

# A COMPUTATIONAL MODEL OF MOTOR COGNITION FOR ICONIC GESTURE PROCESSING

AMIR SADEGHIPOUR

*A doctoral thesis submitted in fulfilment of the requirements for the degree of*  
Doktor-Ingenieur (Doctor of Engineering)

*in the*

Social Cognitive Systems group  
Cognitive Interaction Technology – Center of Excellence  
Faculty of Technology  
Bielefeld University  
Bielefeld, Germany

Supervisor: Prof. Dr.-Ing. Stefan Kopp

June 2014



Berichte aus der Robotik

**Amir Sadeghipour**

**A computational model of motor cognition  
for iconic gesture processing**

Shaker Verlag  
Aachen 2015

**Bibliographic information published by the Deutsche Nationalbibliothek**

The Deutsche Nationalbibliothek lists this publication in the Deutsche Nationalbibliografie; detailed bibliographic data are available in the Internet at <http://dnb.d-nb.de>.

Zugl.: Bielefeld, Univ., Diss., 2014

Copyright Shaker Verlag 2015

All rights reserved. No part of this publication may be reproduced, stored in a retrieval system, or transmitted, in any form or by any means, electronic, mechanical, photocopying, recording or otherwise, without the prior permission of the publishers.

Printed in Germany.

ISBN 978-3-8440-3451-6

ISSN 1434-8098

Shaker Verlag GmbH • P.O. BOX 101818 • D-52018 Aachen

Phone: 0049/2407/9596-0 • Telefax: 0049/2407/9596-9

Internet: [www.shaker.de](http://www.shaker.de) • e-mail: [info@shaker.de](mailto:info@shaker.de)

## ABSTRACT

---

In the present thesis, we propose a computational cognitive model for processing iconic gestures performed with hands and arms. Among different social behaviors, investigating gestures has gained special attention during the last decades. From a phenomenological perspective, gestures comprise mostly unconventional non-verbal behaviors that bring different aspects of social, motor and embodied cognition together. This inherent variability creates considerable challenges during processing, and in particular when the goal is cognitive modeling. In the field of human-computer interaction, advances in motion tracking systems have opened the gateway to the broader application of gesture-based user interfaces. Virtual humanoid agents are being increasingly used in human-computer interaction as user friendly interfaces that allow for intuitive and natural face-to-face communication. Furthermore, they can embody computational models of human social cognition, and provide suitable testbeds for cognitive models and theories.

Against this background, we propose a model that endows artificial humanoid agents with the ability to gesture. This capacity includes, for instance, fast and reliable recognition of highly diverse gesture performances, the ability to learn how to perform gestures through imitation, and to establish gestural alignment during interaction with human users. To this end, we propose a cognitive model of the human sensorimotor system based on neuroscience and psychological empirical evidence, and couched in current cognitive theories. To implement information processing within this cognitive model, we propose two computational approaches with complementary strengths and weaknesses. Empirical Bayesian Belief Update (EBBU) features the fast, incremental and cognitively plausible recognition of gestures during interaction with humans, achieved by combining bottom-up perception with top-down prediction. We also propose a Feature-based Stochastic Context-Free Grammar (FSCFG) to learn discriminative or descriptive grammar models of gestures. Applying this approach to a recorded dataset of iconic gestures demonstrated its considerable generalization capacity over both syntactic structure variabilities, and the statistical spatiotemporal deviations inherent to gesture.



## ACKNOWLEDGEMENTS

---

I would like to express my deepest gratitude to my advisor, Prof. Stefan Kopp, for his excellent guidance, motivation and enthusiasm, and for allowing me the room to develop and explore my own ideas. His guidance helped me to find my way in the multifaceted research field of cognitive science.

I would like to thank my wonderful colleagues for making it possible to work in an excellent research environment. My special thanks go to Hendrik Buschmeier, Ramin Yaghoubzadeh and Farina Freigang, for the stimulating discussions, and for their helpful suggestions to improve this thesis, as well as for the fruitful collaborations and long working nights. I would also like to thank Prof. Louis-Philippe Morency for offering me a summer internship in his group, and for his guidance in creating a gesture dataset.

Many thanks go to the Cognitive Interaction Technology (CITEC) Center of Excellence at Bielefeld University for providing me with the support and equipment I needed. In addition, my most sincere gratitude to the Deutsche Forschungsgemeinschaft (DFG) for funding this project.

I would never have been able to finish – or even start – my Ph.D. research, without the immeasurable support of my family. I thank my parents for their support and for encouraging me with their best wishes during this long period, and despite the great distance between us. Finally, and most importantly, I would like to thank my dear wife, not only for her support through the good times and bad, but also for the very inspiring discussions we had about my research.





## MAIN TOPICS

---

i	STATEMENT OF THE PROBLEM	1
1	INTRODUCTION	3
2	BACKGROUND	9
3	ICONIC GESTURES	31
ii	A CONCEPTUAL COGNITIVE MODEL	71
4	COGNITIVE FOUNDATIONS	73
5	THE COGNITIVE MODEL	109
iii	COMPUTATIONAL APPROACHES	141
6	EMPIRICAL BAYESIAN BELIEF UPDATE	143
7	FEATURE-BASED STOCHASTIC CONTEXT-FREE GRAMMAR	183
iv	CONCLUDING DISCUSSIONS	239
8	CONCLUSION AND OUTLOOK	241
	BIBLIOGRAPHY	257
	ACRONYMS	277
	NOMENCLATURE	279
	INDEX	281



## CONTENTS

---

<b>i</b>	<b>STATEMENT OF THE PROBLEM</b>	<b>1</b>
<b>1</b>	<b>INTRODUCTION</b>	<b>3</b>
<b>2</b>	<b>BACKGROUND</b>	<b>9</b>
2.1	Computational Cognitive Modeling	9
2.1.1	Modeling approaches	10
2.1.2	Goals of modeling and cognitive plausibility	13
2.1.3	Modeling levels	18
2.2	Embodied Conversational Agents	19
2.2.1	Applications of ECA in HCI	21
2.2.2	ECAs as social actors	23
2.2.3	The importance of nonverbal behavior	26
2.2.4	The application of ECAs in this research	27
2.3	Summary	29
<b>3</b>	<b>ICONIC GESTURES</b>	<b>31</b>
3.1	Phenomenology of gestures	31
3.1.1	From actions to gestures, language and thought	32
3.1.2	Gesture taxonomy	35
3.1.3	Gestures' temporal anatomy	41
3.1.4	Gestures in Human-Computer Interaction	42
3.1.5	Gestures in the focus of this research	43
3.2	Analysis of iconic gesture performances	46
3.2.1	Three-Dimensional Iconic Gesture (3DIG) dataset	48
3.2.2	Representational techniques of iconicity in gestures	54
3.2.3	Structural variabilities of gesture performances	61
3.2.4	Spatiotemporal feature variabilities of gesture performances	67
3.3	Summary	70
<b>ii</b>	<b>A CONCEPTUAL COGNITIVE MODEL</b>	<b>71</b>
<b>4</b>	<b>COGNITIVE FOUNDATIONS</b>	<b>73</b>
4.1	Embodied social motor cognition	73
4.2	Empirical evidence of mirroring	75

4.2.1	Neuroscience evidence	75
4.2.2	Psychological evidence	78
4.2.3	Degree of mirroring	79
4.3	Mirroring mechanisms	80
4.3.1	Priming	81
4.3.2	Mimicry	83
4.3.3	Imitation learning	84
4.3.4	Goal emulation	88
4.3.5	Mirroring mechanisms of gestures	88
4.3.6	Social effects and benefits of mirroring mechanisms	90
4.4	Hierarchical representation of motor knowledge	97
4.4.1	Hierarchical representation and perception	97
4.4.2	Hierarchical representation and motor control	99
4.4.3	Abstraction in the sense of memory consolidation	102
4.5	Intertwined processes of motor cognition	103
4.5.1	Predictive perception during motor control	103
4.5.2	Predictive perception of others' behaviors	105
4.5.3	Bottom-up perception and top-down anticipation	106
4.5.4	Anticipation results in recognition or adaptation	107
4.6	Summary	108
5	THE COGNITIVE MODEL	109
5.1	Representation of hierarchical motor knowledge	109
5.1.1	What to represent?	109
5.1.2	How to represent gestural motor knowledge?	111
5.2	Processes of motor cognition	116
5.2.1	Perception module	119
5.2.2	Prediction module	120
5.2.3	Control module	123
5.2.4	Realization module	125
5.2.5	Mirroring mechanisms	126
5.3	Derivation of computational requirements	134
5.3.1	Handling uncertain perception	134
5.3.2	Handling uncertain generation	137
5.3.3	Fast learning and recognition	138
5.3.4	Predictive perception and incremental recognition	138

5.3.5	Generalization and incremental learning	139
5.4	Summary	140
iii	COMPUTATIONAL APPROACHES	141
6	EMPIRICAL BAYESIAN BELIEF UPDATE	143
6.1	Related work	144
6.1.1	Related cognitive models	144
6.1.2	Related technical models	146
6.2	Pre-processing	147
6.2.1	Perception and preparation	147
6.2.2	Segmentation	148
6.3	Graph-based representation	150
6.4	Cognitive processes	152
6.4.1	Bottom-up perception	152
6.4.2	Top-down prediction	158
6.4.3	Acquiring motor knowledge from observation	161
6.4.4	Perception-action link	163
6.5	Evaluation of EBBU based on ECA-human interaction	166
6.5.1	Detecting and learning new gestures	167
6.5.2	Recognizing familiar gestures	170
6.5.3	Perception-action link and mirroring mechanisms	176
6.6	Summary	181
7	FEATURE-BASED STOCHASTIC CONTEXT-FREE GRAMMAR	183
7.1	Related work	183
7.2	A probabilistic framework for FSCFG	185
7.2.1	Parsing	187
7.2.2	Structure learning	192
7.2.3	Parameter learning	195
7.2.4	Handling uncertain input	197
7.3	Application to the problem	199
7.3.1	Segmentation	199
7.3.2	Initializing FSCFG and incorporating samples	201
7.3.3	Supervised and unsupervised learning of motor knowledge	201
7.3.4	Generation of gesture performances	203
7.4	Evaluation of FSCFG based on 3DIG dataset	203
7.4.1	Supervised learning	205

7.4.2	Classification confusions	209
7.4.3	Classification metrics	212
7.4.4	Human judgment as base-line	214
7.4.5	Comparing to other classifiers	217
7.4.6	Analysis of learned grammar models	218
7.4.7	Hyper-parameter exploration	221
7.4.8	Unsupervised learning	227
7.4.9	The grammar behind motor knowledge	234
7.5	Summary	236
iv	CONCLUDING DISCUSSIONS	239
8	CONCLUSION AND OUTLOOK	241
8.1	What has been gained?	242
8.1.1	Hierarchical representation of gestural motor knowledge	242
8.1.2	Modeling the processes of motor cognition	243
8.1.3	Realizing mirroring mechanisms of gestural interaction	244
8.2	Combinations of EBBU and FSCFG	245
8.2.1	Combining processes	245
8.2.2	Common representation	247
8.3	Limitations and outlook	253
	BIBLIOGRAPHY	257
	ACRONYMS	277
	NOMENCLATURE	279
	INDEX	281