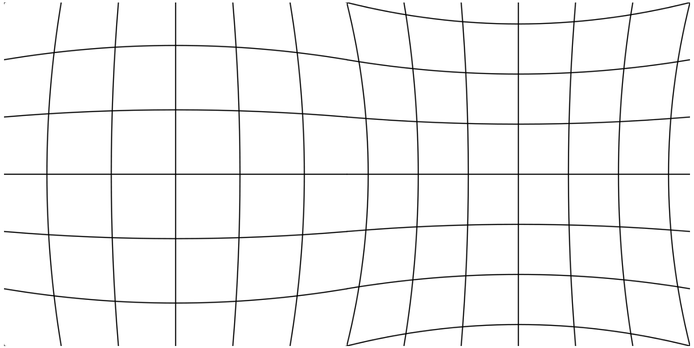
Ricks, D. J., & Colton, M. B. (2010). Trends and Considerations in Robot-Assisted Autism Therapy. In *Robotics and Automation (ICRA), 2010 IEEE International Conference on* (pp. 4354-4359). IEEE.

Tielman, M., Neerincx, M., Meyer, J. J., & Looije, R. (2014). Adaptive Emotional Expression in Robot-Child Interaction. In *Proceedings of the 2014 ACM/IEEE international conference on Human-robot interaction* (pp. 407-414). ACM.

Tivive, F. H. C., & Bouzerdoum, A. (2006). A Shunting Inhibitory Convolutional Neural Network for Gender Classification. In *18th International Conference on Pattern Recognition 2006 (ICPR 2006)* (Vol. 4, pp. 421–424). IEEE.

Excursus: Shunting inhibition

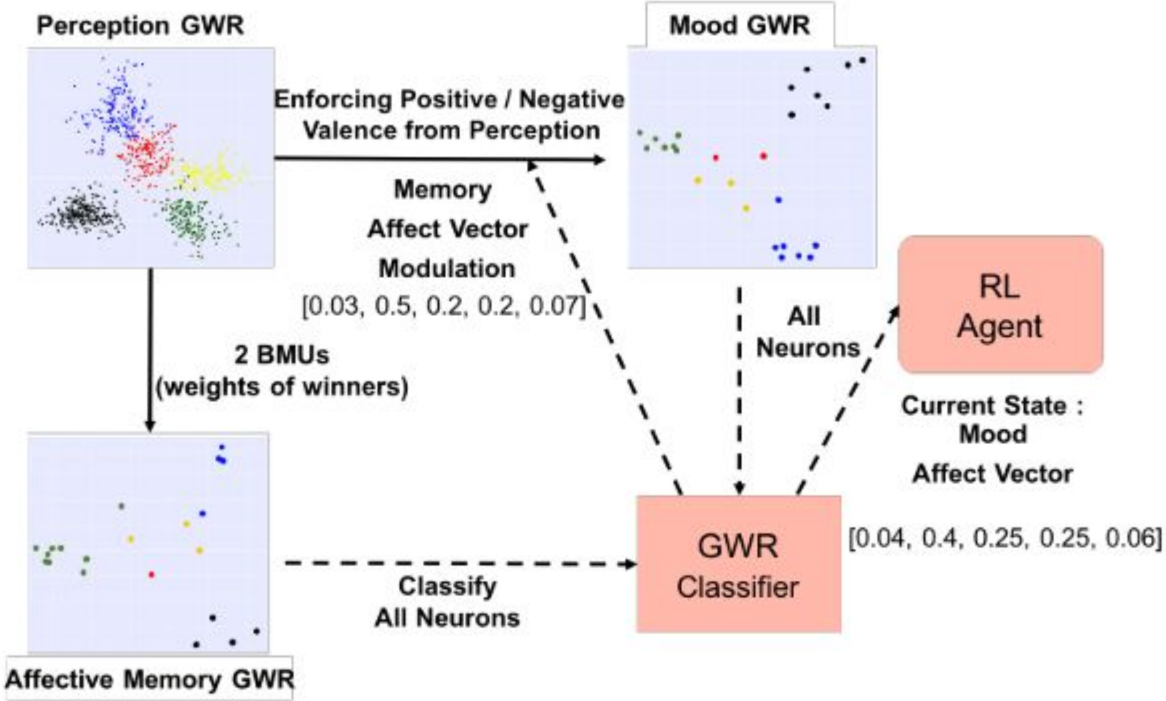
- Neuro-physiological plausible mechanisms present in several visual and cognitive functions
- Improve efficiency of filters when applied to complex cells:
    - increase robustness to geometric distortion
    - learn more high-level features
- Can reduce amount of layers needed
    - less parameters to be trained



https://en.wikipedia.org/wiki/Distortion_(optics)

Barros & Wermter (2016)

Excursus: Intrinsic Emotion



Thank you for listening!

Any questions?