KTH Computer Science
and Communication

Potentially Brain Related Research at TMH

Giampiero Salvi

KTH/CSC/TMH giampi@kth.se

Brain-IT Workshop Jan 2014

Topics at Tal, Musik och Hörsel (TMH)

Speech Technology

- Speech Recognition
- Speech Synthesis
- Dialogue Systems

Topics at Tal, Musik och Hörsel (TMH)

Speech Technology

- Speech Recognition
- Speech Synthesis
- Dialogue Systems

Basic Building Blocks:

- Signal Processing
- Machine Learning
- Computational Linguistics

Outline

Ex1: Phoneme recognition with deep vs cortex-inspired architectures (CB+TMH)

Ex2: Mapping between voices based on topology preserving Self Organizing Maps

Ex3: Word discovery with non-parametric Bayesian methods

Ex4: Word meaning association and grounding with Bayesian Networks (TMH+IST)

Conclusions

Outline

Phoneme recognition with deep vs cortex-inspired architectures

Joint work with CB:

- Pawel Herman
- Tin Franovic [3][4]
- Nizar Gandy Assaf Layouss [2]

---

[3] T. Franovic. "Exploratory Multivariate Search for Spectro-Temporal Associations in Speech Data Using a Biomimetic Framework". MA thesis. KTH, CSC, 2012

[4] T. Franovic, P. Herman, G. Salvi, S. Benjaminsson, and A. Lansner. "Cortex-inspired network architecture for large-scale temporal information processing". In: *Frontiers in neuroinformatics*. Vol. 7. 2013

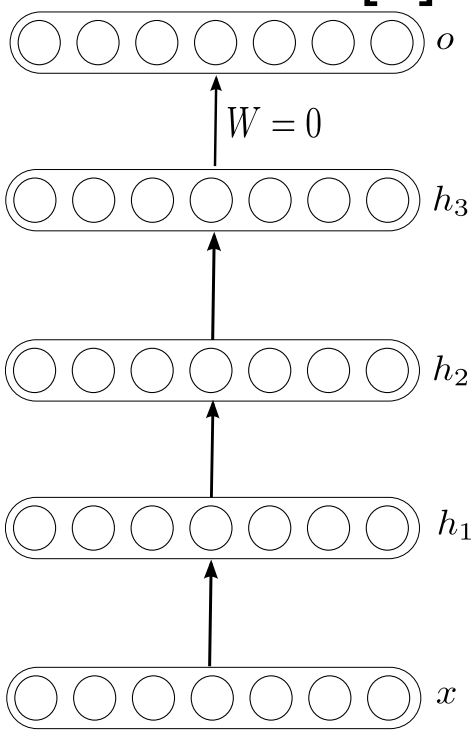
[2] N. G. Assaf Layouss. "A critical examination of deep learning approaches to automated speech recognition". MA thesis. KTH, CSC, 2013

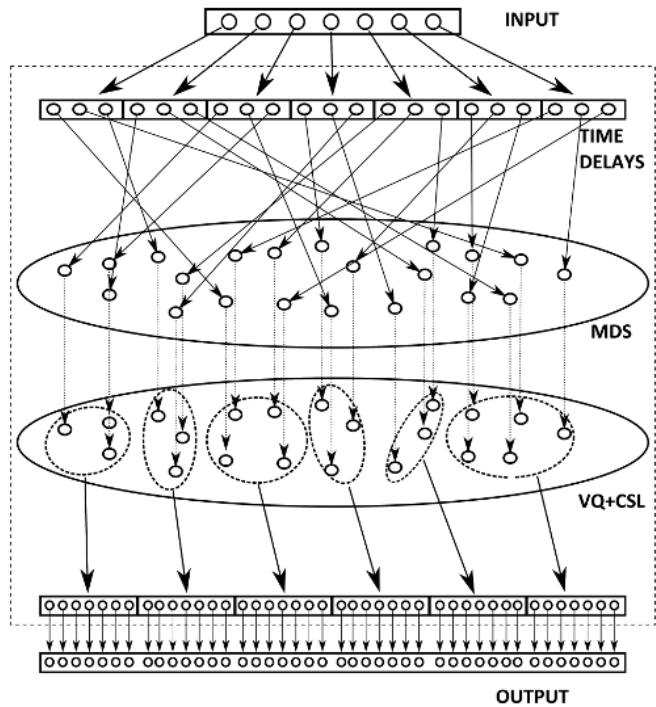
Goal: Compare speech representations learned by:

Deep Neural Networks [6]

and

Cortex-inspired architecture

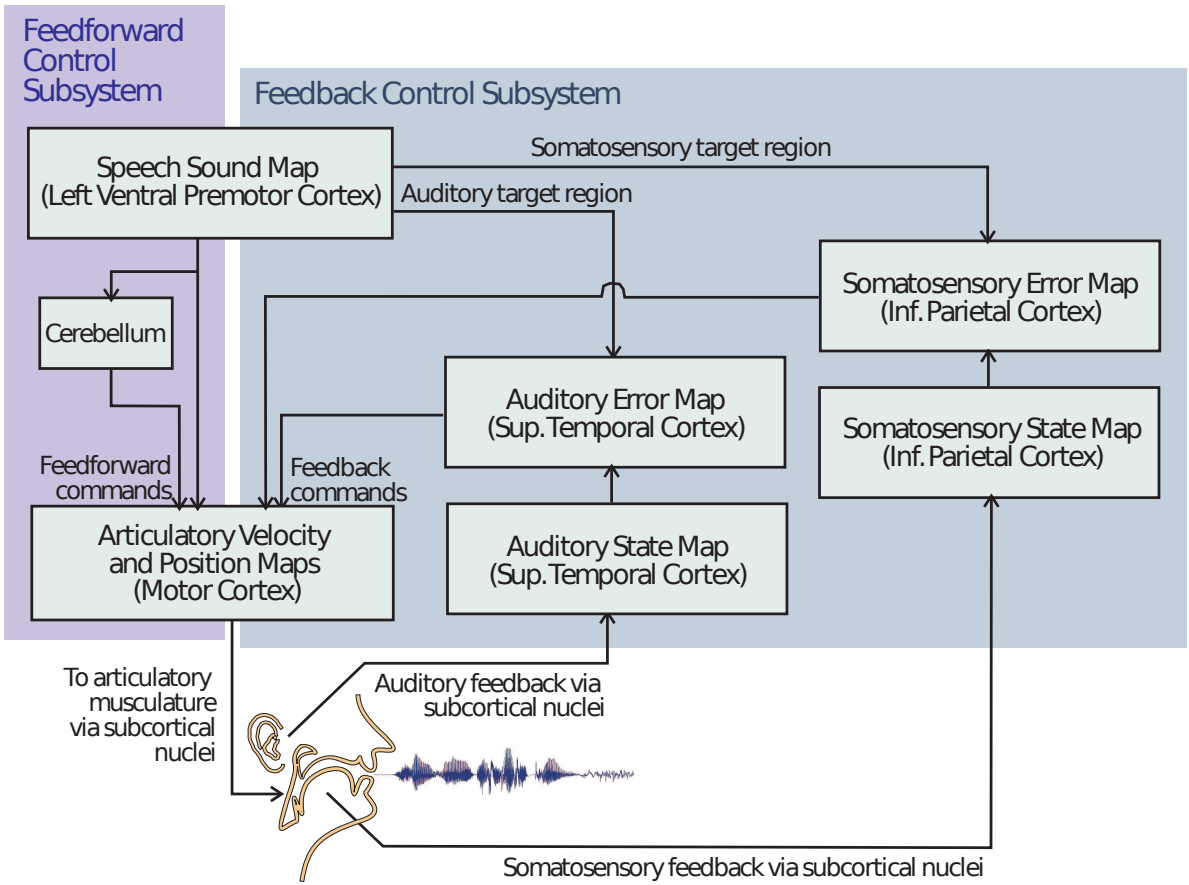


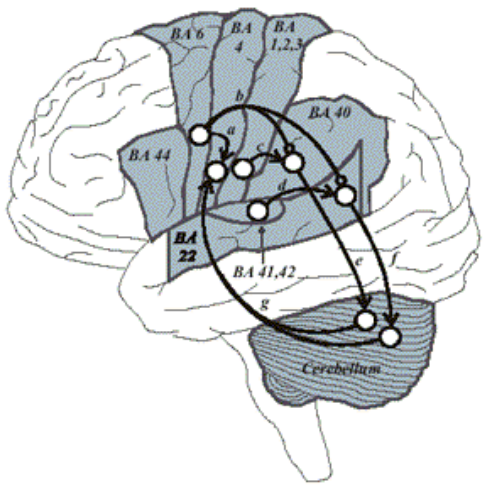


[6] G. Hinton, L. Deng, D. Yu, G. Dahl, A. Mohamed, N. Jaitly, A. Senior, V. Vanhoucke, P. Nguyen, T. Sainath, and B. Kingsbury. "Deep neural networks for acoustic modeling in speech recognition". In: *IEEE Signal Processing Magazine* 29.6 (2012), pp. 82–97

Outline

The DIVA model [5]

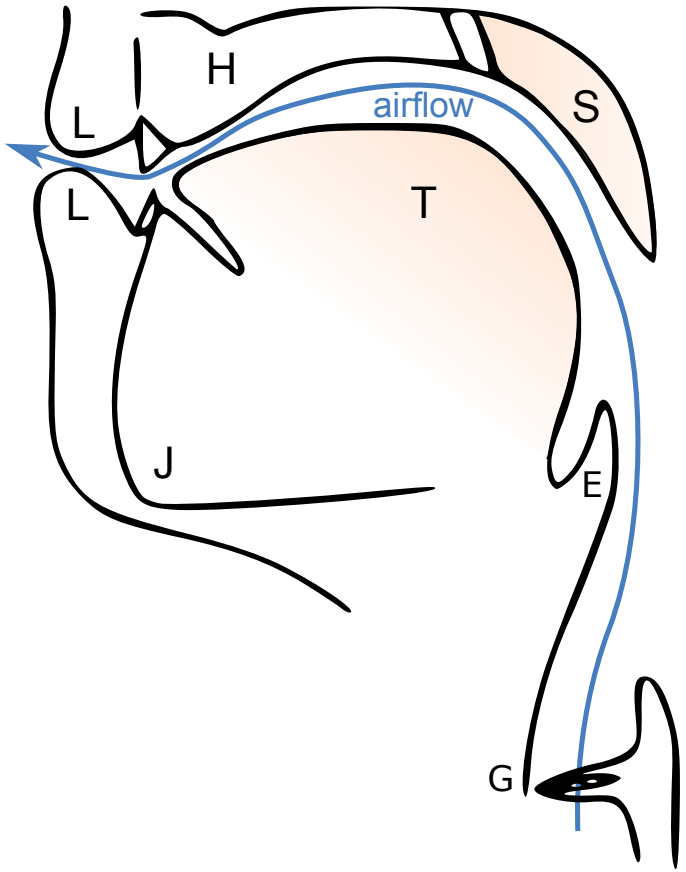




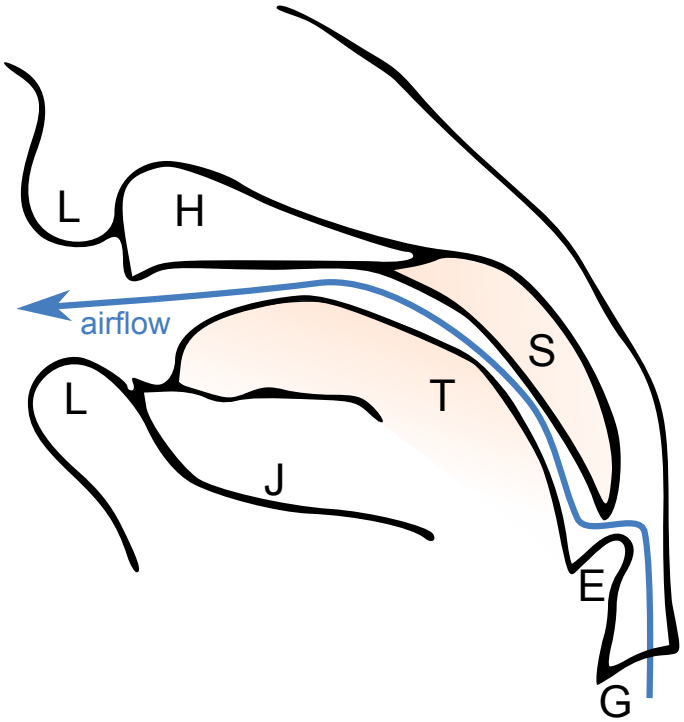
[5] F. H. Guenther, S. S. Ghosh, and J. A. Tourville. "Neural Modeling and Imaging of the Cortical Interactions Underlying Syllable Production". In: *Brain and Language* 96 (2006), pp. 280–301

Imitation Learning: voice mismatch

Adult vocal tract

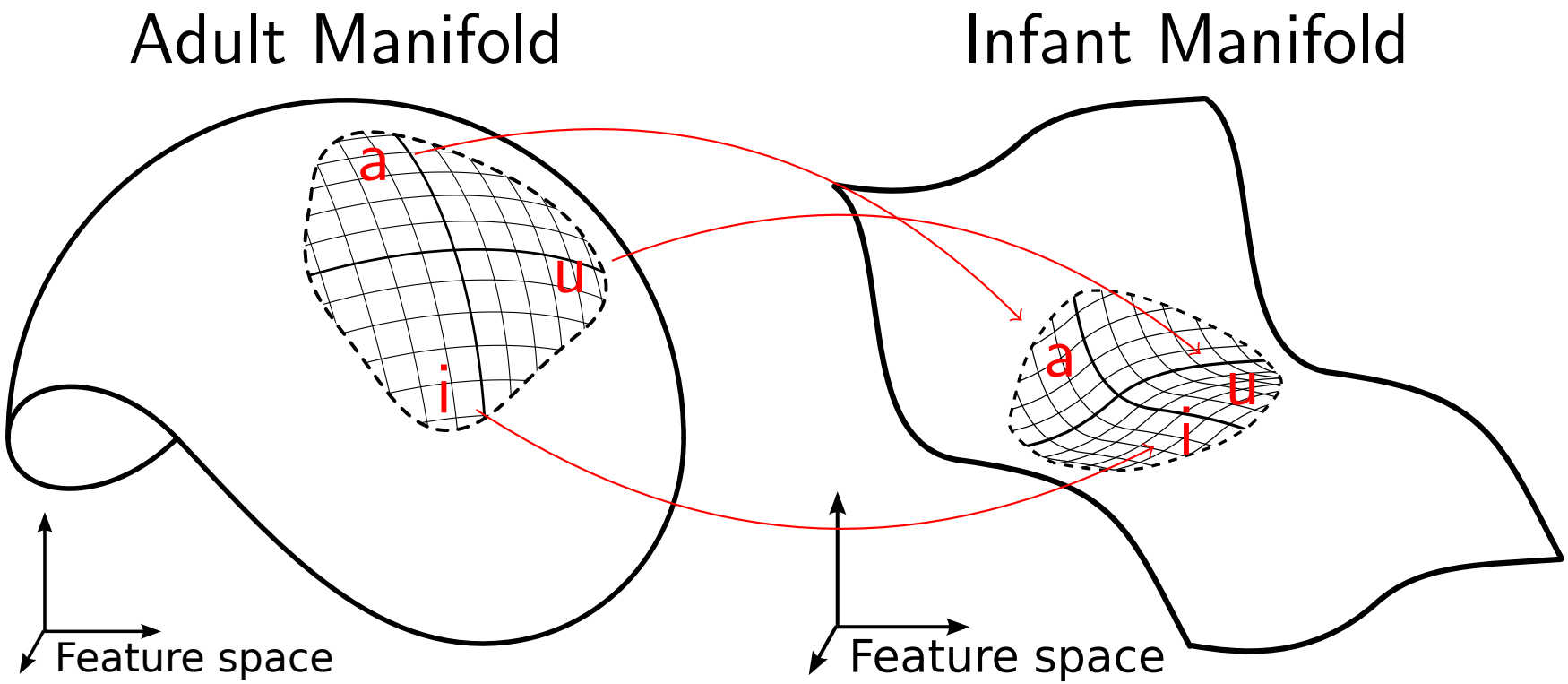


Infant vocal tract



H=hard palate, S=soft palate, E=epiglottis,
G=glottis, T=tongue, J=jaw, L=lips

Imitation Learning: voice mismatch [1]



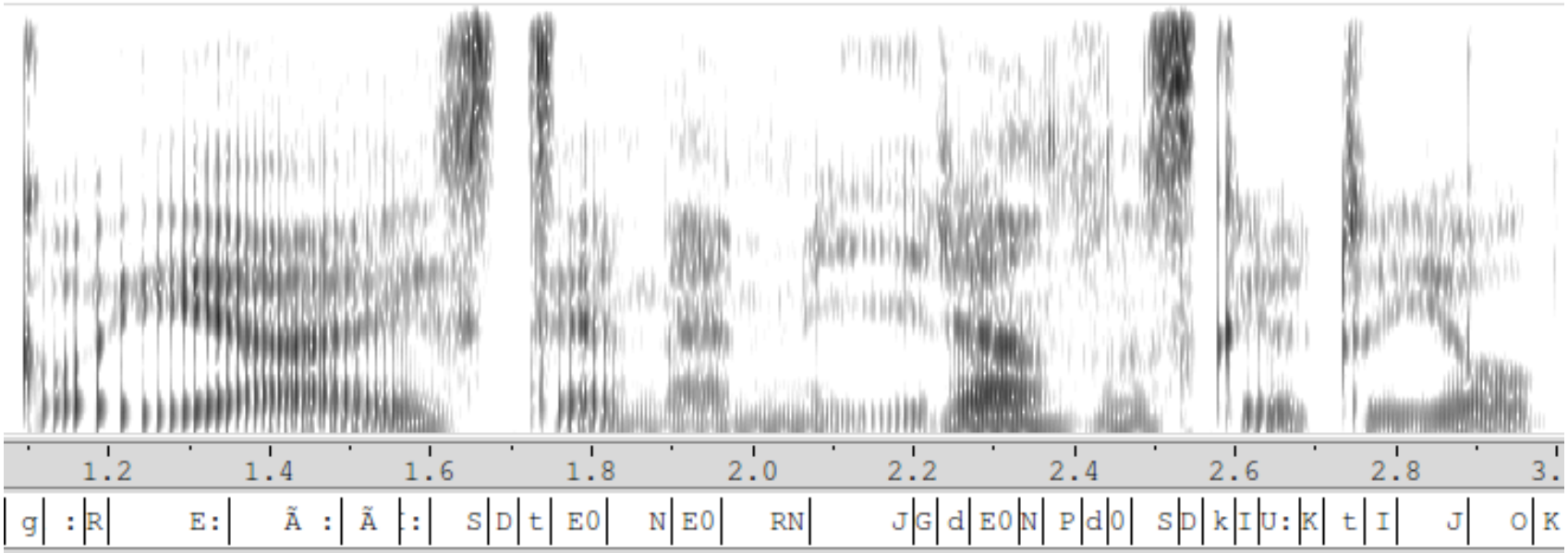
Learning by imitation with Self Organizing Maps

---

[1] G. Ananthakrishnan and G. Salvi. "Using Imitation to learn Infant-Adult Acoustic Mappings". In: *Proc. of Interspeech*. Firenze, Italy, 2011

Outline

Learning words



[7]

---

[7] F. Lacerda, E. Klintfors, L. Gustavsson, L. Lagerkvist, E. Marklund, and U. Sundberg. "Ecological Theory of Language Acquisition". In: *Forth International Workshop an Epigenetic Robotics*. 2004, pp. 147–148

Illustration (10 “words”)

MOIWXMOPOQSMNVQVSEQASDPOEMOASF
MOPOQSMOPOQSMOASFANMO
ANMOMNMXSONMNMOPOQS
ANMOANMOMOPOQSMOASFZSWOS
MOASFANMOANMOMOPOQS
MOASFMOPOQSMNVQVSEMNVQVSE
NSKDFEMXSONMOASFMOPOQSNSKDFE
MNVQVSEMNMXSONNSKDFEMXSON
MNZSWOSMOIWXMNVQVSE
MXSONNSKDFENSKDFEMOASF
MNMXSONQASDPOEANMOMNVQVSEMNVQVSE
MOASFANMOMXSONQASDPOEANMO
NSKDFEMOASFMXSONMOPOQS
MOIWXMOIWXMOASFZSWOSNSKDFE
ANMOANMOMOPOQSMXSONNSKDFE
QASDPOEZSWOSMOASFMNVQVSE
MNMNQASDPOE

Illustration (10 "words")

MOIWX MOPOQS MNVQVSE QASDPOE MOASF
MOPOQS MOPOQS MOASF ANMO
ANMO MN MXSON MN MOPOQS
ANMO ANMO MOPOQS MOASF ZSWOS
MOASF ANMO ANMO MOPOQS
MOASF MOPOQS MNVQVSE MNVQVSE
NSKDFE MXSON MOASF MOPOQS NSKDFE
MNVQVSE MN MXSON NSKDFE MXSON
MN ZSWOS MOIWX MNVQVSE
MXSON NSKDFE NSKDFE MOASF
MN MXSON QASDPOE ANMO MNVQVSE MNVQVSE
MOASF ANMO MXSON QASDPOE ANMO
NSKDFE MOASF MXSON MOPOQS
MOIWX MOIWX MOASF ZSWOS NSKDFE
ANMO ANMO MOPOQS MXSON NSKDFE
QASDPOE ZSWOS MOASF MNVQVSE
MN MN QASDPOE

Our solution

- Based on non-parametric Bayesian methods [10][9]
- not biologically related
- it would be interesting to find similar processing capabilities in the brain

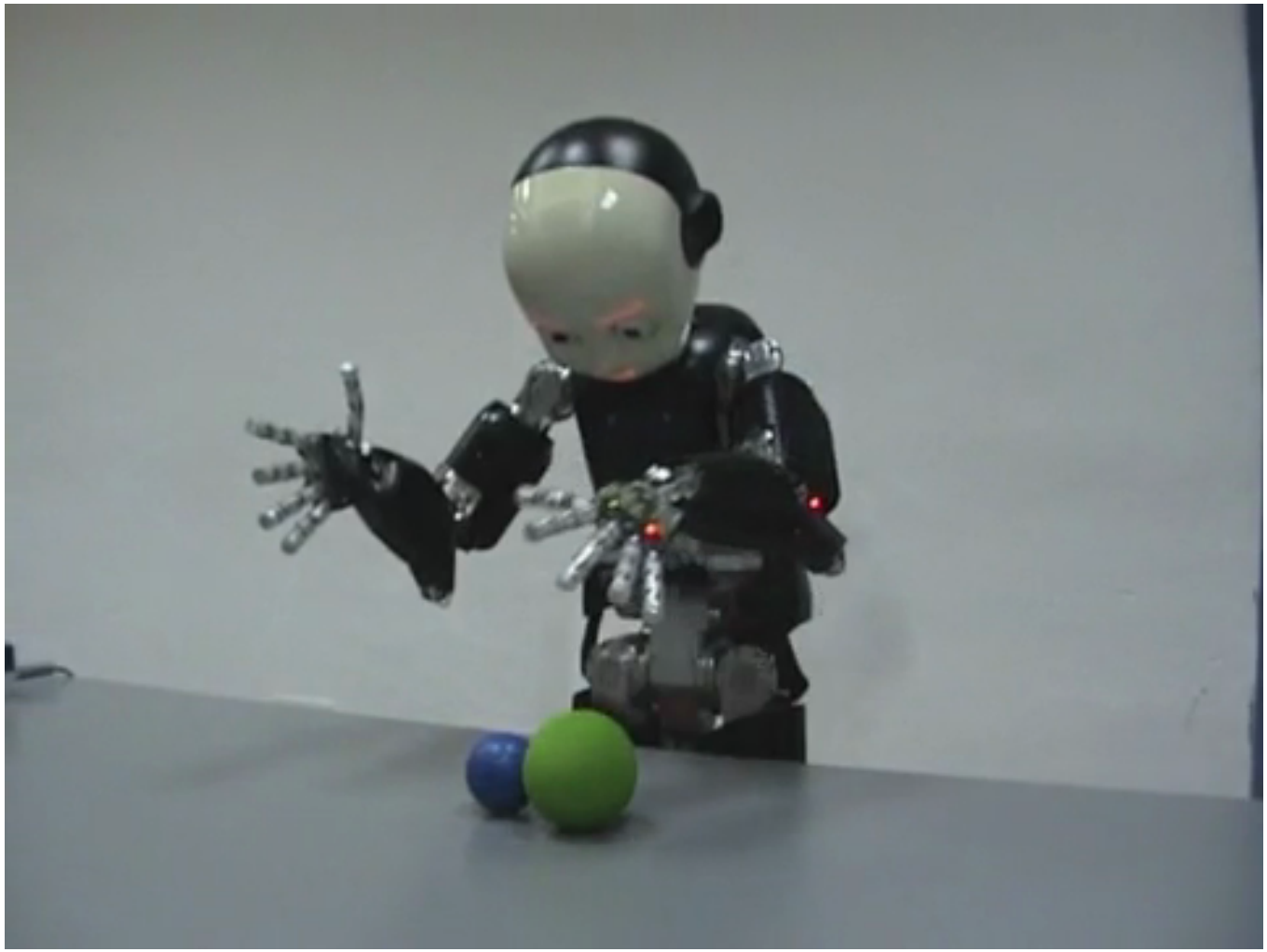
---

[10] N. Vanhainen and G. Salvi. "Word Discovery with Beta Process Factor Analysis". In: *Proc. of Interspeech*. Portland, OR, USA, Sept. 2012

[9] N. Vanhainen and G. Salvi. "Pattern Discovery in Continuous Speech Using Block Diagonal Infinite HMM". In: *Proc. of IEEE ICASSP*. submitted

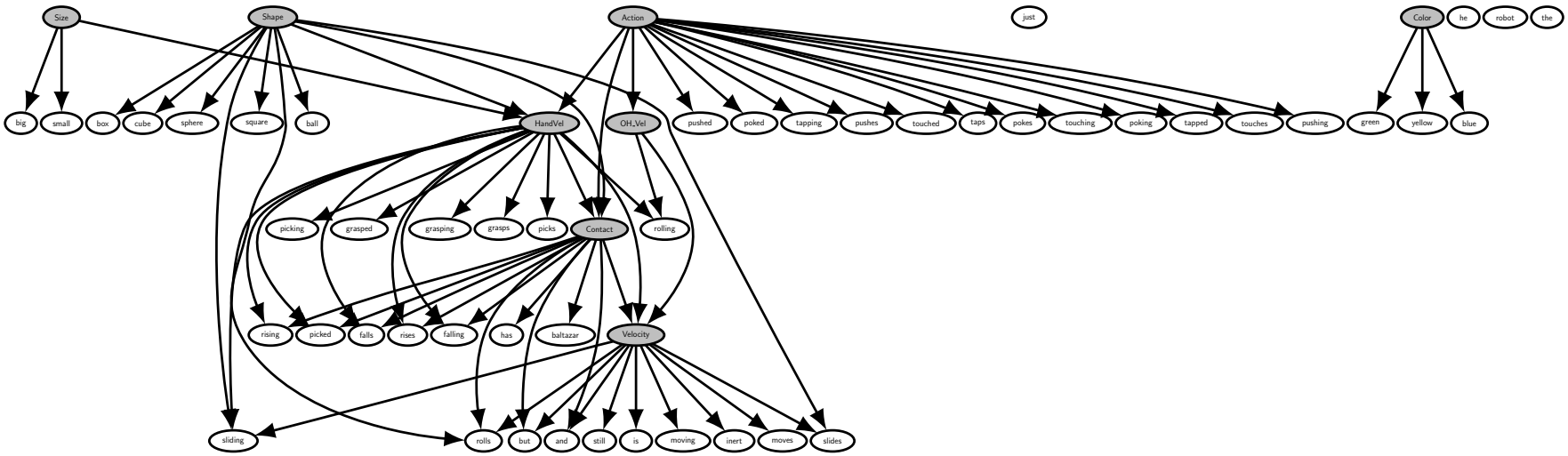
Outline

Words and affordances [8]



[8] G. Salvi, L. Montesano, A. Bernardino, and J. Santos-Victor. "Language bootstrapping: Learning word meanings from perception-action association". In: *IEEE Trans. Syst., Man, Cybern. B* 42.3 (June 2012), pp. 660–671

Words and affordances [8]



the meaning of words is grounded into the robots action/perception world

---

[8] G. Salvi, L. Montesano, A. Bernardino, and J. Santos-Victor. "Language bootstrapping: Learning word meanings from perception-action association". In: *IEEE Trans. Syst., Man, Cybern. B* 42.3 (June 2012), pp. 660–671

Outline

Ex1: Phoneme recognition with deep vs cortex-inspired architectures (CB+TMH)

Ex2: Mapping between voices based on topology preserving Self Organizing Maps

Ex3: Word discovery with non-parametric Bayesian methods

Ex4: Word meaning association and grounding with Bayesian Networks (TMH+IST)

Conclusions

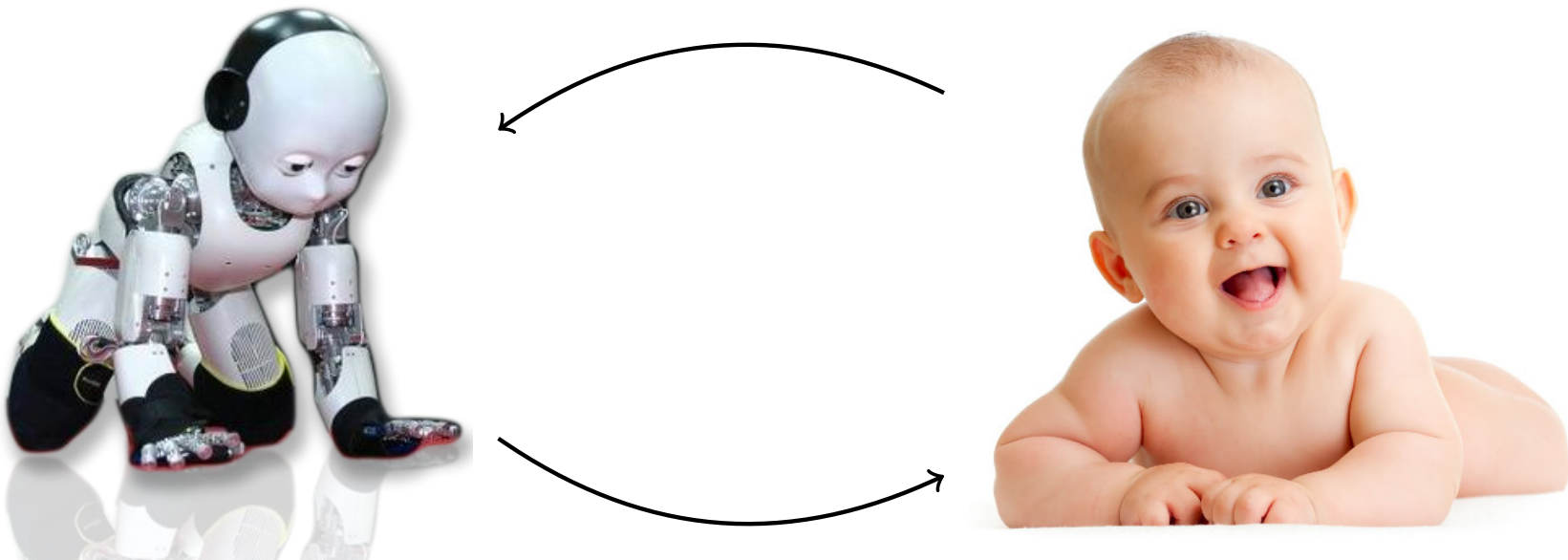
Machine Learning and Biological Systems

Successful because:

- powerful tool to solve complex problems
- can solve aspects of the problems that are not entirely understood ("black box")
- not necessarily similar to human learning

My Interest in Machine Learning

- there is much to be learned about biological systems from ML modelling
- inspiration from biological systems can be very beneficial to ML



Conclusions

We try to model speech related cognitive abilities

Technological goal:

creating talking machines

Scientific goal:

understanding humans (human brain?) better

References

[1] G. Ananthakrishnan and G. Salvi. "Using Imitation to learn Infant-Adult Acoustic Mappings". In: *Proc. of Interspeech*. Firenze, Italy, 2011.

[2] N. G. Assaf Layouss. "A critical examination of deep learning approaches to automated speech recognition". MA thesis. KTH, CSC, 2013.

[3] T. Franovic. "Exploratory Multivariate Search for Spectro-Temporal Associations in Speech Data Using a Biomimetic Framework". MA thesis. KTH, CSC, 2012.

[4] T. Franovic, P. Herman, G. Salvi, S. Benjaminsson, and A. Lansner. "Cortex-inspired network architecture for large-scale temporal information processing". In: *Frontiers in neuroinformatics*. Vol. 7. 2013.

[5] F. H. Guenther, S. S. Ghosh, and J. A. Tourville. "Neural Modeling and Imaging of the Cortical Interactions Underlying Syllable Production". In: *Brain and Language* 96 (2006), pp. 280–301.

[6] G. Hinton, L. Deng, D. Yu, G. Dahl, A. Mohamed, N. Jaitly, A. Senior, V. Vanhoucke, P. Nguyen, T. Sainath, and B. Kingsbury. "Deep neural networks for acoustic modeling in speech recognition". In: *IEEE Signal Processing Magazine* 29.6 (2012), pp. 82–97.

[7] F. Lacerda, E. Klintfors, L. Gustavsson, L. Lagerkvist, E. Marklund, and U. Sundberg. "Ecological Theory of Language Acquisition". In: *Forth International Workshop an Epigenetic Robotics*. 2004, pp. 147–148.

[8] G. Salvi, L. Montesano, A. Bernardino, and J. Santos-Victor. "Language bootstrapping: Learning word meanings from perception-action association". In: *IEEE Trans. Syst., Man, Cybern. B* 42.3 (June 2012), pp. 660–671.

[9] N. Vanhainen and G. Salvi. "Pattern Discovery in Continuous Speech Using Block Diagonal Infinite HMM". In: *Proc. of IEEE ICASSP*. submitted.

[10] N. Vanhainen and G. Salvi. "Word Discovery with Beta Process Factor Analysis". In: *Proc. of Interspeech*. Portland, OR, USA, Sept. 2012.