DISSERTATION

A Bionic Model for Human-like Machine Perception

Submitted at the
Faculty of Electrical Engineering, Vienna University of Technology
in partial fulfillment of the requirements for the degree of
Doctor of Technical Sciences

under supervision of

Prof. Dr. Dietmar Dietrich
Institute number: 384
Institute of Computer Technology

and

Prof. Dr. Peter Palensky
Department of Electrical,
Electronic and Computer Engineering
University of Pretoria

and

Prof. Dr. Wolfgang Kastner
Institute number: 183/1
Institute of Computer Aided Automation

by

Dipl.-Ing. Rosemarie Velik
Matr.Nr.: 0125411
Lassendorf 43
A-9064 Pischeldorf

Vienna, 11.4.2008

Abstract

Machine perception is a research field that is still in its infancy and is confronted with many unsolved problems. In contrast, humans generally perceive their environment without problems. For the work at hand, these facts were the motivation to develop a bionic model for human-like machine perception, which is based on neuroscientific and neuropsychological research findings about the structural organization and function of the perceptual system of the human brain. Having systems available that are capable of a human-like perception of their environment would allow the automation of processes for which, today, human observers and their cognitive abilities are necessary. Potential applications are, among others, security and safety surveillance of public and private buildings and the automatic observation of the state of health of persons in retirement homes and hospitals. Furthermore, autonomous robots and interactive environments would take advantage of more effective mechanisms to perceive their surrounding.

The introduced model is designated for applications in the field of building automation for autonomous monitoring and surveillance systems to observe objects, events, scenarios, and situations in buildings. Therefore, buildings have to be equipped with a huge number of diverse sensors. The challenge is to merge and interpret the information coming from these different sources.

For this purpose, an information processing principle called *neuro-symbolic information processing* is introduced using *neuro-symbols* as basic information processing units. The inspiration for the utilization of neuro-symbols comes from the fact that humans think in terms of symbols, which emerge from information processed by neurons. Neuro-symbols are connected in a modular hierarchical fashion to a so-called *neuro-symbolic network* to process sensor data. The architecture of the neuro-symbolic network is derived from the structural organization of the perceptual system of the human brain. Connections and correlations between neuro-symbols can be acquired from examples in different learning phases. Besides sensor data processing, *memory*, *knowledge*, and *focus of attention* influence perception to resolve ambiguous sensory information and to devote processing power to relevant features.

The introduced model was implemented with AnyLogic. The model proved to be successful in perceiving all test cases specified in the simulation environment. Furthermore, the insights gained during development allowed it to draw certain conclusions about the inconsistency or incompleteness of neuroscientific and neuropsychological theories including issues like the binding problem, the processing and storage of perceptual information, the computation of location information, and stability considerations.

Kurzfassung

"Machine Perception" ist ein junges Forschungsgebiet, das mit vielen ungelösten Problemen konfrontiert ist. In Gegensatz zu Maschinen können Menschen ihre Umgebung im Allgemeinen mühelos wahrnehmen. Diese beiden Tatsachen waren ausschlaggebend, um im Rahmen der vorliegenden Arbeit ein bionisches Modell für menschenähnliche Wahrnehmung zu entwickeln. Dieses Modell beruht auf neurowissenschaftlichen und neuropsychologischen Forschungserkenntnissen über die strukturelle Organisation und Funktion des menschlichen Wahrnehmungssystems. Ein technisches System mit menschenähnlichem Wahrnehmungsvermögen würde es erlauben, eine Vielzahl von Prozessen zu automatisieren, für die bis jetzt immer noch menschliche Beobachter und deren kognitive Fähigkeiten notwendig sind. Potentielle Anwendungsbereiche sind die Sicherheitsüberwachung von öffentlichen und privaten Gebäuden und die Beobachtung des Gesundheitszustands von Personen in Krankenhäusern oder Altenwohnheimen. Abgesehen davon würden autonome Robotersysteme und interaktive Umgebungen von effektiveren Mechanismen zur Wahrnehmung ihres Umfeldes profitieren.

Das entwickelte Modell ist im Bereich der Gebäudeautomatisierung für autonome Überwachungssysteme vorgesehen, um Objekte, Ereignisse, Szenarien und Situationen in Gebäuden zu beobachten. Zu diesem Zweck müssen Gebäude mit einer Vielzahl verschiedener Sensoren ausgestattet werden. Die Herausforderung besteht darin, Information von diesen Quellen zu kombinieren und zu interpretieren. Dafür wird ein Informationsverarbeitungsprinzip genannt *neuro-symbolische Informationsverarbeitung* eingeführt. Dieses verwendet *Neuro-Symbole* als elementare Informationsverarbeitungseinheiten. Die Verwendung von Neuro-Symbolen ist von der Tatsache inspiriert, dass Menschen in Form von Symbolen denken, welche jedoch aus einer neuronalen Informationsverarbeitung resultieren. Um Sensordaten zu verarbeiten, werden Neuro-Symbole zu einem so genannten *Neuro-Symbolischen Netzwerk* verbunden, welches eine modulare und hierarchische Struktur aufweist, die vom Aufbau des menschlichen Wahrnehmungssystems abgeleitet ist. Verbindungen und Zusammenhänge zwischen Neuro-Symbolen können aus Beispielen in einer Reihe von Lernphasen ermittelt werden. Neben der Sensordatenverarbeitung beeinflussen die Mechanismen *Memory*, *Knowledge* und *Focus of Attention* die Wahrnehmung, um zweideutige Sensorinformation behandeln und Rechnerkapazitäten auf relevante Merkmale konzentrieren zu können.

Das vorgestellte Modell wurde mit AnyLogic implementiert und erwies sich als erfolgreich bei der Erkennung aller spezifizierten Testfälle. Des Weiteren erlaubten die während der Entwicklung gewonnenen Erkenntnisse bestimmte Rückschlüsse über die Inkonsistenz oder Unvollständigkeit neurowissenschaftlicher und neuropsychologischer Modellvorstellungen. Diese beziehen sich unter anderem auf das sogenannte Binding Problem, auf die Verarbeitung und Speicherung von Wahrnehmungsbildern im Allgemeinen und Ortsinformation im Speziellen, sowie auf Stabilitätsbetrachtungen.

Preface

The topic of this thesis lies in the field of cognitive science, cognitive computing, and cognitive automation and aims to develop a bionic model for human-like machine perception based on neuroscientific and neuropsychological research findings.

To understand the intricacy of the tasks attempted by these research fields, a brief retrospect of history shall be given first: The fields just mentioned overlap with the research field of artificial intelligence (AI). About 50 years ago, in the fifties, this research field started to evolve with the aim to build intelligent machines. The first years of AI were marked by a strong optimism. It was believed that computers would soon be able to think and reason in a similar effective manner as humans do. However, at the end of the sixties, it got clear that making computers think – even on a childlike level – is an extremely complex problem. Therefore, researchers started to focus on far more simple problems like reacting to situations by using certain rules. Until today, there does not exist any technical system that can even nearly compete with the capacity and capabilities of the human mind. Within the last years, several research groups recognized that the reduced approaches often focused on in current AI projects cannot lead to technical systems with skills and capabilities comparable to human mental abilities. The AI researcher Marvin Minsky postulates that, like at the beginning of artificial intelligence research, findings how natural intelligence works should be the basis for the development of concepts for artificial intelligence [Min06]. This is the aim of the research fields of cognitive science and cognitive computing. Cognitive automation aims to automate cognitive activities that are currently performed by human operators.

At the Institute of Computer Technology of the Vienna University of Technology, in the year 2000, Prof. Dr. techn. Dietmar Dietrich formed a research group of interdisciplinary researches with the aim to model and implement functions of the human brain into technical systems. These approaches are guided by insights from neuroscience, neuropsychology, and neuro-psychoanalysis about the human brain and mind. This thesis is a part of this research project and focuses on the perceptual system and the perceptual capabilities of the human brain.

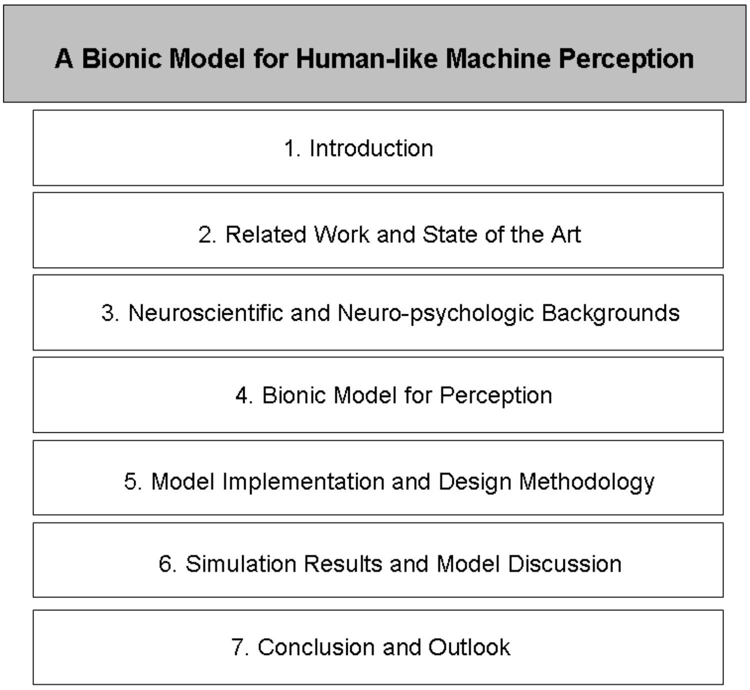

A Bionic Model for Human-like Machine Perception
1. Introduction
2. Related Work and State of the Art
3. Neuroscientific and Neuro-psychologic Backgrounds
4. Bionic Model for Perception
5. Model Implementation and Design Methodology
6. Simulation Results and Model Discussion
7. Conclusion and Outlook

Figure 1: Structure of Work

Figure 1 illustrates the structure and organization of this work graphically. In chapter 1, which is the introductory chapter, the challenges of machine perception are outlined as well as characteristics of human perception are pointed out. The aim of the work is specified in detail, and possible applications and implications are discussed. Chapter 2 gives an overview about related research projects and the state of the art in sensor fusion, neural networks, symbolic systems, and hybrid neuro-symbolic systems, which are related research fields. Chapter 3 describes neuroscientific and neuropsychological research findings about the human brain, which are the basis for model development. Chapter 4 presents the developed bionic model for human-like machine perception. In chapter 5, it is outlined how this model can be implemented in software and the design methodology is summarized. Chapter 6 discusses the developed model based on the requirements identified in chapter 1, the insights gained during the development and the implementation process, and the results of simulation experiments carried out. Additionally, a comparison and demarcation to other existing models is given. Finally, chapter 7 summarizes the issues discussed in the former chapters, gives a conclusion, recommendation for future research work, and an outlook.

Acknowledgements

Writing this dissertation would not have been possible without the support of several mentors. I would like to thank Prof. Dietmar Dietrich, Prof. Peter Palensky, and Prof. Wolfgang Kastner for their guidance, feedback, and comments. Further, I would like to thank all members of the ARS team at the ICT for fruitful subject-specific as well as humorous private discussions. Last but not least, deep gratitude goes to my parents, who always supported me on my way.

Table of Contents

Chapter 1

Introduction

"The best way to have a good idea is to have lots of ideas."

[Linus Pauling]

The work at hand is embedded in the field of cognitive sciences, cognitive computing, and cognitive automation. Cognitive science and cognitive computing are interdisciplinary research fields involving various disciplines including neuroscience, psychology, and computer science. They focus on studying the human mind and the nature of intelligence and on emulating internal information processing mechanisms of the human brain to develop next generation intelligent information and software technologies and new architectures of computing systems [Sta07]. The goal of cognitive automation is to automate cognitive activities such as situation assessment, monitoring, and fault management that are currently performed by human operators [TBM97].

The concrete goal of this thesis is to develop a bionic model for human-like machine perception based on neuroscientific and neuropsychological research findings, which shall lead to more effective artificial autonomous perception systems applicable for example to surveillance systems in buildings or autonomous robots. In section 1.1 of this chapter, the challenges for developing such a model are pointed out. Next, in section 1.2, the aim of the thesis is outlined in more detail. Finally, section 1.3 describes possible applications and implications of the model to be introduced.

1.1 The Challenges for Machine Perception

Over the last decades, automation technology has made serious progress in observing and controlling processes in order to automate them. In factory environments, where the number of possible occurring situations and states is quite limited and well known, observation and controlling of most industrial processes do no longer pose an unsolvable problem. However, the situation changes if we go from the observation of industrial processes to the detection of objects, events, scenarios, and situations in a real world environment. Here, the number of possible occurring objects, events, scenarios, and situations is almost infinite. As research from image processing, audio data processing, and natural language processing shows, for a machine, recognition of real world situations is still a task far from trivial [Vel07]. On the other hand, humans – even small children – can perceive such a "real world environment" very effectively. The human brain reconstructs the environment from the incoming stream of (often ambiguous) sensory information and generates unambiguous interpretations of the world. The challenging question is what gives humans the ability to perform these tasks and how to design machines that perform in a similarly effective and efficient way.

The goal of this thesis is to introduce a bionic model for human-like machine perception, which is inspired from the working principles and organizational structures of the perceptual system of the human brain. In this first section, after giving a definition of the terms perception and recognition, important principles and characteristics that are involved into perception as well as mechanisms that influence perception are identified, and the occurring difficulties are described when trying to transpose them to a technical model.

Defining Perception and Recognition

In literature, there cannot be found a single definition of perception but various. According to the Oxford English Dictionary, the word perception comes from the Latin *perception-, percepio*, meaning *receiving, collecting, action of taking possession, apprehension with the mind or senses.*

The term perception as it is used in this work refers to the whole process of acquiring, selecting, organizing, and interpreting sensory information. In connection with perception, there can also be found the term recognition. In literature, the terms perception and recognition are often used as synonyms. However, several authors consider them as separate processes. According to [Gol07, chapter 1, p. 7], recognition is the ability to place an object in a category that gives it meaning. This can also be expressed as the awareness that something observed has been observed before. Thus, in order to recognize something, it must be familiar. An example would be the recognition of a certain red object as tomato. [Lur73, chapter 8, p. 239] points out that in recognition, memory and cognitive processes influence perception. Objects are classified and recognized by means of prototypic exemplars stored in memory. If the process of recognition is disturbed, a subject, although he perceives the individual cues of a visual image, is unable to synthesize them into a single visually perceived entity. The subject can still clearly see an object or picture shown to him, but is unable to relate it to his past experience. He cannot recognize it. [Gol02, chapter 1] remarks that perception of an object is always associated with its recognition. This statement implies that, even if perception and recognition might be different processes, both processes are needed for a workable system. It also seems to be unfeasible to draw a sharp boarder between perception and recognition. Therefore, in this work, it is not distinguished between the terms perception and recognition per se. However, to avoid misunderstanding for readers with different backgrounds, there is only used the term perception in the following chapters. Perception is considered as the process of acquiring, selecting, organizing, and interpreting sensory information.

One term, which will be often used in this work, is the term perceptual image, for which a definition shall be given in the following. According to [Rös07, p. 20], *"perceptual images are generated due to sensing and evaluation processes reflecting the concrete changes of the environment and the organism itself, forming various images of different sensory modalities."* Adapting this definition for technical purposes, she defines a perceptual image to be *"formed by a perception system due to the sensory input via the sensory architecture. The image represents a snapshot of the environment and/or internal state of the organism or technological system and can contain other, simpler images. On the lowest abstraction level, images contain symbols that* *have been predefined."* Taking this definition as basis, in the work at hand, a perceptual image can be understood as meaningful information derived from sensor data of different modalities. Depending on the hierarchical level of processing they occur, perceptual images can be simple perceptual features like edges and lines or sounds of a certain pitch or more complex issues like objects, persons, or melodies, which are made up of these simpler images. The phrase *"the image represents a snapshot of the environment"* is not understood literally meaning that a perceptual image can only represent perceptive information arriving at one distinct moment in time, but within a certain period of time. That way, perceptual images can also represent activities being carried out over time. In the following, activities being carried out within a circumscribed area and a short period of time in the range of some tenths of a second to some seconds will be referred to as events. Activities taking some more time and consisting of sequences of such events, which can also occur in different spatial areas, will be referred to as scenarios. Another word that is used when referring to perceptual images is the term situation, which is best described as a circumstance or condition caused by the objects being present and events and scenarios going on in the environment.

Characteristics and Requirements of Human Perception

Having found a definition of perception, it shall now be identified what mechanisms and factors make human perception that powerful, and it is discussed what are the challenges of integrating these mechanisms into a technical system. Figure 1.1 shows important factors that form and influence human perception.

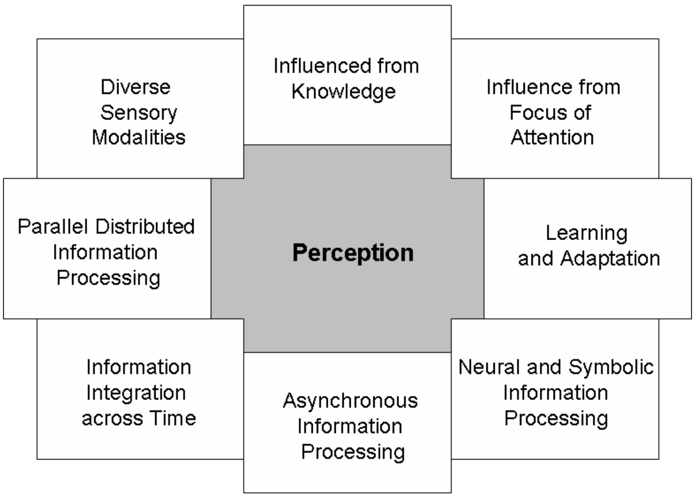

Diverse Sensory Modalities
Influenced from Knowledge
Influence from Focus of Attention
Parallel Distributed Information Processing
Perception
Learning and Adaptation
Information Integration across Time
Asynchronous Information Processing
Neural and Symbolic Information Processing

Figure 1.1: Characteristics and Requirements of Human Perception

The mentioned points were identified by analyzing neuroscientific and neuropsychological research findings about perception and binding being described in the sections 3.3 and 3.4 in more detail. As the human brain is a complex system where *everything seems to be connected to everything*, it is clear that the list might be incomplete and further research may identify additional important factors, which influence perception. Nevertheless, these characteristics and requirements are the starting point for developing the bionic model of human-like machine perception. In the following, each of the mentioned points is described in more detail together with a remark what poses the difficulties of integrating it into a technical system.

Diverse Sensory Modalities: To perceive the external environment, our brain uses multiple sources of sensory information derived from several different modalities, including vision, touch, and audition. Each of these modalities is based on information coming from a huge number of sensory receptors. The combination and integration of multiple sources of sensory information is the key to robust perception [EB04]. This is because no information processing system, neither technical nor biological, is powerful enough to work accurately under all conditions. If a single modality is not enough to come up with a robust estimate, information from several modalities can be combined to complement each other with the effect of increasing the information content. To achieve a coherent and robust percept, all different sources of information have to be efficiently merged. The main problem with data from multiple sensory sources is that different sensor types do not only provide complementary but also partly redundant, contradicting, ambiguous, and inconclusive information [VPL07]. Additionally, for events happening concurrently in the environment, there arises the question how to correctly assign diverse sensory information to different events.

Parallel Distributed Information Processing: As just outlined, for perception, information from various sources is processed. However, the perceptual system is no unitary central unit that processes all information in one step. Instead, sensory information is processed in a distributed fashion. First, information from different sensory receptors is processed separately and in parallel before being merged in later processing steps. The challenge for a technical model is to develop an architecture that allows a similar parallel distributed processing and a combination of the separate processing results to one unified perception. In neuroscience, this problem is referred to as *binding problem* (see section 3.4).

Information Integration across Time: To perceive objects, events, scenarios, and situations in an environment, single-moment snapshots of sensory information provided by different modalities is not always sufficient for unambiguous perception. The course and the succession of sensor signals over time are also of importance. Therefore, not only a binding of information across different modalities is necessary but also a binding of sensory information across time. In the course of time, the brain collects more and more information about perceptual events and finally resolves ambiguities. Again, the tempting question is how to correctly assign over time certain sensory stimuli to one event or situation when different events and situations are going in the environment concurrently.

Asynchronous Information Processing: In the brain, information is processed asynchronously. The term asynchronous information processing as used here can be understood in a physical sense and means that signals and there corresponding characteristics arrive at different points in time. In perception, asynchronous information processing already starts at the lowest sensory levels, because one event happening in the environment may not trigger sensory receptors of different modalities absolutely concurrently. Additionally, information processing and transmission of different sensor data may take different amounts of time. The research question that has to be answered is how it can be made feasible to process the data from different sources arriving asynchronously.

Neural and Symbolic Information Processing: In the human brain, perceptual information from different modalities is processed by interacting neurons. However, humans do not think in terms of action potential and firing nerve cells but in terms of symbols. Examples for symbols are objects, characters, figures, sounds, or colors used to represent abstract ideas and concepts. The challenge for emulating information processing of the brain is how to come from sensory stimuli via neural information processing to symbolic representations and how associations between symbols result in new symbolic representations.

Learning and Adaptation: The perceptual system of the human brain is not fully developed at birth. Although certain patterns need to be predefined by the genetic code, lots of concepts and correlations concerning perception are learned only during lifetime. The challenging question for a technical model is what needs to be predefined before system startup, what can be learned from examples and experience, and how this learning can be preformed.

Influence from Focus of Attention: According to the hypothesis of focused attention, what we see is determined by what we attend to. At every moment, the environment presents far more perceptual information than can be effectively processed. Attention can be used to select relevant information and to ignore irrelevant or interfering information. Instead of trying to process all objects simultaneously, processing is limited to one object in a certain area of space at a time. An important question is how and on what level attention should interact with perception.

Influence from Knowledge: Perception is facilitated by knowledge. Prior knowledge is often required for interpreting ambiguous sensory signals. Much of what we take for granted as the way the world is – as we perceive it – is in fact what we have learned about the world – as we remember it. Much of what we take for perception is in fact memory. A fundamental question for a technical model is how knowledge can be represented and how and on what level interaction with sensory perception should take place.

Looking at the points described above, the challenges for developing a bionic model for human-like machine perception lie in specifying different "functional systems" and mechanisms involved in perception, in representing information to be processed in these systems, and in merging information coming from different systems.

Taking the perceptual system of the human brain as archetype for model development and acquiring neuroscientific and neuropsychological research findings, the problem that has to be faced is that there does not exist a complete, unitary model for perception neither in neuroscience nor in any other area of brain research. There exist many different, sometimes contradicting theories and many blind spots within these theories. This makes it difficult to derive a technical model from the biological archetype. On the other side, trying to develop such a model can also be seen as chance to show up weak points of existing brain theories.

1.2 Aim of this Work

Having defined important mechanisms and functions of perception as well as identified problems and challenges of integrating them into technical systems in section 1.1, there shall now be declared the concrete aim of this work.

The central goal of the thesis is to develop a bionic model for human-like machine perception. This model shall make it possible to build technical systems, which are able to perceive objects, events, scenarios, and situations in a real world environment. Such systems would be very valuable for automatic surveillance systems – for example in buildings – or for autonomous robots.

To perceive objects, events, scenarios, and situations, a large number of sensors of different types is required. The challenge that has to be taken up for such a "real world perception" is the merging and interpretation of the sensory data from these various sources. To solve this problem, a bionic approach is used. The chosen approach bases on neuroscientific and neuropsychological research findings about information processing in the perceptual system of the human brain.

In section 1.1, there were introduced characteristics of and requirements for human perception. These characteristics and requirements shall be the starting point for model development. Therefore, mechanisms shall be found how to

- merge information from a large number of sensors of diverse sensory sources
- perform parallel distributed information processing
- bind sensory information across time
- evaluate how asynchronous information processing can be made feasible
- get from neural to symbolic information processing
- learn correlations between data
- restrict and facilitate information processing by focus of attention
- integrate knowledge into the perception process

As far as possible and known, these mechanisms shall follow the principles of how information is processed in the brain. Neuroscience and neuropsychology present important insights into the theory of perception. However, as already pointed out in section 1.1, these findings are not sufficient for a technical realization. If neuroscience and neuropsychology do not provide an answer to certain problems, the model has to be supplemented by conventional engineering methods.

For the model, automatic surveillance tasks in buildings are envisioned as first application. The model shall be capable of handling sensory information of different sensor types like video data, audio data, tactile information, etc. typically used for such tasks. However, as image processing and audio data processing are huge and complex research fields themselves, developing new image processing and audio data processing mechanisms is out of scope of this thesis. Instead, the model has to be designed in a way that existing image processing and audio data processing methods can be integrated. Information derived from image and audio data processing shall be merged and interpreted together with information from other sources following neuroscientific and neuropsychological principles.

Besides presenting a model for human-like machine perception, it shall also be considered how such a model can be realized and implemented.

1.3 Possible Applications and Implications

As outlined in section 1.2, the aim of this work is to develop a model for human-like machine perception. The field of machine perception concerns the building of machines that sense and interpret their environments [Nev82]. For automatically perceiving objects, events, scenarios, and situations in a real world environment, a huge amount of diverse sensory information has to be processed, condensed, and interpreted to result in “perceptive awareness” of what is going on in the environment. Today's information processing systems are barely capable of handling this task. The aim of the model is to provide methods to handle such problems.

By implementing these information processing methods into technical systems, many tasks could be automated wherefore today human observers and their cognitive abilities are still necessary. This would help to pare down personnel for monotonous observation tasks. Valuable applications would be the surveillance of public and private buildings for safety and security reasons and the increase of comfort of the occupants. The automatic observation of the activities and the state of health of persons in retirement homes and hospitals to detect critical situations would be a further sensible utilization. This would allow it to economize nursing staff. Different scenarios are conceivable to be perceived like e.g., that an elderly person has collapsed and cannot get up any more, or that a confused and disoriented person does not find his room or leaves the building unattended. Such systems could also allow elderly people to live longer independently in their own homes.

Surveillance in buildings is not the only possible application for the model. An artificial perceptual system performing similarly efficiently like humans do would also be valuable for a range of other applications. E.g., for autonomous robots, which are equipped with lots of sensors to navigate in their environment and to interact with it, in many situations, a more effective and efficient method of sensor data processing and interpretation is very desirable. Interactive environments are a further application domain that would benefit from such mechanisms to perceive their surrounding in order to interact with people.

Considering the project carried out in this work not from the technical but from the neuroscientific point of view, designing a model of the perceptual system of the human brain, which actually has to be technically realizable, can identify existing weak points and gaps in the underlying brain theories, because, unlike neuroscientific and neuropsychological models, functioning technical models need to be complete and consistent.

Chapter 2

Related Work and State of the Art

"The next best thing to being clever is being able to quote someone who is."

[Mara Pettibone Poole]

The work described in this thesis is related to several research fields. At the beginning of this chapter, previous research work on which the thesis bases and current projects it is connected to are introduced (see section 2.1). Afterwards, important associated research fields and latest developments in these areas are described. In section 2.2, there is given an overview about the research domain of sensor fusion. In section 2.3, neural networks, symbolic systems, and hybrid neuro-symbolic approaches for the purpose of information processing are outlined.

2.1 Project Environment

This thesis was written at the Institute of Computer Technology (ICT)[1] of the Vienna University of Technology. Several years ago, under the supervision of Prof. Dr. techn. Dietmar Dietrich, an interdisciplinary operating team of scientists was formed at the ICT, which attended all its focus on developing next generation intelligent automation systems [DS00]. One application domain for these systems lies in the field of building automation. Today, building automation is mainly concerned with simple monitoring of the environment (e.g., temperature) and adjusting it to predefined value ranges targeting comfort and energy preservation. However, as outlined in [PP05], [PLD05], and [DLP^{+}06], in future, this will shift towards applications like safety and self-learning environment control, and more and more sensory information will be available for processing. Existing approaches will be challenged by this abundant amount of data and the way in which it shall be responded to them. There will be a need to introduce new concepts for handling the demands of the upcoming future. A second application is designated to the field of autonomous agents. Autonomous agents are systems, which are situated in an environment and have to sense their environment and act on it in pursuit of their own agenda. The agents act on the environment to change it and therefore influence what is sensed in later stages. Again,

[1] www.ict.tuwien.ac.at

perceiving complex environments and reacting on them adequately or "intelligently" challenges existing approaches in this domain.

Inspiration for designing such intelligent automation systems is taken from biology and particularly from human brain research. This section describes research projects in this field carried out at the ICT until now. Each of these projects is described briefly. A special focus is put on the description of the so-called ARS-PC project, which influenced the model developed in this thesis most. A demarcation of the perceptual model introduced in the thesis at hand to the former ARS-PC model as it will be described in this section can be found in section 6.3.1.

2.1.1 The Smart Kitchen Project

The first attempts to apply bionic concepts to building automation were carried out in the Smart Kitchen Project[2], which started in 2000 [SRT00, DRT+01]. The aim of the project was to achieve "perceptive awareness" of what is going on in a building and to react adequately on perceived situations [TF03, Rus03]. A number of small, inexpensive devices were networked to implement functions for increasing comfort, security, safety, and better energy management in the home.

The project got its name from the realization and testing environment of the system – the Smart Kitchen. In the course of the Smart Kitchen project, the institute's kitchen was equipped with different sensors and actuators. The developed model for information processing was based on the idea of the ISO/OSI 7 layer reference model and modular distributed systems and attempted to include biologic concepts [TDDR01, TFR02]. However, it is questionable how much of the model can really be regarded as a bionic instead of a pure engineering approach.

2.1.2 The Project ARS

The follow-on project of the Smart Kitchen project was the project ARS[3], which started in the year 2003. The abbreviation ARS stands for Artificial Recognition System. Similar to the Smart Kitchen project, the aim of the project ARS is to build systems, which are capable of perceiving their environment and reacting adequately on situations going on in the environment. Envisioned applications are automatic surveillance systems in buildings and autonomous agents. Concept development is guided by research findings about the human brain and the human mind. Due to the complexity of the task, the project was split into two sub-projects: the projects ARS-PC and ARS-PA.

The abbreviation PC stands for PerCeption. The aim of the PC part of the project is to perceive objects, events, scenarios, and situations in an environment. The developed model of perception is inspired from neuroscientific and neuropsychological research findings.

The abbreviation PA stands for PsychoAnalysis. The aim of the PA part of the project is to take decisions how to react adequately on certain objects, events, scenarios, and situations perceived in the environment. The developed model of perception is inspired from neuro-psychoanalytic research findings.

The combination of the ARS-PC and ARS-PA approaches in later stages of development shall result in a model for a system able to perceive real world sensory information, to evaluate these data, and to take decisions based on these evaluations.

[2] http://smartkitchen.ict.tuwien.ac.at/project/project.html

[3] http://ars.ict.tuwien.ac.at/

The ARS-PC Model

For the ARS-PC model developed until now, the application is targeted in the field of building automation for automatic surveillance systems. Therefore, relevant information has to be extracted from a huge amount of data coming from various sensor types [PP05]. Building on research results from the Smart Kitchen project, a layered model for sensor data processing was introduced [PLD05, Pra06, PDHP07]. According to this model, sensor data are processed by bottom-up information processing in three layers to perceive different scenarios going on in a building. The layers are labeled as *micro symbol layer*, *snapshot symbol layer*, and *representation symbol layer* (see figure 2.1). In these layers, information is processed in terms of symbols, which are called *micro symbols*, *snapshot symbols*, and *representation symbols*. They all exist simultaneously, but on different levels. A symbol is seen as a representation of a collection of information [Pra06, Göt06]. Since learning is not addressed in the model until now, symbols and correlations between symbols are predefined. This means that the system is only capable of recognizing known information patterns of the sensors.

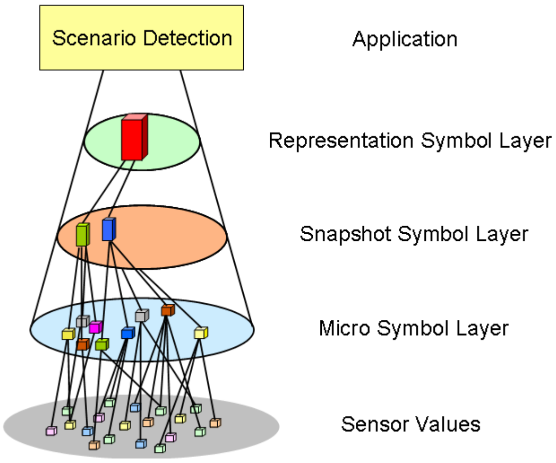


Figure 2.1: Layered Structure of the ARS-PC Model

Symbols can be created, their properties can be updated, and they can be deleted. In figure 2.1, symbols are shown as cuboids of different size, indicating that their level of sophistication increases with each layer. The number of symbols is different at each layer. At the lowest layer, there occur a large number of micro symbols. At the representation layer, there exist only a few symbols, where each symbol represents a lot of information of a higher quality. The three types are defined as follows:

Micro Symbols: Micro symbols are formed from sensory input data. They present the basis of the symbol alphabet and bear the least amount of information. Similar to the many different sensations that the human brain has to process every moment, a micro symbol is created from a few single sensor inputs at a specific instant of time. Micro symbols are created whenever the real world changes and this change causes sensors to trigger. Micro symbols have an event-like character and exist for one instant. In the envisioned application, micro symbols will for example be created delivering basic information like where movements and objects have been detected.

Snapshot Symbols: A group of micro symbols is combined to create one snapshot symbol. These symbols represent a part of the world at a certain moment of time. The combined snapshot symbols represent how the system perceives the world at a given time instant. Whenever the system perceives a situation or an object of interest, it creates an according snapshot symbol. However, it is of utmost importance that snapshot symbols are solely created from information that represents the current state of the outside world. Put in other words, the only information allowed is provided either by the presence of micro symbols or the absence of specific micro symbols. This results in an increased creation of snapshot symbols as there are no associations between new snapshot symbols and previously created ones. For example, if the system perceives a person moving around, multiple symbols at different positions and with different timestamps are created. Each of these symbols only exists in the single instant of its respective detection. Establishing associations between these symbols happens in the next symbol level.

Representation Symbols: The third level of symbolization is the representation of the world. Similar to snapshot symbols, representation symbols are used to present what the system perceives. The fundamental difference is that representation symbols are created and updated by establishing associations between snapshot symbols. Thereby, the representation level contains not only the information how the world is perceived at the current instant but also the history of this world representation. Compared to the lower levels of symbols, there exist only a few representation symbols, and these are seldom created or destroyed. Only their properties are updated regularly. On the representation level, the system has information about the current state of the world together with the history of recent events. Based on the snapshot symbols, the system utilizes all currently available perceptions to create a consistent and continuous representation of the environment. Following the example mentioned above, the representation level is supposed to hold only one person symbol as long as there is only one person physically present. All occurrences of snapshot symbols for this person are – if possible – associated to one representation symbol. As the positions of the different snapshot symbols vary, the representation symbol experiences a series of updates. It is important to note that the world representation does not hold the entirety of all sensory information available but just what is defined as relevant. If for example a person walks around, the world representation does not present information at which exact positions the person has placed its feet. Rather than that, it presents just a position for this person, which may be more or less accurate. Representation symbols are the first level used by applications to obtain information about the world. In contrast, the snapshot symbol layer and the micro symbol layer are not used by applications since the applications represent higher cognitive functions that operate on the resulting set of symbols.

The representation layer can be regarded as the interface to applications. Applications are required to monitor the world representation in order to obtain the information needed to fulfill their specific tasks. The ARS approach relieves applications from handling large amounts of sensory information and provides a condensed and filtered composition of all this information in a highly reusable way. When an application is running, it searches the existing world representation for scenarios that the application knows (e.g., an elderly person has collapsed on the floor). The events that are required for the scenario to take place can be found on the representation level. Therefore, the application augments the representation by noting that it has found a scenario. It does so by creating a *scenario symbol*. This makes it possible to study the output of applications later. Additionally, an application can create higher-level scenarios by linking together lower-level scenarios of other applications. That way, the hierarchy can be even further extended by having lower-level applications looking for simple scenarios and higher-level applications using these scenarios to find more complex scenarios.

To validate the developed model, again the institute's kitchen was taken as test environment [Göt06]. The kitchen was equipped with about hundred sensors of the following types: tactile floor sensors, motion detectors, door contact sensors for the entrance door and the fridge door, one camera, and one shock detector to indicate if the kitchen's coffee machine was activated. In a first implementation, from these sensor values, on the micro symbol level, it was detected if an item's operation status changed, if an object occurred, if a movement occurred, if a face occurred, and if a footstep occurred. On the snapshot symbol level, it was derived from these micro symbols if something entered or left the room, if something used an item, if something walked, and if a person was perceived. The only symbol existing at the representation symbol layer was the symbol "person". This symbol was specified in more detail by properties it comprised. The values of these properties were derived from information coming from the snapshot symbols. In the first implementation, the system was tested only with one scenario, which was called the "coffee detection scenario". In this scenario, the system perceived if a person entered the room to make coffee and whether this coffee was with or without milk. It was determined by hard-coded rules (if-then-rules) what higher-level symbols to create from which lower-level symbols [Göt06]. In a further step, the implementation was extended by introducing more different scenarios to the system and using slightly different lower-level symbols. There were defined the following scenarios: "meeting", "person makes coffee", "person manipulates object", "child near hot stove", and "child makes coffee". In this implementation, the formulations what lower-level symbols correspond to what higher-level symbols were based on fuzzy rules [Ric07]. The model proved to be suitable to detect all defined scenarios.

The model just introduced is based partly on neuroscientific and neuropsychological research findings and partly on engineering methods. The concepts taken from neuroscience are the hierarchal processing of information in different layers and the fact that information is processed in terms of symbols. [Bur07] extended this model by introducing additional neuroscientific and neuropsychological concepts to the model. He adapted symbolic information processing in a way that is more compliant with the neuroscientific model of information processing in the perceptual system of the human brain as described by [Lur73] (see also chapter 3). In [Pra06], it was defined that micro symbols and snapshot symbols can only contain information perceived in one instant of time. As this rule resulted in difficulties in the implementation, [Bur07] softened this rule and allowed the processing of sensory information within a certain time period. Additionally, he suggested to process sensor data from the same sensor type first separately and to combine this information only later with information derived from other sensor types, which is in accordance with the neuroscientific archetype. However, in the implementation, it was not strictly complied with this design rule.

Additionally, in [Bur07] and [BLPV07], a technical model for emotions was introduced, which can influence perception based on the model of [Pan98]. Therefore, so-called e-systems were suggested, which correspond to basic emotions in the mammalian brain. However, the introduced technical model for emotions was not subject to system implementation as it turned out to be difficult to define useful emotions and correlations between emotions and symbols for the control systems of buildings.

As already described, in a first instance, the test platform for evaluating the different generations of ARS-PC models was the institute's kitchen, which was equipped with about 100 sensors of different types. However, in the course time, this test environment turned out to allow only limited testing due to its spatial restriction and the relatively small number of sensors. High costs and assembly effort did not allow it to enlarge the physical test environment. To overcome this problem, there is currently a simulator under development, which shall allow it to generate sensor values based on a virtual environment [HPB05, Har08]. The simulator is developed to simulate sensor values in order to perceive scenarios in a virtual office environment. The reason for simulating the sensor values is on the one hand the already mentioned cost reduction for testing in comparison to real physical installations. On the other hand, the simulator allows it to evaluate which sensors are necessary to detect scenarios most effectively and efficiently and where they should be mounted. There were already performed the first successful tests with this simulator [Bur07].

To sum up, until now, the ARS-PC model is a model for bottom-up information processing of sensory data for scenario detection. Bottom-up information processing means that the starting point for processing are sensor values, which are then processed in several steps. Information processing is performed in terms of symbols arranged in three layers. Associations between symbols of different hierarchical layers are defined by rules and are not subject to learning. The model proved to be successful at least for the detection of a limited number of scenarios. What has to be criticized about the model is the fact that it claims to rely on neuroscientific, neuropsychological, and even neuro-psychoanalytical research findings. However, in [Pra06], who introduced the first version of the model, besides the fact that information is processed hierarchically in different layers and that information is processed in terms of symbols, no such concepts are used for the model. The design decision that symbols up to the snapshot symbol level can only contain information perceived at one instant of time even contradicts neuroscientific findings, as already in the lowest areas of the visual cortex there can be found neurons that respond to movements of objects [Gol02]. In [Bur07], who made certain changes and extensions to the model, it was suggested to process information in a modular hierarchical fashion, which is in accordance to neuroscientific research findings. However, it was not strictly complied with this concept in the implementation of the model.

The ARS-PA Model

The aim of the ARS-PA project is to develop a technical model, which supports the decision making process of an intelligent autonomous system. By making certain decisions, appropriate actions shall be chosen without external supervision of a human operator [DLP$^+$06, PLC07, Rös07, Pal08]. The decision making process is not a straightforward approach but is multi-layered and contains a number of feedback loops. The bases for the ARS-PA model are neuro-psychoanalytic research findings about the human mental apparatus. The model includes concepts like emotions, drives, episodic and semantic memory as well as Sigmund Freud's Ego-Superego-Id personality model. According to the model, decisions are taken based on perceived images of the world, internal states of the system, which are represented by concepts corresponding to emotions and drives, and memory of different kinds.

There have already been made first attempts to apply this model to building automation. However, it turned out to be difficult to apply concepts like emotions, drives, etc. to a building, which consists in fact of dead matter and has no living body with internal states that need to be represented by emotions and drives. Even when ignoring needed embodiment of emotions and drives, nevertheless, there arose the question what emotions and drives could be useful for e.g., in a kitchen. Therefore, to be able to test the concepts developed in the ARS-PA model, the so-called *Bubble Family Game* was developed [DLP$^+$06]. The Bubble Family Game is a virtual simulated environment with virtual autonomous agents called *Bubbles*. These agents can navigate through a two dimensional world. They can perceive their environment through simplified sense organs. They can detect the presence of other agents, energy sources, and obstacles. The current goal of the agents is to survive in the environment by finding energy sources and filling up their energy level. Agents compete in different groups and try to find an optimum strategy in diverse (unknown) situations [Rös07, DGLV08, LZD$^+$08]. A point that has to be criticized about the ARS-PA model is the fact that it mixes concepts of different brain research areas and integrates them at the same level of abstraction. Sigmund Freud's Ego-Superego-Id personality model and drives are pure psychoanalytical concepts [Fre15, Fre23]. In contrast, episodic and semantic memory are terms used in neuropsychology [Tul83]. The concept of emotions occurs in both disciplines. However, in neuropsychology, they refer to processes excluding consciousness, whereas feelings are conscious experiences of emotions [Dam94]. In psychoanalysis, the terms feelings and affects are often used as synonyms for emotions [LP73].

2.1.3 The PAIAS Project

The PAIAS (Psycho-Analytically Inspired Automation System) project, which started in autumn 2007, can be regarded as a next generation approach of the ARS-PA project. Its main focus is on improving the existing ARS-PA concepts. The research shall discover and define the elements that are needed between the two borders of perception and action and to formulate a technically feasible and implementable model of the human mental apparatus. The starting point for model development is Freud's psychoanalytic Ego-Superego-Id personality model. In contrast to former approaches, the most important design premise is to develop the model as a top-down approach starting with the interacting modules Ego, Superego, Id, and perception-consciousness and dividing these modules into further interacting sub-modules.

2.1.4 The Project BASE and the Project SENSE

From the project ARS, several side-projects split off – namely the projects BASE and SENSE. These projects exploit results and mechanisms of the ARS-PC project. However, these projects focus less on sticking to neuroscientific and neuro-psychoanalytic bases but more on practical and feasible technical realizations.

The Project BASE

The project BASE (Building Assistance system for Safety and Energy efficiency) introduces a self-learning system that can learn what is regarded as normality and will alert in the case of deviations. This project was conducted in cooperation with the ARCS Seibersdorf Research GmbH (Geschäftsbereich Informationstechnologien) and was running from 2004 until 2006. In [BSR06], [LBP$^+$07], and [Bru07], it is investigated how statistical methods can be applied to building automation systems to recognize erroneous behavior and to extract semantic information and context information from sensor data. Therefore, a hierarchical model structure based on hidden Markov models is proposed. The lower levels of the model structure are used to observe the sensor values whereas the higher levels provide a basis for the semantic interpretation of what is happening in a building.

The Project SENSE

The project SENSE (Smart Embedded Network of Sensing Entities) was started in September 2006 and is founded by the 6th European Framework Program. The SENSE project will develop methods, tools, and a test platform for the design, implementation, and operation of smart adaptive wireless networks of embedded sensing components [PF07, BKVH08]. The network, which is an ambient intelligent system, shall adapt to its environment, create ad-hoc networks of heterogeneous components, and deliver reliable information to its component sensors and the user. Different sensors (video cameras and microphones) cooperate to build and maintain a coherent global view from local information. Newly added nodes shall automatically calibrate themselves to the environment, and share knowledge with neighbors. The network shall be self-organizing based on the physical placement of nodes and is scalable due to local information processing and sharing. The test platform will be installed in an airport to yield real data and performance goals from a realistic test environment.

2.2 Sensor Fusion

A research field related to the topic of this thesis is the research field of sensor fusion as both have – at least to a certain extend – the same aim, which is the combination of sensor data from diverse sources (and sometimes also other information sources) to achieve a "better perception" of the environment.

There can be found various definitions of sensor fusion differing slightly in the meaning. According to the definition of [Elm02], sensor fusion is *"the combining of sensory data or data derived from sensory data in order to produce enhanced data in form of an internal representation of the process environment. The achievements of sensor fusion are robustness, extended spatial and temporal coverage, increased confidence, reduced ambiguity and uncertainty, and improved resolution."*

The research field of sensor data fusion is relatively recent and dynamic. A standard terminology has not yet been adopted. There have been widely used the terms "sensor fusion", "sensor integration", "data fusion", "information fusion", "multi-sensor data fusion", and "multi-sensor integration" in technical literature to refer to a variety of techniques, technologies, systems, and applications, which use data derived from multiple information sources [van98, BLS06].

[BRG96], [van98], and [Elm02] enumerate the following advantages expected from fusion of sensor data from heterogeneous or homogeneous sensors:

- Extended spatial and temporal coverage
- Improved resolution
- Completeness
- Robustness and reliability
- Increased confidence
- Reduced ambiguity and uncertainty
- Robustness against interference
- Reduced system complexity (at higher abstractive levels)
- Deduction of new meaning or qualities

Applications for fusion are various and range from measurement engineering and production engineering over robotics and navigation to medicine technology and military applications. Examples for application can be found in [LK89], [LYS02], [BLS06], and [RL07].

Data for sensor fusion can come from data of one single sensor taken from multiple measurements subsequently at different instants of time, from multiple sensors of identical types, or from sensors of different types.

Concepts for Fusion

Sensor fusion is generally based on the combination of redundant or complementary information. Among others, [van98], [Elm02], and [RL07] distinguish three types of sensor data fusion, which are not mutually exclusive: complementary fusion, competitive fusion, cooperative fusion.

Complementary fusion is the fusion of incomplete sensor measurements from several disparate sources. Sensor data do not directly depend on each other, but are combined to give a more complete image of a phenomenon under observation.

Competitive fusion is the fusion of redundant sensor measurements from several sources. Each sensor delivers independent measurements of the same property. Competitive sensor configurations are also called redundant configurations.

Cooperative fusion uses the information provided by independent sensors to derive information that would not be available from the single sensors. An example for cooperative sensor fusion is stereo vision. In contrast to complementary and competitive fusion, cooperative fusion generally decreases accuracy and reliability.

Levels of Abstraction

Fusion of data can be performed at different levels of abstraction. According to [RL07], sensor fusion can be performed on three levels of abstraction: on the signal level, the feature level, and the symbol level.

On the signal level, signals of particular sensors are combined directly. A precondition for a fusion on this level is the comparability of measurement signals. On the feature level, signal descriptors (features) derived from signals are combined into meaningful representations or more reliable features. On the symbol level, symbolic signal descriptors are combined at the highest level of abstraction. This information is often used in decision-based systems. Fusion on a higher abstractive level is in the majority of cases more efficient [RL07]. In [NGCV04], additionally to these three levels, a fourth level – the pixel level – is introduced, which is located between the signal level and the feature level. Pixel level fusion is intended to increase the information content associated to pixels of images.

Models for Fusion

Concerning models for sensor fusion, it has to be noted that sensor fusion models heavily depend on the application. Up to now, there does not exist a model for sensor fusion that is generally accepted, and it seems unlikely that one technique or architecture will provide a uniformly superior solution [Elm02]. Therefore, there exist numerous models for sensor fusion in the literature. [Elm07] gives an overview over the most common approaches for sensor fusion models and introduces an attempt to classify them. In his article, he mentions the JDL fusion model architecture, the Waterfall model, the Intelligence cycle, the Boyd loop, the LAAS architecture, the Omnibus model, Mr. Fusion, the DFuse framework, and the Time-Triggered Sensor Fusion Model. These models can be classified into three groups: abstract models, generic architectures, and rigid architectures.

Abstract models serve as a way to think of or explain an aspect of a fusion system without guiding the engineer in its implementation. Abstract models are the Waterfall model and the Boyd control loop.

Generic architectures give an outline how to implement an application, but leave open several design decisions. For example, it is not specified which operating system, hardware, communication system or database should be used. The JDL model and the Omnibus model belong to the group of generic architectures.

Rigid architectures guide the engineer well in its implementation at the cost of flexibility, because several design decisions have already been taken. New systems can be realized quickly by taking advantage of existing hardware designs, tools, and source code, but the cost of migrating a design from one rigid architecture to another is high. Examples for rigid architectures are the LAAS architecture, Mr. Fusion, DFuse, and the Time-Triggered Sensor Fusion Model.

Methods for Fusion

There have been suggested various methods for sensor fusion. According to [RL07], sensor fusion methods can principally be divided into grid based (geometric) and parameter based (numerical) approaches whereby in the case of numeric approaches, he makes a further distinction between feature based approaches (weighted average, Kalman filter), probabilistic approaches (classical statistics, Bayesian statistics, Dempster-Shafter theory of evidence), fuzzy methods, and neural approaches. In contrast, [LYS02] classifies fusion algorithms into estimation methods (weighted average, Kalman filter), classification methods (cluster analysis, unsupervised or self-organized learning algorithms), interference methods (Bayesian interference, Dempster-Shafter evidential reasoning), and artificial intelligence methods (neural networks, fuzzy logic). Similar like for the models of sensor fusion, there also does not exist one sensor fusion method suitable for all applications.

Biological Sensor Fusion

[PWM03] and [VLBD08] point out that it is well appreciated that sensor fusion in the perceptual system of the human brain is of far superior quality than sensor fusion achieved with existing mathematical methods. Therefore, it seems to be particularly useful to study biological principles of sensor fusion.

Such studies can on the one hand lead to better technical models for sensor fusion and on the other hand to a better understanding of how perception is performed in the brain. Sensor fusion based on models derived from biology is called biological sensor fusion. In literature, among others, the following attempts of biological sensor fusion are described:

[Mur96] reviews literature from biological and cognitive science about sensory integration and derives an architecture for intelligent sensor fusion systems. This so-called Sensor Fusion Effects (SFX) architecture is suited for robot navigation and incorporates a concept of two-phase sensor fusion activity. [HH92] present an approach to model biological vision with neural networks and traditional processing. The architecture has two channels: a location channel and a classification channel. The location channel searches for objects in the field of view. The classification channel learns and recognizes objects. The Accurate Automation Corporation (AAC) developed a neural network-based sensor fusion system inspired by how information from multiple sensors is fused by the central nervous system. Based on this model, a system was developed, which merges two or more sensor signals to generate a fused signal with an improved confidence of target existence and position. [Dav97] and [CR04] suggest a neural network for multi-sensory perception. This network processes auditory and visual information separately in the first layers before combining it in the next layers. In [KBS01], a mathematical model of the human perception process is presented. The proposed system's theoretical framework describes the principles of human perception as a concatenation of nonlinear vector mappings. In [HG06] and [GJ07], a concept for so-called hierarchical temporal memory (HTM) is introduced. HTM is a machine learning model that emulates some of the structural and algorithmic properties of the neocortex using an approach related to Bayesian networks.

Although there have already been introduced a number of models for biological sensor fusion, yet success of research efforts incorporating lessons learned from biology into "smart algorithms" has been limited [PWM03]. One reason therefore might be that the use of biological models in actual machines is often only metaphorical, using the biological architecture as a general guideline [KZK97].

Symbolic Processing of Sensory Information

There have been made some attempts to perform sensor fusion by transforming sensor data into symbols. Approaches to process sensor information symbolically have been described by [Pra06], [Göt06], [Ric07], and [Bur07], who suggest a layered architecture for this purpose. [JRCC03] attempt to achieve symbol grounding by adding a sensory concept to an abstract symbol.

Direct and Indirect Fusion

Fusion of sensor data from a set of heterogeneous or homogeneous sensors, soft sensors[4], and history values of sensor data is called *direct fusion*. However, there also exists *indirect fusion*. Indirect fusion uses information sources like prior knowledge about the environment and human input. Furthermore, it is possible to fuse the outputs of the former two [Elm02].

In literature, different models for such hybrid systems are described. [EB04] claim that it is impossible to reconstruct the environment "bottom-up" from the sensory information alone, and that prior knowledge is needed to interpret ambiguous sensory information. Bayesian inference is suggested to combine prior knowledge with observational, sensory evidence to infer the most probable interpretation of the environment. [Cro05] points out that many human activities follow a loosely defined script in which individuals assume roles. A layered, component-based software architecture model is proposed and illustrated with a system for real-time composition of synchronized audio-video streams for recording activities within a meeting or lecture. [GK02] mention

[4] Soft sensor – also called virtual sensor or software sensor – is a common name for software where several measurements are processed together. There may be dozens or even hundreds of measurements. The interaction of the signals can be used for calculating new quantities that need not be measured. Soft sensors are especially useful in data fusion, where measurements of different characteristics and dynamics are combined. It can be used for fault diagnosis as well as control applications [KKvW$^+$06, p. 68].

that different sources of information do not always keep the same relative reliability and that a rational perceptual system should adjust the weights that it assigns to different information sources. A Bayesian approach is suggested to understand how the reliability of different sources of information, including prior knowledge, should be combined by a perceptual system. [BAR04] exploit location information about sound signals to conclude from what source a detected sound originates. For example, a sound originating from the manipulation of dishes is likely to be detected in the kitchen near the sink. [SKSV00] describe a system for the recognition of mixtures of noise sources in acoustic input signals. The problem is approached by utilizing both bottom-up signal analysis and top-down predictions of higher-level models. [Ell96] presents a prediction-driven approach to interpret sound signals. The analysis is a process of reconciliation between the observed acoustic features and the predictions of an internal model of the sound-producing entities in the environment. [DTL^{+}03] propose a scheme where perception crucially involves comparison processes between incoming stimuli and expected perceptions built from previous perceptions.

2.3 Neural Networks, Symbolic Systems, and Hybrid Approaches

In this thesis, an information processing principle called neuro-symbolic information processing is introduced, which unifies advantages of neural and symbolic approaches. Therefore, this section shall give a brief overview about the research field of artificial neural networks, symbolic artificial intelligence, and existing hybrid approaches.

2.3.1 Neural Networks

Artificial neural networks – also referred to as connectionist systems – can be seen as simplified models of neural processing in the brain. An artificial neural network involves a network of simple processing elements (neurons), which can exhibit complex global behavior determined by the connections between the processing elements. The structure and function principle of artificial neurons is derived from biological neurons of the human brain consisting of the four basic elements of dendrites, synapses, cell body, and axon (see section 3.1).

In neural networks, knowledge is represented in a distributed form. An important issue of neural networks is learning from examples achieved by adjusting the weights of connections between neurons by certain learning algorithms during a training phase. Learning from examples allows it to apply neural networks to applications where no algorithmic solutions can be found or where a structure in existing data shall be discovered. Generally, it can be distinguished between supervised and unsupervised learning [CS97, Roj96]. Neural networks are applied to solve various information processing problems. There can be handled problems in the field of pattern classification, function approximation, prediction, etc. In literature, there have been suggested various different types of neural networks each of them being best suited for certain applications. The type most often used is the multi-layer perceptron [MCM96]. An extensive overview about important facts concerning neural networks as well as examples for their applications can be found in [Vel06].

In [Hau98], [Sch97], [Vel06], and [Pal08], there are mentioned the following advantages of neural networks in comparison to other solutions: Neural networks can learn from examples and are therefore applicable in situations where the usage of algorithmic solutions is difficult. Because of the ability to learn and to adapt, neural networks offer a certain degree of flexibility. Neural networks are able to generalize well to unseen cases and are robust and fault tolerant to exceptions, noise, and incomplete data. They can process information in parallel and therefore guarantee high performance. Neural networks can be described mathematically by matrices. This offers the possibility to simulate them on a computer very effectively.

A problem with neural networks is that their effective usage requires a certain amount of experience. There need to be determined many parameters before learning can start. This is no trivial task. For good performance, it is crucial to select appropriate input data for the neural network and to pre-process them. There also has to be chosen a suitable way to represent the output data. There has to be selected a network architecture with a certain number of nodes for the problem at hand. A suitable learning algorithm has to be chosen, and the number of training epochs has to be selected [Vel06]. Besides the problem of parameter selection, one factor often considered as disadvantage is that in neural networks, knowledge is represented implicitly by the weights between neurons. The drawbacks of neural networks lie in the incapacity to provide an explanation for the underlying reasoning mechanisms. Therefore, neural networks are considered as black box models [dGBG01].

2.3.2 Symbolic Artificial Intelligence

According to the theory of symbolic systems, the human mind is a symbol system and cognition is symbol manipulation [Fre96]. An assumption underlying most work in artificial intelligence is that intelligent behavior can be achieved through the manipulation of symbol structures representing bits of knowledge [Caw98].

Symbolic artificial intelligence (symbolic AI) concerns itself with attempting to explicitly represent human knowledge in a declarative form (i.e. facts and rules). Therefore, it is necessary to translate often implicit or procedural knowledge (i.e. knowledge and skills, which are not readily accessible to conscious awareness) possessed by humans into an explicit form using symbols and rules for their manipulation [RN07].

The design process of symbolic artificial intelligence is generally considered as a top-down process. Intelligence is viewed as computations, which in turn are viewed as rule-based manipulations on symbols [Pal08]. In symbolic AI, there has to be fixed a set or alphabet of elementary symbols which is known in advance. The alphabet is finite. The basic symbols can be combined in various ways, but not all combinations are allowed. Rules of syntax are needed to specify which combinations are valid. Besides rules of syntax, there also exist rules of semantics, which specify how the meaning of these combinations depends on the meaning of the component symbols. These syntax and semantic rules need not themselves be explicitly present. Finally, there exist rules for manipulating these symbol combinations, which derive new combinations from old [Pai07].

In [Caw98], there are distinguished three main approaches to knowledge representation in artificial intelligence: frames and semantic networks, logic, and rule-based systems.

Semantic network knowledge is represented as a graph. Nodes in the graph represent concepts, links represent relations between concepts. The most important relations between concepts are instance relations and subclass relations. However, other relations are also allowed. Subclass and instance relations can be used to derive information not explicitly represented. Semantic networks allow it to represent knowledge about objects and their relations in a simple and intuitive way.

Frames are a variant of semantic networks. They are often used to represent facts in an expert system. Information relevant to a particular concept is stored in a single entity called frame.

Logic has a well-defined syntax and semantics and is concerned with truth preserving inference. Therefore, it seems to be a good candidate to represent and reason with knowledge. The most important logic for knowledge representation is predicate logic. With predicate logic, complex facts about the world can be represented and new facts can be derived that are guaranteed true if the initial facts were true.

Rule-based systems represent knowledge in terms of a set of rules. These rules define what shall be done or concluded in different situations. A rule-base consists of a set of rules (if-then-rules), a set of facts, and an interpreter controlling the application of the rules, given the facts.

Symbolic AI had some impressive successes. Artificial systems mimicking human expertise – so-called expert systems – are emerging in a variety of fields, which constitute narrow but deep knowledge domains. Game playing programs (e.g., chess) being written now challenge the best human experts. The big advantage of symbolic systems is that their discrete knowledge representations are explicit and manipulable in an open-ended manner. However, the difficulties encountered by symbolic AI are deep, possibly irresolvable [RN07]. The main problem with symbolic AI is that it can only be used when there is complete information about the part of the world to be modeled. This problem has become known as the c*ommon sense knowledge problem* or *general knowledge problem*. While researchers were aware of the fact that in a symbolic AI system, knowledge has to be explicitly represented, they did not anticipate the vast amount of implicit knowledge we all share about the world and ourselves. Areas, which rely on procedural or implicit knowledge, such as sensory processes and motor processes are much more difficult to handle within the symbolic AI framework. In these fields, symbolic AI has had limited success [RN07]. A further problem referred to as *frame problem* is that the set of relevant features whose changes have to be tracked has to be known in advance in order neither to miss important changes nor to be forced to always evaluate every change occurring somewhere in the system [Pal08]. Beyond simple toy domains, the common sense knowledge problem and the frame problem are rarely resolvable. Furthermore, according to the *symbol grounding problem*, symbolic representations are not grounded in the system's interactions with its environment. [Har90] claims that symbolic representations must be grounded bottom-up in non-symbolic representations to give them meaning. Another problem is that symbolic systems cannot be kept effectively in tune with changing environments. Additionally, they lack generalization ability and fault tolerance. Symbolic representations and operations on them are domain-specific, restricted, static, and time-consuming. A further important point of criticism concerning symbolic AI is that its algorithms are exclusively sequential and centrally controlled.

2.3.3 Neuro-symbolic Integration

In chapter 4, there will be introduced a concept for information processing of sensor data based on so-called neuro-symbols. These neuro-symbols combine certain characteristics of neurons and symbols. Although the approach suggested in this thesis is unprecedented, in AI literature, there can already be found certain attempts to combine artificial neural networks with symbolic systems to hybrid neuro-symbolic approaches to solve diverse tasks. The motivation for these attempts as well as a short overview of existing models and their classification is outlined in the following.

Motivation for Neuro-symbolic Integration

Both neural networks – also called connectionist systems – and symbolic systems have a broad field of applications. According to [HU94], neural networks and symbolic systems are two disparate approaches to model cognitive processes and to engineer intelligent systems. Primary efforts have generally focused always only on one of these two disparate approaches. Both approaches have their specific strengths and weak points.

Connectionist systems are robust and can learn from examples. They are fault tolerant, can handle incomplete information, and are able to generalize to similar input. As they are parallel distributed systems, they also potentially provide increased speed of processing [WS00]. The drawbacks of neural networks lie in the incapacity to provide an explanation for the underlying reasoning mechanisms, wherefore they are considered as black box models [Wer98, Huy99].

Symbolic systems can explain their inference process and use powerful declaration languages for knowledge representation. They allow explicit control, fast initial coding, dynamic variable binding, and knowledge abstraction [WS00]. Problems of symbolic systems are a lack of robustness as well as the inability to handle incomplete information and to generalize. They generally fail to learn new associations between symbols and to do things on their own [Wer98, Huy99]. One basic problem with symbolic systems is the question how symbols get their meanings, because symbols and concept need to be grounded somehow in reality [Har90].

A comparison of the characteristics of neural networks and symbolic systems shows that symbolic systems have certain problems that connectionist systems seem to solve and vice versa. However, although it seems quite obvious that the weaknesses of connectionist and symbolic systems could potentially be overcome through a judicious integration of techniques and tools of both approaches, there has been only little cooperation between these two disciplines until now. By taking up the challenge to combine connectionist approaches and symbolic approaches, a new method could be developed that shows the advantages of both without suffering from their weaknesses [Huy99].

According to [dGGB02, chapter 1], the aim of neural-symbolic integration is to explore and exploit the advantages that each approach presents. Among the advantages of artificial neural networks are massive parallelism, generalization capabilities, and inductive learning. Symbolic systems on the other hand can explain their interference process and use powerful declarative languages for knowledge representation. From the perspective of cognitive neuroscience, a symbolic interpretation of an artificial neural network architecture is desirable since the brain has a neuronal structure and the capability to perform symbolic processing [Wer98]. In [WS00, chapter 1], it is pointed out that cognitive processes are not homogenous but a wide variety of representations and mechanisms are employed. Some parts of cognitive processes are best captured by connectionist models, while others by symbolic models. Therefore, in cognitive modeling, there exists a need for "pluralism", which leads to the development of hybrid models.

[Wer98] points out the following areas of research, which are interested in the design of hybrid systems:

- Integration of symbolic and neural techniques for
  - integrating techniques for language and speech processing
  - integrating different modes of reasoning and inferencing

  - combining different techniques in data mining
  - integration for vision, language, and multimedia
  - hybrid techniques in knowledge-based systems
  - combining fuzzy/neuro techniques
  - neural/symbolic techniques and applications in engineering

- Exploratory research in

  - emergent symbolic behavior based on neural networks
  - interpretation and explanation of neural networks
  - knowledge extraction from neural networks
  - various forms of interacting knowledge representations
  - dynamic systems and recurrent networks
  - evolutionary techniques for cognitive tasks (language, reasoning, etc.)

- Autonomous learning systems for cognitive agents that utilize both neural and symbolic learning techniques

Classification of Integrated Neuro-symbolic Systems

The research field of neuro-symbolic integration is quite recent. Up to now, there does not exist a model for neural-symbolic integration that is generally accepted. The model heavily depends on the application. In [Hil97], [WS00, chapter 1], and [dGBG01, chapter 1], a classification of integrated neuro-symbolic systems as given in figure 2.2 is proposed. According to this classification scheme, neuro-symbolic integration systems can roughly be divided into unified strategies and hybrid strategies. Unified strategies try to attain neural and symbolic capabilities by using neural networks alone. Hybrid strategies combine neural networks with symbolic models such as case-based reasoning systems, expert systems, and decision trees.

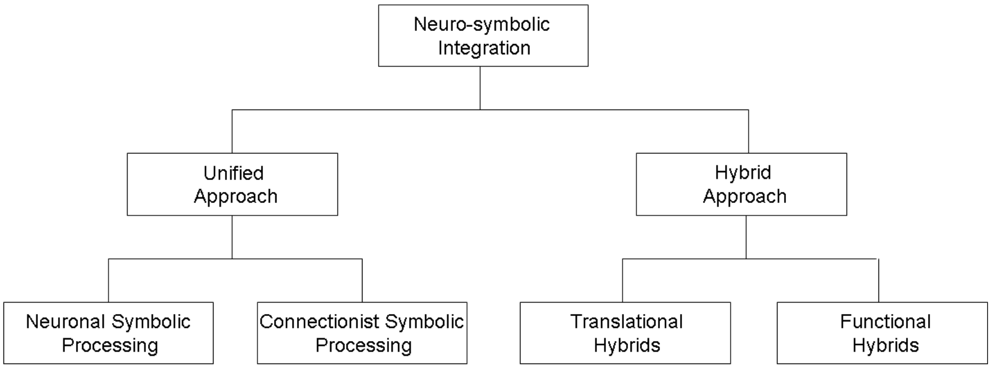

Neuro-symbolic Integration
Unified Approach
Hybrid Approach
Neuronal Symbolic Processing
Connectionist Symbolic Processing
Translational Hybrids
Functional Hybrids

Figure 2.2: Classification of Neuro-symbolic Integration Systems

Unified strategies base on the claim that there is no need for symbolic structures, because full symbol processing functionalities emerge from neural structures alone. Unified strategies can be further subdivided into neuronal symbolic processing and connectionist symbolic processing.

The objective of *neuronal symbol processing* is to model the brain's high-level functions. This approach is a bottom-up approach with the biological neuron as mandatory starting point. As the ambitions of this research field are high, it is still immature, and it may take some time before real world applications can even be envisaged.

In contrast, *connectionist symbol processing* (or neural symbol processing) lays no claim to neurobiological plausibility. Here, artificial neural networks are used as basic building blocks to build cognitive architectures capable of complex symbol processing. Connectionist symbol processing can be further divided into localist, distributed, or combined localist/distributed architectures [Hil97, WS00, chapter 1]. Localist architectures contain one distinct node for representing each concept. Distributed architectures comprise a set of non-exclusive, overlapping nodes to represent each concept. To incorporate prior knowledge into a system, it is generally easier to use localist models, because their structures can be made to directly correspond to that of symbolic knowledge. In contrast, neural learning usually leads to distributed representations.

Hybrid approaches rest on the assumption that the full range of cognitive and computational powers can only be attained by synergetic combination of neural and symbolic models. Here, it is distinguished between translational and functional hybrids.

Translational hybrids (or transformational models) represent an intermediate class between unified and functional hybrids. Similar to unified models, they rely only on neural networks as processors. However, they can start from or end with symbolic structures. Their objective is to transform symbolic structures into neural networks before processing or to extract symbolic structures from neural networks after processing. Most often, the symbolic structures used are rules. The key point is that symbolic structures are not processed in translational systems.

Functional hybrids comprise complete symbolic and connectionist components. Besides neural networks, they comprise both symbolic structures and their corresponding processors (rule interpreter, parsers, case-based reasoners, theorem provers, etc.). Functional hybrids achieve effective functional interaction and synergy among the combined components. Functional hybrids can be subdivided depending on their integration mode, which refers to the way in which the neural and symbolic components are configured in relation to each other. [WS00, chapter 1] and [Hil97] distinguish between loosely coupled architectures, tightly coupled architectures, and fully integrated architectures considering the degree of integration as a quantitative criterion. Loosely coupled architectures have separate symbolic and neural modules. The control flow is sequential. Only one module is active at any time. Processing has to be finished in one module before the next module can begin. Communication between modules is unidirectional. Tightly coupled architectures have separate symbolic and neural modules. Control and communication takes place via common shared internal data structures in each module. The common data structures allow bidirectional exchange of knowledge between modules. Fully integrated architectures show no discernible external difference between symbolic and neural modules. The modules have the same interface and are embedded in the same architecture. The control flow can be parallel and communication can be bidirectional between modules.

Potentials of Neuro-symbolic Systems

Neuro-symbolic systems are often reported in connection with the implementation of human cognitive capabilities, especially natural language processing [SA97, dGBG01]. In recent years, the research field of neuro-symbolic integration has seen a remarkably active development. However, neuro-symbolic models have often been criticized to have no common foundation, to be purely empirical and only designed to solve one precise symbolic or connectionist limitation, to be only applicable to elementary toy problems, and to be not able to deal with real world applications including perceptual or behavioral components. Nevertheless, neuro-symbolic systems begin to be a mature domain, and it has been shown that they can often perform better than purely symbolic or connectionist approaches [SA97, chapter 20].

Chapter 3

Neuroscientific and Neuropsychological Backgrounds

"Brain: an apparatus with which we think we think."

[Ambrose Bierce]

The human brain is a highly complex system. Different research areas are involved in its research, including pedagogics, psychology, neuroscience, psychiatry, psychoanalysis, neuro-psychoanalysis, and cognitive sciences. Different disciplines have diverse, sometimes also conflicting theories and models about the structural organization and function of the human brain. The concepts developed in the work at hand will be guided by the research findings of neuroscience and neuropsychology. Neuroscience is devoted to the scientific study of the nervous system. Neuropsychology is an interdisciplinary branch of psychology and neuroscience that aims to understand the cerebral organization of human mental processes and how the structure and function of the brain relate to specific psychological processes and behaviors [Lur73]. In this chapter, an overview about research findings relevant for the model to be developed is presented. First, in section 3.1, a distinction between the terms "brain" and "mind" is drawn, and the basic information processing units of the brain and the mind are discussed. In section 3.2, the functional units of the brain are discussed. Section 3.3 describes the perceptual system of the brain. Finally, section 3.3 gives an overview about latest research findings concerning the so-called binding problem to answer the question how information being processed in different areas of the brain is merged.

3.1 Basic Information Processing Units of the Brain and Mind

This section describes the basic processing units of the brain and the mind. However, before proceeding to this description, the difference between the terms "brain" and "mind" is pointed out. The brain is the physical, biological matter contained within the skull, responsible for electro-chemical neuronal processes. In contrast, the mind consists in mental attributes, such as beliefs, desires, perceptions, and so on. There exist scientifically demonstrable correlations between mental events and neuronal events. Information processing in the brain as well as in the mind takes place by information exchange of many quite simple information processing units. On the physiological basis of the brain, information processing is regarded as being performed by neurons[1]. On the more abstract level of the mind, information processing can be considered in terms of symbols.

Neural Information Processing

Brain researchers commonly agree that information processing in the brain is performed by interacting neurons or nerve cells. The brain consists of approximately 100 billions of such active elements, which are massively interconnected. These neurons can be regarded as basic processing units of the brain. Neurons are specialized cells that can generate an electro-chemical signal. The basic function of a neuron is to transfer information. There exist different types of nerve-cells. However, the basic structure of a nerve cell and its function principle is always the same.

A neuron collects input information via dendrites. This information is received from other nerve cells through specific points of contact – the synapses. The axon of the neuron links up with a dendrite of other neurons. Each dendrite of a neuron can accept many axon terminals. This allows multiple interconnections. The cell body reacts on input stimuli and can transmit an output signal to other neurons through the axon. This transmission only takes place when the total aggregate strength of the input signals from the dendrites exceeds a certain threshold. The total strength results from a weighted sum of all input signals. The weighting is achieved by the synapses. Basically, two different types of synapses can be distinguished: excitatory and inhibitory synapses. An excitatory synapse results in a positive weighting and an inhibitory synapse results in a negative weighting of the input signal. By altering the weights of the synapses, the brain has the ability to learn and to adapt to new situations [Vel06].

To exchange information between neurons, spike trains are used. However, the code by which this information is transmitted – the *neural code* – is not yet well understood [SZ96]. The traditional view in system physiology is that it is the mean firing rate alone that encodes the signal and that variability about this mean is noise. An alternative view, which has recently gained increasing support, is that it is the variability itself that encodes the signal [Zad98].

Symbolic Information Processing

Cognitive information processing in terms of interacting neurons as just described is evident as it has a physiological foundation. Nevertheless, because of the complexity of mental processes, until now, mental states could not be captured by such low-level explanations. For this reason, information processing in the brain is often described in terms of symbols. In the theory of symbolic systems, processes are not considered on a neural basis but on the more abstract level of symbols. Symbols are regarded as the basic information processing units of the mind. According to the theory of symbolic systems, the mind is a symbol system and cognition is symbol manipulation. Examples for symbols are objects, characters, figures, sounds, or colors used to represent abstract ideas and concepts. Each symbol is associated with other symbols. Symbol manipulation offers the possibility to generate complex behavior [Fre96]. Instead of the term "symbol", other authors use the labeling "image" [Dam94].

[1] A second type of information processing in the brain bases on chemical substances including neurotransmitters, hormones, and peptides. However, chemical information processing will not be considered in the model proposed in chapter 4.

3.2 Functional Units of the Brain and Mind

In the last section, the basic processing units of the brain and the mind were described. The information processing performed by such a basic unit seems to be quite simple. However, what makes the understanding of the function of the brain so difficult, maybe even impossible, is the fact that there exist various interconnections between billions of these basic processing units. As their interconnections are that complex, their true interaction is still not deciphered, and the function of the brain as a whole still remains a mystery up to a certain degree [Gol02, chapter 1, p. 2].

Besides the attempt to explain brain functions by such bottom-up methods starting with the basic processing units, there also exist approaches to consider it in a top-down manner. According to the Russian neurologist and neuropsychologist Aleksandr Romanovich Luria [Lur73, chapter 2], three principal functional units of the brain can be distinguished whose participation is necessary for any type of mental activity. They can be described as the *unit for regulation tone and waking of mental states*, the *unit for receiving, analyzing, and storing information arriving from the outside world*, and the *unit for programming, regulating, and verifying mental activity* (see figure 3.1). The three units cannot carry out a certain form of activity completely independently. Each form of conscious activity is a complex functional system. It takes place through the combined working of all three brain units. Each of them makes its own contribution.

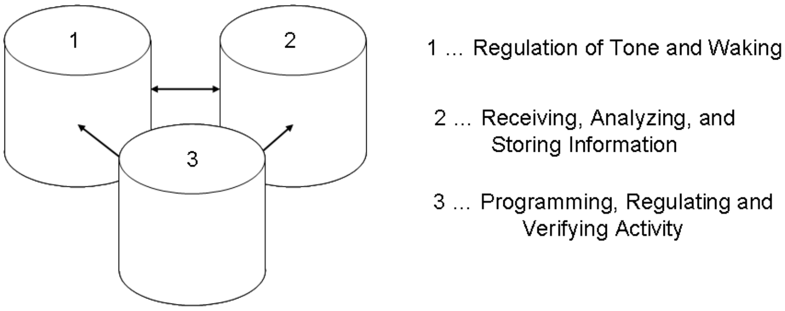

1
2
3
1 ... Regulation of Tone and Waking
2 ... Receiving, Analyzing, and Storing Information
3 ... Programming, Regulating and Verifying Activity

Figure 3.1: Three Basic Functional Units of the Brain

Each of these basic units has a hierarchical structure and consists of at least three cortical zones built one above the other. They are referred to as *primary*, *secondary*, and *tertiary area* (see figure 3.2). The primary (projection) area receives impulses from or sends impulses to the periphery. In the secondary (projection-association) area, incoming information is processed or programs are prepared. The tertiary area (zone of overlapping) is the latest system of the cerebral hemispheres to develop and is responsible for most complex forms of mental activity requiring the concerted participation of many cortical areas. In the following, each of the three principal functional units is described briefly summarizing research findings outlined in [Lur73] and [ST02].

The Unit for Regulating Tone and Waking of Mental States

For human mental processes, the waking state is essential. Only under optimal waking conditions, information can be received and analyzed. Precise regulation of mental processes is impossible during sleep. Organized, goal-directed activity requires maintenance of an optimal level of cortical tone. The reticular formation of the brainstem is a powerful mechanism for maintaining cortical

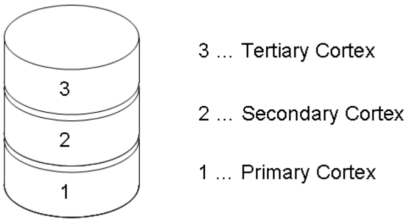


Figure 3.2: Hierarchical Structure of a Basic Functional Unit

tone and regulating the functional state of the brain and is a factor determining the level of wakefulness. At least three principal sources of activation can be distinguished. The first source is the metabolic process of the organism. Metabolic processes leading to maintenance of the internal equilibrium of the organism in their simplest forms are connected with respiratory and digestive processes, with sugar and protein metabolism, with internal secretion, and so on. More complex forms are connected with metabolic processes organized in certain inborn behavior systems. These systems are widely known as systems of instinctive food-getting and sexual behavior. The second source of activation is of completely different origin. It is connected with the arrival of stimuli from the outside world in the body. It leads to the production of completely different forms of activation, manifested as an orienting reflex. The third source of activation plays the most intimate part. The fulfillment of plans or the achievement of a goal requires a certain amount of energy and is only possible if a certain level of activity can be maintained.

The Unit for Receiving, Analyzing, and Storing Information Arriving from the Outside World

This unit is also referred to as afferent system and is located in the back half of the forebrain. The component parts of this unit are adapted to the reception of visual, auditory, vestibular, or general sensory information. The system also incorporates the central systems of gustatory and olfactory reception. Information coming from the visual, auditory, or other sense organs is first processed in different, separated parts of the brain and is merged only later on in the processing.

The basis of this unit is formed by the primary areas of the cortex. They consist mainly of neurons, which possess extremely high specificity. The neurons of the cortical visual systems, for example, only respond to the narrowly specialized properties of visual stimuli like shades of color, the character of lines, or the direction of movement. Neurons of the primary auditory cortex only respond to highly differentiated properties of acoustic stimuli. The primary zones of the individual cortex regions also contain cells of a multimodal character, which respond to several types of stimuli. Additionally, there exist cells, which do not respond to any modally-specific type of stimuli and evidently retain the properties of non-specific maintenance of tone. These multimodal and non-modally-specific cells, however, form only a very small proportion of the total neuronal composition of the primary cortical areas. The primary areas are surrounded by systems of secondary cortical zones. The cells of the secondary cortical zones have much lower degree of specificity. The tertiary zones are responsible for enabling groups of several analyzers to work concertedly. This means that the information coming from the different sense organs having been processed separately until now in the primary and secondary zones of each modality are merged in these zones. The tertiary zones are specifically human structures. They are almost entirely concerned with the function of integrating excitation arriving from different analyzers. The great majority of neurons of these zones are multimodal in character. They respond to general features to which neurons of the primary and secondary cortical zones are unable to respond. The tertiary zones are also responsible for the transition from direct, visually represented syntheses to the level of operations with word meanings, with complex grammatical and logical structures, with systems of numbers and abstract relationships. They play an essential role in the conversion of concrete perception into abstract thinking.

The unit for receiving, analyzing, and storing information can also be considered as perceptual unit. As this unit is of great interest for model development in this work, a more detailed description of this unit is given in section 3.3.

The Unit of Programming, Regulation, and Verification of Mental Activity

Reception, coding, and storage of information are only one aspect of human cognitive processes. Another is the organization of conscious activity. Man creates intentions and forms plans and programs of his actions. He inspects their performance and regulates his behavior so that it conforms to these plans and programs. He verifies his conscious activity by comparing the effects of his actions with the original intentions and corrects any mistakes he has made. This task is linked with a third fundamental functional system, responsible for programming, regulation, and verification. This system is also referred to as efferent system and is located in front half of the forebrain.

The zones of the unit of programming, regulation, and verification of activity are governed by the same principles of hierarchical organization and diminishing specificity like the system for reception, coding, and storage of information. The main difference to the second, afferent system, where processes go from the primary to the secondary and tertiary zone, is that in the efferent system, the processes run in a descending direction, starting at the highest levels of the tertiary and secondary zones. There, motor plans and programs are formed. The primary area sends the prepared motor impulses to the periphery. The second feature distinguishing the efferent system from the afferent system is that the efferent system does not contain a number of different modally-specific zones.

Sequential Development of the Primary, Secondary and Tertiary Brain Areas

According to [Lur73, chapter 2], the “localization” of higher mental processes in the human cortex is never static but moves about essentially during development of the child and at subsequent stages of training. This is expressed by the *law of the hierarchical structure of the cortical zones*, which he proposed to govern the working structure of the unit for receiving, analyzing, and storing information as well as the unit of programming, regulation, and verification of activity. This law describes the relationships between the primary, secondary, and tertiary cortical zones, responsible for increasingly complex synthesis of incoming information. The relationships between these primary, secondary, and tertiary cortical zones change in the course of ontogenetic development.

In the young child, the formation of properly working secondary zones could not take place without the integrity of primary zones. The proper working of the tertiary zones is impossible without adequate development of the secondary cortical zones. A disturbance of the lower zones in infancy must therefore lead to incomplete development of higher cortical zones. The main line of interaction between the cortical zones runs “from below upward”. In the adult person, the higher cortical zones have assumed the dominant role. When he perceives the world around him, he organizes (codes) his impressions into logical systems. The highest tertiary zones of the cortex begin to control the work of the secondary zones. If the secondary zones are affected by pathological lesions, the tertiary zones have a compensatory influence on their work. The main line of interaction runs “from above downward”. E.g., in the initial stage, writing depends on memorizing the graphic form of every letter. With practice, writing is converted into a single “kinetic melody”, no longer requiring the memorizing of the visual form of each isolated letter. This fact was already noticed by the neuroscientist Sigmund Freud [Fre91].

In summary it can be said that the hierarchical arrangement of perception and memory reverses during the maturational process. For small infants, everything depends on the senses, and cognition is driven by concrete perceptual reality. In adults, abstract knowledge derived from these early learning experiences comes to govern the perceptual process. We see what we expect to see and are surprised or fail to notice when our expectations are contradicted [ST02, chapter 5].

3.3 The Perceptual System of the Brain

The focus of this thesis is on the development of a model for human-like machine perception. The primary goal of the perceptual system of the brain is to inform the individual about characteristics of the environment which are important for life [Gol02, chapter 1]. This section gives an overview about research findings concerning the cerebral organization and function of the perceptual system of the human brain.

3.3.1 Bottom-up and Top-down Processes in Perception

Generally, perceptual processes can be considered as the result of *bottom-up processing* and *top-down processing* working together (see figure 3.3). Bottom-up processing, also called data-based processing, is based on incoming data from the receptors of our sense organs. Incoming data are always the starting point for perception. Without incoming data, there is no perception [Gol07, chapter 1, p. 9]. Top-down processing, also called knowledge-based processing, is based on knowledge. This knowledge can be factual knowledge about objects, pre-experience, knowledge about the context in which the objects occur, and expectation. According to [Gol07, chapter 1, p. 9], knowledge is not always involved in perception but it often is and sometimes without the individual even being aware of it. In contrast, [EB04] report that it is impossible to reconstruct the environment “bottom-up” from the sensory information alone. Prior knowledge is needed to interpret ambiguous sensory information.

3.3.2 Cerebral Organization of Perception

As outlined in section 3.2, Aleksandr Luria suggests the brain to be made up of three principal functional units, each of them consisting of at least three hierarchical areas. The three hierarchical areas are called primary cortex (projection zone), secondary cortex (association zone), and tertiary cortex (zone of overlapping). The unit, which Luria calls the unit for receiving, analyzing, and storing information, can be regarded as perceptual system responsible for perceiving the outer world.

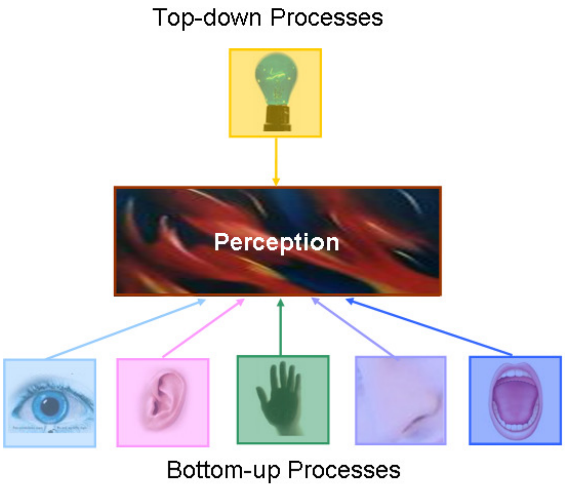


Figure 3.3: Bottom-up and Top-down Processes in Perception

The process of outer world perception actually involves five different perceptual systems – the systems of visual perception, auditory perception, somatosensory perception, olfactory perception, and gustatory perception. Each of these systems is localized in a separate brain region. As pointed out in [Lur73, chapter 2], the information from different sense organs is first processed separately in different brain areas and their outcome is combined and merged not until later in further processing steps. Visual information is processed mainly in the occipital cortex, auditory information is handled to a great extend in the temporal cortex, somatosensory information processing happens mostly in the parietal cortex, and olfactory and gustatory information processing can be assigned to activities in the cortex of the insula and the structures inside the temporal lobe [ST02, chapter 1].

Each of these unimodal sensory processing units have a primary and a secondary cortex that is responsible mainly only for the processing of information of the particular modality. The different senses are then merged in the zones at the boundary between the occipital, temporal, and postcentral regions of the hemisphere where the cortical areas for visual, auditory, vestibular, cutaneous, and proprioceptive sensation overlap [Lur73, chapter 2]. These zones are labeled as tertiary zones. As outlined in [Gol02, chapter 3], the primary and secondary cortices of all perceptual systems are organized in largely the same manner, which facilitates the further processing in higher levels.

Figure 3.4 illustrates graphically what has just been described about the hierarchical organization of the perceptual unit. In this model, the five senses mentioned are first processed separately in a two-layered architecture representing the primary and secondary cortex of each region. The third layer represents the zones of overlapping in the perceptual cortex. Here, the different unimodal sensory perceptions are associated and merged with each other to result in a unified multimodal perception.

Visual Perception

The processing of visual information is mostly carried out in the occipital cortex. Based on experiments in the visual cortex of cats, David H. Hubel and Torsten N. Wiesel, the Nobel Prize

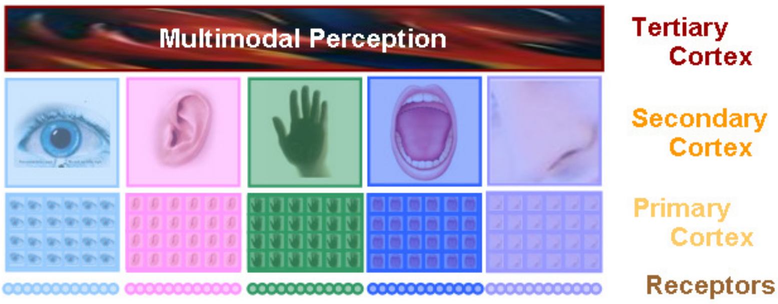


Figure 3.4: Modular Hierarchical Organization of the Perceptual System of the Brain

Winners in Physiology and Medicine 1981, suggested a model for visual information processing in the cortex with a "simple-to-complex hierarchy" – a feed forward sequence of more and more complex and invariant neuronal representations [HW62]. According to this model, simple cells have a small receptive field. Complex cells have a bigger receptive field. Complex cells receive input from several simple cells. These finding conform to the descriptions of [Lur73] about the hierarchical organization of perception in a primary, secondary, and tertiary cortex.

The primary areas of the occipital cortex are those where fibers from the retina terminate. These projection zones have a *somatotopic (topographic)* structure. This means that there is a correspondence between the position of the receptors in the retina activated by a stimulus and the area of the cerebral cortex that is activated by it. Nerve cells located in the primary cortex react to stimuli of high specificity in circumscribed areas of the visual field. For example, there exist neurons, which respond exclusively to shades of color, the character of lines, edges, angles, balks of a specific length, orientation, or the direction of movement [Gol02, chapter 3, p. 89]. These types of cells are also called *feature detectors*. Observations showed that a stimulation of the primary zones of the occipital cortex by an electric current evoked the appearance of elementary visual hallucinations in patients like the perception of flashes of lights, tongues of flame, and colored spots. These phenomena appeared in strictly defined parts of the visual field [Lur73, chapter 3].

The secondary zones of the occipital cortex are placed over the primary zones. The visual association zone is distinctly larger than the primary visual cortex. The secondary cortical zones loose the character of somatotopical projection of the corresponding sensory structures. The degree of specificity of neurons is much lower than in the primary cortex. The secondary zones of the visual cortex play the role of synthesizing visual stimuli, coding them, and forming them into complex systems. The function of the secondary zones is to combine the features detected in the primary cortex into complete forms. These zones play a decisive role in the provision of a higher level of processing and storing of visual information. There have been found neurons that respond to the presence of objects independent of their size, their location, and their orientation [Gol02, chapter 4]. E.g., there have been found cells in the secondary cortex that are specialized for the analysis of faces. Stimulation of a point in the secondary zones of the visual cortex gave rise to complex recognizable visual hallucinations like images of flowers, animals, familiar persons, etc. The stimulation could even cause appearances of complex sequences [Lur73, chapter 3].

Acoustic Perception

The processing of auditory information is mostly carried out in the temporal cortex. Similar to the visual system, this cortex is divided into primary and secondary auditory zones.

The primary zones of the auditory system contain so-called tonotopic as well as topographic maps. A tonotopic map contains points on a neural structure that correspond to frequencies of a sound stimulus. This means that neurons that respond best to low-frequency tones are located elsewhere than neurons that respond best to high frequency tones [Gol07, chapter 11, p. 256]. A topographic map contains points on a neural structure that correspond in a systematic way to locations in space. There exist neurons that respond specifically to sound signals coming from a certain direction in space [Gol07, chapter 12, p. 272]. There also exists a second method for localization of sound sources in space based on time differences. As the ears are arranged in a certain distance from each other, a sound signal commonly does not reach them at the same time. The receptors in one ear receive it with a relative time delay in relation to the receptors in the other ear. This time difference between the arrivals of the signal at both ears allows a localization of sound sources in space. There have been found neurons in the auditory cortex that respond to specific time differences [Gol02, chapter 11].

Neurons in the primary cortex can be activated by simple sounds, such as pure tones. Neurons of the areas outside the primary cortex require more complex sounds, such as auditory noise that contains many frequencies or human vocalizations [Gol07, chapter 11]. The secondary zones of the auditory cortex are also concerned with the analysis of temporal series of sound signals [Gol02, chapter 10].

Somatic Sensation

Somatic information is processed in the parietal cortex. The somatosensory system actually comprises a whole group of sensory systems. First, it is responsible for cutaneous sensations, which are based on the stimulation of receptors in the skin responsible for tactile sensation, vibration sense, temperature sense, and pain sense. Secondly, it also includes proprioception – the "body sense" that leads to a perception of the body. This system processes sensory information coming from the skin, the muscles, the tendons, and the vestibular system. Thirdly, the somatosensory system accommodates the sense of position and movement of the limbs called kinesthesis [Gol07, chapter 14]. Each of these senses is served by a specific type of receptor and projects separately to the brain. The somatosensory system is of equal complexity as the visual system and has the same functional organization as in the visual and auditory cortex. It is based on primary zones of topographic organization. E.g., there exist neurons that respond specifically to tactile movements over the skin surface in a certain direction [Gol02, chapter 14]. Above these primary zones of the parietal cortex are superposed its secondary zones. Their stimulation leads to the appearance of more complex forms of cutaneous and kinesthetic sensation [Lur73, chapter 2].

Olfactory and Gustatory Sensation

In contrast to the other senses, the olfactory and gustatory senses are chemical in nature [ST02, chapter 1, p. 21]. Science is just beginning to understand the processing of these senses, especially in the cortex. If the olfactory system follows the pattern observed in the other senses, higher cortical areas will be involved in perception of complex odors and tastes. Olfactory and gustatory senses are also closely linked to emotions [Gol02, chapter 15, p. 337 and p. 570].

Multimodal Perception

The secondary zones of the cortex can be regarded as systems responsible for the highest levels of special mental forms of modality-specific activity. However, the most complex forms of information analysis depend on the combined working of several analyzers. Perception is multisensory. Many events occurring in our surrounding are registered by the sense organs of more than one modality. If a single modality is not enough to come up with a robust estimate, information from several modalities can be combined. The coordination and integration of information derived from different sensory systems is essential for a unified perception of our environment. The zones of the cortex, which lie between the occipital, temporal, and central regions, are tertiary in function and play a basic role in the organization of complex simultaneous (spatial) syntheses. They constitute the specifically human portions of the brain. They mature later than all other zones of the posterior regions of the cortex. They do not become fully operative until the seventh year of life. They play a special role in inter-analyzer syntheses. The great majority of neurons of these zones are multimodal in character. They respond to general features to which neurons of the primary and secondary cortical zones are unable to respond. The tertiary zones fit together individual elements of incoming impressions into a single structure.

Much of the history of perceptual research can be characterized as a “sense-by-sense” approach where researchers have focused on the functional properties of one sensory modality at a time [CSS04]. Until now, very little is known about how information from different modalities combines to form a single multisensory representation. [New04] argues that in order for information to be shared across modalities, the information must be encoded in a similar manner for all modalities. This assumes a functional equivalence among the modalities.

In general, multimodal perception yields to a more robust perception than could be achieved with only one sense. However, there sometimes occur problems when cues of different sensory modalities are incongruent. [KC01] give an example for the occurrence of incongruent visual and auditory cues by dubbing one syllable onto a movie showing a person mouthing a different syllable. The listener typically reports hearing a third syllable that represents a combination of what was seen and heard. This audio-visual illusion has become known as the McGurk effect or McGurk illusion. It is a striking demonstration of the combined (bimodal) nature of speech understanding [Gol02, chapter 9, p. 365].

[CSS04] report that when conflicting signals are presented via the different sensory modalities, the emergent percept is typically dominated by the most persuasive sensory cue in the particular context. An important question is how the brain weights the inputs it receives from the different senses in producing a final perceptual output or experience. Vision has traditionally been viewed as the dominant modality [SKS04]. Vision generally dominates our perception of space, because visual spatial information is exceptionally reliable and precise. Spatial information of other sensory systems is almost always less precise and less reliable. If the localization of a stimulus based on non-visual information is ambiguous or conflicts with visual localization of the same stimulus, the non-visual percept of location is sometimes drawn to the visually identified location, a phenomenon called “visual capture”. A traditional view is that visual dominance is an inherent physiological advantage of visual over other sensory connections in the brain. However, [WK05] propose an alternative hypothesis that suggests that visual dominance results form the statistically optimal integration of auditory and visual information. An optimal integration of bimodal signals requires taking into account the reliability of the encoded stimuli. Vision does not dominate under all circumstances about the other sense. For instance, for temporal processing, vision is far less precise than audition. Audition dominates over vision in the time domain [EB04].

3.3.3 Influence of Knowledge on Perception

According to [Gol02, chapter 1], object recognition is facilitated by knowledge. This knowledge can be factual knowledge about objects, pre-experience, knowledge about the context in which the object occurs, and expectation. Much of what we take for granted as "the way the world is" – as we perceive it – is in fact what we have learned about the world – as we remember it. Much of what we take to perception is in fact memory. Studies have shown that objects are recognized more easily if the context is available to which the objects normally belong. [ST02, chapter 5] mention that we frequently see things that are not there simply because we expect them to be there. Adults project their expectations onto the world all the time. They largely construct rather than perceive the world around them. [WC99] propose that the same sensory stimuli could produce different patterns of errors depending on a subject's expectations. [Gre97] points out that perception requires intelligent problem solving based on knowledge. Errors of perception, which result in phenomena of illusions, can be due to knowledge being inappropriate or misapplied. It might be that progress in artificial intelligence has been delayed, because it was not recognized that artificial potential intelligence of knowledge is needed to be compared to brains. This knowledge is derived from past experience. Thus, perception is largely based on the past. The majority of studies about object identification in scenes have found that consistent scene context can facilitate object identification. However, this opinion is not shared by all scientists [HH98]. [Hen05] discusses the involvement of sensory and cognitive processing of visual information in real world scene perception. He points out that the use of knowledge about the laws of physics and the functions of an environment can help to identify objects in a scene more easily.

Knowledge of different types influences perception in a top-down process. A fundamental question is, on which level they interact with sensory perception. The answers to this question are controversy. [HH99] discuss this topic in connection with visual object identification and knowledge about scene context and summarize the three principal models of object identification in scenes that exist. Therefore, they describe visual object identification to consist of three component processes: First, the retinal images are translated into a set of visual primitives (surfaces, edges). Second, these primitives are used to construct structural descriptions of so-called object tokens in the scene. Third, these constructed descriptions are matched to stored long-term memory descriptions. When a match is found, identification has occurred, and semantic information stored in the memory about that object type becomes available. The first two stages can be considered as perceptual. The matching state can be seen as an interface between perception and cognition. In this state, perceptual information must make contact with memory representations. Based in these three levels, models of object identification in scenes can be divided into three groups. The difference of the models is the stage of object identification at which scene context is proposed to exert an influence. The first group proposes that expectations derived from scene knowledge interact with the perceptual analysis in the first two stages. The second group suggests that the locus of interaction is at the matching stage where the perceptual descriptions are matched to long-term memory representations. The third group proposes that object identification is isolated from scene knowledge.

3.4 Merging of Information – The Binding Problem

In the last sections, different units and organizational structures of the brain have been explained that have different functions and are specialized to the processing of distinct information. However, these modules do not work isolated and separated, but interact with each other in order to form functional systems. For abstracting a technical model from neuroscientific and neuropsychological research findings, a fundamental question is how all these units can interact to result in a "unified experience". This is the so-called *binding problem*. In the last years, there were published many articles concerning the binding problem in the brain. However, these articles generally focus on special aspects and do not offer an extensive overview about the research area as a whole. The aim of this section is to provide such an overview in order to derive mechanisms of binding for the model introduced in chapter 4.

3.4.1 The Binding Problem as Key Question to Brain Function

The binding problem concerns our capacity to integrate information across time, space, attributes, and ideas. Language comprehension and thinking depend on correct binding of syntactic and semantic structures. Binding is also required when we select an action to perform in a particular context or when we perceive the world around us [Tre99]. The question how the brain solves the binding problem has puzzled and intrigued physiologist, psychologists, and theoreticians for decades [GHT96]. According to [Ros99] and [RP99], the binding problem as a theoretical problem was originally formulated by Christoph von der Malsburg in 1981 in his article "The Correlation Theory of Brain Function" [von81]. However, the term "binding" itself never occurs in this article. In fact, von der Malsburg did not formulate the binding problem but suggested a theorem for a solution to the binding problem which is based on neural signal synchrony. Binding by neural signal synchrony has already been mentioned in literature before, for example by C. Legendy in 1970 and P. Milner in 1974 [von99, Gra99]. However, neural signal synchrony is not the only possible solution suggested to the binding problem (see section 3.4.5). The solutions of the binding problem proposed in literature until now are controversial and hotly debated in neuroscience. Years of research have shown that the binding problem cannot be solved easily. [Ros99] considers the binding problem to be *"one of the most puzzling and fascinating issues that the brain and cognitive sciences have ever faced"*. [TvdM96] regard the binding problem as one of today's key questions about brain function. [Tre96] points out that a solution to the binding problem may also throw light on the problem of the nature of conscious awareness.

3.4.2 A First Simplified Explanation of the Binding Problem

The problem of binding is maybe best illustrated by a classical example originally described by Frank Rosenblatt in 1961 in his book "Principles of Neurodynamics: Perceptrons and the Theory of Brain Mechanisms" [von95, von99]. Rosenblatt introduced a hypothetical perceptual system based on a classical neural network. This system consists of four neurons. Two neurons are able to respond to the presence of objects. One responds to the presence of a triangle (1), the other to the presence of a square (2). They both generalize over position. The other two neurons indicate the position of the objects. One responds to the upper half (3) of the image, the other to the lower half (4). They both generalize over the nature of the object (see figure 3.5).

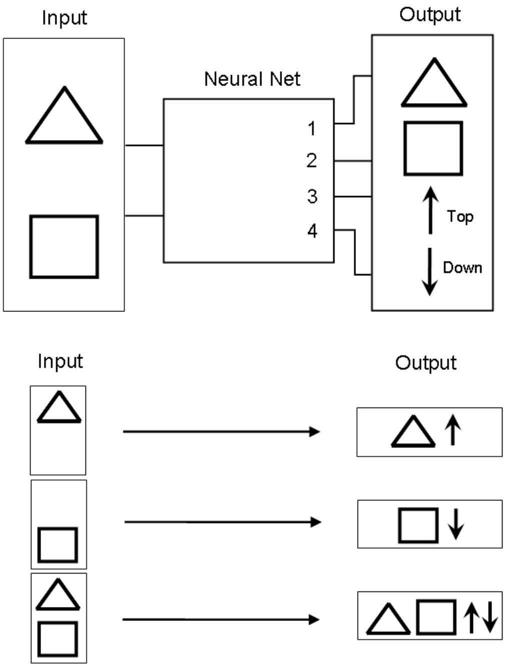


Figure 3.5: Example from Rosenblatt to Explain the Binding Problem

When showing a single object to the network, it responds adequately. For example, if the system is to diagnose the presence of a triangle in the upper half, the neurons (1) and (3) are activated. If a square in the lower half is presented to the system, the neurons (2) and (4) are activated. A problem arises in situations where two objects are present simultaneously. If there occurs a square in the upper half of the image and a triangle in the lower half of the image, all four neurons are activated. The example of Rosenblatt shows that representing the presence or absence of features alone is not sufficient to represent multiple objects simultaneously [SH02]. It is not clear whether the triangle or the square is in the upper position. The neural data structure has no means of binding the proposition triangle to the proposition top or the proposition bottom to the proposition square. This is the binding problem. According to [TG80], such a "conjunction error" can be induced in humans if they are given insufficient viewing time. These misperceptions are labeled "illusory conjunctions".

3.4.3 Evidence for a Binding Problem in the Brain

There exist conflicting opinions about the question whether binding is really a problem within the brain. According to [Gra99], binding is not a problem for nervous systems, as evolution has sculpted their organization to solve the problem efficiently and effectively. It is only a problem for those of us trying to understand how the nervous system achieves the task. [IAMN03] states that the binding problem is a problem naturally processed in our brain, but its mechanism has been yet unknown. [Ros99] points out that even if the brain usually does not appear to have a problem in correctly binding signals, scientist still lack an understanding of how information variously distributed in patterns of neural firing results in coherent representations. [RP99] report that the binding problem may be only a problem in the eye of the beholder, but is not necessarily a problem for all object recognition devices and perhaps may not be one for the brain.

In contrast, [TG80] and [WC99] mention errors such as illusory conjunctions as striking evidence for a binding problem in the brain. Illusory conjunctions can occur when perceptual features (e.g., location or form of objects) become unbound from their original objects and are recombined to form a new object representation. An example for an illusory conjunction has already been given in section 3.4.2 when mentioning the example introduced by Frank Rosenblatt to explain the principle of the binding problem. Illusory conjunctions occur in the normal brain when there are temporal or capacity limitations, when spatial attention is diverted, when displays are briefly presented, and/or in peripheral presentation where spatial resolution is decreased. [RD99] mention that the number of potentially erroneous feature conjunctions increase exponentially with the number of objects in a larger receptive field. The receptive fields become larger and larger at each processing stage of the ventral stream. Therefore, an increasing number of erroneous feature bindings has to be ruled out. The binding problem emerges as a necessary consequence of the large receptive fields found in higher-order areas.

A binding problem in the brain can also occur in pathological cases and in cases of brain damage. The most extreme examples of binding errors occur when both parietal lobes are damaged. This results in a condition known as Balint's syndrome, which describes the inability to perceive more than one object at a time. Binding in synaesthesia, in many respects, represents an inverse problem from that of the Balint's syndrome [Rob03]. In synaesthesia, there are perceived features together although one feature in not present in the stimulus. Typical forms of synaesthesia are phonemic/chromatic, in which sounds induce the perception of color, or graphemic/chromatic, in which shapes induce the perception of color. An example would be that the written letter A might induce blue and the digit 7 might induce green. An induced color can appear in different places relative to the inducer. Synaesthesia is automatic and consistent throughout life. Synaesthetic experience represents direct connections between cortical feature maps, maybe through synaptic connections that fail to undergo normal synaptic pruning during development [Rob03]. Most people appear to experience odor-taste synaesthesia. Certain odors are constantly associated with certain tastes. For example, the odor of vanilla is consistently reported as smelling sweet although sweetness is normally associated with the stimulation of the sense of taste. [SB04] suggest that this odor-taste synaesthesia is learned.

3.4.4 Classes of Binding Problems

The term "binding problem" does not refer to a unitary problem but to a whole class of problems. According to [Sch01], binding can be spatial or temporal. It can occur within a single modality, across modalities, or may even require sensory-motor integration. [Ros99] distinguishes between *perceptual binding* and *cognitive binding.* Perceptual binding problems involve unifying aspects of percepts. Cognitive binding problems include relating a concept to a percept. [Sch01] subdivides perceptual and cognitive binding in further categories. In connection with perceptual binding, he mentions the terms visual binding, auditory binding, binding across time, and cross-modal binding. Cognitive binding includes sensory-motor binding, cross-modal identification, relating a concept to a percept, and memory reconstruction.

Reading the many different existing types of binding, it becomes obvious that binding is required almost everywhere in the brain and in all processing levels. However, only very few systems of the brain have been investigated and discussed in connection with binding. Until now, binding has been most extensively discussed in the visual system. *Visual binding* is the process of linking together the attributes (color, form, motion, size, and location) of a perceptual object. Seeing seems to be a deceptively simple process. We perceive objects, symbols, movement, persons, and other aspects of a visual scene without any effort or awareness of the mechanisms that process visual information. However, for this continuous and seamless visual perception, our brain has to cope with a huge mass of information that continuously streams to us through our eyes. Concerning other classes of binding, only very view publications can be found. [Rob03] briefly mentions the problem of *cross-modal binding*, because the information from different sensory receptors is registered in diverse areas of the brain and therefore requires binding. [Hom04], [SH99], and [SH02] describe the binding problem in *action control*. Like in perception, the simultaneous representation of multiple actions requires a mechanism for coding which motor features belong together. The multiplicity of action components, action features, and action control systems point to an integration or binding problem. [HM07] report about the binding problem *across perception and action* arguing that the things we perceive and the actions we perform must be systematically related to each other. Some integration spanning perception and action is necessary, and this integration somehow needs to be tailored to the current task and context. [ST02], [Sin01], and [Mas04] write about binding in connection with *consciousness*. A property of consciousness is that it is generally a unified experience. Each one of us has the impression of being a single entity, experiencing an integrated perceptual world at each particular moment in time. The neuro-anatomical structures involved in generating these obviously connected perceptions are located in different places on the brain. The problem of how all this information comes together to form an ordinary unified experience of consciousness is also a binding problem. [Ros99] regards the problem of consciousness as the most mystifying binding problem of all. He arises the question how something as simple and mechanistic as neural firing can add up to subjectivity, raw feelings, and self. He asks whether the mechanisms that allow us to attribute the correct color and shape to an object are the same ones that lead to the unity of phenomenal experience and if the solution to the binding problem will be the solution to the mystery of consciousness.

3.4.5 Potential Solutions to the Binding Problem

As described in section 3.4.4, binding occurs in many different kinds of brain processes. To explain how coherent representations can be formed from information that is distributed throughout the brain, different binding mechanisms have been hypothesized. According to [Ros99], it is likely that most different forms of binding problems are solved via common mechanisms. On the other side, he takes in consideration that something as complex as binding may not have one single mechanistic solution. A number of possible mechanisms have been investigated for the binding problem which will be discussed in the following. They are not mutually exclusive [Tre96, Bau05]. Potential solutions to the binding problem can be approached from both cognitive and neurobiological levels of analysis [Rob03]. The most extensive discussion so far has focused on the problem of binding in visual perception. The binding problem is usually discussed in the context of visual perception, presumably because the visual system is particularly well studied, and because the multiplicity of specialized feature maps responsible for the representation of color, form, size, and motion it apparently houses makes the consequences of distributed processing particularly obvious. However, there is no reason to believe that binding problems are restricted to visual perception [HM07]. In the following, different possible solutions to the binding problem suggested in literature are described. It must be mentioned that the solutions of the binding problem proposed until now are controversial and hotly debated in neuroscience [SH02]. The study of binding is in its infancy. Conclusions will surely change as more data are collected [Rob03].

Combination Coding

As described by [HW62], the visual cortex is organized hierarchically (see section 3.3.2). This fact inspired the model of convergent hierarchical coding, also called combination coding or specificity coding, which, at first sight, may be the seemingly simplest approach. According to this model, incoming information is integrated and more and more condensed at higher processing stages. Each level in this hierarchy carries out operations depending on the input from earlier levels. So-called combination coding cells react only to combinations of features. For example, a cell reacts only to an object of a particular shape and color at a particular retinal position. The hierarchical processing leads to an increase in the complexity of the neuronal representations. That way, increasingly complex features are represented by higher levels in the hierarchy. A small group of neurons or a single neuron receives convergent input from populations of neurons at lower levels in the hierarchy.

In 1941, in his book "Man on His Nature", Sir Charles Sherrington introduced the notion of one ultimate pontifical nerve cell as the climax of the whole system of integration and immediately rejected the idea in favor of a concept of the mind as a cooperation of many cells. [Bar72] agreed with Sherrington that the "pontifical cell" should be replaced by a number of "cardinal cells". Among these many cardinal cells, sometimes also referred to as "grandmother cells" or "connector cells", only a few fire at once to represent what is currently going on in the environment and the body.

Concerning visual perception, evidence for this theory of combination coding is revealed by gradual decrease of retinotopic specificity, the increase in receptive field size, and the dependence of neuronal responses on increasingly complex stimulus features. In [RP99] and [RP02], such a hierarchical feed-forward architecture is suggested for the recognition of simple forms on a model "retina" composed of 160x160 pixels. The model consists of layers of units. Two types of operations, selection and template matching, are combined in a hierarchical fashion. That way, a complex, invariant feature detector is built up from small, localized, simple cell-like receptive fields in the bottom layer.

The concept of convergent hierarchical coding surely plays a role in the binding of features, but cannot be the complete solution to binding problem [Bau05]. One open question is how cells with these specificities should be created. If they were to be learned, many examples of objects of all colors and shapes in all possible locations would have to be given to the system before it would work. If they were prewired, most of the connectivity patterns of the network would have to be stored in the genes. This is an unlikely proposition [TvdM96]. The other problem of the convergent hierarchical coding theory is that it would require as many binding units as there are distinguishable objects. If a cardinal cell is able to represent a whole class of objects, the individual object cannot be represented in detail, because the signal of a single cardinal cell is too crude. Or there has to be a cardinal cell for each pattern which would quickly lead to a combinational explosion of the number of needed cells. This problem is called *combinatorial problem.* It would be impossible to recognize new patterns, which differ from familiar ones merely in detail (for example, a known person with a new facial expression). Additionally, a cardinal cell would have to be silent until its pattern appeared again (possibly for decades) [von81].

Population Coding

The combinatorial problem can be overcome by a simple modification of the just mentioned combination coding. Instead of representing the integration of features by the activity of a few neurons or even single neurons at specific cortical locations, complex feature combinations could be represented by the activity of a population of neurons distributed within and across levels of the cortical hierarchy [Gol02, chapter 4]. Each stimulus pattern could be represented by a distinct pattern of firing in a distributed population of cells. Therefore, this coding type is either called population coding or distributed coding. Such a coding scheme would greatly increase the firing in a representational capacity of the cortical network. The number of distinct patterns of activity far outstrips the number of neurons available to represent the stimuli. The combinatorial complexity of the sensory world would no longer pose a problem.

This theory meshes well with many aspect of the anatomical and physiological organization of the visual cortex. There is strong physiological evidence from both sensory and non-sensory cortical areas that stimuli are represented by distributed populations of cellular activity.

Attractive though it may be, this model of sensory representation is again incomplete. The situation becomes far more complex if two objects are presented in the same scene. Then two populations of neurons would fire. This might particularly pose a problem if the objects are close to or overlapping with one another. The unanswered question is how one or more distinct patterns can be identified from the many others that are present in the same network at the same time. This is the so-called *superposition problem.* The essence of this problem is the question how members of a representation are identified as belonging to one representation and how interference between simultaneously representations is avoided.

Temporal Binding

To overcome the superposition problem of population coding and the combinatorial problem of convergent hierarchical coding, the temporal binding hypothesis was suggested. The basic concept for temporal binding, also called binding by synchrony or temporal correlation hypothesis, was formulated independently by C. Legendy in 1970, P. Milner in 1974, and C. von der Malsburg in 1981 [von99]. They proposed that neurons responding to the same object could be grouped into assemblies by invoking a temporal dimension to the responses of cells. This seems plausible because neurons are sensitive to the timing of their synaptic input. Psychophysical observations suggest that the temporal precision of cortical neurons is important for guiding behaviors. Humans are sensitive to timing differences as small as 150 µs [GM99]. The basic idea of temporal binding is that the binding problem can be solved by temporal correlation or synchronization of neuronal firing. Signals of neurons representing features of the same object are mutually correlated in time. Signals of neurons representing features of different objects are not correlated or anti-correlated in time. Synchrony serves as a signature of relatedness. Synchrony reinforces the interactions among the members of the same assembly. Different assemblies are distinguished from one another by the independence of their firing patterns. Thus, multiple distributed signals can coexist in the same network of cortical area at the same time. Each signal provides an independent representation of grouped features. Binding by synchrony has the advantage that it is flexible and dynamic [Tre96].

At the time these theories were developed, very little evidence was available for the existence of cortical neurons engaged in synchronous firing. However, later on, several groups reported experimental evidence from the visual system in support of this theory [GM99]. Numerous electrophysiological experiments have demonstrated that cortical neurons engage in synchronous firing on a millisecond time scale. [Gra99], [Gab04], and [von99] estimated that synchronous firing should occur on a time range of 1–10 ms. [Sin03] discovered oscillatory modulation of cell firing in a frequency range between 30 Hz and 90 Hz which is called γ-frequency range.

Nevertheless, the role played by synchronous firing in feature binding is still highly controversial and far from being understood. [SM99] articulate their doubts and concerns about the temporal binding hypothesis. They mention a number of problems with the idea of binding by a temporal code. The first is that there is no biophysical evidence that cortical neurons can respond selectively to synchronous input of such a precision that would be needed. Second, reports of cortical activity with synchrony of such high precision are rare. Criticisms were also raised about the observations of synchronized neurons. The data were obtained from anesthetized animals and the correlations might have been a consequence of anesthesia. Another problem is that the hypothesis is not a theory about how binding is computed. It is only a theory of how binding is signaled. The theory proposes that the result of the binding computation is represented by synchronous neuronal activity. This begs the question how synchrony is achieved [GM99]. In [SM99], it is pointed out that binding by synchrony might not be computed in the primary cortex. It might be that synchrony is imposed by feedback connections form the higher cortical areas in which the computation is done. However, it is unclear what might be the utility of feeding back information of this kind. In [SGP01], the limits upon the amount of information that an ideal observer is able to extract from a synchrony code are studied. They try to determine whether the available amount of information is sufficient to support computational processes such as feature binding. However, they do not draw a final conclusion.

Temporal binding was suggested to overcome the combinatorial explosion of convergent hierarchical coding. However, [GM99] arise the question whether the combinatorial explosion is really a problem. Considering the visual binding problem, the critical question is not how many different objects might occur, but rather how many our visual system allows us to distinguish from one another. Objects that are not seen as distinct need not have different representations in the nervous system. They estimate that people can distinguish 100,000 different types of objects. To consider also items that are not counted as object, like text, scenarios, and differences in such low-level attributes as orientation, brightness, and color, they multiply that number by the factor 100. This calculation results in a final value of 10,000,000 distinguishable items, which is well below the number of neurons in the visual cortex. This suggests that there is no compelling need for binding by temporal correlation. Instead, visual performance could depend on the existence of small groups of neurons with highly specialized response properties. At last, there is to mention that we can of course see differences between objects presented, even if they do not appear distinct at first glance. However, this discrimination is a sequential process of evaluation, which would require binding across time. Because there is no need for the visual system to access any of these different representations simultaneously, a binding of the activity of distributed neurons with synchronization may not be compelling.

[GM99] point out that, even if there is no absolute need for binding by temporal correlations, it might play an important role in information processing. It might have a role in recognition learning. [von95] mentions the limited bandwidth of neural signals as disadvantage that arises from temporal binding. Only stereotypical tasks have short reaction times. The flexibility required in unfamiliar situations is obtained only for the price of considerable delays. Therefore, it is quite conceivable that time-consuming temporal binding is used by the brain only for novel situations. As soon as a certain binding structure has shown its long-term value, it is frozen into less flexible but faster and more capacious special circuitry. These structures could be combination coding or connector cells. [GHT96] point out that for synchronization to provide a binding mechanism, it would have to occur very close to the stimulus onset. However, synchronization begins at a variable time after stimulus presentation and is not phase locked to stimulus onset. Therefore, the establishment of synchrony is simply too slow to account for normal object perception. For that reason it seems unlikely that synchronization plays a crucial role in binding during everyday perception of familiar object. However, it might have a role in recognition learning, which may have a longer time course. [Sin03] supports the theory that the connections between neurons achieved through temporal binding can be stabilized by learning so that a familiar object activates always its corresponding neurons.

Binding by Attention

[TG80] propose a hypothesis about the role of focused attention in solving the binding problem. The hypothesis of binding by focused attention suggests that instead of trying to process all objects simultaneously, processing is limited to one object in a certain parrot of space at a time.

In [CW01], the following explanation of focused attention – in this case about visual attention – is given: What we see is determined by what we attend to. At every moment, the environment presents far more perceptual information than can be effectively processed. Attention can be used to select behaviorally relevant information and to ignore irrelevant or interfering information. Attention can modulate or enhance the selected information according to the state and goals of the perceiver. Active attentional selection occurs over space and time. The spotlight has been a favorite metaphor for spatial attention. As attention, it can be deployed like a beam of mental light to reveal what was hidden in the world. Attention can also be split into multiple spots. Attention can be allocated to regions of different size – the spotlight has a variable width of focus.

The problem with the temporal correlation hypothesis is that the information about the spatial information of the combined features is lost. Spatial organization of features is reintroduced by restricting a hypothetical attentional spotlight to a single object or location in space, so that only those feature codes that belong to this object or location get activated [HM07]. According to [TG80], focal attention provides the “glue” that integrates the initially separable features into unitary objects. They assume that a visual scene is initially coded along a number of separable dimensions (color, orientation, brightness, direction of movement, etc.). These features are then related to each other by means of focused attention. Through focal attention, stimulus locations are processed serially. All features which are present in the same central “fixation” of attention are combined to form a single object. Visual attention can be used over a small area with high resolution or spread over a wider area with some loss of detail. Once the features have been correctly registered, the compound objects continue to be perceived and stored as such. The authors claim that attention is necessary for the correct perception of conjunctions. Nevertheless, unattended features are also conjoined prior to conscious perception. However, in the absence of focused attention, conjunctions could be formed on a random basis. These unattended couplings can give rise to “illusory conjunctions”. Illusory conjunction may also occur if attention is diverted or overloaded. [RD99] mention that the number of potentially erroneous feature conjunctions increases exponentially with the number of objects in a larger receptive field. The receptive fields become larger and larger at each processing stage. Therefore, an increasing number of erroneous feature bindings have to be ruled out. The binding problem emerges as a necessary consequence of the large receptive fields found in higher-order areas. Attention solves the binding problem by increasing the effective spatial resolution of the visual system. That way, even neurons with multiple stimuli inside their large receptive field process information only about stimuli at the attended location.

Consistent with this interpretation, damage to the parietal lobes, the region which is considered to be involved in allocation attention, can result in illusory conjunctions during free viewing [RD99].

Focusing attention can be considered as a top-down approach. However, this top-down approach is difficult to reconcile with the apparent speed with which object recognition can be proceed and is incompatible with reports of parallel processing of visual scenes. How can attention be capable of serially searching all the possible feature combinations in a reasonable amount of time? Object recognition does not seem to depend only on explicit top-down selection in all situations [RP99]. There have to exist mechanisms that act prior to attention and also serve to attract it [Gra99].

Bundling and Binding of Features

Some hypotheses hold that there exists a pre-attentive stage of processing at which features are identified and represented completely independently of location. According to them, detection of features and localization of features are separate operations. However, other studies found only weak evidence for identification without localization. They found that in many trials, subjects either reported both the color and shape correctly or got them both wrong. [GHT96] point out that it would be surprising if the brain did not make use of spatial information freely available to it at least partially to solve the binding problem. [WC99] propose that there are an unattended and an attentive answer to the binding problem. In the early stages of visual processing (in the primary visual cortex), visual information is represented within spatially organized maps of the visual field (see section 3.3.2). Features are represented as occupying a fairly specific location, and thus location serves as a means for linking all of the features belonging to a single object. Although the representations at this early level contain all the information necessary to determine the relationships between the features in an object, those relationships are not explicitly represented at this level. Features of this early level can be considered as being loosely “bundled” together rather than tightly “bound”. In the absence of visual attention, the spatially organized maps of the visual field prevent features from becoming truly “free floating”. However, without attention, explicit representation of the relationships among features might not be recoded into memory. The processes of object recognition require that features are tightly “bound” rather than loosely “bundled”, as they were in the earlier levels. The simple spatial association used in early vision may not help in later stages of object recognition, because specific information about the location of each feature is no longer available. If information from multiple visual fields were represented simultaneously at a higher level, it would be difficult to determine which features belong to which objects. Selective attention is the apparent solution of this aspect of the binding problem. In summary, objects are held together by the spatial organization of the early visual system. At later stages, a recognized object is held together by the explicit binding of a selected set of features. Working in tandem, these processes of bundling and binding deliver a coherent perceptual world. This account neither requires nor contradicts to the other concepts of binding.

Binding by Knowledge

In section 3.3.3, it was mentioned that knowledge of different forms influences perception. These top-down processes also influence binding. [TG80] suggest that, besides focused attention, contextual information and past experience play a role in binding of features. Even when attention is directed elsewhere, subjects are unlikely to see a blue sun in a yellow sky. Utilizing past experience and contextual information are considered as top-down-processes. Through top-down processing, in a familiar context, likely objects can be predicted. In misleading context, this route to object recognition can give rise to errors. [EB04] point out that prior knowledge is often required for interpreting the sensory signals. [WC99] mention that the same stimuli could produce different patterns of errors depending on subjects' expectations. Until now, it has not been discovered at which level knowledge influences the binding process.

The Feature-integration Theory of Attention

[TG80] suggest a model for binding, which they call "the feature-integration theory of attention". In their model, features (color, orientation, brightness, direction of movement, etc.) are registered early, automatically, and in parallel across the visual field, while object are identified separately and only at a later stage. We become aware of unitary objects in two different ways – through focal attention and through top-down processing by utilizing contextual information and past experiences. In normal conditions, these two processes operate together. In extreme conditions, they may work almost independently of each other.

The True Solution to the Binding Problem?

Looking at the different solutions suggested for the binding problem, it turns out that each theory presents certain reasonable aspects of how binding could be performed in the brain, but none of them gives a complete explanation. However, by combining the different methods and by making certain supplementations, the binding problem could be overcome. Suggestions for how to combine and supplement different mechanisms of binding in order to get a feasible bionic model of human-like machine perception will be made in chapter 4.

Chapter 4

Bionic Model

"A theory should be as simple as possible, but not simpler."

[Albert Einstein]

In section 1.1, it was outlined that – up to now – machine perception still shows a lot of deficits. In contrast, humans are generally capable of perceiving real world scenes without problems. For the work at hand, these facts were the motivation to develop a bionic model for human-like machine perception, which is based on neuroscientific and neuropsychological research findings about the structural organization and function of the perceptual system of the human brain. This chapter presents the developed model. The term "human-like perception" as it is used in this thesis means that for defining the structure of the bionic model and its information processing principles, the structural organization of the perceptual system of the human brain and the way it processes information is taken as archetype. When developing a model of human-like perception, it has to be clear that a full understanding of how the brain works is still missing. There exist many blind spots and contradicting theories. Due to complexity, research works often only focus on very particular and circumscribed topics and problems and leave out a consideration of more global coherences and an explanation of how the investigated results fit into the big picture. This makes it difficult to abstract a unified technical model from the variety of incomplete and contradicting neuroscientific and neuropsychological models. As the model to be developed shall be actually implementable, it cannot leave parts and functions just open or undefined. Therefore, if neuroscience does not provide a clear, comprehensible description of a certain part that is needed for the system, it has to be supplemented by considerations and design decisions taken by the system engineer.

4.1 Neuro-symbols as Basic Information Processing Units

In section 3.1, there were introduced neurons and symbols as basic information processing units of the human brain and the mind. In this section, so-called *neuro-symbols* are introduced, which shall serve as principle information processing units of the proposed model. These neuro-symbols combine characteristics of neural as well as symbolic information processing.

In principle, neural and symbolic information processing are two different approaches to describe one and the same thing on different levels of abstraction. Neurons have a physiological basis and can be regarded as basic processing units of the brain. Symbols are detached from the physiological basis and emerge somehow from neural information processing. Therefore, symbols can be regarded as basic processing units of the mind. An interesting question is if there is a connecting link between these two different levels of abstraction. In section 3.3.2, it was outlined that a stimulation of neurons in the secondary visual cortex gave rise to hallucinations of images of flowers, animals, persons, etc. Additionally, there have been found neurons in the visual cortex that specifically respond to the detection of faces or neurons in the auditory cortex that specifically respond to a certain melody [Gol02, Lur73]. Faces and melodies can be regarded as perceptive symbols. These results can be considered as evidence for an existing bridge between neural information processing and symbolic information processing. Inspired from these facts, a so-called neuro-symbolic information processing concept was developed for the model of human-like machine perception using interacting neuro-symbols for information processing. Neuro-symbols represent *perceptual images*, which can be features, objects, events, scenarios, and situations. Concrete examples for perceptual images will be given later on in this chapter.

The structure of a neuro-symbol has its paragon in the structure of biological neurons (see section 3.1) and is depicted in figure 4.1. A neuro-symbol has several inputs and one output. A neuro-symbol can receive input information from several other neuro-symbols which corresponds to the function of dendrites of neurons. The input information contains – among others – the activation grade of the symbol it originates from. The activation grades from all incoming neuro-symbols are summed up likewise in the cell body of a nerve cell. If this sum exceeds a certain threshold, the neuro-symbol is activated. The information about the activation is passed to other neuro-symbols it is connected to in analogy to the axon of a neuron. If necessary, the input information representing the activation grades from connected neuro-symbols can be weighted. This corresponds to the synaptic connections between neurons. Like in the brain, where many neurons are active at the same time, different neuro-symbols can process information in parallel.

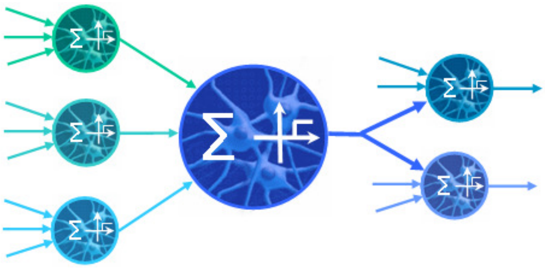


Figure 4.1: Structure of a Neuro-symbol

The just given description of the structure and function of a neuro-symbol closely resembles the structure and function of artificial neurons used in neural networks. However, there also exists a number of differences from this concept.

The first and maybe most important difference is that in artificial neural networks, there can generally not be assigned an interpretable meaning to single neurons in these networks. Therefore, neural networks are considered as black box models. From input information presented to the network, output information is calculated. It is generally not intuitively comprehensible for humans how this information is determined, because this information results from weighted connections between neurons distributed over the whole network. In contrast, when using neuro-symbols, each neuro-symbol stands for a certain perceptual image and therefore has a concrete meaning. The activation of a neuro-symbol means that the perceptual image it stands for has been perceived in the environment. Besides this, the function of weights of connections between neuro-symbols – if used – is different from the function of weights between artificial neurons. In neural networks, weights are altered by a learning algorithm to achieve that certain input values are mapped to certain output values. In neuro-symbolic networks, weights of connections are used when different sensor modalities deliver information of different reliability (see section 4.4.2).

A second difference lies in the fact that neuro-symbols can contain so-called properties. Neuro-symbols cannot only pass their activation grade to other neuro-symbols but also the current values of their properties. The utility of properties will be discussed in section 4.2.4 after having introduced methods how to structure and connect neuro-symbols to neuro-symbolic networks.

A third difference is the way how information is exchanged between neuro-symbols. For common neural networks, to calculate output values from certain input values, all the necessary information has to be present at the inputs of the network and its succeeding layers always at one instant of time. To process signals changing over time, there can be used time delay elements. However, the function of these elements is just the storage of former values in a delay chain to make available the whole needed signal information at the input neurons of the network concurrently. In contrast, information exchange between neuro-symbols is event-based, which means that information is only processed if a new input signal is received. That method allows it to reduce the communication and information processing effort. The concept of event-based information exchange will be presented in more detail in section 4.3.1. Neuro-symbols also comprise methods to handle information arriving asynchronously and certain sequences of events (see section 4.4.4).

The fact that neuro-symbols can contain properties and handle data arriving asynchronously and/or in a certain temporal succession makes clear that neuro-symbols actually do not correspond to the function of single neurons but to the function of a whole group – also called population – of neurons.

For a description of further differences between neural networks and neuro-symbolic information processing as well as for a demarcation to symbolic systems see section 6.3.2.

4.2 Neuro-symbolic Networks for Perception

In the last section, neuro-symbols were introduced as basic information processing units for the model. To become a powerful information processing tool, neuro-symbols have to exchange information and therefore have to be interconnected in a suitable manner. By structuring neuro-symbols to perform a certain task, so-called *neuro-symbolic networks* emerge. This section described how neuro-symbols need to be structured to interact and to exchange information in order to extract relevant information from sensory raw data provided by diverse sensor types. The architecture that serves as archetype for information processing is the structural organization of the perceptual system of the human brain.

4.2.1 Architecture for Modular Hierarchical Processing of Sensory Information

In section 3.2, the three principal functional units of the brain were introduced. The unit, which is of most interest for the model to be developed is the unit for receiving, analyzing, and storing information, which can be regarded as perceptual system of the brain. As described in section 3.3.2, the perceptual system of the brain has a modular hierarchical structure and consists of at least three cortical zones built one above the other. They are referred to as primary, secondary, and tertiary area. Human perception does not rely on a single modality but involves different perceptual systems: visual perception, auditory perception, somatosensory perception, olfactory perception, and gustatory perception. The somatosensory system actually comprises a whole group of sensory systems, responsible for cutaneous sensations, proprioception, and kinesthesis. Each of these senses is served by a specific type of receptor and projects separately to the brain. Each sensory modality has its own primary and secondary area located in a specific area of the brain. The primary areas receive impulses from the periphery. They consist mainly of neurons which have extremely high specificity. The cells of the secondary cortical zones have a much lower degree of specificity. In the tertiary zones, the information coming from the different sense organs being processed separately and in parallel until now in the particular primary and secondary zones is merged.

From this description, a model architecture for sensory information processing is derived. Figure 4.2 illustrates the suggested architecture for modular hierarchical information processing graphically.

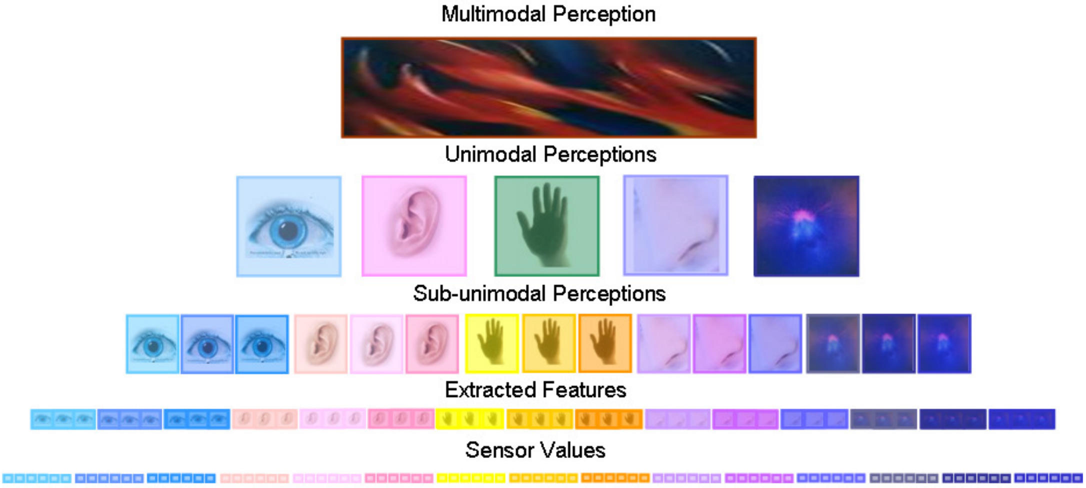


Figure 4.2: Modular Hierarchical Architecture for Sensory Information Processing

According to the model, sensors of different types, which have their analogy in sensory receptors of the human body, deliver sensor values. In a first processing stage – similar to information processing performed in the primary cortices of the different sensory modalities – the sensory raw data are pre-processed to extract *features* suitable for further processing. In section 3.3.2, it was mentioned that the primary areas of the brain have a topographic structure, which means that there is a correspondence between the position of a receptor and the area of the cerebral cortex that is activated by it. Accordingly, extracted features have a strong correspondence to the position of the sensors they are triggered by. The feature layer can in fact consist of a group of layers.

A more detailed description of the structure of this layer and its information processing principle will be given in section 4.2.6. After having extracted features from sensory raw data, in the next two stages, which correspond to information processing performed in the secondary cortices, the extracted features are combined to result in *unimodal* perceptions. As outlined in section 3.3.2 when describing the somatosensory system, unimodal perception can result from an integration of information coming from different *sub-unimodal* perceptive systems. Sub-unimodal and unimodal perceptions do no longer have a topographic projection of the corresponding sensory structures. The processing of information of each layer of the first three levels is performed separately and in parallel for each sensory modality. In a fourth stage – analogous to the processing in the tertiary cortex – the information coming from the unimodal perceptive systems is combined and merged to result in a unified *multimodal* representation of the environment.

With this modular hierarchical architecture, sensor information from various types can be processed and merged. As described in section 3.3.2, human perception involves five different perceptual systems: visual perception, auditory perception, somatosensory perception, olfactory perception, and gustatory perception. For a monitoring system that shall perceive objects, events, scenarios, and situations in a building, the use of similar modalities – at least to a certain extend – is recommendable. Additionally, it might be useful to utilize information from sensor types, which have no analogy in sense organs of the human body.

To achieve visual perception – in figure 4.2 symbolized as square with an eye – normal video cameras, stereo video cameras, infrared cameras, or retina sensors could be used. With these sensors[1], different features like the form, size, color, location, or the motion of objects can be determined. The splitting of visual information processing into a sub-unimodal layer followed by a unimodal layer becomes important when more than one of these sensor types (or more than one entity of one of these sensor types) is used. In this case, information coming from each of these sensors is first processed separately and yields to independent sub-unimodal visual perceptions. In the next stage, the information from these sub-systems is combined yielding to a unified visual perception of the environment.

For perception of acoustic information – in figure 4.2 depictured as square with the symbol of an ear in it – microphones or arrays of microphones mounted at different positions can be used. The use of more than one microphone is useful as through differences in signal runtime and signal amplitude, the location of a sound source can be determined. Again, the information recorded from each microphone or array of microphones is first processed separately in the sub-unimodal layer before being merged in the unimodal layer to a unified acoustic representation of a situation.

The squares in figure 4.2 with the picture of a hand inside stand for somatosensory perception. For surveillance systems in buildings, sensors with similarities to the receptors of the tactile sensation of somatosensory perception are particularly useful. Therefore, in the following, the somatosensory sense will be referred to as tactile perception. Sensors utilizable for the purpose of monitoring systems are, among others, tactile floor sensors, motion detectors, light barriers, door contact sensors, pressure sensors, or distance sensors[2]. The principle of processing sensory information in a sub-unimodal and a unimodal stage as describe for visual and auditory perception is also applicable to the tactile system.

[1] Cameras are considered as an array of sensors and are therefore also labeled as sensors in this text.

[2] If the target system would not be a surveillance system but a robot, additionally, sensors similar to proprioception – the "body sense" – would be useful, which could be angle transmitters and displacement sensors to determine the position and/or orientation of joints, actuators, end effectors, and wheels and the velocity or acceleration of their motion if they are currently moving.

The olfactory and gustatory senses of the human body are, in contrast to the other senses, chemical in nature. For an application in building automation, there might probably not be the need for a gustatory sense. However, sensors comparable to the olfactory receptors could be smoke detectors or chemical sensors, which detect the presence of certain substances in the air. In figure 4.2, the symbol for olfactory perception is a square with the image of a nose inside.

In measurement engineering, there also exist sensor types, which find no correspondence in a sensory receptor of the human body. However, for the purpose of surveillance systems, the utilization of some of these sensor types might be advantageous. One example are measurements of electrical power consumption. That way, it could be detected if there are plugged in electric appliances in sockets and how much power they consume, which could allow to conclude what kind of electric appliance is plugged in. In figure 4.2, such additional perceptual modalities are depicted as squares with an abstract symbol inside.

Concept Clarification

To make the proposed architecture more comprehensible, it is applied to a first simple, concrete example, which will be reused and extended for explanations in further sections. There shall be detected different activities in a room wherefore a room is equipped with a number of sensors: a motion detector, two tactile floor sensors, two light barriers, a camera, and a microphone. The sensors are mounted at different positions as depicted in figure 4.3. They have the property to have partly overlapping sensory fields of perception and to provide partly redundant information.

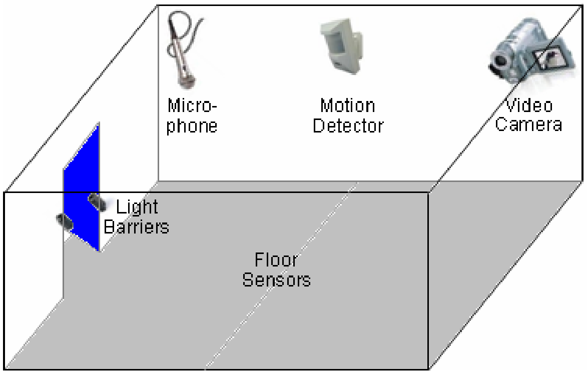


Figure 4.3: Environment Equipped with Different Sensors

Figure 4.4 illustrates the modular hierarchical information processing structure for this example. Multimodal perception is achieved by combining data from three unimodal perceptive systems, which are – in analogy to their biological archetypes – referred to as visual perception, auditory perception, and tactile perception. The tactile perceptive system integrates information coming from three tactile sub-systems corresponding to the three different sensor types used. The visual and the auditory perceptual systems have no further sub-modalities, because only one camera and one microphone are used, respectively.

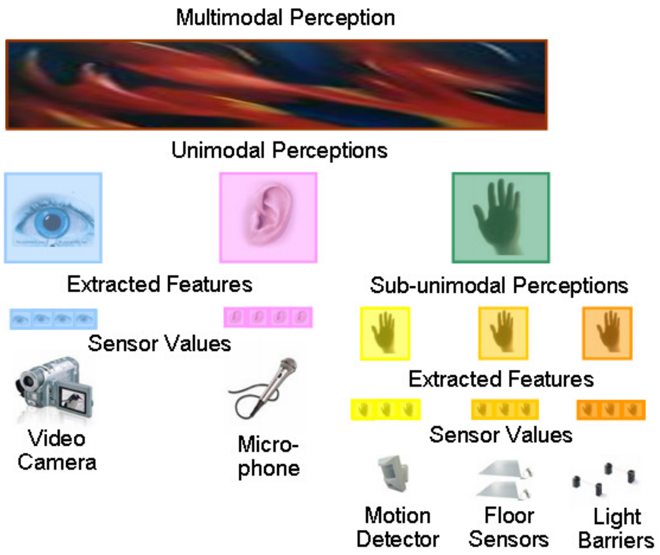


Figure 4.4: Architecture for Modular Hierarchical Information Processing for a Concrete Example

4.2.2 Modular Hierarchical Arrangement of Neuro-symbols

Following the strategy of modular hierarchical information processing described in section 4.2.1, neuro-symbols – in the following due to ease of writing also simply referred to as symbols – can occur in different hierarchical levels. According to the level and the modality in which they occur, they represent different information. Therefore, they are named differently. In figure 4.5, an overview about the labeling of the symbols of the different hierarchical levels is given.

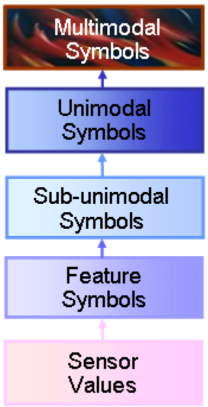


Figure 4.5: Neuro-symbol Hierarchy

To process sensor information neuro-symbolically, sensor values have to be transformed into neuro-symbolic information. In similarity to information processing in the primary cortex of the brain, relevant features have to be extracted from the sensory raw data in a first pre-processing step. These extracted features are represented by *feature symbols*. In the next stage, the feature symbols of each sensory domain are either directly merged to *unimodal symbols* if the modality has only sensors of one single sensor type or – if different sensor types are used – they are first combined to *sub-unimodal symbols*, which are then processed to unimodal symbols. In a further processing stage, unimodal symbols are merged to *multimodal symbols*. Depending on the sensory modalities used in a system, unimodal symbols can be further classified for example as visual unimodal symbols, acoustic unimodal symbols, tactile unimodal symbols, olfactory unimodal symbols, as well as other unimodal symbols that do not have an analogy in the human body. In a similar way, sub-unimodal symbols and feature symbols can be specified in more detail.

Figure 4.6 comprises a schematic representation how neuro-symbols are structured in different levels and different modalities according to the modular hierarchical architecture introduced. Sensor values are represented by colored squares. Sub-unimodal, unimodal, and multimodal symbols are pictured as colored circles with an image inside. Feature symbols are depicted as colored squares with an image inside. Features symbols have another form than higher-level symbols, because they are halfway between sensor data and symbols. In contrast to the other symbol levels, they have a topographic structure and there can exist more than one level of feature symbols (see section 4.2.6).

Each neuro-symbol represents a certain perceptual image. Concrete examples for neuro-symbols in the different layers and modalities will be given later on in this chapter. The number of neuro-symbols used in a certain level and a certain modality is principally not constraint and depends on the requirements of an application. Before initial system startup, a set of neuro-symbols is defined. During a learning phase, correlations between neuro-symbols are determined, additional neuro-symbols can be added, and redundant neuro-symbols can be removed (see section 4.5).

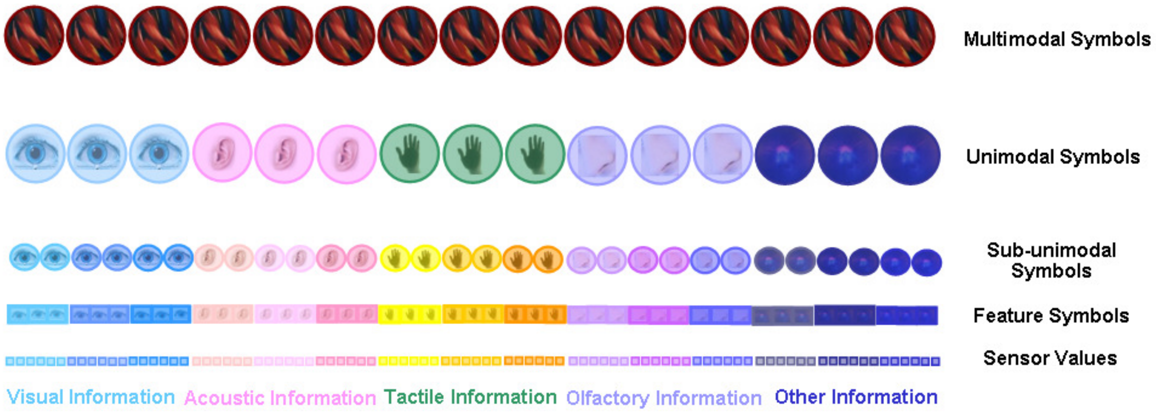


Figure 4.6: Modular Hierarchical Arrangement of Neuro-symbols

4.2.3 Information Flow between Neuro-symbols

In order perform complex tasks, neuro-symbols of different modalities and hierarchical levels need to interact and to exchange information. In neuroscience and neuropsychology, the question how information from different distributed sources is merged is called the binding problem, which was explained in section 3.4. Years of research have not yet given a final answer to the binding problem. For the model developed in this thesis, there also needs to be solved such a binding problem. Within this chapter, mechanisms and concepts will be introduced step by step to overcome the binding problem for neuro-symbolic information processing in perceptive tasks. The current section discusses the flow of information between neuro-symbols of and within different modalities and hierarchical levels. In section 3.3, it was mentioned that bottom-up and top-down processes influence perception. Accordingly, it is distinguished between an information flow that is directed bottom-up and one that is directed top-down. Additionally, in the human brain, there exist feedbacks between neurons [MJS07]. Analogously, in the model, information can be exchanged through so-called feedback-loops.

Bottom-up Information Flow

As outlined in section 3.3.1, in the brain, incoming sensor data are always the starting point for perception. Figure 4.7 illustrates the principal of bottom-up flow of information for neuro-symbolic networks. Bottom-up information flow always goes from lower levels to higher levels. The colored lines show what lower-level symbols can principally send information to which higher-level symbols. Sensor information of one sensor type can only be passed to the corresponding feature symbols and the feature symbols can only send information to the according sub-unimodal symbols. In the unimodal level, each modality can receive information from all its sub-modalities. Finally, on the multimodal level, information from all unimodal sources is processed.

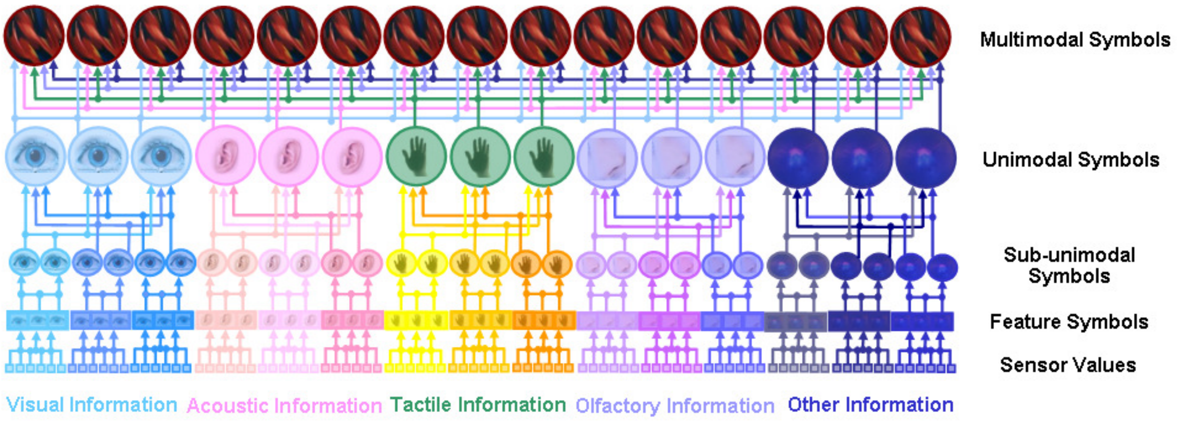


Figure 4.7: Bottom-up Information Flow

Feedback Loops

As already mentioned, in the brain, there do not only exist forward connections but also feedbacks between neurons. However, the function of these feedback connections is largely unknown [MJS07]. One useful function that can be performed by feedbacks between neuro-symbols is the inhibition of undesired concurrent activations of neuro-symbols within one modality. To suppress such concurrent activations, feedbacks from the outputs of neuro-symbols of one modality to the inputs of neuro-symbols of the same modality are necessary (see figure 4.8). For a more detailed explanation see section 4.2.5. In the model, feedback connections can exist in the sub-unimodal, the unimodal and the multimodal layer[3]. In the feature symbol layer, which in fact can consist of a group of layers, the problem of undesired activations of symbols is handled without feedbacks (see section 4.2.6).

[3] Additionally, they can exist in the scenario symbol level, which will be introduced in section 4.4.4

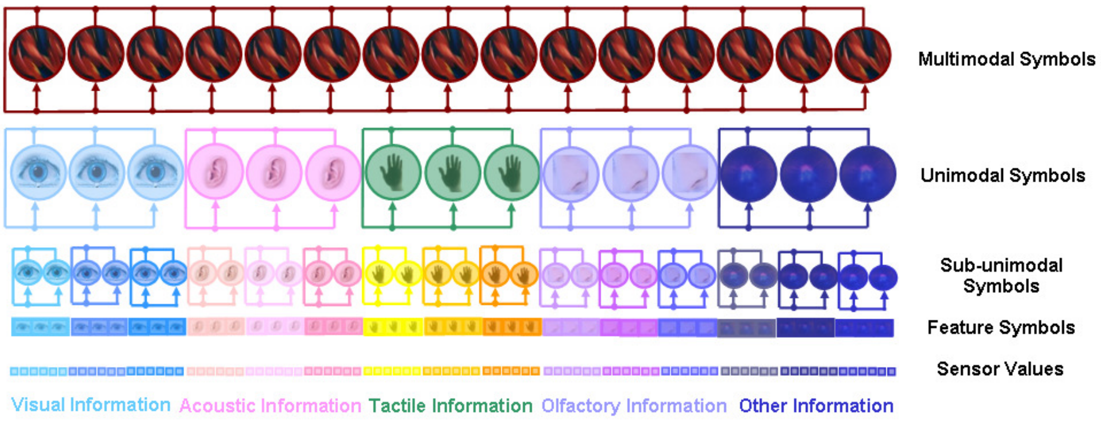


Figure 4.8: Feedback Loops

Top-down Information Flow

In the brain, besides processing information coming from sensory receptors, perception is also influenced by higher cortical processes labeled as top-down processes. Top-down processes that have effect on perception are knowledge (see sections 3.3.3 and 3.4.5) and focus of attention (see section 3.4.5). Accordingly, in the model of neuro-symbolic information processing, neuro-symbols can receive information from such sources, which are labeled as "cognitive information" in figure 4.9. They are referred to as cognitive information, because higher cortical processes are involved in the generation of this information. Cognitive information can principally influence perception on different hierarchical neuro-symbolic levels. For a more detailed description about the integration of top-down processes in perception see sections 4.4.5, 4.4.6, and 3.3.3.

In the figures 4.7 to 4.9, it was illustrated what neuro-symbols and what higher-level modules can principally be connected and exchange information. Due to ease of illustration, there was always depicted only one connection line between different units. However, this connection line represents in fact potential point to point connections between the different units. In a neuro-symbolic network, which has already been configured to perceive certain objects, events, scenarios, and situations in the environment, there do not exist point to point connections between all of these units but only between a subset of them.

Concept Clarification

To clarify the concept of neuro-symbolic information processing, a first very simple example shall be given, which only considers bottom-up information processing. Explanations including feedback loops and top-down information processing will be given in later sections.

In the given example, it shall be detected whether a person enters a room. For this purpose, a room is equipped with five different sensors: a video camera, a microphone, a motion detector, a tactile floor sensor, and a light barrier. Figure 4.10 shows where in the room the sensors are mounted. The detection range of all sensors is directed towards the door.

Figure 4.11 illustrates the symbol hierarchy for detecting that a person enters the room through a door. To perform this task, there have been defined the symbols "person enter", "person", "steps", "object enters", "motion", "object present", "object passes", as well as a number of

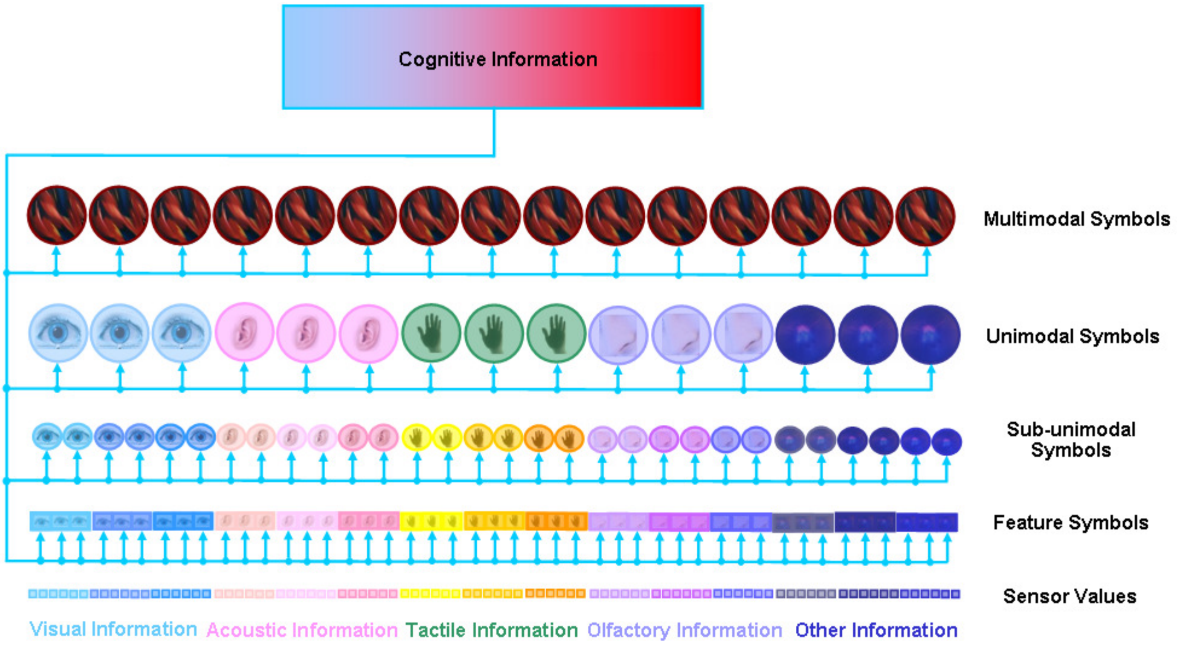


Figure 4.9: Top-down Information Flow

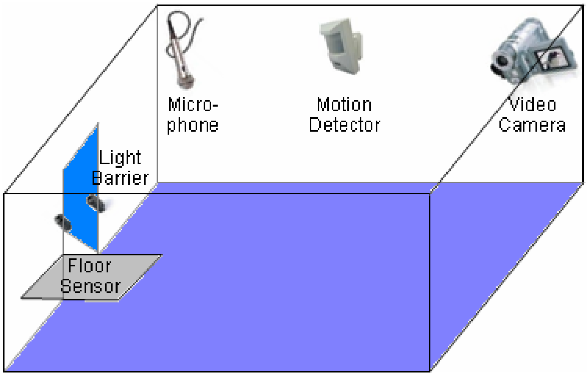


Figure 4.10: Environment Equipped with Different Sensors for Detecting that a Person Enters the Room

feature symbols. To indicate that the system has perceived that a person entered the room, the symbol “person enters” needs to be activated.

Information processing starts with sensor values. In a first processing stage, feature symbols – pictured as squares – are activated if certain sensors are triggered. In the presented example, for the case of the sensors of the tactile modalities (floor sensor, motion detector, light barrier), the associations between the sensor values and the feature symbols are very simple and intuitively clear due to their binary sensor output values and the fact that there exists only one sensor of each type. Coherences get more complicated if the number of sensors increases. The feature symbols for the visual and the auditory modality are more complex. Feature symbols of the visual modality are for example edges, lines, curves, colors, forms, etc. derived from pixel information. Feature symbols for auditory processing are for instance spectral components of a sound signal.

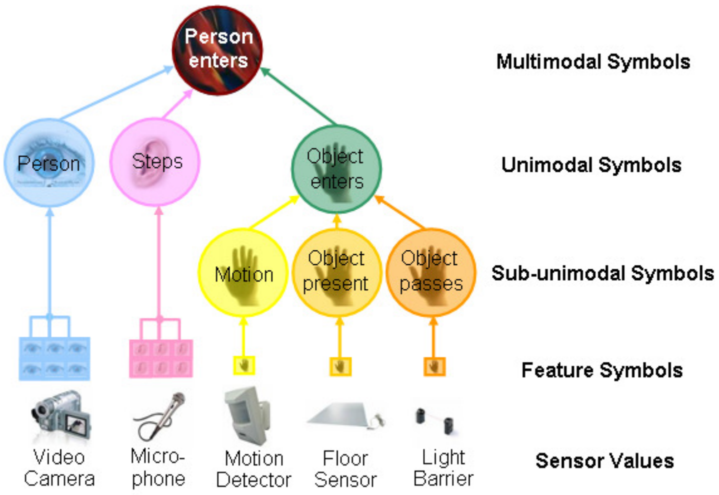


Figure 4.11: Symbol Hierarchy for Detecting that a Person Enters a Room

For the tactile modalities, sub-unimodal symbols are activated when certain feature symbols are active. The activated sub-unimodal symbol corresponding to the motion detector indicates that there is perceived motion near the door. From the tactile floor sensor it is derived if an object is present near the door. From the light barrier the information is extracted whether an object passes the door. The information of these sub-unimodal tactile systems is combined to a unified tactile perception, which indicates if an object enters the room. The sub-unimodal and unimodal symbols of the tactile perceptive system represent states and events of objects. They are not directly associated with states and activities of a person, because the sensors could also be triggered by something else like an animal moving in the room or an object positioned in the room. In case of the visual and auditory sense, only always one camera and one microphone are used. Therefore, the unimodal symbols can be directly extracted from feature symbols without a sub-unimodal processing stage. Due to ease of illustration, there is depicted only one connection line from the feature symbols to the unimodal symbols. In fact, there exists always a point to point connection between the output of feature symbols and the input of the corresponding higher-level symbols. From the visual feature symbols it is detected whether a person is present in the room. From the auditory feature symbols it is extracted whether the characteristic noise of steps is perceived. It has to be mentioned that visual image processing and auditory data processing are huge research fields. There might already exist workable solutions to recognize persons directly from images or to detect the noise of steps directly from audio data. If this is the case, it is recommendable to use these existing solutions to generate unimodal symbols and skip the step of explicitly generating feature symbols. However, implicitly, these algorithms also extract features from the raw data. For further information about the integration of such workable solution to solve different sub-problems see section 4.7.

Finally, in the multimodal layer, the information coming from the visual, acoustic, and tactile unimodal layers is combined and results in the activation of the neuro-symbol “person enters”.

4.2.4 Properties of Neuro-symbols

In neuro-symbolic networks, neuro-symbols can have properties. Properties comprise information that specifies the neuro-symbols in more detail. The concept of properties can reduce the number of necessary different neuro-symbols to detect an object, event, or situations. Examples for properties are the location where objects, events, or situations occur, the size of an object, or the velocity and direction of a moving object. Properties can have certain values. The information what value a property currently has needs to be passed to other neuro-symbols. In section 4.1, it was described that neuro-symbols pass a signal to the other neuro-symbols they are connected with to "inform" them about their current activation state. Besides indicating their activation, there can also be transmitted information about the values of properties. Neuro-symbols with properties can be compared to a group of neurons in the brain, which interact to represent congeneric perceptual images. One property that is of special importance for merging information of different sensory sources when events are happening concurrently in the environment is the property about the location where a certain perceptual image is perceived (see section 4.4.3).

Concept Clarification

The use of properties shall be explained by extending the example described in section 4.2.3. It shall now not only be detected if a person enters the room, but also if he/she leaves the room. A simple way to distinguish these two cases is to mount a second light barrier near the first one. By combining the information of both light barriers, it can be detected which of them is triggered first when an object passes. This calculation is carried out by the feature symbol layer. Out of this, the direction of such a pass can be determined. For the sub-unimodal layer, which processes information coming from the light barrier, two possible strategies exist for symbolic representation. The first strategy, which is depicted in figure 4.12a, is to distinguish the two cases by using two different sub-unimodal symbols: "object moves in" and "object moves out". This corresponds with the neuroscientific combination coding theory described in section 3.4.5. However, if not only two but more cases have to be distinguished (for example the velocity of a moving object with the fuzzy values very slow, slow, medium, fast, and very fast), the number of possible neuro-symbols would increase drastically. In neuroscience, this problem is referred to as combinatorial problem. To overcome this problem, neuroscience proposes population coding according to which perceptual images are coded by a group of neurons. Properties of neuro-symbols have a similar function. A neuro-symbol with a property corresponds to a population of biological neurons, which work concertedly.

Figure 4.12b illustrates how the neuro-symbols "object moves in" and "object moves out" can be reduced to one sub-unimodal symbol "object passes" with the property "direction d", which can be "in" or "out". In the unimodal tactile layer, by this single symbol "object passes", either the symbol "object enters" or "object leaves" can be activated depending on the value of the property "direction d". It would also be possible to use only one symbol instead of these two and add a property to it to distinguish these cases. Both methods are possible in the model. The use of properties becomes especially valuable if they do not have only two but more different values. The decision to model a certain symbol with a property or to generate separate symbols for each case is taken by the system engineer.

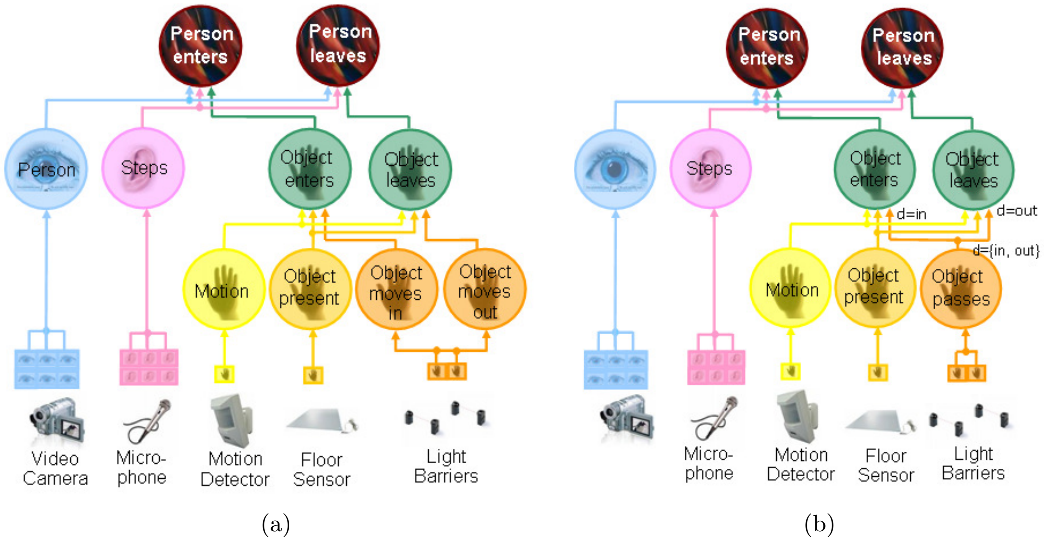


Figure 4.12: Usage of Properties to Reduce the Number of Neuro-symbols

4.2.5 High-level Neuro-symbolic Information Processing

In the perceptual system of the brain, neurons and groups of neurons of different cortical layers show different characteristics (see section 3.3.2). Similarly, in the model, there exist certain differences in neuro-symbolic information processing of higher levels and lower levels. In the model, information processing from the sub-unimodal level upwards is regarded as high-level neuro-symbolic information processing (see figure 4.13). Information processing from sensor data up to the sub-unimodal layer is considered as low-level information processing (see figure 4.17). Characteristics of high-level neuro-symbolic information processing are outlined in this section. Characteristics of low-level information processing are described in section 4.2.6.

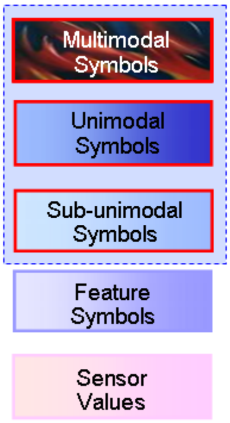


Figure 4.13: High-level Neuro-symbolic Information Processing

Representation of Location Information

High-level neuro-symbolic information processing corresponds to information processing performed in the secondary and tertiary cortex of the perceptual system of the brain (see sections 3.2 and 3.3.2). In these cortical zones, neurons generally respond to perceptual images independent of the location of these images. E.g., there have been found neurons in the secondary visual cortex, which respond to faces independently from the distance of the face and the position where it is perceived in the visual field [Gol02]. According to this, in the model, neuro-symbols in the sub-unimodal, unimodal, and multimodal layer can be activated independently of the position where in the environment the perceptual images they represent have been perceived. Therefore, neuro-symbols of these layers contain location information only as a property. Location information is of urgent importance for merging information from different modalities when different events happen in the environment in parallel. A more detailed discussion about the utility of location information for different purposes is given in section 4.4.3.

Learning of Correlations

At birth, the brain is not completely hardwired. Lots of connections between neurons are only set in later development stages based on learned concepts and correlations. In section 3.2, it was pointed out that in the perceptual system of the brain, higher cortical levels can only evolve if lower ones have already developed. Therefore, in the model, connections between neuro-symbols from the sub-unimodal layer upwards are not predefined at initial system startup but have to be learned during a learning phase. Learning is performed based on examples. A detailed description about neuro-symbolic learning is given in section 4.5.

Interaction with Knowledge

In section 3.3.3 and section 3.4.5, it was described that knowledge can have influence on perception and the way how information from different sources is merged. However, up to now, scientists from neuroscience and neuropsychology do not agree about the question at which level knowledge interacts with perception. For the model presented in this thesis, it was decided that knowledge can principally interact with perceptual images – which are equivalent to neuro-symbols – not on the level of feature symbols or even sensor values but on higher neuro-symbolic levels. For a more detailed discussion see section 4.6.

Feedback of Information and the Single Symbol Activation Principle

When describing the flow of information between neuro-symbols in section 4.2.3, it was mentioned that information cannot only be directed bottom-up and top-down, but that there can also exist feedback loops between neuro-symbols. In the brain, the function of feedback connections between neurons is largely unknown. In the model of neuro-symbolic information processing, a useful function that can be performed by feedbacks between neuro-symbols is the inhibition of undesired concurrent activations of neuro-symbols within one modality. In the model, the employment of feedback connections is reserved for the higher neuro-symbolic layers from the sub-unimodal level upwards. A detailed description of the utility of feedback connection is given in the following:

In a neuro-symbolic network, each neuro-symbol represents a certain perceptual image. The activation of a neuron-symbol means that the perceptive image it stands for has been perceived in the environment. One important question for system design is if for one single object, event, scenario, or situation happening in the environment, one or more neuro-symbols shall be activated in each modality of each hierarchical level. This decision gets especially important if more than one object, event, or situation occur concurrently. In such a case, activated lower-level symbols have to be assigned to the correct higher-level symbol. To minimize the amount of incorrectly assigned symbols, a design strategy called *single symbol activation* is applied, which is illustrated in figure 4.14.

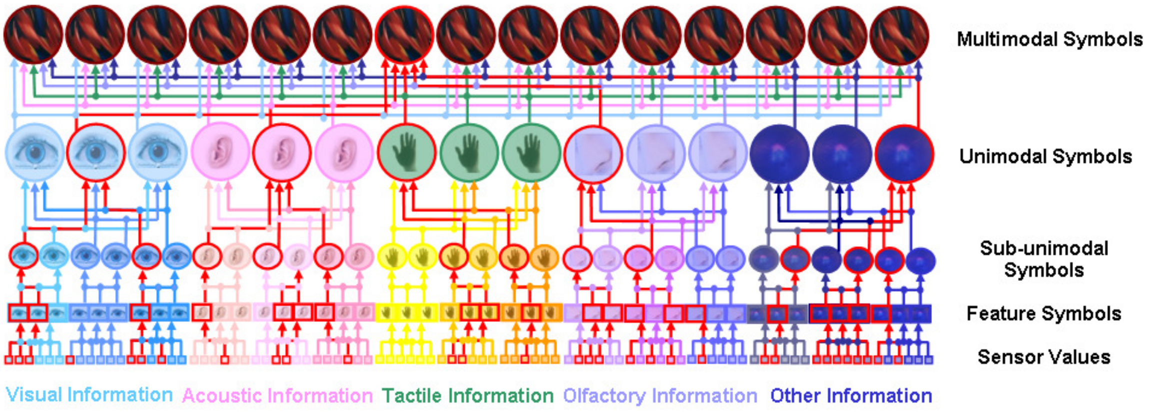


Figure 4.14: Concept of Single Symbol Activation in Neuro-symbolic Information Processing

According to this principle, one event only triggers one symbol of each sub-modality and modality and only one neuro-symbol in the multimodal layer. On the multimodal level, one situation or event is always correlated with one and only one multimodal neuro-symbol. On the unimodal symbol level, for this event or situation, no more than one unimodal neuro-symbol can be activated in each existing modality. It may also occur that not all existing modalities are involved in the activation of a neuro-symbol corresponding to a certain event or situation. Viewed from another point, it could also be said that from the activated neuro-symbols of the different unimodal systems, only one multimodal symbol shall be triggered. Like on the multimodal and unimodal level, on the sub-unimodal levels, one single object, event, or situation shall only activate one single neuro-symbol in each sub-modality. The principle of single symbol activation for each level and modality is particularly useful if connections between symbols are not predefined but learned from examples, which will be outlined in section 4.5.

In contrast to the higher symbol levels, on the feature symbol level and the sensor level, the number of feature symbols and sensors activated by one event is not limited in the different modalities and sub-modalities. The reason why several feature symbols and sensor values can be assigned to one event lies in the fact that they have a topographic representation which means that they have a strong correlation to the position of the sensors they are derived from. Besides this, correlations between these layers are predefined and the problem of learning, which is facilitated by the single symbol activation principle, does not occur in these layers (see also section 4.2.6).

To assure that only one symbol is activated in a certain modality at higher layers, feedbacks between the symbols of this modality need to be introduced. Feedbacks can occur in all hierarchical levels from the sub-unimodal layer upwards. Principally, feedback connections could have inhibitory as well as excitatory influence on the activation grade of a neuro-symbol. However, for the concept of single symbol activation, only inhibitory feedbacks are used. These inhibitory feedbacks are necessary if there exist higher-level neuro-symbols, which are activated by a number of lower-level symbols that are a subset of the lower-level symbols of another symbol of the same level and modality. In such a case, more than one symbol would be activated by one single situation in the particular hierarchical level and modality, which contradicts the principle of single symbol activation. In such a case, inhibitory feedbacks need to suppress unwanted concurrent activations of more than one symbol.

Concept Clarification

The concept of single symbol activation and feedbacks shall be clarified by the example already mentioned in section 4.2.1. It shall be perceived if a person enters a room, leaves a room, walks around in the room, or stands in the room. Therefore, a room is equipped with a video camera, a microphone, a motion detector, two tactile floor sensors, and two light barriers (see figure 4.3). The detection range of the video camera, the microphone, and the motion detector cover the whole room. The light barriers only detect passes of objects between the door jambs at a certain height. One tactile floor sensor covers the left half of the room and the area around the door of the room. The other tactile floor sensor covers the right half of the room. To keep the explanation simple, it is assumed that there is always only one person in the room. Figure 4.15 shows the neuro-symbol structure that is used to perceive the different situations. The visual and the auditory perceptive systems have the same structure as already shown in section 4.2.4. Also the sub-unimodal tactile layer is quite similar. The only difference is that the neuro-symbol "object present" now also has a property "location l", which can be "left" or "right" depending on which of the two floor sensors is triggered. The unimodal tactile layer now has four different symbols: "object stands", "object moves", "object enters", and "object leaves". The symbols "object stands" and "object moves" can be triggered no matter what value the property "location l" currently has. The symbols "object enters" and "object leaves" are only activated if the property "location l" has the value "left". Depending on the activation of the unimodal symbols from the visual, auditory, and tactile systems, one of the four multimodal symbols is activated.

Taking a closer look at the unimodal layer of the tactile system in figure 4.15, it attracts attention that the neuro-symbols "object stands" and "object moves" are connected to a subset of the sub-unimodal symbols that trigger the symbols "object enters" and "object leaves". This means that whenever the symbol "object enters" or "object leaves" is activated from the corresponding event in the environment, there are also activated the symbols "object present" and "object moves", because they result from a combination of a subset of the same sub-unimodal tactile symbols. For the same reason, the symbol "object stands" is also triggered whenever the symbol "object moves" is activated. To overcome this undesired activation of more than one symbol at a certain moment, inhibitory feedbacks are inducted. Their task is to inhibit that more than one neuro-symbol of a modality is triggered by one and the same event. In figure 4.16, these inhibitory feedback connections are depicted as dotted lines. If a certain symbol is activated, the output signal comprising its activation grade is not only transmitted to the next higher symbol level but also to symbols of the same level it is connected to via feedback connections. For an activated neuro-symbol that receives such an inhibitory feedback signal, the activation grade is decreased in a way that it falls below the threshold value and the symbol is deactivated.

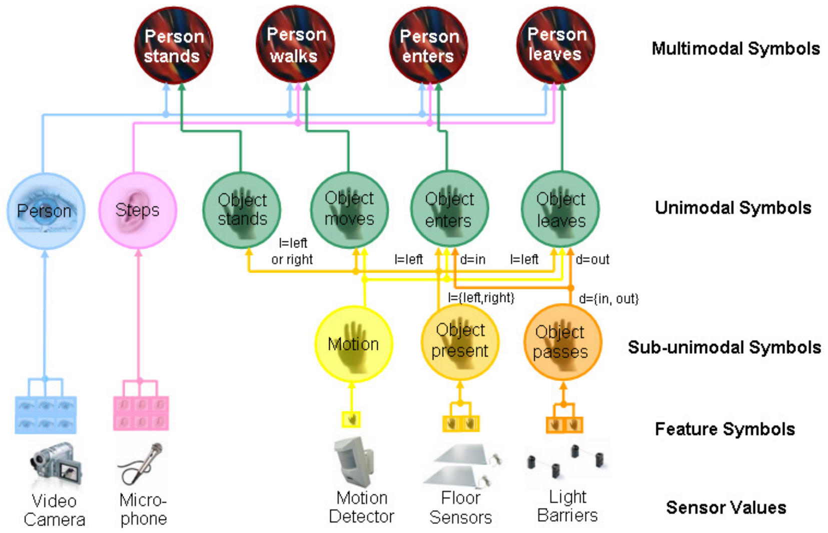


Figure 4.15: Neuro-symbol Hierarchy without Feedbacks

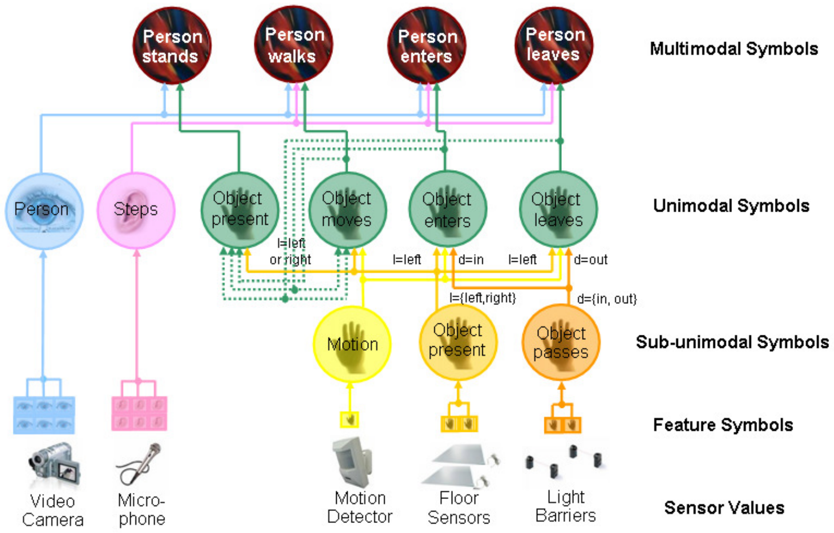


Figure 4.16: Neuro-symbol Hierarchy with Feedbacks

4.2.6 Low-level Neuro-symbolic Information Processing

In the model, information processing from sensor values up to the sub-unimodal level is regarded as low-level neuro-symbolic information processing (see figure 4.17)[4]. Neuro-symbolic information processing in the lower layers of a neuro-symbolic network differs from higher-level neuro-symbolic information processing in some characteristics. These characteristics are outlined in the following.

[4] For the case that there does not exist a sub-unimodal layer for a certain modality, feature symbols are directly merged to unimodal symbols. In this case, the mechanisms normally taking place between the feature symbol level and the sub-unimodal layer occur between the feature symbol level and the corresponding unimodal level.

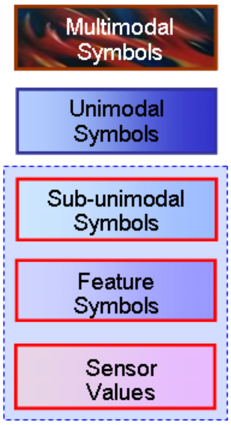


Figure 4.17: Low-level Neuro-symbolic Information Processing

Representation of Location Information

Low-level neuro-symbolic information processing corresponds to information processing performed in the primary and partly in the secondary cortex of the perceptual system of the brain. As described in section 3.3.2, the primary cortices of the different sensory modalities have a topographic structure. This means that there is a correspondence between the position of a receptor and the area of the cerebral cortex that is activated by it. In the primary cortical zones, neurons generally respond to simple perceptual images located at a certain position. These types of cells are also called feature detectors [Gol02]. Accordingly, in the model, feature symbols have a strong correspondence to the position of the sensors they are derived from. The location dependency of neurons gets weaker in higher neural layers until it vanishes completely. In the secondary cortex of the visual system, there have been found neurons that respond specifically to faces independently of their location, orientation, or size. In the model, the sub-unimodal level already corresponds to information processing in the secondary cortex of the brain. Therefore, from the sub-unimodal layer upwards, location information is handled only as property and a symbol can be activated independently of the location where the perceptual image that it represents is perceived. For that reason, there has to be performed a transition from location dependent perceptual images to location independent images. How such a transition can look like is described by means of a concrete example in appendix A. Besides this, a more detailed discussion about the usage of location information is given in section 4.4.3.

Predefinition of Correlations

As outlined in section 3.2, at birth, the brain is not completely hardwired. However, there have to exist already certain lower-level connections to allow learning in higher cortical levels [Lur73]. Therefore, in the model, lower-level connections between sensor data and feature symbols are predefined before system startup. Depending on the desired task, correlations between feature symbols and sub-unimodal symbols can either be predefined or can be learned. However, if the correlations are learned, the learning methods might differ from learning methods applied in higher layers (see section 4.7). The setting of feedback connections between sub-unimodal symbols is already subject to neuro-symbolic learning. The transformation of sensor data into feature symbols is the first important step for efficient information processing. As outlined in section 3.3.2, in different sensory modalities, there exist very different features that neurons respond to in the primary cortex. Accordingly, it strongly depends on the particular sensor type that is used how an efficient transformation of sensor information into feature symbols has to look like. However, certain principles are always the same. In appendix A, it is explained for one concrete modality how to get from sensor values over features symbols to sub-unimodal symbols. The basic information processing principles can be taken over for other sensory modalities.

Different Levels of Feature Symbols

As mentioned in section 3.3.2, there exist nerve cells in the primary visual cortex that respond exclusively to shades of color, the character of lines, edges, angles, balks of a specific length, orientation, or the direction of movement. The starting point for all these different images is information coming from only two kinds of visual receptors – rods and cones – which have different properties. [HW62] suggested a model of a "simple to complex" hierarchy for information processing in the primary visual cortex. According to this model, the primary cortex consists of a feed forward sequence of more and more complex and invariant neuronal representations. Complex cells receive input from several simple cells. Simple cells have a small receptive field. Complex cells have a bigger receptive field. Inspired from these research findings, the feature symbol level of the proposed model consists in fact not only of a single layer but comprises a whole group of layers with a hierarchical structure. Certain feature symbols are derived directly from sensor values of a certain type. Other feature symbols, which are located in higher levels process information coming from lower-level feature symbols. Feature symbols of higher hierarchical layers are generally more complex and cover a bigger receptive field. The principle of information processing in the feature symbol level is further clarified by the explanation given in appendix A.

Concurrent Activation of Feature Symbols

Unlike for neuro-symbols from the sub-unimodal layer upwards, there exist no feedback connections between feature symbols. This fact is in accordance with the description of the primary visual cortex as a feed forward sequence of more and more complex neural representations mentioned in section 3.3.2. In contrast to higher layers, there can be activated more than one feature symbol by one single object, event, scenario, or situation. Therefore, there exists also no necessity for a mechanism of feedback as described in section 4.2.5. How to get from concurrent activations of feature symbols to single symbol activations in the sub-unimodal level is outlined in appendix A for a concrete example.

Interaction with Focus of Attention

In section 3.4.5, it was pointed out that focus of attention can help to correctly assign information coming from different sources if perception is overloaded. An overloading of perception means that there are too many perceptual stimuli present at a moment to integrate them all at once into a unified perception. Similar like in human perception, in the model, it can also happen that lower-level symbols cannot be correctly combined to higher-level symbols if too many events occur concurrently. For such a case, there has been integrated a focus of attention mechanism into the model. By this mechanism, the spatial area within which perceptual images of different modalities are merged is constrained. In the model, focus of attention interacts with perception on the feature symbol level. A more detailed description about the interaction mechanism is given in section 4.4.5.

4.3 The Neuro-symbolic Code

In section 4.1, neuro-symbols as basic information processing units were introduced. Section 4.2 comprised a description how to structure and interconnect these neuro-symbols to neuro-symbolic networks. To perform the desired tasks, these connected neuro-symbols have to exchange information. The signals occurring between biological neurons to exchange information are referred to as neural code (see section 3.1). Analogously, the information transmitted between neuro-symbols is labeled *neuro-symbolic code*. How this neuro-symbolic code looks like is described in the following.

4.3.1 Event-based Neuro-symbolic Information Exchange

In section 3.1, the working principle of neurons was explained. To interact, neurons have to exchange information with other neurons they are connected with. It is generally agreed that neurons transmit information about their synaptic inputs through spike trains. However, the code by which this information is transmitted – the neural code – is not yet well understood. The traditional view in systems physiology is that it is the mean firing rate alone that encodes the signal, and that variability about this mean is noise. An alternative view, which has recently gained increasing support, is that it is the variability itself that encodes the signal. As the principles of the neural code are not yet clear, it does not make sense to take it as archetype for information exchange in the suggested model. Besides this, an information exchange by spike trains may be necessary or advantageous between neurons, because transmitted signals are electrochemical in nature. However, neuro-symbols are intended for the usage in technical systems and are therefore detached from a chemical basis. There might be no need for modeling signals by spike trains. Furthermore, specially if the model is not realized in a hardware structure but simulated on a computer – which will be certainly the case during the test and evaluation phase of the model but quite probably also for a range of applications – an exchange of information between units every instant of time would require a lot of computational resources. In the case that a system is not just equipped with a small number of sensors like in the concrete examples presented until now but with a vast amount of different sensors, lots of calculations would have to be performed every instant to process all the sensory information from the feature level up to the multimodal level. Therefore, to pare down computational expenses for information exchange between neuro-symbols, a method of event-based information exchange is used, which is described in the following.

In section 4.1, it was outlined neuro-symbols receive information about the activation grade of other neuro-symbols via their input. Unlike in its biological archetype, this activation grade is not represented by spike trains but by an analogous signal with a value between *zero* and *one*. Zero stands for deactivation, one stands for full activation of the neuro-symbol. The activation grade is calculated as normalized sum of the activation grades of all input signals. Therefore, the sum of input activations is divided by the number of inputs. In case of weighting the inputs (see section 4.4.2), this weighting has to be considered in the calculation additionally. If the activation grade of a neuro-symbol exceeds a certain threshold, the symbol is activated. The activation of a neuro-symbol indicates that the perceptual image it represents has been perceived in the environment.

Information about the activation grade of neuro-symbols is not transmitted continuously but only if changes in sensor values, in activation grades of neuro-symbols, or in the values of their properties occur. This strategy of information exchange is referred to as *event-based neuro-symbolic information exchange*. Every time that at least one input value of a neuro-symbol changes, its activation grade is recalculated and data are sent to neuro-symbols connected to the output. By using event-based information exchange, computational power can be pared down, because a transmission and recalculation only needs to be performed if changes in the environment are perceived. The usage of event-based neuro-symbolic information processing makes it necessary that every neuro-symbol can store the information about the value it received via an input until this value is overwritten by new incoming data, which is easily realizable in a technical implementation.

Concept Clarification

To make the principle of event-based neuro-symbolic information processing even better comprehensible, a concrete example is given. Similar like in section 4.2.3, it shall be detected if a person enters a room. Therefore, information from a video camera, a microphone, a motion detector, a tactile floor sensor, and a light barrier is used. In the example, all neuro-symbols have a threshold value of 0.95. Figure 4.18a shows the situation that there have already been triggered certain sensors and accordingly, there have already been activated the neuro-symbols “person”, “steps”, and “object present”. Figure 4.18b now shows the case that the motion detector begins to detect motion. In the figure, connections between which information is currently exchanged are depicted as red lines. The sensor data coming from the motion detector activate the corresponding feature symbol, which in turn sends information via its output and activates the sub-unimodal symbol “motion”. This symbol sends a message to the symbol “object enters”, which now recalculates its activation grade. Although its sum of activations is below the threshold value, a message is sent to the symbol “person enters”, which also recalculates its activation grade, but is also not activated. Figure 4.18c now shows the situation that the light barrier is activated additionally. Accordingly, the corresponding feature symbol is activated, which in turn activates the sub-unimodal symbol “object passes”. From this symbol, a message is sent to the unimodal symbol “object enters”, which recalculates its activation grade and is now activated. By sending a message to the multimodal level, the symbol “person enters” is also activated.

4.3.2 Neuro-symbolic Activation Grades and Thresholds

As outlined in section 4.3.1, whenever a change in input values of a neuro-symbol occurs, the neuro-symbol transmits information about its recalculated activation grade via its output. The information sent depends on whether the neuro-symbol is currently active or not.

In section 3.1, when explaining the function principle of neurons, it was described that for one instant of time, neurons can only be either activated or deactivated. They send an action potential via their axon if the sum of their inputs is above a threshold value and they remain silent if it is below the threshold. However, neurons do not only fire once but fire in spike trains with a certain frequency. The intensity of their activation can be coded in the frequency of firing and the variability about this mean, respectively.

For the presented model, information about the activation grade is not coded in spike trains but in analogous values between zero and one. Similar as in artificial neural networks, which can have different transfer functions, different strategies are possible what information to transmit dependent on the current activation grade of symbols. One possibility is to transmit the value

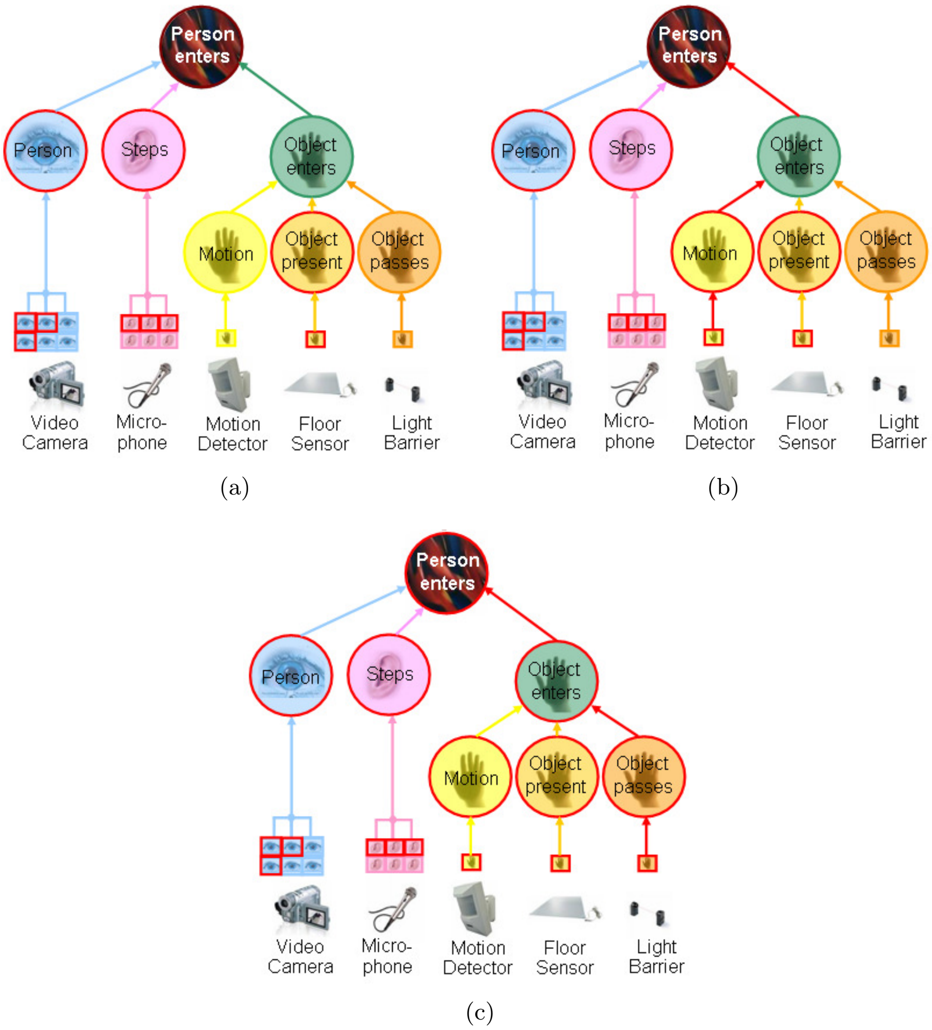


Figure 4.18: Illustration of the Concept of Event-based Neuro-symbolic Information Processing

zero if the symbol is not activated and the value one if it is activated. Such a behavior correlates to calculations performed in artificial neurons of a McCulloch-Pitts network [Vel06, Roj96]. A second strategy is to always transmit the value of the normalized input sum no matter if it exceeds the threshold value or not. A similar strategy is applied in neural networks, which have neurons with a linear transfer function [DB05]. A third possibility would be to transmit the value zero if the normalized input sum is below the threshold and the value of the calculated normalized input sum itself if it is above the threshold. The fourth option is to transmit the activation grade if the sum is below the threshold and to transmit the value one if it is above the threshold. In many situations, all four mentioned strategies will lead to the same result, which means that the same highest-level neuro-symbols will be activated. However, there may also appear differences in certain cases. What strategy is chosen for a system depends on the used sensor data, their reliability, the error-proneness of sensory modalities, and the used symbol hierarchy. Besides the chosen strategy, the selection of the threshold value is also of importance.

Concept Clarification

To illustrate the difference between the four mentioned strategies, a concrete example is given. For the explanation, the same example is taken as in section 4.2.4, where it was detected if a person enters or leaves a room. The figures 4.19a to d show the differences between the four mentioned methods. It is illustrated what value corresponding with the activation grade is transmitted by each neuro-symbol from the sub-unimodal level up to the multimodal layer when a person enters the room. Neuro-symbols that are activated in these layers are marked red. For all four cases, the value of the threshold of all symbols is 0.95. The weights of all input signals are one. The results of the given examples would change, if the values of the thresholds were different. Figure 4.19a shows the case that there is transmitted the value one if an activation grade is above the threshold and the value zero if it is below the threshold. Figure 4.19b illustrates the second mentioned possibility where the transmitted value is equal to the activation grade no matter if it is above or below the threshold. In figure 4.19c, the case is covered that there is transmitted the value zero if the activation grade lies below the threshold and the value of the activation grade itself if it lies above the threshold. Finally, figure 4.19d depicts the case that there is transmitted the activations grade itself if it is below the threshold and the value one if it is above the threshold.

4.3.3 Handling Static Information

As mentioned in section 4.3.1, in the used concept of event-based neuro-symbolic information processing, there is only processed and transmitted information if changes in the environment occur. However, this strategy leaves one question open: How can sensory information be handled that is already present at initial system startup and does not change later for a certain amount of time? One example therefore would be the information from a door contact, which indicates whether a door is open or closed. This information is important as a person can for example not enter or leave a room when the door is closed. If the door is open at initial system startup and remains open, the system would never get information about the status of the door although this information would be important. One way to overcome this problem is to scan all sensors at initial system startup and to activate certain neuro-symbols, which are important for later system operation, on the base of these sensor values. Besides activating neuro-symbols, it is also possible to set so-called memory symbols for this purpose (see section 4.6).

A related question is how to handle objects that are always present in the environment or that are most certainly present in the environment. Examples for objects always present would be walls, doors, and windows of a building. Objects that are most certainly present would be furniture like tables and chairs. For objects that are always present, there is generally no need to perceive them by sensors. It is in many cases more effective to just let the system "know" that these elements exist in the environment. In section 4.6, there is described how knowledge can interact with perception in the introduced model. For objects that are most certainly present, the same strategy can be applied. If desired, the simple declaration to the system that certain objects exist in the environment can be complemented by a check if they are actually present, which is based on sensor data. For example, the presence of a table in a room, which normally always stands at the same position, can be checked by analyzing if certain tactile floor sensors the table stands on are active. A similar verification could be performed by the vision system. Such examinations are generally performed at system startup.

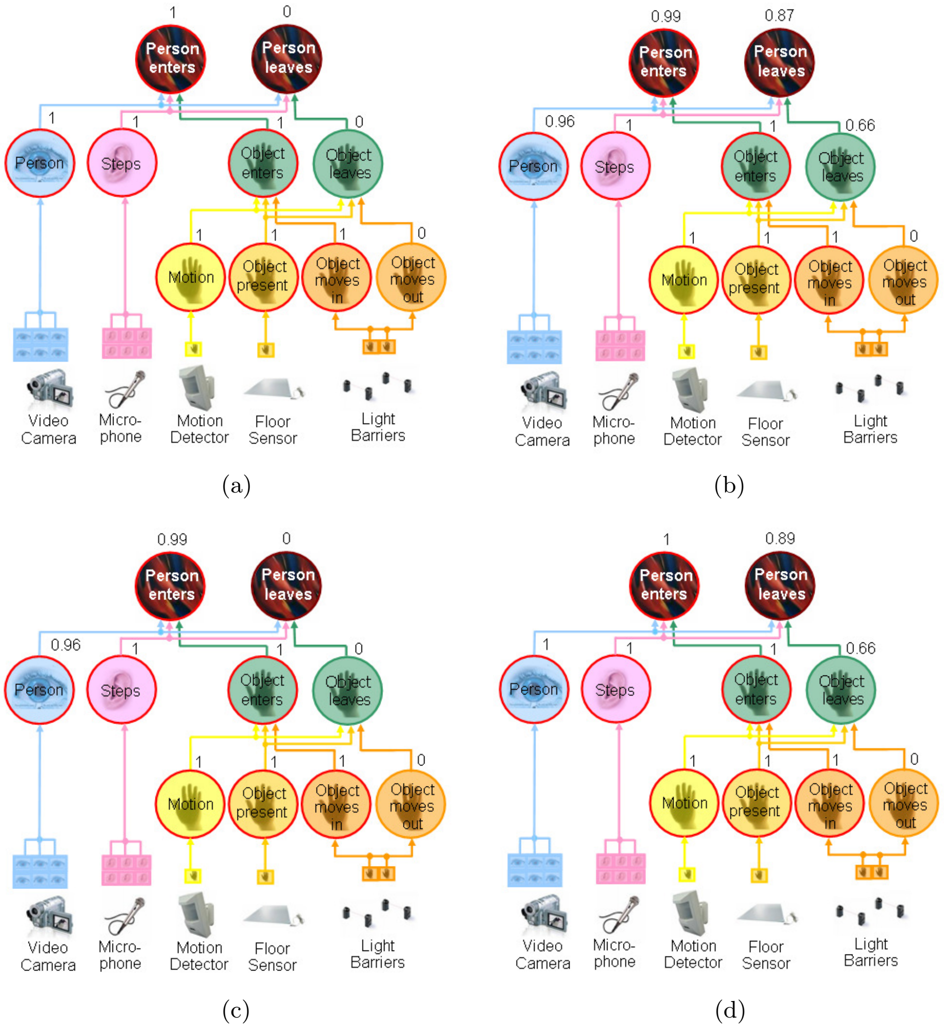


Figure 4.19: Different Strategies for Transmitting Information Correlating with the Activation Grade of Neuro-symbols

4.4 Binding of Neuro-symbolic Information

Neuro-symbolic information processing is a method to process information coming from various distributed sources. To get a unified representation of the available data, the information from these different sources has to be merged. In neuroscience and neuropsychology, the question how to integrate information from different sources is called the binding problem (see section 3.4). The binding problem concerns our capacity to integrate information across time, space, attributes, and ideas. To bind information in perception, binding must be performed within single modalities, across modalities, across space, and across time. Knowledge influences the binding process. There have been suggested different potential solutions to the binding problem. However, each of these theories suffers from certain weak points. Years of research have shown that the binding problem cannot be solved easily. Up to now, there has not been found a satisfying solution. The binding problem is regarded as one of today's key questions about brain functions.

In this section, it is described how to solve the binding problem for the proposed bionic model for human-like perception. Therefore, different suggested solutions mentioned in neuroscientific literature serve as archetype and are supplemented by additional considerations. First, it is explained how to perform binding within a modality. Next it is outlined how to merge information coming from different modalities. There is performed binding of information across space and time. Processing of events happening in parallel is considered. Finally, it is discussed how knowledge can facilitate the binding process.

4.4.1 Binding within a Sensory Modality

The first step in binding of information in perception is binding within a sensory modality. The term “one sensory modality” refers to the processing of information coming from sensors of one particular sensor type. In section 4.2.6, it was explained how to process and merge information from sensors of one sensor type to achieve a sub-unimodal perception. This task is performed by activating feature symbols based on triggered sensor values and activating sub-unimodal symbols based on feature symbols. The mechanism chosen for combining sensor data to get feature symbols is inspired from the principle of combination coding that is one proposed solution to the binding problem outlined section 3.4.5. The sensor level and the feature symbol level have a topographic structure. Information from sensors at neighboring spatial areas is combined to feature symbols and the resulting feature symbols neighboring in location are again combined to higher-level feature symbols (see appendix A for a concrete example).

4.4.2 Binding across Sensory Modalities

Besides binding of information within one modality, for multimodal perception, information also has to be bound across modalities. This type of binding can also be referred to as cross-modal or multimodal binding. In the presented model, binding across modalities is performed to get from sub-unimodal symbols to unimodal symbols and from unimodal symbols to multimodal symbols. Principally, lower-level symbols can be “bound” to higher-level symbols they are connected to. Therefore, the connections set between symbol levels comprise the information relevant for binding. How adequate connections are determined is described in section 4.5. One lower-level symbol can have connections to different higher-level symbols. Therefore, lower-level symbols can be regarded as *symbol alphabet* for higher levels. If there is not only one situation going on in the environment but different events are happening concurrently, there arises the question how to correctly assign a lower-level symbol to a higher-level symbol. As mentioned in section 4.4.1, for binding sensory information, in the lowest levels, a principle inspired by combination coding is used. However, following this principle of combination coding up to the highest level could lead to a combinational explosion when lots of different images have to be perceived. There would have to exist a separate neuro-symbol for each different location of a certain perceptual image. Furthermore, in the brain, there exist neurons in higher cortical levels that react to certain perceptual images independently of their concrete location, which is not consistent with a solution only building on combination coding. To overcome the problem of combinatorial explosion, from the sub-unimodal layer upwards, properties were introduced to neuro-symbols. By introducing a location property to neuro-symbols from the sub-unimodal layer upwards instead of using topographic neuro-symbols, the number of needed neuro-symbols can be reduced drastically. The location property represents the information where in the environment a particular perceptual image is perceived. Therefore, neuro-symbols from the sub-unimodal layer upwards can be activated more or less independently from the location of the perceptual image they represent. Besides location information, there can also be represented other information by properties, which are also involved in the process of binding information (see also section 4.2.4 and appendix A). Neuro-symbols with properties can be compared to a group of neurons in the brain, which interact to represent congeneric perceptual images. In section 3.4.5, this type of coding was referred to as population coding, which is a second proposed solution to the binding problem.

When considering information from diverse sources, there can occur the case that the data coming from different modalities have different reliabilities. As outlined in section 3.3.2, for the human perceptive system, visual information is in many situations considered to be most reliable and therefore visual information often has more influence on perception than other modalities. This effect is called visual capture. However, under certain circumstances, also other modalities can be dominant. In the introduced model, to allow the consideration of information of different reliability, information coming from particular modalities can be weighted. Therefore, in neuro-symbols receiving information from more than one modality or sub-modality, the input information about the activation grade of symbols of the different modalities is multiplied with a certain value (weight), which corresponds to the influence that the particular input values shall have. The weighted inputs are then summed up and can activate the neuro-symbol if they exceed a certain threshold value.

4.4.3 Binding across Space

As already indicated in the last two sections, location information about the spatial position of perceived images plays an essential role in integrating and binding of information coming from different sensor types corresponding to the same object or event. Location information gets especially important if more than one event is happening in the environment concurrently. Unfortunately, neuroscience and neuropsychology do not provide a unified theory about how location information is handled in the brain. Some hypotheses hold that the detection of perceptual images and their location are separate operations. There has been found certain evidence that the visual system contains two functionally distinct visual pathways: a ventral “what” pathway directed towards the temporal lobe that is involved in object recognition and a dorsal “where” pathway directed towards the parietal lobe that is involved in spatial localization and the visual control of action [Tod04]. In contrast, other studies found only weak evidence for identification without localization. In many trials, subjects either reported both the color and shape of an object at a certain location correctly or got them both wrong. [GHT96] point out that it would be surprising if the brain did not make use of spatial information freely available to it at least partially to solve the binding problem.

In fact, if information about object identity and object location were really coded separately, location information could not be used for binding processes. Additionally, there would arise the puzzling question how to merge this information again if necessary. This section describes the utility of location information for the introduced model of human-like perception and shows how it can be acquired.

As already outlined, up to the feature layer, location information is coded in form of topographic maps. In contrast, from the sub-unimodal layer upwards, location information is coded as properties of neuro-symbols. Information about the location of neuro-symbols can be sent to other neuro-symbols they are connected to. Neuro-symbols can only be activated from other neuro-symbols if the perceptual images they represent are located within a certain spatial area.

From the different solutions suggested to the binding problem, the method of using location information as feature for binding is most closely related to the theory of temporal binding. According to the theory of temporal binding, signals of neurons representing features of the same object are mutually correlated in time. Signals of neurons representing features of different objects are not correlated or anti-correlated in time. Synchrony serves as a signature of relatedness in location.

In the model, location information is important in different processes of information binding. If several events happen concurrently in the environment, location information is absolutely necessary to correctly assign sensory information of various sources to the events they belong to. In the case that only one event is happening in the environment at a certain time, location information can be used for fault detection in binding. The underlying principles will be outlined in the following. Additionally, if many events happen in parallel, a mechanism called focus of attention can be used which also makes use of location information and is described in section 4.4.5.

Location Information for Binding between Different Neuro-symbolic Levels

To form neuro-symbolic networks, neuro-symbols of different levels have to be interconnected. Without considering the location where a perceptual image was perceived, a neuro-symbol of a lower-level can principally contribute to the activation of all neuro-symbols it is connected to in the next higher level. However, when different situations happen in parallel in the environment, this can lead to a merging of neuro-symbols, which in fact do not belong together. That way, an activation of inadequate higher-level symbols can occur. Location information of neuro-symbols can help to resolve this problem. There are only bound together lower-level symbols to a higher-level symbol that lie within a certain spatial area[5].

Concept Clarification

The principle of binding neuro-symbols by location information shall be illustrated by means of a concrete example. Therefore, a room is equipped with different sensors. By a concurrent triggering of different sensors, the unimodal symbols “person”, “steps”, and “object moves” are activated. Similar to the example given in section 4.2.5, in the example, these three neuro-symbols can activate the multimodal symbol “person walks”. However, as illustrated in figure 4.20, the unimodal symbols are perceived in different spatial areas. Therefore, it is very likely that they origin from different events, and that their binding to the symbol “person walks” would be incorrect. To avoid an undesired activation of neuro-symbols, it is useful to define how much certain lower-level symbols may deviate in location to be bound to a higher-level symbol. This information can either be predefined, which requires knowledge from the system designer, or it can be learned from examples presented to the system (see section 4.5).

Handling Location Information of Restricted Spatial Resolution

In realistic situations, location information provided from sensors is not of infinite spatial resolution. Additionally, the resolution is not the same for all sensor types. Therefore, one event detected by different sensors may cause the generation of neuro-symbols with slightly different

[5] Location values of a neuro-symbol can change during the activation time of the symbol. This happens when certain objects in the environment change their position (e.g., a person walking through a room).

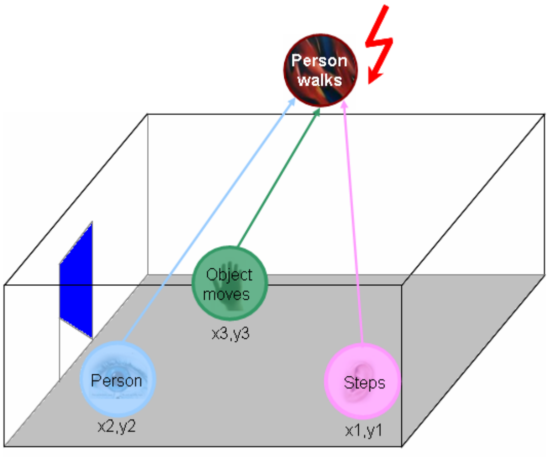


Figure 4.20: The Role of Location Information in Binding between Different Layers

location values. This is the reason why not only neuro-symbols at one very specific location need to be merged but neuro-symbols occurring within a certain spatial area. A finite spatial resolution also has an influence on the determination of the location value of higher-level neuro-symbols. Higher-level neuro-symbols receive information coming from lower-level symbols. Accordingly, the values of the location properties of higher-level symbols have to be derived from location information provided from lower levels. However, due to imprecise mounting of sensors and due to their limited resolution for determining location information, the location values of all lower-level symbols might not always correspond perfectly. Therefore, a way must be found to calculate the location values of higher-level symbols out of the location information from all lower-level symbols they were activated from. Different strategies are possible to calculate this information. One method is to calculate the average of all information received. However, as mentioned in section 3.3.2, in the brain, there seems to be always one modality for determining positions of perceptual images that dominates over the others for certain cases, because it is of greatest reliability. For locating objects, situations, and events, the dominant modality in human perception is most often vision. The domination of this modality is called visual capture. If a dominant modality exists, its location information is weighted more in the calculation of the location. Inspired from the existence of a dominant modality in the brain, for determining the location of neuro-symbols, there can also be used location information from all sources, which is however weighted differently corresponding to their reliability and resolution. Furthermore, there can also be used only one lower-level modality to determine the location of a higher-level symbol. This method simplifies the calculation process and makes sense if this modality provides by far more reliable location information than the other modalities or location information of a better spatial resolution. Which of the mentioned strategies is actually chosen depends on the concrete sensor types used and on their reliability and spatial resolution.

Location Information for Feedbacks between Neuro-symbols

Besides connections from lower-levels neuro-symbols to higher-level neuro-symbols, there also exist feedback connections between neuro-symbols of the same level (see section 4.2.5). Without using location information, an activated neuro-symbol would influence the activation of all neuro-symbols to which it is connected by feedback loops. However, it may occur that such an influence is not wanted, because the symbol that is influenced is needed to code another situation happening concurrently. To avoid an unwanted interaction, it is useful to check whether the values of the location property of the two activated symbols are within the same spatial area. If this is the case, one of the two symbols was activated because it is made up by a subset of lower-level symbols of the other symbol. In such a case, its activation should be inhibited by a feedback. If the locations do not match, an inhibitory feedback is not desired. Again, the location area within which feedbacks shall have an influence can either be predefined or learned from examples (see section 4.5).

Location Information for Fault Detection

Location information is important to correctly bind neuro-symbolic information from different sources. There were already introduced examples how correct binding is facilitated by location information. Besides the function of binding, spatial information can also have the function of fault detection. Reasons for errors can be faulty sensors, an incorrect data transmission, or incorrect binding. There might exist certain objects, events, and situations that occur only at certain locations. If the location property of the perceptual images they are represented by in lower layers has a value that is not common, there might have occurred an error in the perception process. Examples would be that lower-level symbols connected to the symbols "object enters" or "person enters" are perceived at a location where no door exists. If there is assigned a location value to a certain lower-level perceptual image that is very atypical for a higher-level symbol it is connected to, a further processing can be inhibited. This process could be considered as a process where factual knowledge interacts with perception and could therefore be handled as top-down process (see section 4.6). However, in the model, there also exists the possibility to handle the inhibition directly and locally. What are allowed areas for certain perceptual images and which areas are forbidden can either be predefined or learned from examples (see section 4.5).

4.4.4 Binding across Time

So far, in the description of the model, it has always been assumed that binding of sensory information is performed for sensor values (or lower-level neuro-symbols) occurring concurrently or quasi concurrently. This means that during a certain time interval all sensors triggered from one and the same event are activated, which then leads to an activation of certain neuro-symbols. Again, in the higher layers, neuro-symbols of lower layers have to be activated concurrently to activate higher-level symbols.

Concept Clarification

Figure 4.21 illustrates the concept of concurrent activation by means of a concrete example, which is already known from section 4.2.3. The picture shows that the unimodal symbol "object enters" is only activated when all three lower-level symbols it is connected to are active, because only within this time interval, the sum of input activations exceeds the threshold value[6]. In the same way, the unimodal symbols "object enters", "person", and "steps" would have to be active concurrently to activate the symbol "person enters".

[6] In the given examples of this section, it is assumed the activation threshold to be 0.85.

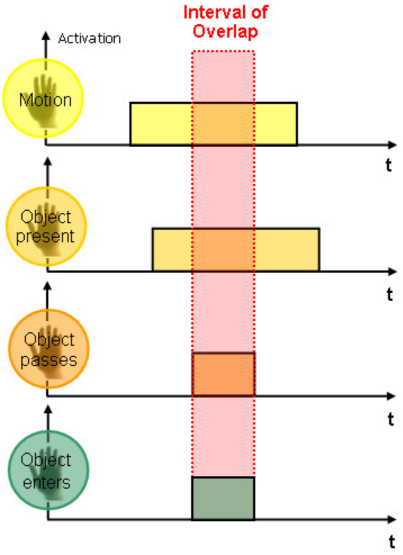


Figure 4.21: Temporal Overlapping of Neuro-symbolic Activations

In reality, however, there might also occur the case that not all sensor values belonging to one event are triggered concurrently. In the example of the object entering, it could for instance happen that the motion detector shows a certain reaction delay time. In this case, the light barrier might already be deactivated again although the motion detector has not yet been activated (see figure 4.22). As there does not exist an interval of temporal overlapping of the activation of the different symbols, the symbol "object enters" would not be activated although the corresponding event occurred in the environment.

From the description just given, it gets clear that binding of information in the brain as well as in the introduced model does not only have to occur across space and different modalities. It also has to occur across time. An interesting question is how the brain "knows" that events happening one after the other belong together and form a certain scenario. The answers to this question coming from neuroscience are still very vague. [Eic06] mentions that the hippocampus – a part of the limbic system located in the forebrain – might be involved in the process of binding sequential events across time. In [Pra06], certain considerations are made about how time-dependent events like the perception of speed or successions of events could be experienced in the brain. It is suggested that physical values like speed and time are represented symbolically and that processing of physical values by symbols requires an abstraction from an actual object (or person) to a generic object. Objects moving with different velocities are represented by different symbols.

The theory of speed being represented symbolically is supported by the fact that there exist neurons in the brain that respond exclusively to the movement of objects. However, it might not be one-hundred percent correct that for perceiving movement of a certain speed, the actual objects are abstracted to generic objects as certain objects show characteristic movement profiles, which give important clues for identifying them. Besides this, in the primary cortex of the brain, there were found different neurons firing depending on the direction of movement and the form and size of the object. Therefore, it should rather be supported the hypothesis that movement is represented symbolically – or in the model introduced in this thesis neuro-symbolically – and

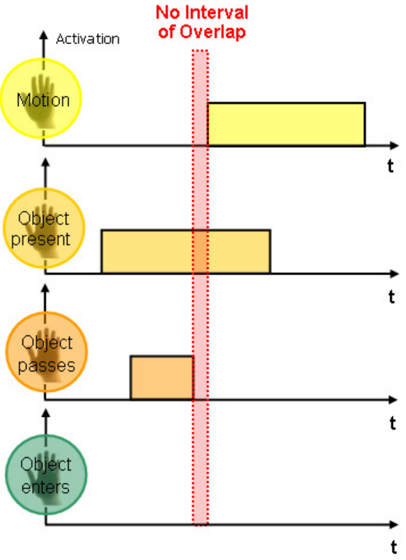


Figure 4.22: No Temporal Overlapping of Neuro-symbolic Activations

not only the speed but also the direction and the motion profile and maybe to a certain extend also the form and the size of the object have influence on the categorization into “speed classes”. Additionally, the estimation of speed might not only be based on visual impressions but also be influenced by the sound being produced by the object and the surrounding within which the object occurs.

Similar as for perceiving velocities, [Pra06] suggests physical relations like time instants relative to the current time being represented symbolically. He points out that instead of a time difference measured in time units, the symbolic representations of “before” and “after” are applied. There is not made a proposition how duration of time could be represented.

The hypothesis that there does not exist an absolute time base is supported by the fact, that different situations lasting the same time can be subjectively experienced as taking different amounts of time. However, it does not seem probable that on a pure perceptual level, there exist symbols like “before” and “after” that are bound together with other perceptual symbols to higher symbolic representations. Instead, there might exist particular neurons or groups of neurons that bind together events having occurred one after the other being responsible for the coding of these sequences. That way, a binding of perceptual data across time could be achieved. A suggestion how this could be performed is described in the following using the concept of neuro-symbolic information processing. Two different solutions are conceivable for binding information across time depending on whether the temporal succession in which different sensors are triggered is of importance or not.

Usage of Time Windows

In case the temporal succession is not always the same, not known, or of no importance, a time window of a certain length can be defined for each neuro-symbol. Sensor values (or activated neuro-symbols) occurring within this time window are considered in the merging process of a certain neuro-symbol. The usage of such time intervals for binding is in accordance with reports from [Sta04], where it is described that for two events occurring very close together in time, it can happen that we are aware that they occurred at different times, but we cannot say which one occurred first.

The neuro-symbols for processing information within certain time windows differ slightly form the neuro-symbol type described until now. They have the additional capability to hold input values active although the sensory source they came from is already inactive again (see figure 4.23). The principle works as follows: If a neuro-symbol receives information from a certain activated neuro-symbol of a lower level, it can hold this signal active for the duration of its time window. Therefore, even if the input signal is inactive again, it can still contribute to the calculated sum of activations. The prolonged activation is reset either if the time window has expired or if the symbol has been activated because the sum of all input values exceeded the threshold value. In the second case, there is also reset the time window[7].

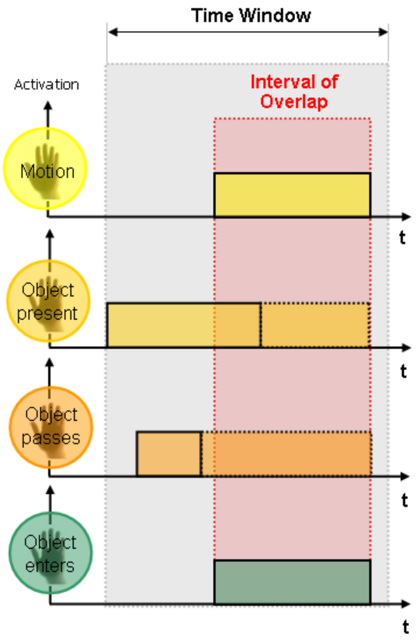


Figure 4.23: Prolongation of the Activation Time by the Usage of Time Windows

On a neural level, a prolonged activation of certain inputs could be achieved by feedback loops between neurons and synaptic connections of different type and number having the function of weighting information and also of causing delays in signal transmission. How time windows were implemented for model simulation is described in section 5.4.1. The length of time windows of certain neuro-symbols can either be predefined or learned from examples (see section 4.5). Neuro-symbols with time windows are compatible with the neuro-symbols without time windows used until now. For representing neuro-symbols without time window, there can be used neuro-symbols with a time window by just setting the length of the time window to zero. Compatibility between the neuro-symbol types is necessary if correlations are not predefined but learned, because in this case, it is not fixed at the beginning which symbol type will be used to code a certain perceptual image.

[7] The reset of the time window is not illustrated in figure 4.23.

Perceiving Successions of Events

Using time windows is sufficient for a range of events and situations and is one possibility to achieve binding across time. However, sometimes not only the occurrence of certain sensor values and active lower-level symbols within a certain time interval is of importance but also the temporal succession in which they occur. Taking the example from section 4.2.3 where an object enters a room, it is conceivable that the floor sensor is always triggered first followed by the light barrier and afterwards by the motion detector (see figure 4.24).

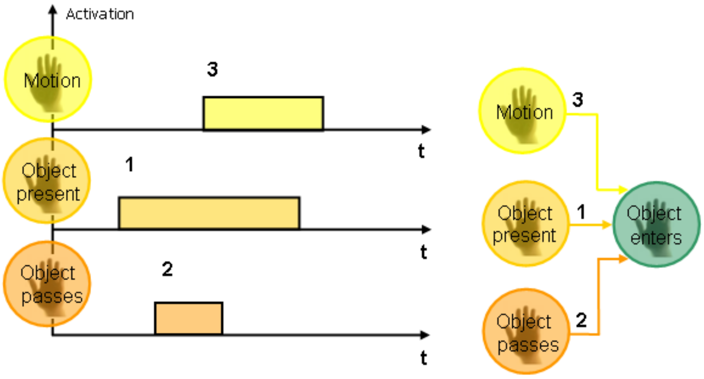


Figure 4.24: Neuro-symbol Activated by a Succession of Events

The succession of data is important if a triggering of the sensor values in another succession is assigned to another event (e.g., to the event that an object leaves the room). On a neural level, such successions of events could be realized by using a number of neurons and exploiting feedbacks. A possible neural structure for this purpose is depicted in figure 4.25. Within this structure, it is followed the principle of neural information processing: If the sum of the input signals exceeds a certain threshold, which is in the given example a little below one, a signal with the value one is transmitted via the output to all units it is connected to. The numbers next to the connections indicate the weights of the connections, which correspond to synapses of biological neurons. By the positive feedbacks of the neural structure, the activation of a neuron can be held active even if the input signals coming from the lower level that caused the activation have already vanished. By negative feedbacks, such a prolonged activation can be annulated. The negative connections of the neurons b_1, c, and e can inhibit the activation of their neighbors a, b_2, and d when being activated first. This measure assures that signal activations arriving in an incorrect succession do not lead to an activation of units in the next higher layer.

The figures 4.26a to c show three different examples of information processing performed with this neural structure. The red arrows indicate effective inhibitory connections. In the figures 4.26a and b, the symbols "object present", "object passes", and "motion" are activated in the correct chronology and therefore, the symbol "object enters" is activated. The difference between these two cases is that in figure 4.26a, the activations of the three lower-level symbols show temporal overlapping and in figure 4.26b they do not. In figure 4.26c, the order is incorrect and the symbol "object enters" therefore remains inactive[8]. For a technical realization, it is no problem to consider successions of incoming events even without using feedbacks (see section 5.4.1).

[8] What is not depicted in figure 4.25 is that the activation of the highest level unit "object enters" has to be inhibited again after a certain time after its activation by an additional input or by a measure where the activation grade of the unit is decreased successively until is falls below the threshold value.

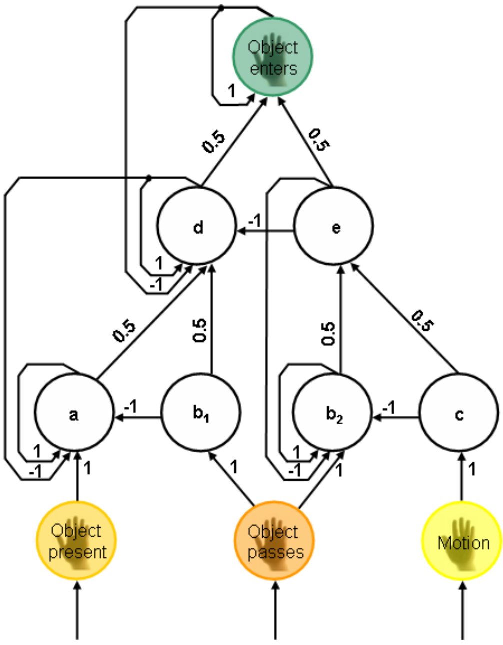


Figure 4.25: Neural Structure with Feedbacks for Detecting a Certain Succession of Events

Introducing Scenario Symbols

In the descriptions about handling data across time given until now, there were always merged sensor values or neuro-symbols from a lower level to neuro-symbols of the next higher layer. The typical time span within which data were merged lay in the range of a few seconds. However, there can also occur the case that a situation consists of a succession of events whereby each of these events is already represented by a multimodal symbol. Besides this, these situations may take more time than a few seconds. Reutilizing the example from section 4.2.5, it could be necessary to detect if a person enters the room, goes to the window, and then stands near the window. In this case, the multimodal symbols "person enters", "person walks", and "person stands" would be triggered one after the other[9]. To handle such successions of events – in this thesis labeled as scenarios – a new symbol layer is introduced, which is called *scenario symbol layer* (see figure 4.27). Scenario symbols are activated if certain multimodal symbols are activated within a certain time span or in a defined succession (see figure 4.28 for the scenario "person goes to window"). The scenario symbol level could be regarded as an implicit form of a short time memory.

The underlying methods for forming scenario symbols are similar to the concepts for binding across time in lower levels. The only difference is that now, symbols of the same type – multimodal symbols – are merged and that the time span within which a merging can occur can be longer. Again, correlations between multimodal symbols and scenario symbols can principally either be predefined or learned from examples (see section 4.5).

[9] The value of the location property of each multimodal symbol also has to be considered for the binding process.

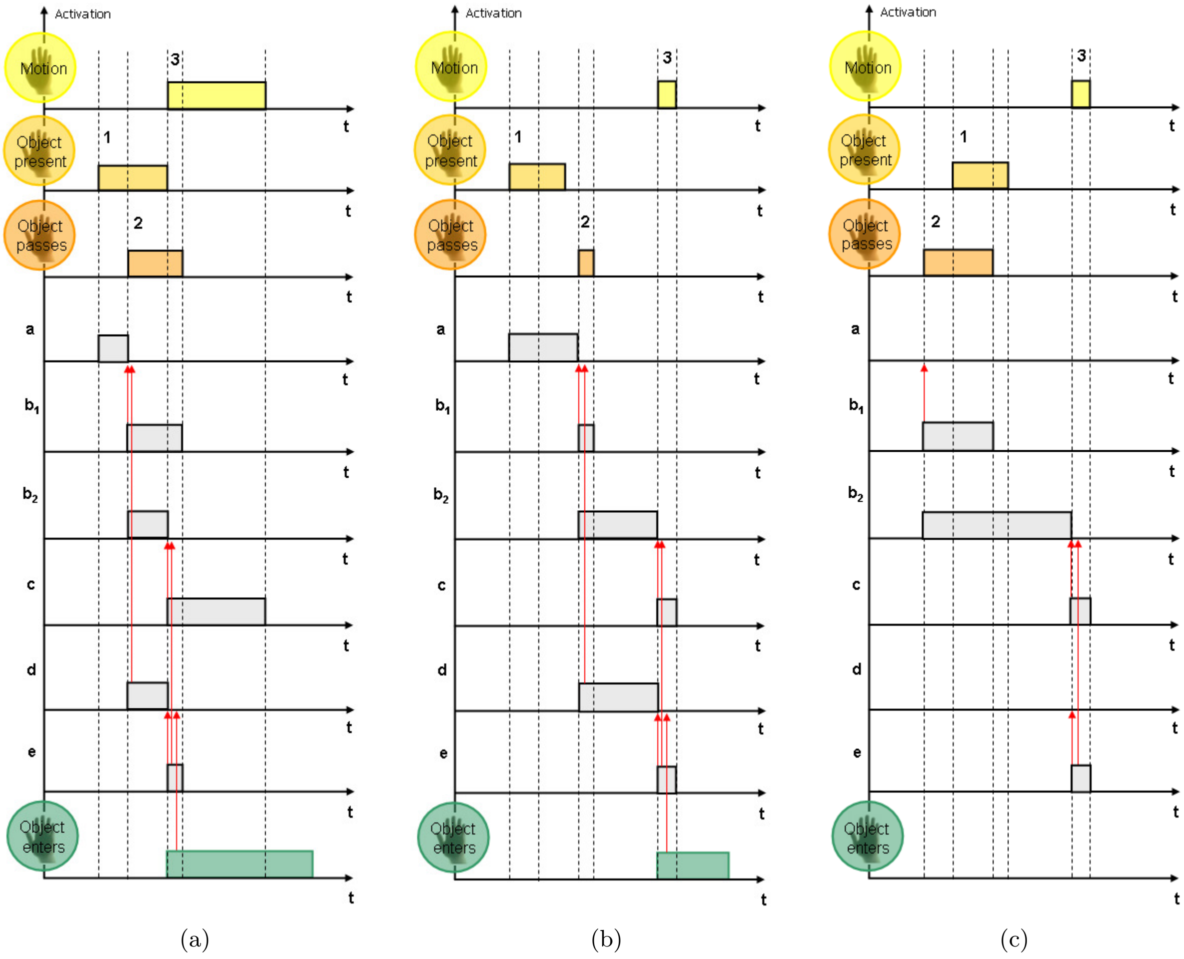


Figure 4.26: Examples for Detecting a Certain Succession of Events

4.4.5 Parallel Processing versus Focus of Attention

Until now, in the given examples of neuro-symbolic information processing, it was mainly focused on how to process information origination from only one particular object, event, scenario, or situation. According to the principle of single symbol activation, in such a case, from the sub-unimodal layer upwards, only always one neuro-symbol is active for each sub-modality and modality and only one single symbol is activated at the multimodal layer. This is assured by inhibitory feedbacks between neuro-symbols (see section 4.2.5). In reality, however, many objects can be present at the same time in the environment and diverse events and activities can happen in parallel. Therefore, many different symbols of each level can be active at once belonging to different events. As mentioned in section 4.4.3, by introducing location information into the system, it can be determined what activated lower-level symbols belong together to form one particular higher-level symbol. However, if every possible perceptual image is represented only by one single neuro-symbol, a problem occurs if different activities shall be perceived in parallel, which are based on the same lower-level symbols.

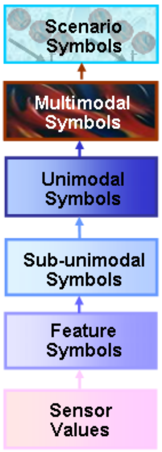


Figure 4.27: Introduction of a Scenario Symbol Level

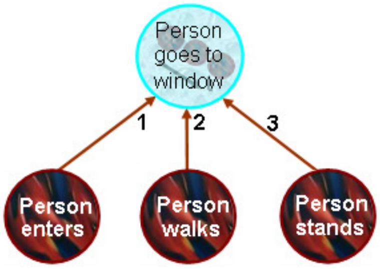


Figure 4.28: Example for the Binding of Multimodal Symbols to a Scenario Symbol

Concept Clarification

To clarify the occurring problem, it is supposed that two persons are present in a room at the same time. One is standing in the room and the other one is walking around. Concerning the symbol structure, the same symbols, properties, and correlations between them are used like in section 4.2.5. To perceive these two situations, two multimodal symbols – “person walks” and “person stands” – need to be triggered. Therefore, the unimodal symbols “person” and “object stands” have to be active to activate the symbol “person stands”. To trigger the multimodal symbol “person walks”, the unimodal symbols “person”, “steps”, and “object moves” have to be active. There occurs a problem if both situations are going on in the room at the same time. In such a case, for both activities there would have to be activated the visual unimodal symbol “person”. However, in the used symbol hierarchy, there exists only one visual unimodal symbol “person”.

To overcome the problem just described, different strategies are conceivable. The first possibility is to use *parallel symbol representations*, which allow it to represent certain perceptual images by more than one neuro-symbol. The second possibility is to introduce so-called *group activity symbols*. The third possibility is to use a method called *focus of attention*.

Parallel Symbol Representations

In the concept of parallel symbol representations, for certain perceptual images, more than one neuro-symbol can exist if it is likely that this image occurs in the environment several times at the same moment. Generally, at lower hierarchical neuro-symbolic levels, there will exist more such parallel symbol representations than at higher hierarchical levels, because lower-level symbols can be regarded as symbol alphabet for higher symbol levels. Each lower-level symbol that stands for a perceptual image can potentially contribute to the activation of different higher-level symbols. As described in section 4.2.6, on the feature symbol level, which is the lowest neuro-symbolic level, all perceptual images are represented by parallel feature symbols located at different topographic positions.

Concept Clarification

To overcome the already mentioned problem of concurrently perceiving a person walking and another person standing, there can be integrated a second symbol "person" into the unimodal layer of the visual system. Based on the location of the neuro-symbols, they are assigned to different higher-level neuro-symbols. If three persons are present in the room, there have to exist three symbols "person". The same principle needs to be applied to other modalities. If there are three persons standing in the room, there have to exist three sub-unimodal symbols "object present", three tactile unimodal symbols "object stands" and three visual unimodal symbols "person".

Such a duplication of symbols a certain number of times might not always be the best solution, especially not for large configuration with lots of different perceptual images being possible to occur, because it would require a huge number of symbols. This does not make sense if it is unlikely that a perceptual image occurs several times at the same moment. Therefore, it is advantageous to also have available other methods to handle concurrently happening events.

Group Activity Symbol Representations

With the principle of group activity symbol representation, situations can be aggregated and represented by a group activity symbol. This makes sense if too many activities are going on in the environment to represent them all by different neuro-symbols.

Concept Clarification

Taking again the example of several persons being present in a room, this situation can be represented by a group activity symbol "group of persons", which is added to the visual modality. In a similar way, there must be added group activity symbols for the other modalities, like a symbol "various objects present" in the tactile modality.

A disadvantage of group activity symbols can be that the information represented by them is of less detail. Besides this, it may be difficult to correctly combine such group activity symbols to higher-level symbols. They can only be combined with other symbols with such a group activity character.

Focus of Attention

To overcome the problems mentioned in the description of parallel symbol representation and group activity symbol representation, the principle of focus of attention is introduced. In section 3.4.5, it was mentioned that in the brain, focus of attention restricts the spatial area that is considered when binding sensory information. In [Jov97], it is described that focus of attention inhibits the further processing of information, which has no relevance in the current situation. In the model, only neuro-symbols are bound to higher-level neuro-symbols that lie within the focus of attention. The activation of neuro-symbols is inhibited if the perceptual images they represent lie outside the spatial area that is currently covered by the focus of attention. With this method, the number of concurrently activated symbols can be reduced which eases the binding process and makes it unnecessary to have parallel representations of the same symbols.

An important question is on which level focus of attention influences the binding of information. Statements from neuroscience to this question are very vague. In the model, focus of attention influences perception on the feature symbol level. As outlined in section 4.2.6, feature symbols are topographic in structure which means that they have a strong correlation to the position of the sensors they are derived from. On the feature symbol level, different events happening concurrently can be represented by feature symbols of different locations. From the sub-unimodal level upwards, location information is contained only as property of symbols and the number of concurrent events that can be coded is restricted. By focus of attention, the transmission of information from feature symbols corresponding to areas outside the focus is inhibited and therefore, the information coming from them is just not further processed in higher layers. Figure 4.29 illustrates this principle for feature symbols of three sensor types. Feature symbols can only transmit information to higher layers if they lie within the focus of attention. These feature symbols are marked red in the pictures. Activated feature symbols outside the focus cannot activate symbols on the sub-unimodal level, because the input of the focus of attention has an inhibitory influence and therefore decreases the activation grade of the symbols[10].

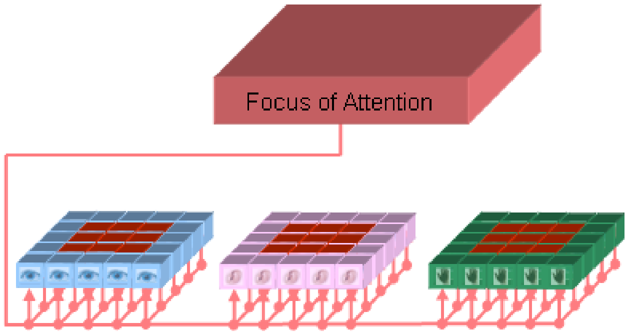


Figure 4.29: Influence from Focus of Attention on Feature Symbol Activation

After information within a certain spatial area has been processed, the focus of attention has to switch to another area to also bind other symbols present at the same time (see figure 4.30). It

[10] If more than one feature symbol layer exists, a transmission of data from one feature symbol layer to the next higher one shall be possible independent from the current focus of attention. This can be achieved by setting the threshold for activation of feature symbols to a lower value than the threshold of sub-unimodal symbols. The processing of feature symbols even without focus of attention is important for later processing stages when the focus of attention is directed to these areas.

might happen that information is lost when symbols are only active for a very short period of time and the focus of attention is not directed to them within this time period. However, the same problem also occurs in human perception. The size of focus of attention can be altered depending on how many situations are going on concurrently in the environment and how large the spatial areas are where they happen. Principally, also the form of the focus could be changed.

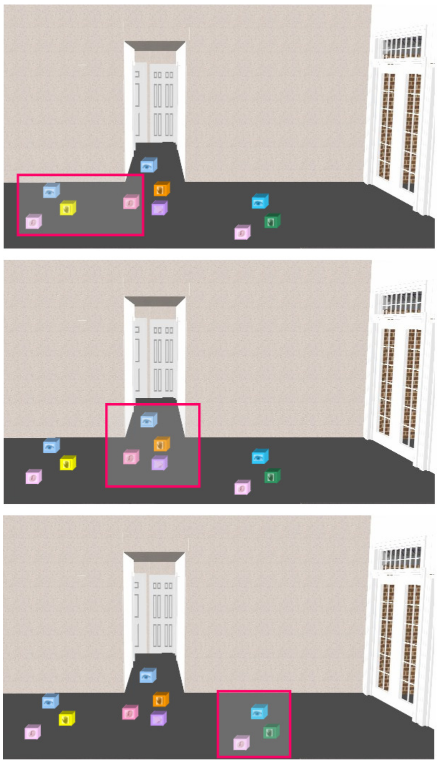


Figure 4.30: Altering the Size and the Center Position of the Focus of Attention

During perception, the beam of focused attention needs to be steered and controlled. The direction of focused attention is for sure to a certain extend influenced by perceived images. In the model, group activity symbols could be one factor that influences the focus. As mentioned, if situations cannot be associated to neuro-symbols because there are too many sensors and lower-level symbols active at the same time, so-called group activity symbols are activated in the different modalities. They indicate that various activities are going on at the same time and provide location information about the approximate spatial area where this location is going on. This location information can be used to direct the focus of attention to a sub-part of this spatial area. When changing the location the focus of attention is directed to, it is often useful that the new focus of attention overlaps with the location of the old focus to a certain degree, because there might have remained lower-symbols that could not be bound to higher-level symbols, because other correlating lower-level symbols were not within the spatial processing area. They might be bound with other symbols that lie within the focus of attention of the next step. In the process of directing the focus of attention, there are most certainly also involved diverse mental processes not directly assigned to perception. Emotions, knowledge, expectation, etc. might play a role. However, their influence has not yet been considered in the model.

4.4.6 Binding by Knowledge

As outlined in the sections 3.3.3 and 3.4.5, knowledge plays an important role in perception and binding of information and facilitates perception in ambiguous situations. Its influence on perception can be regarded as a top-down process. An interesting question is on what hierarchical level the interaction with bottom-up processes of perception takes place. Answers coming from neuroscience and neuropsychology are controversy. In the model, knowledge can principally influence perception from the sub-unimodal level upwards and decrease or increase the activation grade of neuro-symbols. A more detailed discussion about the influence of knowledge is given in section 4.6.

4.5 Adaptability of Neuro-symbolic Structures

To increase flexibility of systems, it is desirable to give them the ability to adapt and to learn. How these mechanisms can be integrated into the proposed model is described in the following. In section 4.5.1 it is discussed what parts should be subject to learning and what has to be predefined. In the sections 4.5.2 and 4.5.3, the used learning mechanisms are introduced.

4.5.1 Predefining versus Learning

Neuro-symbolic networks are basically made up of neuro-symbols with certain properties and of connections between these neuro-symbols. It has to be noted that important information – the "perceptual intelligence" – of a neuro-symbolic network is not only comprised in the neuro-symbols and their properties themselves but also and especially in the connection between the neuro-symbols. An important question is how to determine appropriate neuro-symbols, neuro-symbolic properties, and neuro-symbolic connections and correlations. Principally, they could either be predefined by the system engineer and therefore already exist at system startup, or they could be learned from examples. Both approaches have certain advantages and disadvantages.

By using predefinitions, the "overhead" of the system can be kept small, because no methods for learning and adaptability need to be included into the system. The "overhead" are mechanisms and functions, which are necessary during system learning and adaptability but are not needed any more afterwards during system operation of the already configured neuro-symbolic network. The disadvantage of predefining everything before system startup is that the system designer needs to have a very deep understanding about correlations between objects, events, scenarios, and situations and the sensors and neuro-symbols they activate. Besides this, it is a quite time consuming and monotonous task to define all neuro-symbols, properties, and connections for all possible situations that can occur in the environment. Additionally, predefining allows little flexibility of the design. For making changes in the system, reconfigurations need to be done "by hand". However, in spite of many disadvantages of using predefined correlations, such a predefinition cannot always be avoided. Learning from examples cannot be applied if there shall be defined dangerous events or situation like for instance a fire in a room. Additionally, the lowest-layer correlations always need to be predefined because – similar as in the brain – a system cannot learn higher-level correlations if lower-level correlations do not already exist [Lur73]. There cannot be learned "everything from nothing".

The advantage of using learning in neuro-symbolic networks is that the configuration effort for the system engineer is decreased. Learning offers flexibility and adaptability of the design. The designer does not need such a deep knowledge and understanding about correlations between objects, events, and situations and the sensors and neuro-symbols they trigger.

The learning strategy proposed in this work is based on learning from examples. Although the introduction of learning to the system offers many advantageous, the usage of learning does not decrease the configuration effort to zero. To allow learning from examples, a certain number of representative examples for all necessary objects, events, scenarios, and situations need to be provided.

4.5.2 Supervised Learning for Neuro-symbolic Networks

As just outlined, learning increases the flexibility of the model. However, the fact that not all correlations and concepts can be “learned from nothing” makes it necessary to predefine certain items before system startup. This principle is in accordance with research reports from neuroscience and neuropsychology (see section 3.2). At birth, the brain is not completely hardwired. There have to exist already certain connections to allow further learning. Unfortunately, the answers coming from neuroscience and neuropsychology about what has to be predefined in the brain and what can be learned are controversy and quite vague. Principally, it is reported that learning takes place by setting neural connections and that lower cortical levels must already be developed before higher cortical levels can evolve. In accordance to this, in the presented model, connections between lower levels are predefined and correlations in higher levels can be learned. The learning method used in the model is based on examples and can be regarded as a supervised learning process[11]. This section describes how to determine adequate connections between neuro-symbols including considerations of property values and location and timing information.

At initial system startup, a neuro-symbolic network looks as depicted in figure 4.31[12]. Besides connections between sensors and neuro-symbols of the features symbol level, there generally do not exist connections between neuro-symbols. In certain cases and for certain sensory modalities, there can also already exist higher-level connections, especially connections between feature symbols and sub-unimodal symbols (see section 4.7).

The process of learning applied in the model is divided into four different phases starting with the determination of correlation between the lower levels (see figure 4.32)[13]. Each of the four phases is divided into a *phase A* and a *phase B*. In phase A, examples are presented to the system and forward connections between the hierarchically lower and the hierarchically next higher layer are determined and set. In phase B, the same examples are presented to the system again and feedback connections within a certain layer and modality are set.

Before learning can start, certain items have to be predefined. It has to be defined what sensor types shall be involved in the perception process and how to extract suitable features from the provided sensor values. Additionally, it has to be defined how sensors of different types are taken together to certain modalities.

[11] In the human brain, learning also strongly depends on unsupervised learning processes. However, it is not yet clear how these learning processes work. Therefore, for the model, at the current state, learning strategies are only inspired from supervised learning methods.

[12] Top-down connections from knowledge and focus of attention are not considered in this description.

[13] In section 4.6, for learning correlations between neuro-symbols and memory symbols, a fifth phase needs to be introduced.

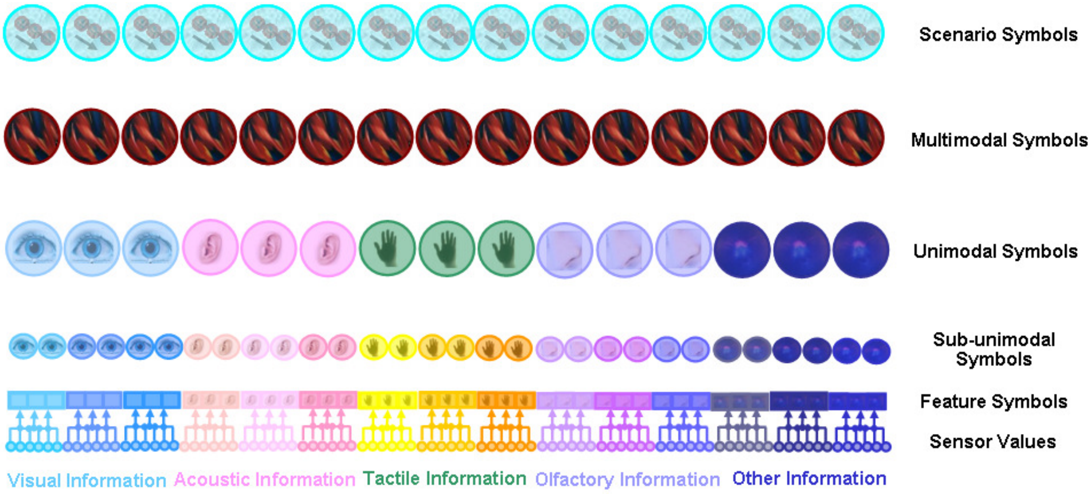


Figure 4.31: Connections between Neuro-symbols at Initial System Startup

In the first learning phase, it is learned from examples what combinations of feature symbols form which sub-unimodal symbols and how sub-unimodal symbols shall influence each other through feedbacks. Accordingly, the connections between the symbols of these layers are set. The learning and setting of connections is performed for each sub-unimodal modality separately. For some modalities, it might occur that there do not exist sub-unimodal symbols. In this case, there are formed connections directly from features symbols to unimodal symbols based on the presented examples. For some modalities, it may also be advantageous and less effort to define correlations "by hand" or to use already existing workable solutions to get from sensor data via feature symbols to sub-unimodal or unimodal symbols. The modular hierarchical structure of the model allows it to integrate such workable solutions (see section 4.7). In phase 2, it is determined based on examples how to best connect sub-unimodal symbols with unimodal symbols and unimodal symbols among each another. Again, learning and setting of connections is done for each unimodal system separately. After having available fully functioning unimodal levels, the relations between unimodal and multimodal symbols and among multimodal symbols themselves are learned in phase 3. Finally, correlations are set between multimodal symbols and scenario symbols and scenario symbols are linked to each other in learning phase 4.

As already mentioned, it is learned from examples what symbols need to be connected to perceive certain objects, events, scenarios, or situations. The learning strategy that is followed is to a certain extend comparable to how supervised learning is performed in artificial neural networks. In the model, learning is also referred to as *training*. During the learning phases, examples for different objects, events, scenarios, and situations are presented to the system. Each object, event, scenario, and situation – which is regarded as perceptual image – is assigned to one neuro-symbol. Consequently, only classes of objects, events, scenarios, and situations can be perceived later by the system, which have been learned before. For each object, event, scenario, and situation, a number of examples is presented to the system to give it the ability to generalize. An example comprises input data and target data. The input data are the values of the sensors that are triggered when a certain object, event, scenario, or situation occurs in the environment. As the lower neuro-symbolic levels are already connected, based on the sensor data, certain low-level symbols are activated and serve as actual input data for the learning procedure. These lower-level symbols can be active concurrently or can be activated sequentially within a certain time period.

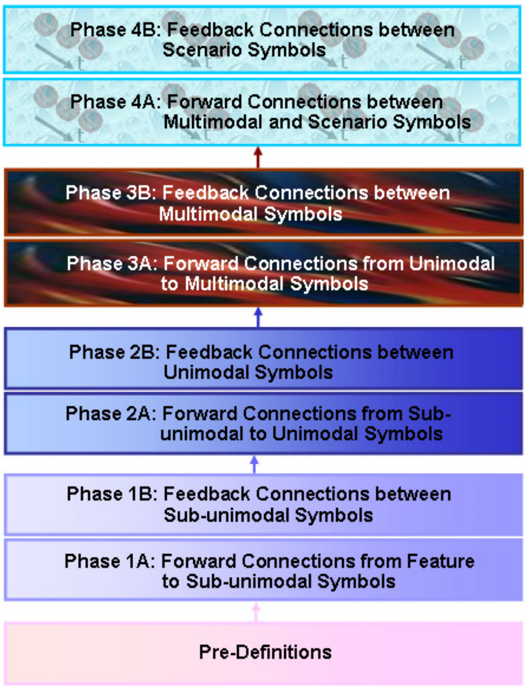
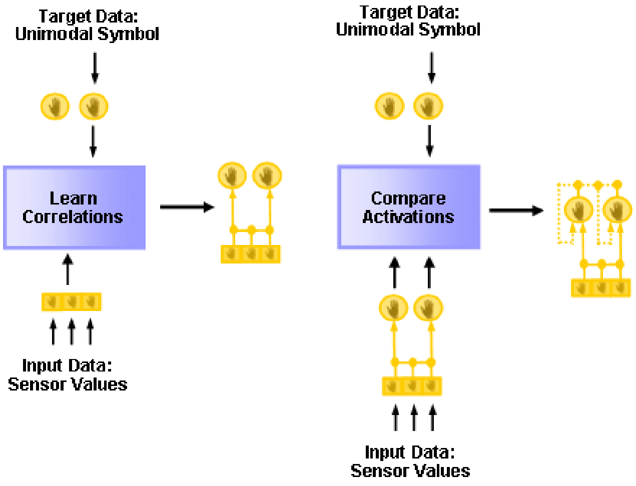

Phase 4B: Feedback Connections between Scenario Symbols
Phase 4A: Forward Connections between Multimodal and Scenario Symbols
Phase 3B: Feedback Connections between Multimodal Symbols
Phase 3A: Forward Connections from Unimodal to Multimodal Symbols
Phase 2B: Feedback Connections between Unimodal Symbols
Phase 2A: Forward Connections from Sub-unimodal to Unimodal Symbols
Phase 1B: Feedback Connections between Sub-unimodal Symbols
Phase 1A: Forward Connections from Feature to Sub-unimodal Symbols
Pre-Definitions

Figure 4.32: Learning Phases for Determining Correlations between Neuro-symbols

The target data are the higher-level neuro-symbol that shall be activated when the perceptual image occurs. During the learning phase, the system can memorize a number of examples and can extract coherences between lower-level symbols and higher-level symbols (or symbols of the same level) from these examples and set connections accordingly. For learning of correlations, different methods can be applied. Similar to training algorithms in neural networks, a learning method needs to offer the possibility to generalize. For a first implementation, statistical calculations were used to determine what correlations between data are frequent and how these correlations look like over time. However, other methods are also conceivable. Especially for learning correlations between feature symbols and sub-unimodal symbols, the usage of neural networks may sometimes lead to good results (see section 4.7).

In figure 4.33, the principle just described is illustrated schematically for the learning phase 1 for one tactile sub-unimodal system. Exactly the same principle can also be applied to all other sub-unimodal systems. Input data for the system are available in form of sensor data. Similar to neural networks, features are extracted out of these sensory raw data before the learning phase 1 starts. These features act as actual input data for learning. Similar to supervised learning in neural networks, there are also presented target values to the system for the desired output. The target value for a certain example is the sub-unimodal symbol that shall be activated when the sensor values of the example occur together. By presenting a certain amount of such *input data - target symbol pairs* to the system, a learning algorithm can determine which neuro-symbols need to be connected and connections are generated.

Figure 4.33a shows the principle how forward connections between two layers are generated based on examples. However, as described in section 4.2.5, there might also be a need for feedback connections to inhibit the activation of certain symbols that are triggered in parallel. For this purpose, the same examples as used in phase 1A are presented to the system a second time



(a) Learning Phase 1A (b) Learning Phase 1B

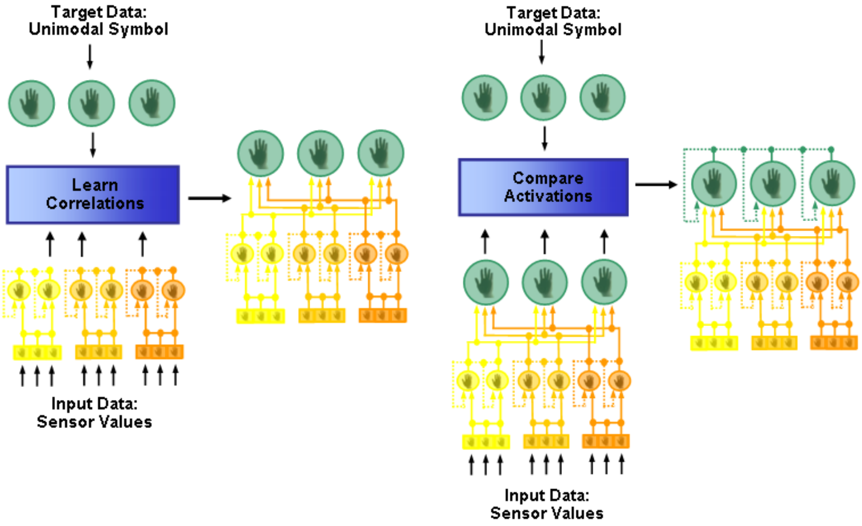
Figure 4.33: Learning Phase 1 for a Tactile Sub-modality

in learning phase 1B (see figure 4.33b). As there do already exist connections between feature symbols and sub-unimodal symbols, one or more sub-unimodal symbols are activated as reaction to the presented examples. The activated sub-unimodal symbols are now compared to the sub-unimodal target symbol that was intended to be activated. There are generated feedback connections between sub-unimodal symbols to prevent such undesired activations.

After the forward and feedback connections between feature symbols and sub-unimodal symbols have been set for each sub-unimodal system, sub-unimodal symbols have to be connected with unimodal symbols based on examples presented to the system. The learning of these coherences is very similar to the former level. Figure 4.34a illustrates the training phase 2A for the tactile system. Again, input data in form of sensor values are presented to the system. This time, not only sensor values of one sub-unimodal system are used but sensor values coming from all tactile sensors. Feature symbols are extracted from these sensory raw data and are further passed to the sub-unimodal layers, because there do already exist the connections between the feature layer and the sub-unimodal layer. According to the presented sensor data, certain sub-unimodal symbols are activated. The system also gets the information which unimodal tactile symbol shall be activated when certain sensor values occur. Based on this information, the connections between the sub-unimodal tactile symbols and the unimodal tactile symbols are set. In the training phase 2B, similar to phase 1B, feedback connections are set between unimodal tactile symbols to inhibit the undesired activation of symbols that do not correspond to the target symbol. In figure 4.34b, the principle is illustrated graphically.

The training phases 3A and 3B follow the same principle like the two former training phases. This time, the presented examples include sensor values coming from all modalities and their sub-modalities. Forward connections between unimodal symbols and multimodal symbols as well as feedback connection between multimodal symbols are formed according to the presented examples. The figures 4.35a and b illustrate the learning concept of training phase 3 graphically.

In the figures 4.36a and b, the learning phase 4, in which correlations between multimodal symbols and scenario symbols as well as feedback connections between scenario symbols are determined,



(a) Learning Phase 2A (b) Learning Phase 2B

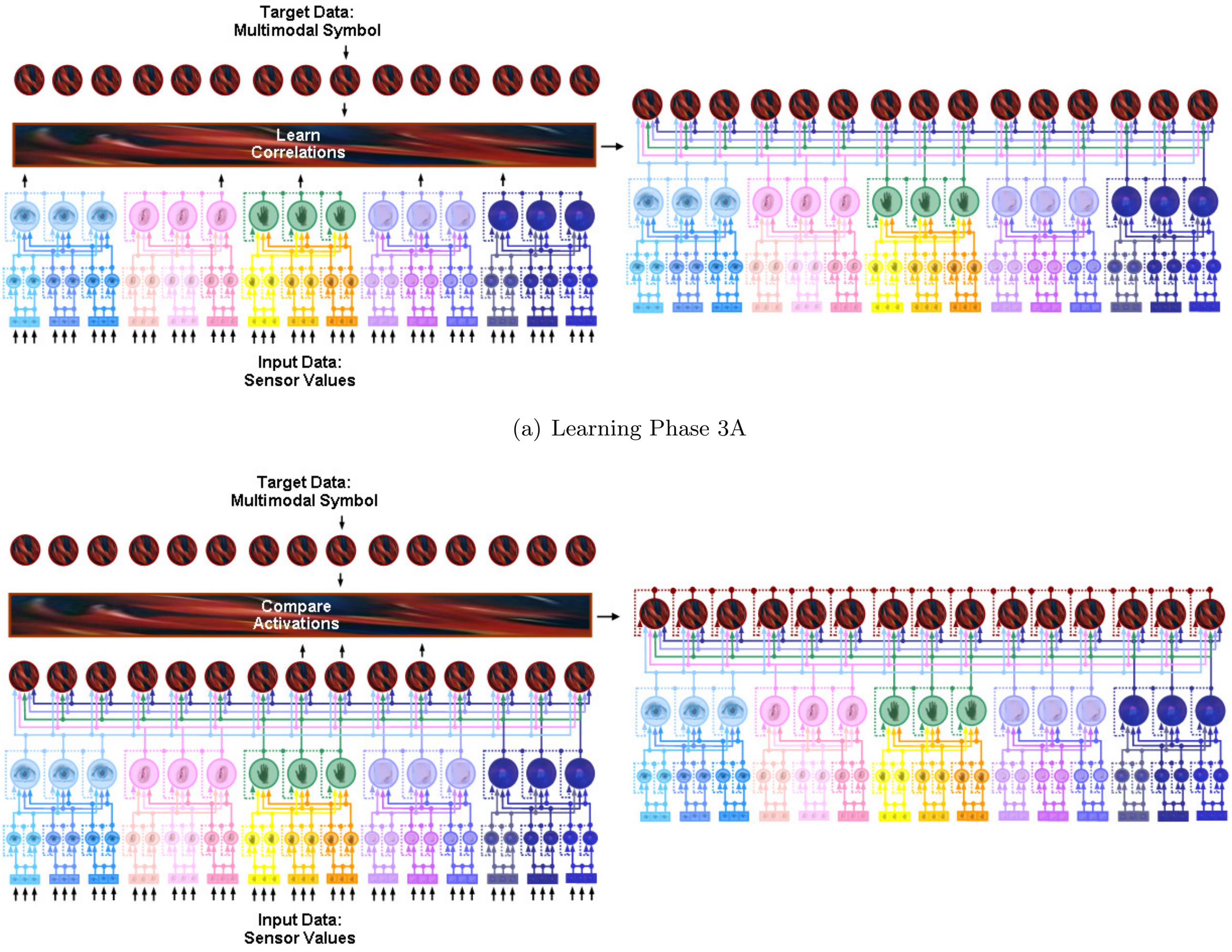
Figure 4.34: Learning Phase 2 for the Tactile Modality

is illustrated. Again, sensor data of all modalities contribute to the training process.

As mentioned in section 4.2.4, neuro-symbols can have properties, which specify a neuro-symbol in more detail. Depending on the current value of the property of a neuro-symbol, it can be assigned to different higher-level neuro-symbols. Therefore, if necessary, the value of properties needs to be considered in the training process. Values of properties are generally learned in the training phase A[14].

A special property of neuro-symbols is the location property, which is derived from the spatial position of triggered sensory receptors in a first instance and from location property values of lower-levels neuro-symbols in subsequent stages. The location property represents the position or spatial area where the perceptual image the neuro-symbol stands for has been perceived (see section 4.4.3). Similar to other property values, the activation of higher-level neuro-symbols can also depend on the position where the lower-level symbols they are connected to were perceived. Like for other properties, it is determined in the training phase A of each level what are allowed spatial areas for lower-level neuro-symbols to activate certain higher-level neuro-symbols. It can be learned how much different lower-levels symbols belonging to a certain event generally deviate in position. Additionally, it can be learned in this phase what are principally allowed positions or areas for certain lower-level neuro-symbols to be bound to a particular higher-level neuro-symbol. In the training phase B of the different levels and modalities, it can be determined how close the positions of neuro-symbols need to match to allow an inhibitory influence by feedback signals. To achieve good results in learning, it is important to have representative examples that cover all different places where an object, event, or situation can occur. However, due to the effort for generating examples, it will not be possible to cover really all locations where certain objects, events, and situation can take place by examples. The system needs to have the ability to generalize over the presented location information. One method how to implement these principles is described in section 5.5.

[14] The values of location properties have to be considered in the training phases A and B.



(a) Learning Phase 3A

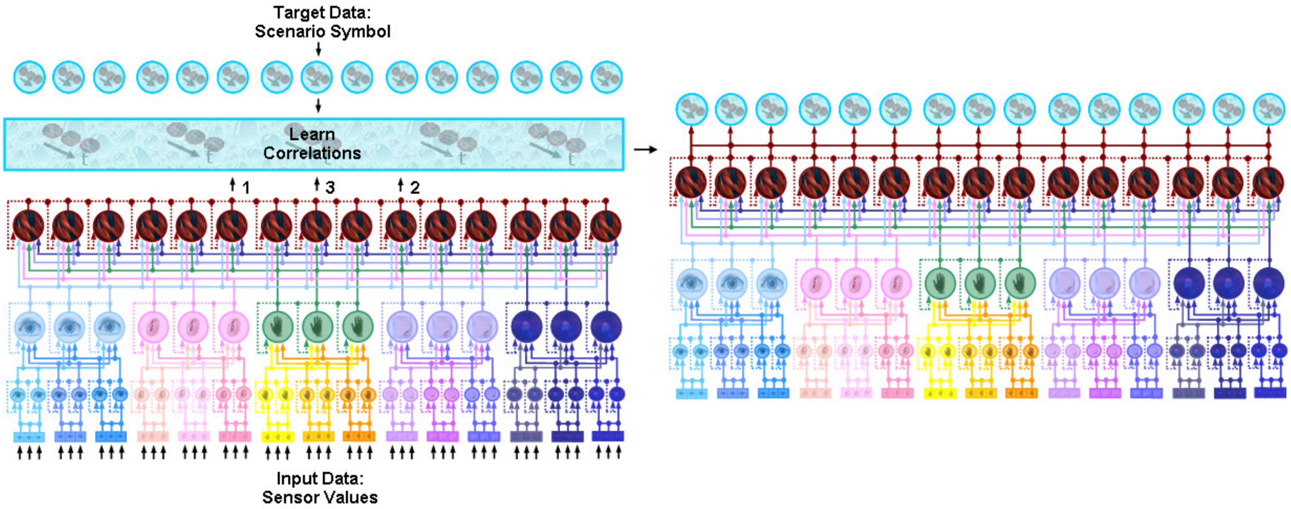
(b) Learning Phase 3B

Figure 4.35: Learning Phase 3

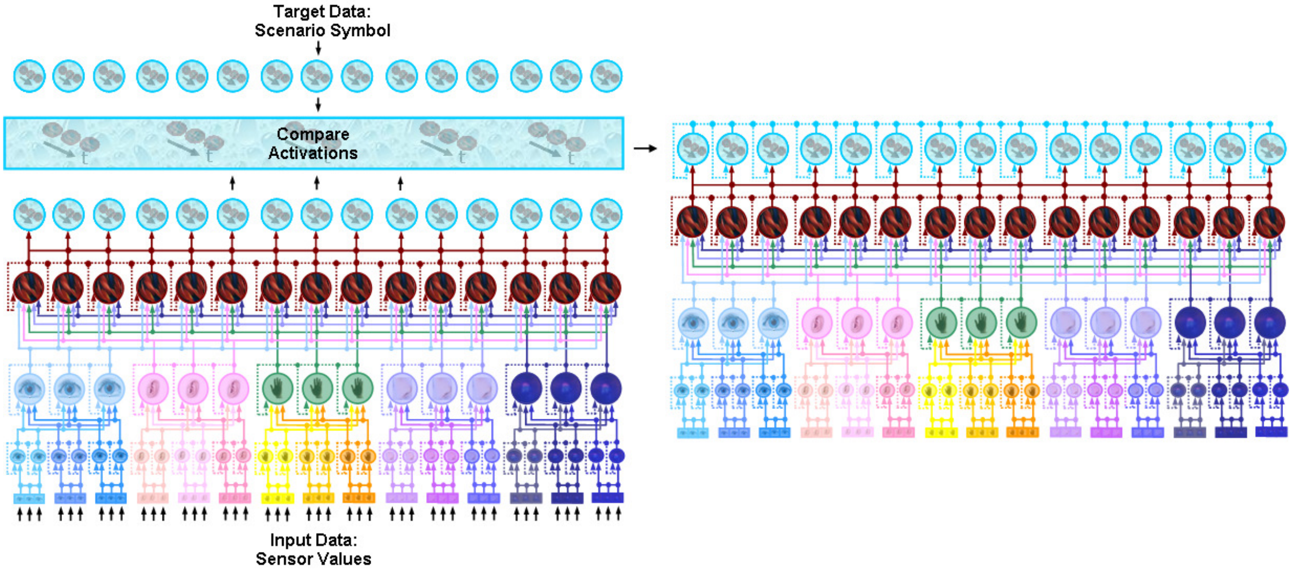
Similar to learning spatial correlations, there can also be learned temporal correlations between neuro-symbols. As described in section 4.4.4, there is the need to bind neuro-symbols occurring within a certain time window or to consider certain sequences of symbol activations. This information can also be extracted from examples and is calculated in the training phase A. A method how to implement these principles is also given in section 5.5.

Concept Clarification

To make the concept of learning even clearer, it shall be illustrated by means of a concrete example. The example uses the same sensor configuration and the same symbol hierarchy as already described in section 4.2.5. It shall now be shown how to set connections in training phase 2A and how to set feedback connections in training phase 2B. In training phase 2A, the connections between sub-unimodal and unimodal symbols are learned and it shall be determined what values properties of sub-unimodal symbols must have to activate a particular unimodal symbol.



(a) Learning Phase 4A

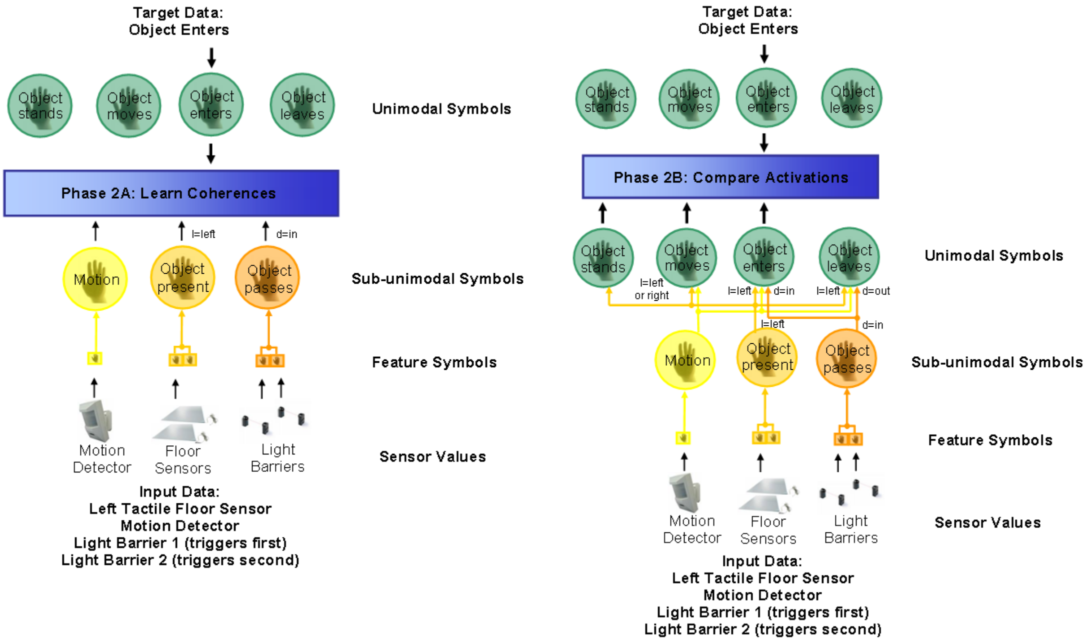


(b) Learning Phase 4B

Figure 4.36: Learning Phase 4

Therefore, examples are presented to the neuro-symbolic network. Input data for the learning module are real world sensor data, which have already been pre-processed and transformed into sub-unimodal symbols. The target data tell the system what signification the sensor values have. For instance, to train the system what the symbol "object stands" means, different objects (table, shelf, person, dog, etc.) are placed at different positions in the room. Because of the few sensors the system is equipped with for this example, the tactile system cannot distinguish between common objects and living beings like persons or animals. Therefore, in this modality, persons and animals are handled like other objects. A further specialization of the object type can be made when using additional information from the camera or the microphone. The values of the sensors that are activated by these object placements serve as input data in the training phase. The meaning of the situation represents the target data.

Figure 4.37a illustrates what sensor activations can trigger the symbol "object enters". To get these sensor activations, different examples are presented to the system in which an object enters the room and the triggered sensor data are recorded. For this situation, there is always activated only the left floor sensor. Additionally, the motion detector is activated and the two light barriers are triggered in a certain succession. In the special case of the situation "object enters", due to the small number of available sensors, there are always triggered the same sensors. However, in other situations or in configurations with more sensor types and sensors of higher spatial resolutions, there can occur differences in what sensors are triggered in different examples of the same situation. In such a case, the learning algorithm needs to have the ability to generalize over examples. To bind sub-unimodal symbols to the unimodal symbol "object enters", there also have to be considered properties of sub-unimodal symbols.



(a) Learning Phase 2A

(b) Learning Phase 2B

Figure 4.37: Example for Input Data, Target Data, and Neuro-symbol Activation in Training Phase 2

The result of setting the forward connections in learning phase 2A corresponds to the connections set in figure 4.15. After the connections between sub-unimodal and unimodal symbols have been set in training phase 2A, a certain combination of sensor values that corresponds to one of the four possible situations defined in the tactile system should trigger exactly one tactile unimodal symbol. However, as mentioned in section 4.2.5, more than one symbol is activated at the same time if there exist symbols that are made up of connections from the lower layer, which are a subset of the connections of other symbols. To avoid these undesired activations, inhibitory feedbacks are set in the training phase 2B. Therefore, the same "input data - target data pairs" like in phase 2A are presented to the system a second time. The learning module 2B memorizes the examples and compares what symbol should be activated in a particular situation and what symbols are actually activated. Accordingly, feedback connections are set between the unimodal tactile symbols. Figure 4.37b illustrates what symbols are activated when presenting an example of the event "object enters". Feedback connections are set to overcome the problem of undesired symbol activation. Figure 4.16 shows the result of training phase 2B after the connections have been set.

4.5.3 Architectural Changes by Flexible Addition and Elimination of Neuro-symbols

Determining neuro-symbolic connections, property values, location data, and timing data based on examples greatly increases the flexibility of the presented system compared to "hardwired" neuro-symbolic processing. A very useful supplement to these mechanisms would be to detect redundancy or incompleteness in the data processing structure and to change the neuro-symbolic structure accordingly to achieve greater efficiency.

Redundancy occurs if different perceptual images of a certain level and modality, which are assigned to different neuro-symbols in the training phase A, are in fact triggered by the same sensors and therefore also by the same lower-level neuro-symbols. In such a case, these neuro-symbols would always be triggered concurrently if one of the corresponding perceptual images occurs. Therefore, these symbols are redundant and all except one can be eliminated. This is done in training phase B. As already described, in training phase B, inhibitory feedbacks are set between symbols if one symbol is activated by a subset of lower-level symbols that activate another symbol. However, this is not the only function that this training phase has. Additionally, in this phase it is checked whether two symbols are always activated concurrently, because they are made up of the same set of lower-level symbols. If such a case occurs, one symbol is eliminated by removing all its connections to other symbols. Different algorithms are possible to check a symbol layer for redundant symbols. One suggestion for an algorithm is made in section 5.5.

Concept Clarification

How redundant symbols can be eliminated shall be explained by means of an example in the tactile system. The same principle can also be applied in the other modalities and the other layers. For the explanation, the example already described in section 4.5.2 is extended a little. For the tactile system, there are now defined five unimodal tactile symbols. Additionally to the former symbols, there is added the symbol "object placed". The correlations between the sub-unimodal and the unimodal symbols are learned from examples in training phase 2A. For the examples presented to the system, the connections set in this phase look as depicted in figure 4.38.

When comparing the connections of the symbol "object stands" with the connections of the symbol "object placed", it attracts attention that they are exactly the same. This is due to the fact that in the examples, the same sensor values were triggered for both examples. There now exist two symbols with different names that both represent the same situation. Therefore, one symbol can be eliminated. This is done in training phase 2B by disconnecting the symbol "object placed" from the sub-unimodal layer. As without connections the symbol will no longer be activated, it will not be considered in later learning phases and has no influence on later processing stages. Principally, the symbol can now also be assigned to another perceptual image if additional symbols are needed in the tactile unimodal layer. In such a case, for being able to better interpret the system, the labeling of the symbol can be changed.

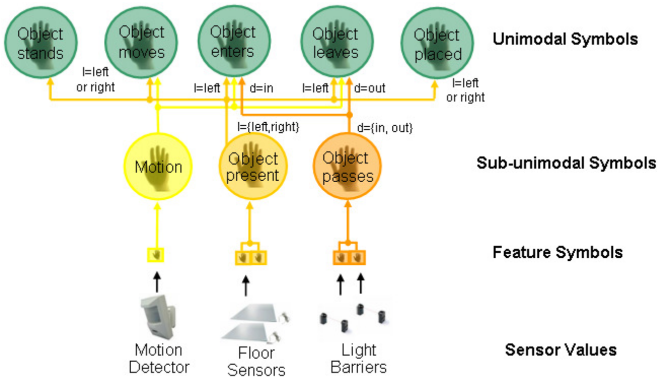


Figure 4.38: Connections after Training Phase 2A

Opposite to the case that neuro-symbols are redundant and should therefore be eliminated, there might also exists only one neuro-symbol for situations that should better be further distinguished and assigned to more than one neuro-symbol. An example would be to split the situation “person opens door” into the two perceptual images “person opens door from outside” and “person opens door from inside”. The differentiation whether the door is opened form inside or from outside makes sense if these two situations have different meaning to the system and correlating perceptions. A splitting of neuro-symbols is done in training phase A of the different levels. The system can detect a potentially desired splitting if in different examples of the same situation, different lower-level symbols or different properties of lower-level symbols are triggered (see also section 5.5 and section 6.1).

4.6 Influence of Knowledge on Perception

In the former sections of this chapter, there have been mainly considered bottom-up aspects (also called data-driven or data-based processes) of perception, which are based on incoming sensory information. Incoming data are always the starting point for perception. Without incoming data, there is no perception. However, as outlined in section 3.3.3, in the later stages of processing, perception is often affected by knowledge of different forms. Such processes are said to be top-down processes (also labeled as hypothesis-driven, expectation-driven, or knowledge-based processes). Knowledge can be factual (semantic) knowledge, pre-experience, knowledge about the context in which the objects and events occur, and expectation. Researchers do not yet agree if knowledge is always involved in perception or not. However, there is broad consensus that prior knowledge can help to interpret ambiguous sensory information and facilitate perception. Therefore, this kind of top-down processing shall also be integrated into the perceptual model proposed in this thesis.

Knowledge integration becomes especially important when not only one out of a few different objects, events, situations, or scenarios has to be detected but one out of thousands. Under these circumstances, it may happen that a situation or scenario cannot unambiguously be perceived from the current sensor values, because two or more situations or scenarios might trigger the same or very similar sensor values. In such a case, knowledge in different forms as well as “awareness” of what has happened before may lead to an unambiguous decision. As outlined in section

3.2, the hierarchical arrangement of perception and knowledge reverses during the maturational process. For small infants, everything depends on the senses, and cognition is driven by concrete perceptual reality. In adults, abstract knowledge derived from these early learning experiences comes to govern the perceptual process. For the model, this means that connections from lower to higher levels have to be formed first before knowledge can influence perception in a top-down manner. In pure bottom-up processing, certain neuro-symbols are activated depending on the sensor inputs. In the model, bottom-up information processing is the starting point for symbol activation. However, knowledge can influence the activation or non-activation of these neuro-symbols. Knowledge can have inhibitory or excitatory influence. This means that it can increase or decrease the activation grade of neuro-symbols.

A fundamental question is on what level knowledge influences perception. Unfortunately, the answers coming from neuroscience to this question are controversy. In the model of neuro-symbolic processing, it does not seem to make sense to let knowledge already interact with the lowest levels of perception like the sensor level or the feature level, because these levels correspond to brain structures, which are already predefined at birth. Knowledge, however, evolves only later on during lifetime. Therefore, it makes more sense to let interact knowledge with layers where correlations are also only learned during system operation. In the model, an interaction can principally take place from the sub-unimodal level upwards. Which level or levels are best suitable and most effective for such an interaction is best determined by trying out the different possibilities and evaluating the results. The results may also depend on the concrete application. In the following, concepts are introduced how knowledge of different forms can influence perception. For these concepts, there is also needed memory to make available information about objects, events, and situations perceived in the past, which are of importance for current or future perception. Therefore, before giving a more detailed description about knowledge integration, the concept of storing relevant information from past events is discussed.

Past Perceptions and Memory

In section 4.3.1, it was outlined that information processing in neuro-symbolic layers is performed event-based. This means that neuro-symbols are activated when certain sensors are triggered and deactivated when the sensory information disappears. In other words, up to the scenario symbol layer, symbol activations change when sensor values change. Information disappears when the sensors are no longer triggered. However, sometimes it might be necessary to store certain information derived form activated neuro-symbols for longer time. This is the case if certain events that already happened and are no longer represented by activated symbols influence the probability that certain other events can occur. In the brain, for storing information, memory is needed. It is not yet fully understood how memory is coded and retrieved in the brain and on what cortical levels it influences perception. It does not seem to be located in a small circumscribed area but to be distributed over large areas of the brain.

In a real world environment, the sensory system can receive lots of information, and many objects, events, or situations can be perceived every instant of time. However, not all information might be useful or necessary for future perceptions. Besides this, the huge amount of incoming data makes it impossible or at least ineffectual to store all information about the past. Therefore, a selection process has to take place, which has the aim to only store information that is relevant for future events. For example, when one multimodal symbol is activated, it might not make sense to store all information coming from all sub-modalities. The information delivered by the multimodal symbol might be sufficient. Aside from this, in many situations, it is not necessary to store the event that occurred itself but abstract from it what influence the event may have on future perceptions.

Concept Clarification

The described circumstances get clearer by means of a concrete example. The example used is again the example already introduced in section 4.2.5. There, it was detected if a person enters the room, leaves the room, stands in the room, or walks around in the room. The event "person enters" can now have an influence on the other three events. There cannot stand a person in the room, walk around, or leave the room if he/she has not entered before. Therefore, it should somehow be memorized if a person entered the room. However, the process of entering itself is not the relevant information. Instead, the information to be stored should be that a person is now present in the room.

Up to the scenario symbol layer, symbols are activated and deactivated when sensor values change. In these layers, there exists no explicit memory that stores past states. In the proposed model, for storing events or the consequences of events happened in the past, a new symbol layer is introduced. This is the *memory symbol layer* (see figure 4.39). This layer comprises so-called *memory symbols*, which can – similar to neuro-symbols – also have properties. The difference lies in the "activation" of memory symbols. Unlike neuro-symbols, memory symbols are not activated and deactivated based on objects and events appearing and disappearing, but they are rather *set* and *reset*. The setting and resetting is triggered by different neuro-symbols. Memory symbols are set after certain neuro-symbols have been activated and they are reset after other symbols have been activated.

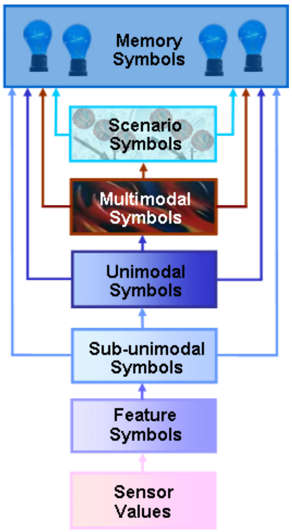


Figure 4.39: Introduction of a Memory Symbol Layer to Memorize Important Past Perceptions

One important question is from neuro-symbols of what level memory symbols can be set and reset. From neuroscience and neuropsychology, there does not yet come a distinct answer to this question. In the model, memory symbols can principally be triggered from information coming from neuro-symbols of the sub-unimodal level upwards. However, it might make little sense for many situations to trigger them from a level as low as the sub-unimodal symbol level, because one single symbol of this level does not comprise much information and by itself will not have much influence on future events. Additionally, lower layers are more often subject to failures and might deliver wrong information.

To set and reset memory symbols, they need to be connected to other neuro-symbols. One question is how these connections can be set. One solution would be to let the system designer choose these connections. However, as already described in section 4.5, for more flexibility, it is useful to learn the adequate connections from examples. The learning method used for this purpose is very similar to the methods presented in section 4.5.2. However, there exists only a learning phase A, and there have to be determined connections that set a memory symbol and other connections, which have the function to reset it. Therefore, memory symbols have two independent inputs. Taking over principles from neuroscience as described in section 3.2, to be able to learn the connections of this layer, the connections between symbols of lower levels they are set and reset from must already have evolved. An example for the usage of memory symbols is given further below.

Integration of Semantic Knowledge

As already mentioned in section 4.4.6, knowledge can influence the activation grade of neuro-symbols. The influence of knowledge can be inhibitory or excitatory. In the following, the influence of semantic knowledge on perception is described. As outlined in [ST02, chapter 5, p. 150], semantic knowledge represents basic knowledge of the world. It is stored in the form of third-person information of the kind that one might find in an encyclopedia. It comprises bits of objective information about the world and its workings-facts such as “objects fall down” and “persons do not walk through closed doors”. As already outlined, neuroscience has not yet answered the question on what level knowledge interacts with perception. In the model, similar like memory symbols, knowledge can principally interact with perception in levels equal to or higher than the sub-unimodal level. The interaction of semantic knowledge with perception can also require the storage of past events as just described.

It is not yet very well understood how semantic knowledge is coded and retrieved in the brain. Semantic knowledge might – at least to a certain extend – be represented and stored in other parts than the perceptual system of the brain. Therefore, in the presented model, semantic knowledge is not represented by perceptive neuro-symbols but is considered to originate from sources outside the neuro-symbolic network and to only interact with perceptive neuro-symbols. As semantic knowledge comprises facts about the world, it can be represented by rules. These rules decide if the activation of symbols is increased or decreased. An important question is how the rules representing semantic knowledge can be acquired. One variant is to let the system designer explicitly define and formulate these rules. As semantic knowledge represents world knowledge, for humans, it is quite simple to define these rules. The second possibility would be to learn and extract these rules from examples. This possibility is far more complex to implement than the first one and will not be considered in the current model.

Concept Clarification

In the following, an example is given which clarifies the usage of memory symbols and semantic knowledge in the perception process. To illustrate this fact, the example of section 4.2.5 is extended a little by introducing a door contact sensor as additional sensor type to the system and some additional neuro-symbols like the sub-unimodal symbol "door status", the unimodal tactile symbols "door opens" and "door closes", the unimodal acoustic signal "door click", and the multimodal symbols "person opens door" and "person closes door". In this concrete example, two memory symbols are set or reset depending on the activation of multimodal symbols. The memory symbols have the purpose to store information, which is – together with top-down knowledge – important for the perception process of objects, events, scenarios, or situations. In the example depicted in figure 4.40, the system shall "know" that a person can only walk around in the room, stand in the room, or leave the room if he/she entered before. This means that the symbols "person walks", "person stands", and "person leaves" can only be activated if the symbol "person enters" was activated before. A second example would be to memorize if the door was opened or closed. The knowledge that a person can never enter or leave a room if the door is closed can help to lead to a resolution of ambiguous scenarios where a person comes close enough to the door to trigger the light barriers but does not enter or leave the room. Furthermore, there cannot be detected a "person closes door" or "person opens door" situation if the door is already closed or open. Such rules can be stored as semantic knowledge. A utilization of information of that kind also requires a sort of memory to store important information from past events. Therefore, the two memory symbols "person present" and "door open" are introduced. In the example, the states of the memory symbols are set or reset when certain multimodal symbols are activated. The memory symbol "person present" is set after the multimodal symbol "person enters" was activated. It is reset after the symbol "person leaves" occurred. The symbol "door open" is set by the symbol "person opens door" and reset by the symbol "person closes door".

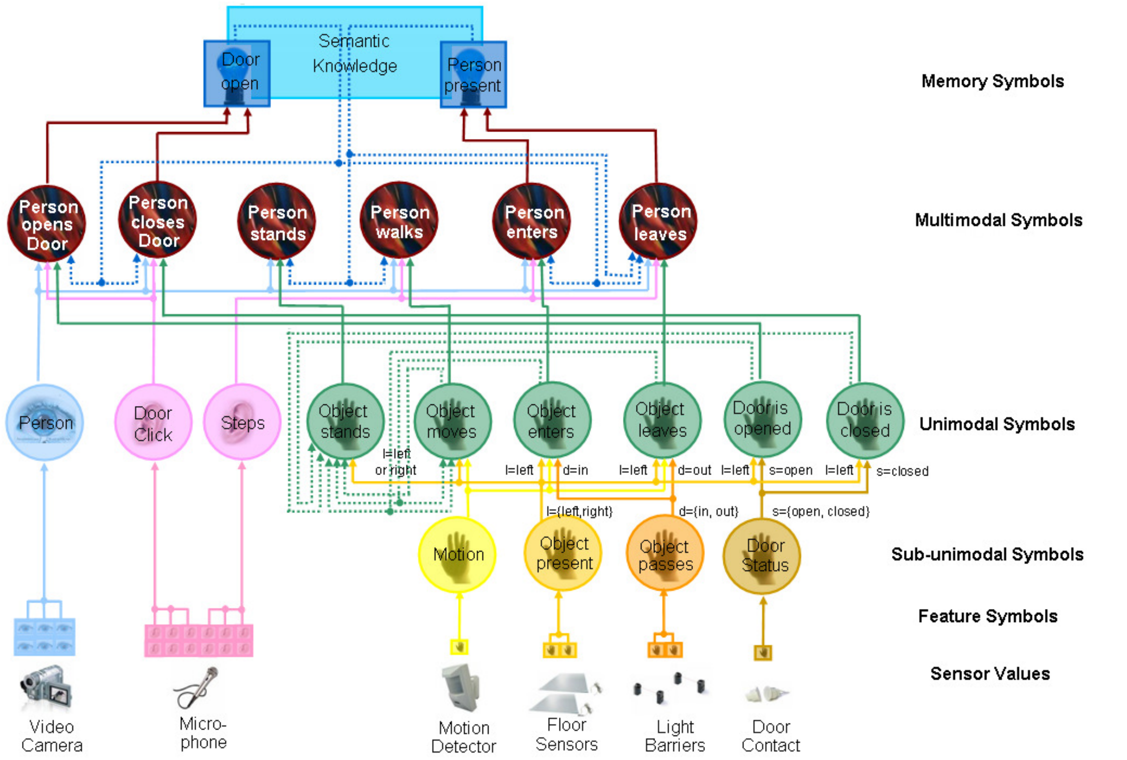


Figure 4.40: Interaction of Memory Symbols and Semantic Knowledge with Neuro-symbols

In figure 4.40, the blue dotted lines illustrate how knowledge influences the activation of multimodal symbols. In this example, semantic knowledge can have inhibitory influence on the activation of symbols. If the memory symbol “door open” is not set, which indicates that the door is closed, the activation of the multimodal symbols “person enters” and “person leaves” is inhibited, because a person cannot pass through a closed door. Additionally, the symbol “person closes door” cannot be activated, because the door is already closed. In contrast, if the symbol “door open” is set, only the symbol “person opens door” is inhibited, because an already opened door cannot be opened a second time. When the memory symbol “person present” is reset, it has inhibitory influence on the multimodal symbols “person stands”, “person walks”, and “person leaves”.

Integration of Context Knowledge

As outlined in section 3.3, besides pure factual (semantic) knowledge, context knowledge also has a not negligible influence on perception. The definitions of context knowledge given by different authors often differ from each other. A detailed discussion about context and its definitions is given in [DA99]. In this article, context is defined as *“any information that can be used to characterize the situation of an entity. An entity is a person, place, or object that is considered relevant to the interaction between a user and an application, including the user and application themselves.”*

Examples for context information that can be used do facilitate perception in the proposed model for the envisioned applications are knowledge about the location and the environment where an object or event is perceived or knowledge about the time of day, season, or temperature.

An example for using the information about the time of day would be to decide if a person is preparing breakfast or supper in a kitchen. Even if the same food is prepared and almost the same sensor values are triggered, knowledge about the context of time will lead to the decision what is going on. There might also exist other activities that always take place at a certain time. For example, the cleaning staff in a building might always work within a certain time period. This shows that knowledge about the context can greatly facilitate object and event perception from inconclusive sensor data. The principle how context knowledge influences perception works the same way as already described above. There could exist a memory symbol “day of time”, which in this case cannot be only set (time is known by the system) and reset (time is not known by the system) but has a property, which contains the information about the current time. Stored rules about which situations are likely to happen at what time can influence the activation of neuro-symbols in a top-down manner.

In the model, context knowledge about the location of events or objects plays a special role. Principally, it would be possible to handle location context like other kinds of knowledge and context only in a top-down manner. However, the information about the location of perceived images is accessible in all hierarchical levels. Therefore, the contextual location restrictions can also be performed directly on lower hierarchical levels. This method was already described in more detail in section 4.4.3.

Influence of Expectation

As described in section 3.2, in adults, abstract knowledge derived from early learning experiences comes to govern the perceptual process. We see what we expect to see and are surprised or fail to notice when our expectations are contradicted. Therefore, expectation is also considered in the model. Expectation can be considered as a type of knowledge and therefore, similar to factual knowledge and context knowledge, it influences perception in a top-down manner. The interaction of expectation with neuro-symbols works similarly to the other types of knowledge already described. Expectation can increase or decrease the activation of neuro-symbols. In the brain, the determination what is likely to happen next is not exclusively performed in the perceptual system of the brain. How expectations are formed in the brain is a question that has not yet been answered. For the model, it could be determined what events are more probable to happen at certain locations, at a certain time, or after certain other events happened either by the system designer or by statistical or other learning methods. Again, the second possibility is more complex than the first one and will not be considered in the current model.

4.7 Modularity and Hybrid System Design

In the former sections of this chapter, a modular hierarchical model for multisensory information processing was presented. This model is based on information processing of sensory data in different modalities and over several hierarchical layers. Principally, information is processed by so-called neuro-symbols, which are the basic information processing units of the system. Except for the lowest level, symbols of a higher layer are “formed” by connections of neuro-symbols from lower layers (or feedback connections from the same layer). On the lowest layer, neuro-symbols are “extracted” from sensor values. However, especially in the lowest levels where neuro-symbolic learning cannot be applied, it can require a lot of design effort to define correlations between lower-level symbols and higher-level symbols. Therefore, in certain cases, it might be easier to use already existing, workable solutions to extract higher-level symbols directly from sensor data instead of determining these neuro-symbols by connections over several layers of lower-level symbols.

The modular hierarchical structure of the model allows it to substitute for certain modules the neuro-symbolic information processing by other methods. This is recommendable if these methods achieve the same or a better result with less effort for the designer. When using such solutions, the otherwise purely neuro-symbolic information processing structure becomes a hybrid system. The usage of existing solutions is especially recommendable for the visual and the auditory modality. Visual image processing and auditory data processing are huge research fields. There might already exist workable solutions to detect and classify various objects and sounds. By using these existing solutions, sub-unimodal or unimodal visual and acoustic neuro-symbols can be determined directly from sensor data. The feature symbol level (and in many cases also the sub-unimodal level) are skipped in processing. However, implicitly, most existing algorithms used also extract features from raw data.

Concept Clarification

To make the concept of hybrid systems design just outlined even clearer, it shall now be explained by means of a concrete example (see figure 4.41). Due to the modularity of the model, for certain modalities or sub-modalities, neuro-symbolic information processing can be substituted by other strategies. For the explanation, the same example is used that was introduced in section 4.2.5 when explaining the usage of feedback connections.

In the research field of image processing, there do already exist methods and algorithms in image processing to identify persons from pixel data of a camera. To derive the information whether

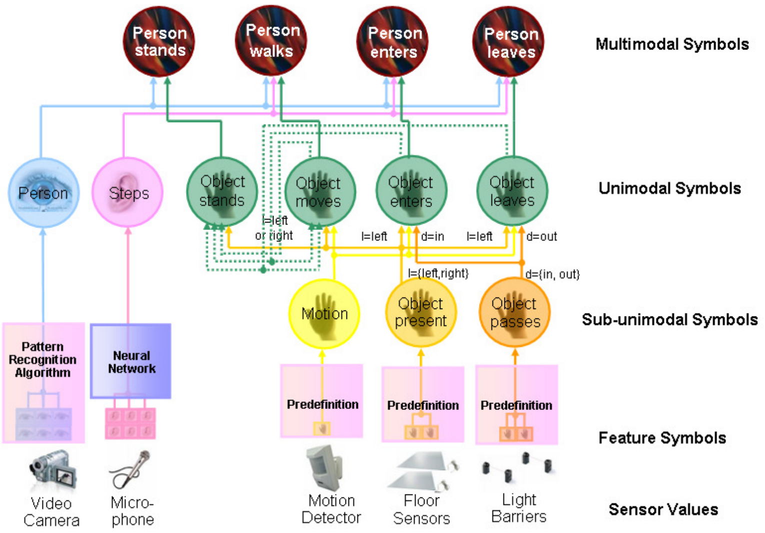


Figure 4.41: Example for Hybrid System Design

a person is present in the room from camera data and activate the unimodal visual symbol "person" in case of the presence of a person, such a pattern recognition algorithm can be applied. Generally, for image processing, special characteristics are calculated from pixel images. The calculation of these characteristics corresponds to the determination of feature symbols. The pattern recognition algorithms to identify persons based on these features correspond to the connections between feature symbols and sub-unimodal as well as unimodal symbols.

Similar to the research field of image processing, there do already exist solutions for lots of situations where certain kinds of sound signals need to be classified. Besides statistical methods, in literature, lots of articles can be found where sound signals are classified by artificial neural networks. Therefore, in the example, to detect the characteristic noise of steps, it is suggested to use neural networks. The advantage of neural networks is that they can learn correlations between data from examples. Methods how to design and train neural networks for this purpose can be found for example in [Roj96, CS97]. Again, there are generally extracted certain features from auditory data before presenting them as input data to neural networks. This corresponds to the calculation of feature symbols. During the training phase of neural networks, weights of connections between neurons are set according to the presented examples, which is related to the connections set during the neuro-symbolic training phase 1A and 2A, respectively[15].

In the example, for the tactile sub-modalities, only very few sensors are used. The correlations between sensor values, feature symbols, and sub-unimodal symbols are quite straightforward. It does not make sense to learn these very simple correlations. Therefore, they are predefined before system startup.

[15] When using only one microphone, connections between features and unimodal symbols are set directly.

Chapter 5

Design Methodology and Implementation

"Nothing shocks me. I'm a scientist."

[Harrison Ford as Indiana Jones]

In chapter 4, a bionic model for human-like perception based on neuroscientific and neuropsychological research findings was introduced. To evaluate and test the model, it has to be implemented. This chapter aims to give an overview about the model implementation[1]. In section 5.1, the used tools for implementation and simulation are outlined. Section 5.2 describes into what modules the model is divided for the implementation and section 5.3 describes the interfaces between the module blocks. How model building blocks are actually realized is drafted in section 5.4. Section 5.5 gives an overview how learning and adaptation is performed, and section 5.6 finally summarizes the system design methodology.

5.1 Tool Selection

Principally, the model could be realized in hardware or in software. The advantage of a hardware realization would be the capability of real parallel processing of data. In contrast to a realization in hardware, a software simulation offers more flexibility in changing design parameters for the evaluation process. For test purposes, the proposed model was simulated on a computer. This section briefly describes the tools used for simulation.

[1] The aim of this chapter is to give a design overview of the whole system and does not intent to get bogged in too specific implementation details. For more specific information, it is recommended to take a look at the source code and its documentation.

5.1.1 Tool for Model Simulation

To test the model of human-like perception introduced in chapter 4, the design is implemented and simulated in software. Therefore, a suitable simulation tool has to be chosen. One requirement for the software tool is to allow parallel processing (or simulated parallel processing) of information, because sensory information has to be processed separately and in parallel before being combined and merged. Furthermore, the introduced concept of neuro-symbols, the interconnection of neuro-symbols, and their information exchange should be easy to implement. Additionally, it is desirable to have a graphical programming interface where the neuro-symbols can be placed and connections can be set to improve design clarity. The simulation tool chosen for this purpose is AnyLogic[2].

AnyLogic has proven to be successful in the modelling of large and complex systems. The main building block of the AnyLogic model is the *active object*. Active objects can be used to model very diverse objects of the real world: processing stations, resources, people, hardware, physical objects, controllers, etc. AnyLogic supports the programming language Java. Active object classes map to Java classes. To implement the neuro-symbolic information processing concept, besides active objects, *state variables*, *interface variables*, *ports*, *connections*, *messages*, *timers*, and *state charts* are used.

Each of the design elements just mentioned is now briefly described to understand how it can be used to realize the proposed model. For further information about the modelling language AnyLogic see [Any04]. Figure 5.1 illustrates how the different design elements are represented graphically in AnyLogic.

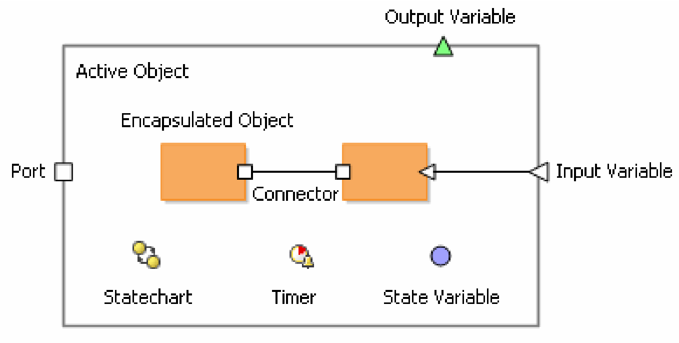


Figure 5.1: Design Elements of AnyLogic used for Implementation

Active Objects: As already mentioned, active objects are the main building blocks of the AnyLogic model. Active objects are instances of active object classes. When developing an AnyLogic model, there are actually developed classes of active objects and their relationships are defined. Active object classes map to Java classes and therefore allow inheritance. Once an active object class is defined, multiple active object instances can be created in the model. It is also possible to model class hierarchies.

State and Interface Variables: To define some data unit, a variable can be used. Variables may be either internal (state variables) or public (interface variables). Variables can appear in differential and algebraic equations and model values changing continuously over time. There can also be declared Java member variables. Interface variables can be shared with other active

[2] Version 5.2

objects. When two interface variables of different objects are linked, changes of one variable are immediately propagated to the other variable. The value of the variable defined as output will be passed to the variable defined as input.

Messages: A message is a data packet, which is passed between active objects. Messages can model various objects of the real world. Usually, a message carries some data. To define a message, a message class needs to be defined with necessary member variables.

Ports: Ports play the central role in message passing inside objects and between objects. Messages are sent and received via ports. Ports are bidirectional and serve both for input and output.

Connections: To establish inter-object interaction, interface elements of active objects (ports, interface variables) have to be connected with connectors. A connector is a line connecting two ports or two variables. Connecting two ports means that messages will be passed between them. Connecting variables means that they will have the same value at any moment of time.

Timers: The activities within the active object can be defined using timers. Timers are used to schedule user-defined actions.

State Charts: Active objects perform operations in response to external or internal events and conditions. The existence of a state within an active object means that the order in which operations are invoked is important. For some objects, this event- and time-ordering of operations can best be characterized in terms of a state transition diagram – a state chart. A state chart is used to illustrate the state space of a given algorithm, the events that cause transitions from one state to another, and the actions that result from state changes.

5.1.2 Tool for Sensor Data Generation

To test the model and its implementation, sensors are necessary that are triggered when certain activities are going on in the environment. These sensor data are the starting point for information processing in the model. For simulation and test, objects, events, scenarios, and situations in a building shall be perceived. However, in principle – with certain modifications – the model could also be applied to other applications (see section 1.3).

As described in section 2.1.2, in the ARS-PC project, for test purposes, a room – the institute's kitchen – has been equipped with about 100 sensors. Among the sensor types are tactile floor sensors, motion detectors, door contact sensors for the entrance door and the fridge door, a camera, and a shock detector to indicate if the kitchen's coffee machine is activated. This environment can be used to get test data. However, as it turned out, this sensor configuration only allows limited testing due to the relative small number of sensors, the spatial restriction of the test environment, and the fact that the configuration is fixed and cannot be changed to perform different experiments. In contrast, the developed simulator mentioned in section 2.1.2 offers more flexibility for testing and allows the consideration of lager areas and whole virtual buildings. With the simulator, the number, types, spatial resolution, and location of sensors can be varied. For this reasons, for implementation and test, simulated sensor data of sensors typically occurring in a building are used. The simulator is capable of calculating from simulated activities going on in the virtual environment what sensor values would be triggered from these activities. The simulation on sensor basis is possible for tactile floor sensors, motion detectors, light barriers, and door contact sensors. For audio and video data, a fragmentation down to the sensor level has not been implemented until now. For these two modalities, there can be generated symbolic information like for example that a person is perceived by a video camera or that the sound of steps or a voice was detected. For the simulation of the model presented in this work, this fact is not necessarily considered as a disadvantage. The provision of symbolic data for the visual and acoustic modality allows it to skip the lowest-level processing steps for these modalities which reduces the design effort without significant loss of testability of the model. It also is in accordance with the principle of modularity and hybrid system design of the model (see section 4.7), according to which for certain modalities – specially for video and audio data processing – workable solutions should be used to extract higher-level symbolic information directly from raw data if this reduces the design effort.

5.2 Model Modularization

To implement and simulate the model introduced in chapter 4, the design is spilt into two main building blocks, both modelled as active objects (see figure 5.2). The first block is the *sensor value generator* and the second is the *perceptual system*. The perceptual system comprises an implementation of the model introduced in chapter 4. The sensor value generator is responsible for producing sensor values or in certain cases also symbolic information to test the model. As outlined in section 5.1.2, for evaluation purposes, sensor data are simulated. This allows flexible changes in the sensor configuration and the environment, which is an advantage during testing and evaluation versus a fixed hardware installation of sensors. The perceptual system receives information from the sensor value generator via ports. Due to the modular splitting into the two blocks, the sensor value generator can be easily substituted by a system providing data from a real world configuration suited for a certain application.

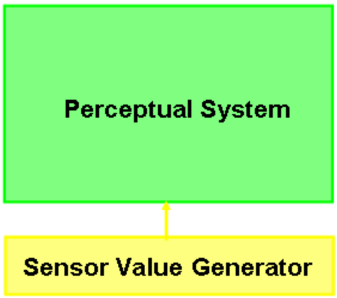


Figure 5.2: Two Main Building Blocks for Simulation

How data provided by the sensor value generator can be generated is not subject of this thesis. A discussion of this topic can be found in [Har08]. What shall be discussed in more detail in this chapter is how to implement the perceptual system. The perceptual system can be further divided into the modules *neuro-symbols*, *memory symbols*, *knowledge*, and *focus of attention* depicted in figure 5.3. The modules “neuro-symbols” and “memory symbols” are in fact not single modules but are here represented in place of all single neuro-symbols and memory symbols existing in the perceptual system (see figure 5.5). The modules knowledge and focus of attention can directly communicate with single neuro-symbols and memory symbols, respectively, without having to interact through a separate interface.

As outlined in section 4.2.2, there exist different types of neuro-symbols for different hierarchical levels and modalities. As described in section 5.1.2, for testing, it is not necessary for all modalities

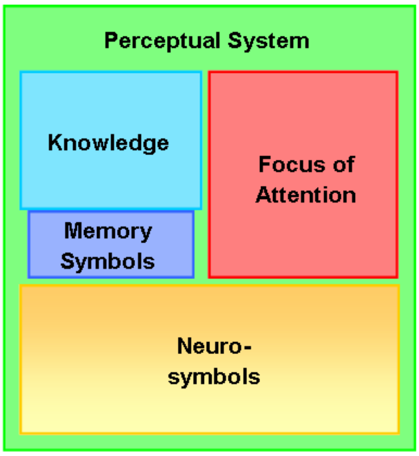


Figure 5.3: Division within the Perceptual System

to process perceptive information from the sensor data up to the scenario level. This would lead to a lot of design effort for defining correlations between the lowest levels without significant gain of insights. The sensor value generator provides the possibility not only to provide sensor data but also symbolic data by just not splitting its perceptive information down to the sensor level.

Figure 5.4 shows from which level upwards information processing has been implemented for the different hierarchical levels and modalities. The sub-modality of tactile floor sensors is the only modality where information processing can actually start from sensor data. Therefore, the first information processing level is the feature symbol level. For the other tactile sub-modalities, the sensor value generator is expected to deliver already sub-unimodal symbols. In the visual and the acoustic modality, there shall even be provided unimodal symbolic information.

As already mentioned, the perceptual system is modeled as active object. Similarly, neuro-symbols, memory symbols, the knowledge module, and the focus of attention module are also modeled as active objects. These model building blocks can now be placed in the perceptual system (see figure 5.5). For the neuro-symbols and the memory symbols, more than one instance of their corresponding active object is used.

5.3 Interfaces between Model Building Blocks

In the last section, the model was divided into different modules. To perform the desired perception task, these modules have to interact and communicate with each other. To allow such a communication, so-called ports and connectors are used in AnyLogic. Ports are graphically depicted as squares and connectors as lines.

Figure 5.6 shows the interface between the sensor value generator and the perceptual system. Both the sensor value generator and the perceptual system have seven ports being interconnected by connectors. The flow of information is directed from the sensor value generator to the perceptual system. Therefore, the ports of the sensor value generator are configured as output ports and the ports of the perceptual system are configured as input ports. To exchange information via ports, messages are used. How messages for different purposes have to look like will be explained later on in section 5.4.

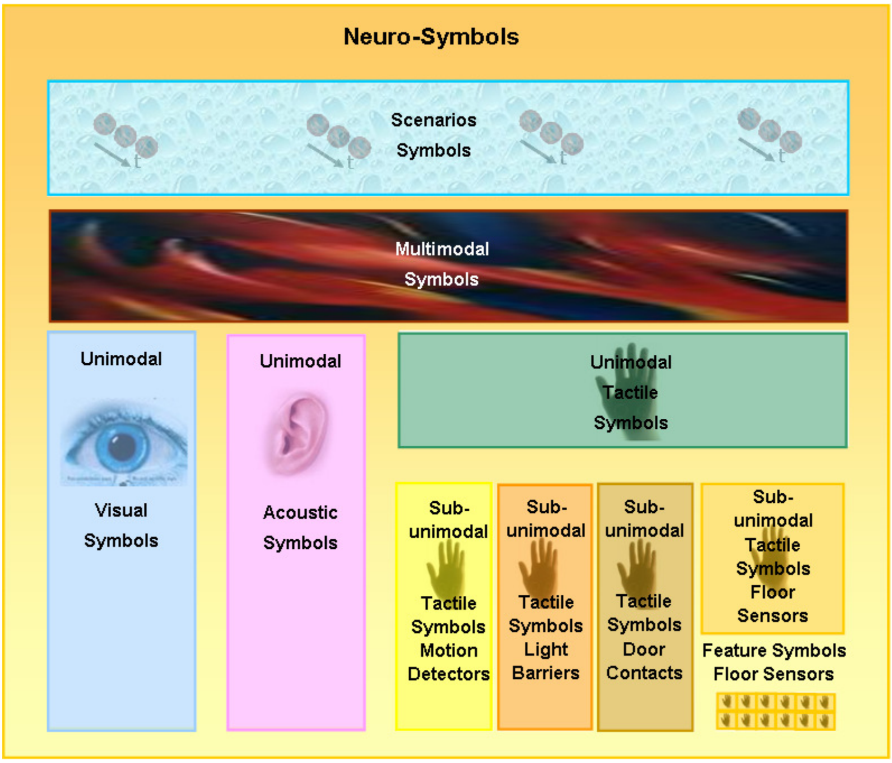

Neuro-Symbols
Scenarios Symbols
Multimodal Symbols
Unimodal Visual Symbols
Unimodal Acoustic Symbols
Unimodal Tactile Symbols
Sub-unimodal Tactile Symbols Motion Detectors
Sub-unimodal Tactile Symbols Light Barriers
Sub-unimodal Tactile Symbols Door Contacts
Sub-unimodal Tactile Symbols Floor Sensors
Feature Symbols Floor Sensors

Figure 5.4: Implemented Levels and Modalities of Neuro-symbolic Information Processing

Taking the sensory receptors of the human body as archetype, there would have to exist as many ports as different sensors. However, as already outlined, for all modalities and sub-modalities except for the tactile floor sensors, the sensor value generator already provides symbolic data. Therefore, it is sufficient to use only one port for each of these modalities and sub-modalities, respectively. To preserve straightforwardness of the design, for the sub-modality of the tactile floor sensors, there is also used only one single port in the simulation. The information from what sensor data originate is coded in the messages sent via the port. Therefore, in sum, there exist six ports for information exchange of sensor data and symbolic data. The seventh port of the sensor value generator and the perceptual system is used to exchange control data. Control data are not necessary after the system is once configured correctly. However, control information is necessary at startup during the learning phases where correlations between sensor values are learned from examples.

As depicted in figure 5.5, within the perceptual system, there exist different model building blocks. To communicate, these blocks also have to be equipped with ports. Memory symbols and the knowledge module additionally use so-called interface variables. Their practical effect will be explained in section 5.4.2. Figure 5.7 depicts the building blocks including their communication interfaces. As can be seen in the picture, there exist additional so-called learning ports, which are needed during the training phases of the system. After training, they are no longer required.

For information exchange between connected units, messages are sent via the ports. There exist different message types for different building blocks. Neuro-symbols use so-called neuro-symbol messages to transmit information. Neuro-symbol messages will be introduced in section 5.4.1. For information exchange between memory symbols and knowledge, there are not used messages

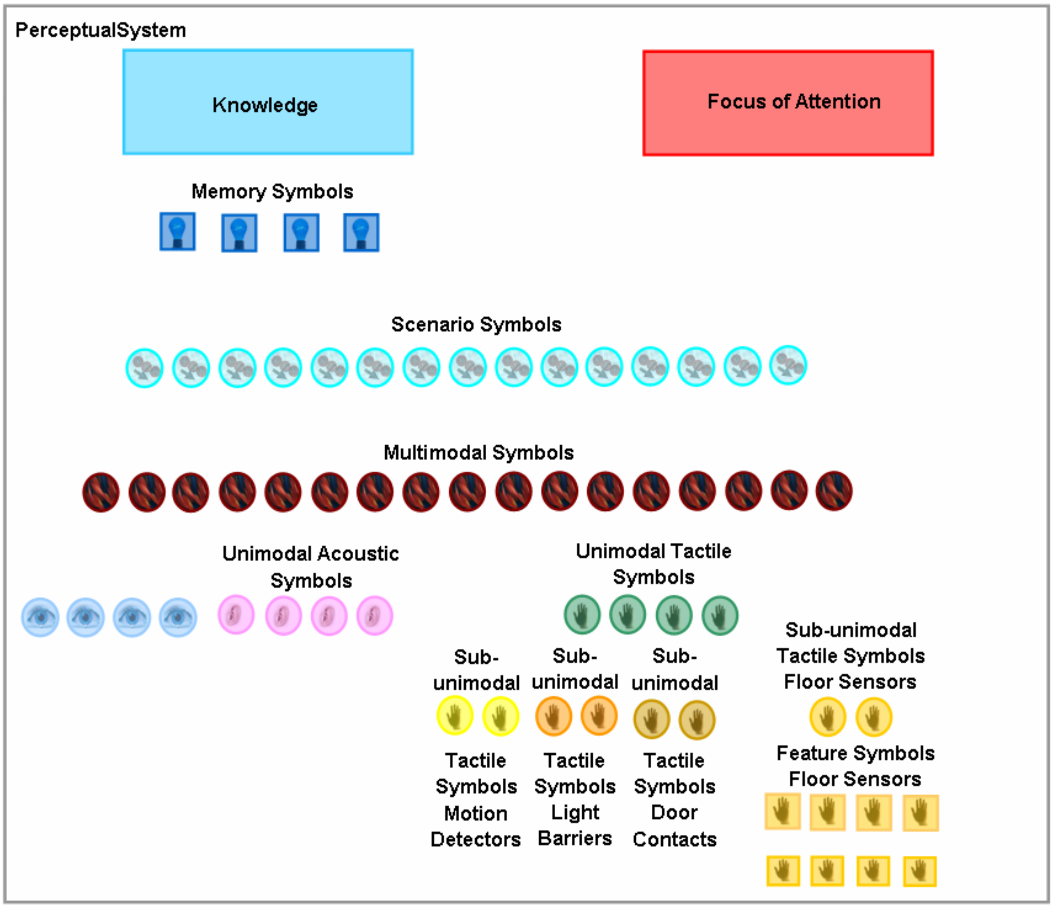

PerceptualSystem
Knowledge
Focus of Attention
Memory Symbols
Scenario Symbols
Multimodal Symbols
Unimodal Acoustic Symbols
Unimodal Tactile Symbols
Sub-unimodal Tactile Symbols Floor Sensors
Sub-unimodal
Sub-unimodal
Sub-unimodal
Tactile Symbols Motion Detectors
Tactile Symbols Light Barriers
Tactile Symbols Door Contacts
Feature Symbols Floor Sensors

Figure 5.5: Placement of Model Building Blocks in Perceptual System

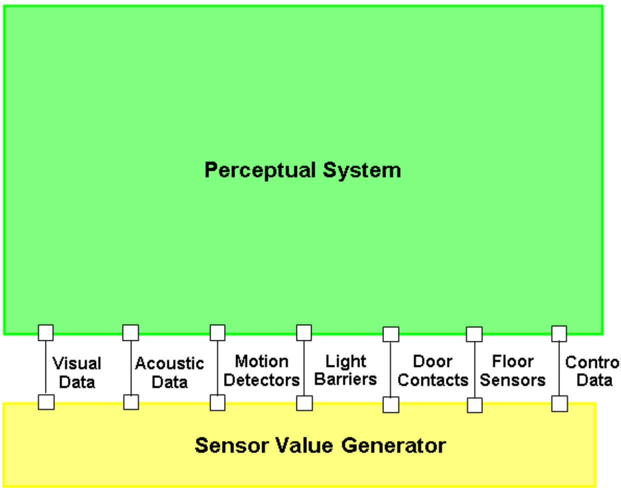


Figure 5.6: Ports and Connectors for Information Exchange between Sensor Value Generator and Perceptual System

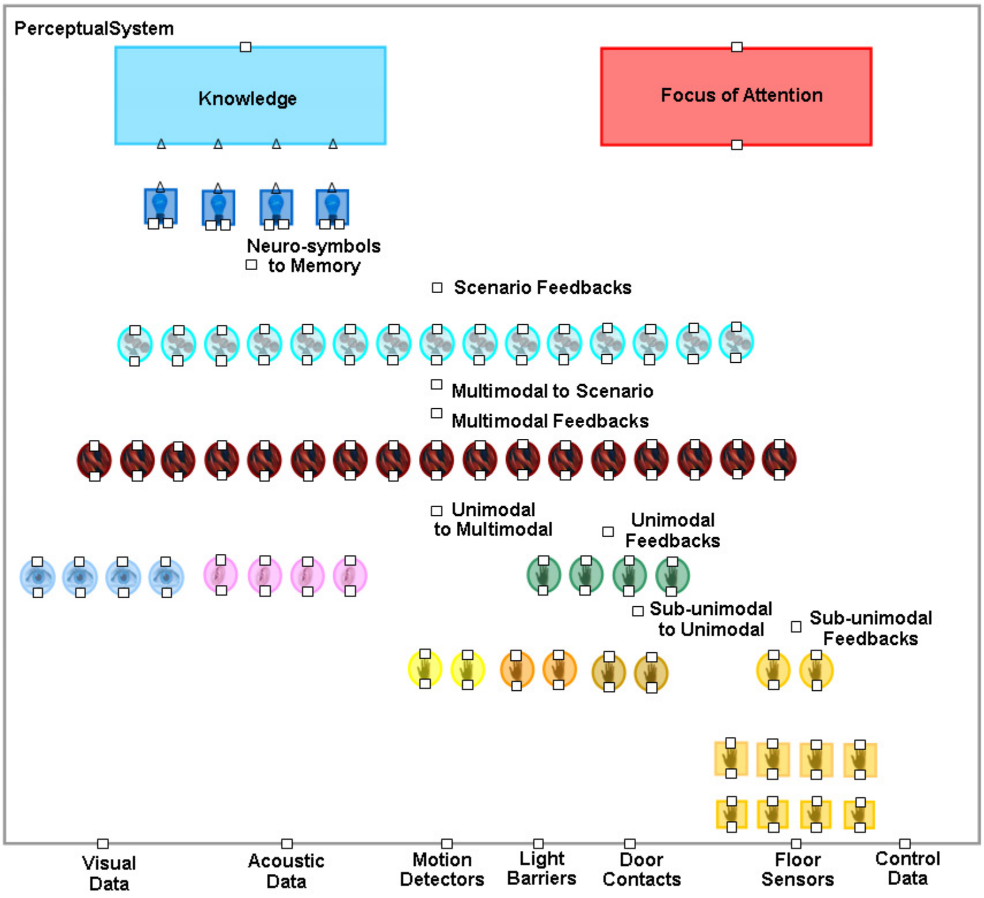


Figure 5.7: Ports and Interface Variables for Information Exchange between Building Blocks within the Perceptual System

but communication takes place by so-called interface variables depicted as triangles (see section 5.4.2). For the focus of attention module and the knowledge module, there also exist special message types (see sections 5.4.2 and 5.4.3). Besides these messages for information exchange between building blocks responsible for the perceptive task within the perceptual system, there also exist messages, which are generated from the sensor value generator and are passed via the input ports of the perceptual system to the units they are connected to. For this purpose there exists a message type to transmit sensory data. For sending symbolic data, the same message type is used like for the information exchange between neuro-symbols. Additionally, there exist so-called control data messages comprising control information in addition to sensor data and symbolic data necessary during the different learning phases.

5.4 Realization of Model Building Blocks and their Communication

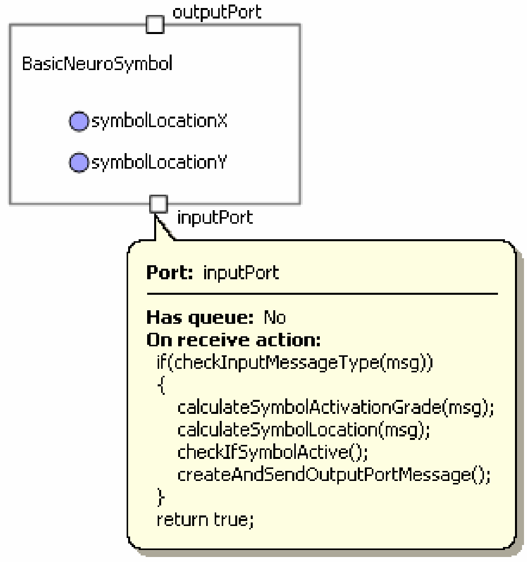
In the figures 5.5 and 5.7, different model building blocks necessary for the implementation were depicted. This section comprises a description how to realize these model building blocks and their communication in AnyLogic. The main building blocks are neuro-symbols of different types (see section 5.4.1), memory symbols (see section 5.4.2), knowledge (see section 5.4.2), and focus of attention (see section 5.4.3).

5.4.1 Neuro-symbols

In chapter 4, the information processing principle in terms of neuro-symbols was introduced. Neuro-symbols are the basic information processing units of the system. This section starts with a description of how neuro-symbols can principally be realized in AnyLogic and continues with an outline about implementation particularities for neuro-symbols of different hierarchical levels.

Principal Realization of Neuro-symbols and their Information Exchange

Basic Neuro-symbols and their Information Exchange: In AnyLogic, neuro-symbols can be modelled as active objects (see figure 5.8). To simulate neuro-symbols and their information exchange with other neuro-symbols – in the following also simply referred to as symbols – besides active objects, the following AnyLogic design elements are used: state variables, ports, connections, and messages.



Port: inputPort

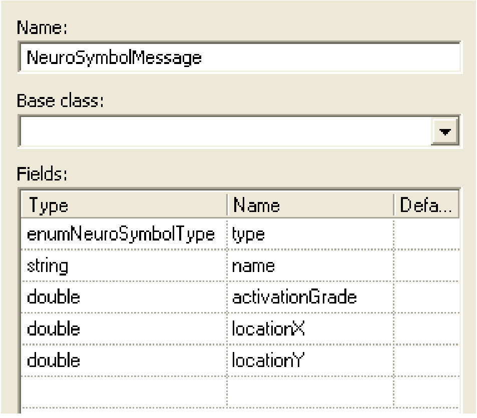
Has queue: No
On receive action:

```
if(checkInputMessageType(msg))
{
    calculateSymbolActivationGrade(msg);
    calculateSymbolLocation(msg);
    checkIfSymbolActive();
    createAndSendOutputPortMessage();
}
return true;
```

Figure 5.8: Realization of a Basic Neuro-symbol in AnyLogic

Each active object has an input port and an output port. Through the input port, the symbol can principally receive information in form of messages from the sensor value generator, from other symbols, from the knowledge module, or the focus of attention module. Through the output port, messages containing the actual activation grade of the symbol and the values of properties are passed to other symbols to which this port is connected. Each neuro-symbol has a location property, which is represented by state variables. For a first implementation, there are considered the x- and the y-position of a neuro-symbol represented by the state variables *symbolLocationX* and *symbolLocationY* both of the type double. Principally, it would also be possible not to represent the position by single coordinate points but by areas within which a perceptual image was perceived. Neuro-symbols can also comprise additional properties being represented as state variables of various types[3].

As just mentioned, information between neuro-symbols is exchanged by messages. A basic message contains information about the type and the name of the neuro-symbol it was sent from, about its activation grade, and the x- and y-location of the perceptual image the neuro-symbol represents. In a message, the type, the name, the activation grade, and the property values are represented as member variables of a message class (see figure 5.9).



Name: NeuroSymbolMessage

Base class:

Fields:

Type	Name	Defa...
enumNeuroSymbolType	type	
string	name	
double	activationGrade	
double	locationX	
double	locationY	

Figure 5.9: Message for Information Exchange between Neuro-symbols in AnyLogic

Whenever a message is received at the input port of a feature symbol, certain calculations are performed. The functions that shall be executed are defined in the *on receive action* section of the input port (see figure 5.8). The functions themselves as well as additional variables can be defined in the *additional class code* section of the active object representing the neuro-symbol. When a message is received via the input port, first the message type is checked by the function *checkInputMessageType(msg)*. If the message is of a valid type, its member variables are extracted and further processing steps are performed. If the message comes from another neuro-symbol, based on the message and the values of the activation grade and the location properties, the current activation grade and the location values of the neuro-symbol are calculated by the functions *calculateSymbolActivationGrade(msg)* and *calculateSymbolLocation(msg)*. These two functions can also consider information of different reliability coming from different modalities. If the message comes from the knowledge module or the focus of attention module, the step of recalculating the location information can be skipped. Next, it is checked by the function *checkIfSymbolActive()*, which is defined in the *additional class code* section, whether the sum of incoming activations exceeds the threshold value of the symbol. If the threshold is exceeded, the symbol is activated. The function *createAndSendOutputPortMessage()* is responsible for creating an output message based on the incoming data as well as for sending this message to other symbols via the output port.

Neuro-symbols with Time Window: As described in section 4.4.4, neuro-symbols reacting to input activations all occurring concurrently are not sufficient in all cases. In certain situations, it is also necessary to consider activations occurring within a certain time window. To cover these cases, the basic symbol type has to be extended. Therefore, a timer with a certain

[3] The assignment of symbols to spatial areas and additional properties are left out in the description of the basic neuro-symbol type given here.

expire time is added to the active object that represents the neuro-symbol (see figure 5.10). The necessary length of the time window can be derived from examples during the learning phase. In comparison to the basic neuro-symbols described in the last paragraph, the method *calculateSymbolActivationGrade(msg)* has to perform some additional functions. The principle for considering activations within a certain time window has already been explained in section 4.4.4. If the symbol received input information from an activated neuro-symbol the first time, the timer with an expire time corresponding to the length of the time window is started. There are now considered activations coming from all inputs until the time window is expired or the activation grade exceeds the threshold value of the symbol. If one of these two conditions is fulfilled, the timer is stopped and its value is reset and not started until the next message of an activated neuro-symbol is received via the input port. The stopping and resetting of the timer is performed in the function *checkIfSymbolActive()*.

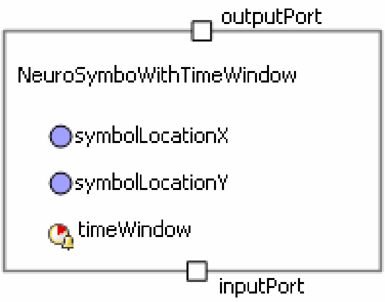


Figure 5.10: Realization of a Neuro-symbol Considering Activations within a Time Window

Neuro-symbols Considering Temporal Successions: In some cases, it is necessary to consider temporal successions of received symbol activations. The principle for handling successions of events was already outlined in section 4.4.4. In AnyLogic, there can be used state charts including timers for handling such successions (see figure 5.11).

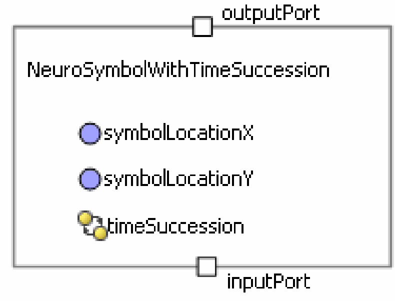


Figure 5.11: Realization of a Neuro-symbol Considering Temporal Successions of Activations

Within the state chart, the necessary temporal succession of events can be stored. If the succession of incoming activations shall be learned from examples, a state chart structure as depicted in figure 5.12 can be used. Between the particular states, transitions with transition conditions exist which can be configured according to the values derived from the presented examples. The forward connections between states depicted as short lines and labeled as “go to next state” represent the transition conditions for getting from one state to the next. A transition takes place whenever the next neuro-symbol necessary for the temporal succession is activated. Transitions from a state to the initial state – here labeled as “go to initial state” – are run through if a timeout occurs. A timeout takes place when there was not made a transition from one state to the next for a certain time, because there was not activated the neuro-symbol necessary for this transition. In such a case, neuro-symbols activated until now have to be activated again to run through the start chart. The time that can pass before there is made a transition to the initial state can be extracted from examples presented to the system. The transitions from the different states to the final state labeled as “go to final state” are necessary, because the number of states needed for a particular neuro-symbol is not known at initial system startup but only after the training phase. For a neuro-symbol considering temporal successions, there are reserved in advance a number of states, which will be sufficient in any case. If it turns out during the training phase that only fewer states are necessary, there is set the transition condition from the last state needed to the final state to true. By this measure, the states coming after this state are skipped. If the final state of the state chart is reached, the corresponding neuro-symbol is activated. For neuro-symbols including state charts, the function *calculateSymbolActivationGrade(msg)* is responsible for extracting information from incoming messages and for setting variables necessary for state transitions.

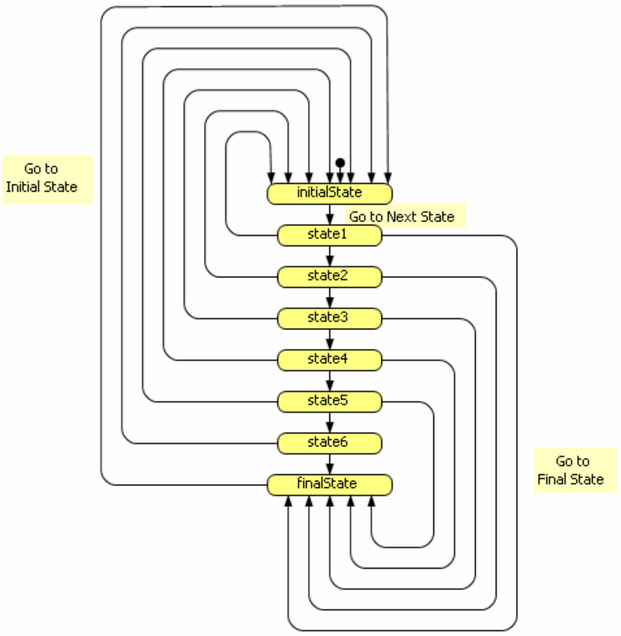


Figure 5.12: State Chart for Learning Temporal Successions of Activations of a Neuro-symbol

General Neuro-symbols: To allow the determination of correlations by learning from examples, it shall not have to be fixed at initial system startup if a neuro-symbol shall handle only concurrent activations, activations occurring within a time window, or temporal successions of activations. Therefore, it is desirable to have available one general neuro-symbol type that can principally handle all three possibilities. The decision which of the three possibilities is needed shall be determined only during the learning phase. As can be seen from figure 5.13, such a general symbol type comprises both a timer for considering activations within time windows and a state chart for considering temporal successions of events. During the learning phase, it is determined which neuro-symbol type is actually needed and the parameters of the timer and the state chart are set accordingly. If the timer or the state chart or both are not needed, the parameters are set in a way that they make no contribution in the process of determining the activation grade of the neuro-symbol.

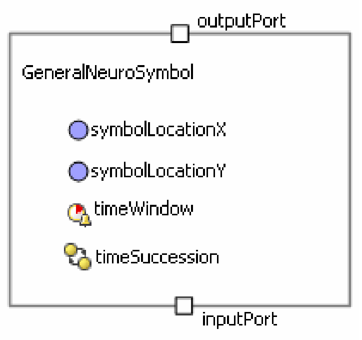


Figure 5.13: Realization of a General Neuro-symbol

Realization of Neuro-symbols of Different Hierarchical Levels

As mentioned in chapter 4, in the model, there exist neruo-symbols of different types: feature symbols, sub-unimodal symbols, unimodal symbols, multimodal symbols, and scenario symbols. The requirements for neuro-symbols in different hierarchical levels and modalities differ slightly. In the following, these different neuro-symbol types are discussed. Principally, the neuro-symbols being described now inherit the variables and functions of the neuro-symbols just introduced. Additionally, there are added further necessary variables and functions. Inherited functions can also be overloaded.

Feature Symbols: Feature symbols are a special kind of neuro-symbols, which have a close connection to sensory raw data and to the topographic arrangement of the sensors. Each feature symbol has a location value represented by the variables *symbolLocationX* and *symbolLocationY*. Unlike to neuro-symbols of higher layers, the values of these two variables are fixed and correspond to the location of the sensors the neuro-symbols are correlated with. Therefore, the call of the function *calculateSymbolLocation(msg)* can be skipped. The feature symbol layer is also the layer where focus of attention interacts with perception (see section 4.4.5). The spatial area the focus of attention is currently directed to as well as the size of the focus of attention is circularized to the feature symbols by messages sent from the focus of attention module (see section 5.4.3). The processing of this information is performed by a sub-function called by the function *calculateSymbolActivationGrade(msg)* of each feature symbol. The function *calculateSymbolActivationGrade(msg)* is also responsible for determining if the value of the location property of a feature symbol lies within the spatial area of the current focus of attention or not. If it lies outside the focus of attention, information about the activation grade of the feature symbol is not transmitted to neuro-symbols of the next higher hierarchical layer, which is generally the sub-unimodal layer. For transmission of information to higher layers, the function *createAndSendOutputPortMessage()* is responsible. In the model, depending on the sensor type the feature symbol is derived from, the appearance of feature symbols can differ slightly as they can have different properties. However, all feature symbol types can be derived from the basic feature symbol type just described.

Sub-unimodal Symbols: Sub-unimodal symbols receive information from feature symbols and the knowledge module. For sub-unimodal symbols, the topographic location information of the feature symbols is transformed into location information only contained as properties. Besides the location property, sub-unimodal symbols can comprise additional properties. The additional properties that sub-unimodal symbols can have strongly depend on the sensor type and the feature symbols they are derived from. For the simulation, there are implemented sub-unimodal symbols extracted from sensor data of tactile floor sensors and their feature symbols. For the sub-modalities correlating with motion detectors, light barriers, and door contacts, there are already used symbolic data.

Unimodal Symbols: Unimodal symbols receive information from different sub-unimodal symbols and from the knowledge module. Like the other symbol types, they can also comprise other properties additionally to their location property. In the implementation, there are realized visual, auditory, and tactile unimodal symbols. Visual and auditory neuro-symbols are directly activated by symbolic information from the sensor value generator. For the tactile modality, unimodal neuro-symbols are activated as a result of sub-unimodal symbol activations. These sub-unimodal symbols comprise location information. As outlined in section 4.4.3, location information is an important factor in the process of binding. Therefore, there have to be defined additional variables and sub-functions to calculate, store, and check such location data. The calculation of these data is performed during learning (see section 4.5.2). In the operation phase, the function *calculateSymbolLocation(msg)* is responsible for checking matches of lower-level symbols based on location information as well as for calculating the location of the current symbol.

Multimodal Symbols: Multimodal symbols receive information from different unimodal symbols and from the knowledge module. Again, multimodal symbols can comprise properties additionally to their location property. Similar as for tactile unimodal symbols, for multimodal symbols, location information plays an important role for binding of information. The handling of location information is equivalent to the handling in the unimodal tactile layer.

Scenario Symbols: Scenario symbols can receive information from a succession of different multimodal symbols and from the knowledge module. Scenario symbols can represent situations taking place within longer time periods, and the multimodal symbols that trigger them can be spread over different locations. Therefore, for scenario symbols, a location property is not obligatory. However, it often makes sense to locate a scenario symbol within a certain spatial area. Again, scenario symbols can contain different additional properties.

5.4.2 Memory Symbols and Knowledge

In section 4.6, it was described how memory and knowledge can influence perception. Integration of memory and knowledge is important for resolving ambiguous sensor data, which occur if the system shall be capable of perceiving many different objects, events, scenarios, and situations. Memory symbols are used to store important consequences of past events. For realizing memory symbols in AnyLogic, again, active objects are used (see figure 5.14). Memory symbols can be set and reset by different activated neuro-symbols from the sub-unimodal level up to the scenario symbol level. Memory symbols have two distinct input ports labeled as *inputPortSet* and *inputPortReset.* Neuro-symbols, which are responsible for setting a certain memory symbol, are connected to *inputPortSet.* Neuro-symbols that shall reset it are connected to *inputPortReset.* Similar like neuro-symbols, memory symbols can have properties. Unlike for neuro-symbols, a location property is only optional. At system startup, memory symbols are generally neither set nor reset, but their state is undefined. The current state of a memory symbol is represented by the variable *symbolActivationState*. This variable is of the type enum and can have three different states: set, reset, and undefined. The variable is declared as interface output variable and is connected to an interface input variable of the knowledge module via a connector (see figure 5.14). This means that the corresponding variable of the knowledge module is changed whenever the variable of the memory symbol gets a new value. The setting and resetting of the variable is performed in the *on receive action* sections of the two input ports.

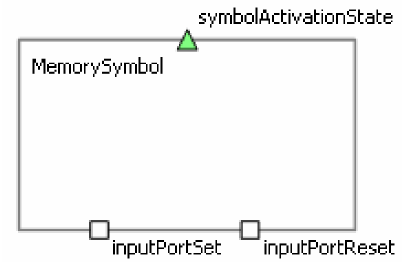


Figure 5.14: Realization of a Memory Symbol in AnyLogic

Figure 5.15 illustrates the flow of information from memory symbols to the knowledge module and back to neuro-symbols. Due to easier interpretation and overview, the connections from neuro-symbols to memory-symbols are not depicted. Activated neuro-symbols from the sub-unimodal level up to the scenario level can set and reset memory symbols. Depending on the set memory symbols, knowledge interacts with neuro-symbols by increasing or decreasing the activation grade of certain neuro-symbols. This interaction can take place at neuro-symbol levels higher than the feature symbol level.

The task of the knowledge model is to determine based on set memory symbols and correlating rules what neuro-symbols shall be influenced and in what way their activation grade shall be modified. Accordingly, messages are sent via the output port to these neuro-symbols. The determination what messages shall be sent is performed by the function *determineRetroaction()* defined in the *additional class code* section of the knowledge module. The sending of messages is carried out by the function *createAndSendOutputPortMessage()*. These functions are called whenever an interface variable of the knowledge module changes. To register these changes, for each interface variable, a state chart with two states is used in which the value of the interface variable serves as transition condition between the states (see figure 5.16). The reason for using interface variables and state charts instead of the communication via a port is due to "stability reasons" to avoid multiple circulations of messages caused by feedback connections to lower processing stages (see section 6.2). In the *action* section of each transition, the mentioned functions *determineRetroaction()* and *createAndSendOutputPortMessage()* are called. A message sent from the knowledge module to neuro-symbols has the format depicted in figure 5.17. It contains information about the type of the target neuro-symbol and its name. Additionally, it is indicated by the variable *activationGrade* in what way the activation grade of the target symbol shall be modified.

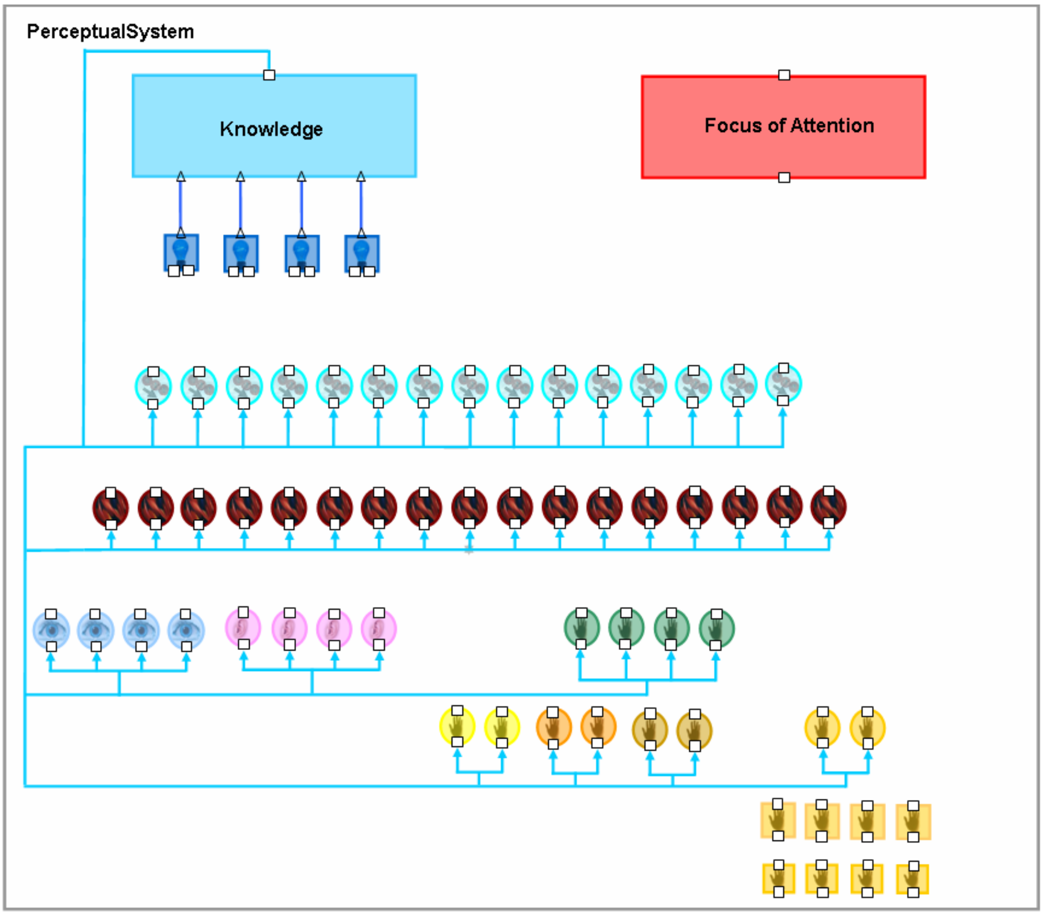


Figure 5.15: Interaction between Memory Symbols, Knowledge, and Neuro-symbols

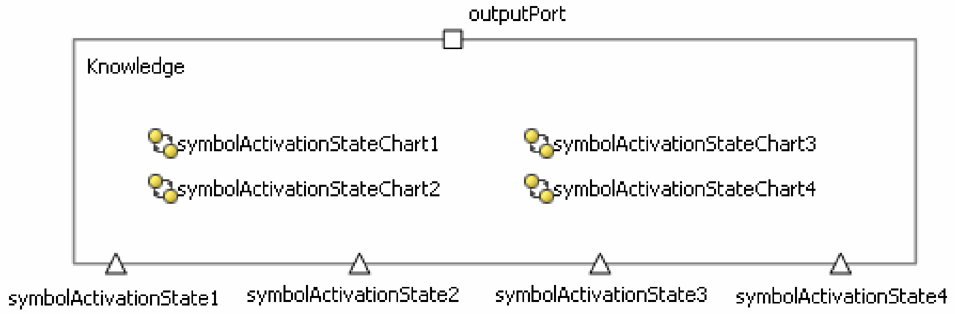
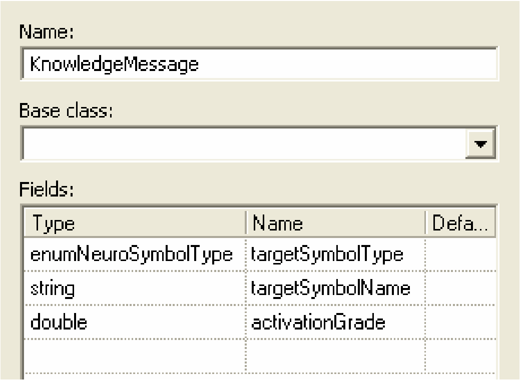

outputPort
Knowledge
symbolActivationStateChart1
symbolActivationStateChart2
symbolActivationStateChart3
symbolActivationStateChart4
symbolActivationState1
symbolActivationState2
symbolActivationState3
symbolActivationState4

Figure 5.16: Realization of Knowledge Module with Interface Variables and State Charts

5.4.3 Focus of Attention

In section 4.4.5, it was explained how focus of attention can resolve the binding problem in perception in case of multiple events happening concurrently. For the model, it was suggested that focus of attention should interact with perception on the feature symbol level. This decision



Name:
KnowledgeMessage

Base class:

Fields:

Type	Name	Defa...
enumNeuroSymbolType	targetSymbolType	
string	targetSymbolName	
double	activationGrade	

Figure 5.17: Message for Information Exchange between Knowledge Module and Neuro-symbols

was taken due to the topographic structure of the feature symbol level.

Figure 5.18 illustrates graphically how focus of attention interacts with neuro-symbolic perception on the feature symbol level. For feature symbols corresponding to spatial areas lying outside the focus, their activation grade is decreased to inhibit further information processing in sub-unimodal level. A further processing of the information in the next-higher feature symbol level is however possible. For other modalities, which already receive symbolic information, there is implemented a filter, which filters out symbols with location values lying outside the current focus of attention. However, this is not depicted in this figure.

As described in section 4.4.5, the beam of attention needs to be guided somehow. This process might be influenced from perceived images as well as from knowledge, expectation, but also from other processes like emotions. It is very probable that there are involved processes, which are not explicitly assigned to the perceptual system of the human brain. At the current state of the model, it is not considered by means of which processes the beam of attention is directed. It is assumed that the coordinates where the beam is directed as well as the size of the beam are provided from an external source and there has only been programmed the interface to perception based on the given coordinates and the size of the beam. The focus of attention interface has one input port and one output port and contains variables representing the position and size of the beam (see figure 5.19). The input port has the function to collect information from perception and other brain processes. In the current implementation, the input port is not connected.

The variables *attentionCenterLocationX* and *attentionCenterLocationY* indicate at what location the center of the focus of attention lies. The variables *attentionWidthX* and *attentionWidthY* determine the size of the beam. The variables are of the type double. In the implementation, the beam has a rectangular form. However, with slight modifications in the implementation, it could also have other shapes. Via the output port, the values of the variables *attentionCenterLocationX*, *attentionCenterLocationY*, *attentionWidthX*, and *attentionWidthY* are sent as message to the feature symbol level. The member variables of the messages sent from the focus of attention module are depicted in figure 5.20. A message is sent whenever the size or the center of the beam of attention changes. The values of the member variables are set by the function *directFocusOfAttention(double attentionCenterLocationX, double attentionCenterLocationY, double attentionWidthX, double attentionWidthY)* defined in the *additional class code* section. The same function is also responsible for sending the message via the output port. The message is sent to

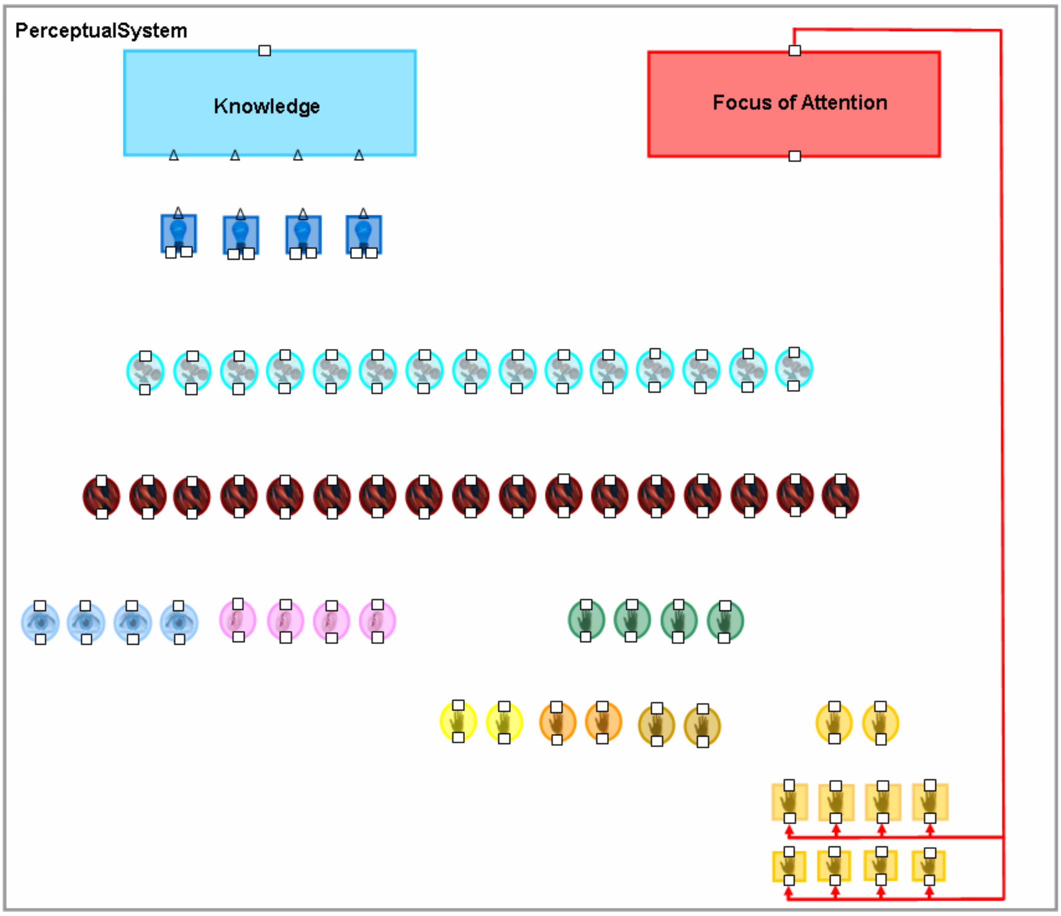


Figure 5.18: Interaction of Focus of Attention with Feature Symbol Level

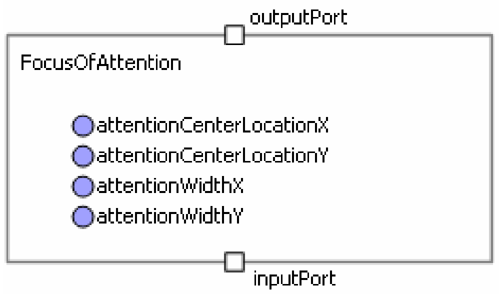
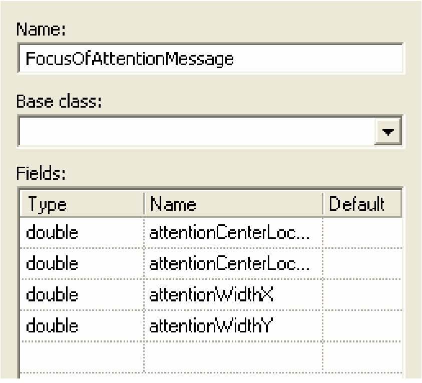


Figure 5.19: Realization of the Focus of Attention in AnyLogic

every feature symbol that is connected to the output port of the focus of attention module. At system startup, the beam covers the whole spatial area.



Name:
FocusOfAttentionMessage

Base class:

Fields:

Type	Name	Default
double	attentionCenterLoc...	
double	attentionCenterLoc...	
double	attentionWidthX	
double	attentionWidthY	

Figure 5.20: Message Format sent from Focus of Attention to Feature Symbols

5.5 Methods for Learning and Adaptation

In section 5.4.2 and section 5.4.3, it was already explained how focus of attention and knowledge interact with neuro-symbols and memory symbols and how these units are interconnected. These connections are fixed at initial system startup. In contrast, connections between neuro-symbols – at least in higher hierarchical levels – as well as connections between neuro-symbols and memory symbols need not be predefined but can be learned from examples. Therefore, for these symbols, there do not exist interconnections at initial system startup. Instead, neuro-symbols are connected to so-called learning ports. In figure 5.21, it is illustrated in what way neuro-symbols are connected to different learning ports at initial system startup.

In the described implementation, there can be learned feedback connections between neuro-symbols of the sub-unimodal tactile modality of the tactile floor sensors (1). There can be determined the connections between the different sub-unimodal tactile neuro-symbols and unimodal tactile neuro-symbols (2). On the unimodal tactile level, there are also learned feedbacks between unimodal tactile neuro-symbols (3). In a further processing stage, there can be determined connections between different unimodal symbols and multimodal symbols (4) as well as feedbacks between multimodal symbols (5). The learning port (6) is responsible for setting the connections between multimodal symbols and scenario symbols and the port (7) for the feedback connections between scenario symbols. Port (8) allows it to determine connections between neuro-symbols of different levels and memory symbols. For the learning port responsible for learning correlations between neuro-symbols of different levels and memory symbols, the connections are only adumbrated with arrows of the colors of the corresponding neuro-symbols to preserve the clarity of the graphic.

As can be seen from figure 5.21, to each learning port, neuro-symbols of certain types as well as the control port of the perceptual system are connected. Via the control port, control data from the sensor value generator containing information about the current training phase and the meaning of the data currently sent via the other six ports are transmitted. After training is completed, all necessary connections between neuro-symbols and memory symbols are set, the connections to the learning ports are cancelled, and the ports no longer have an influence on information processing.

As already outlined in section 4.5, learning is based on examples and is carried out in different phases starting at the lowest level and continuing with the next higher level. As the system

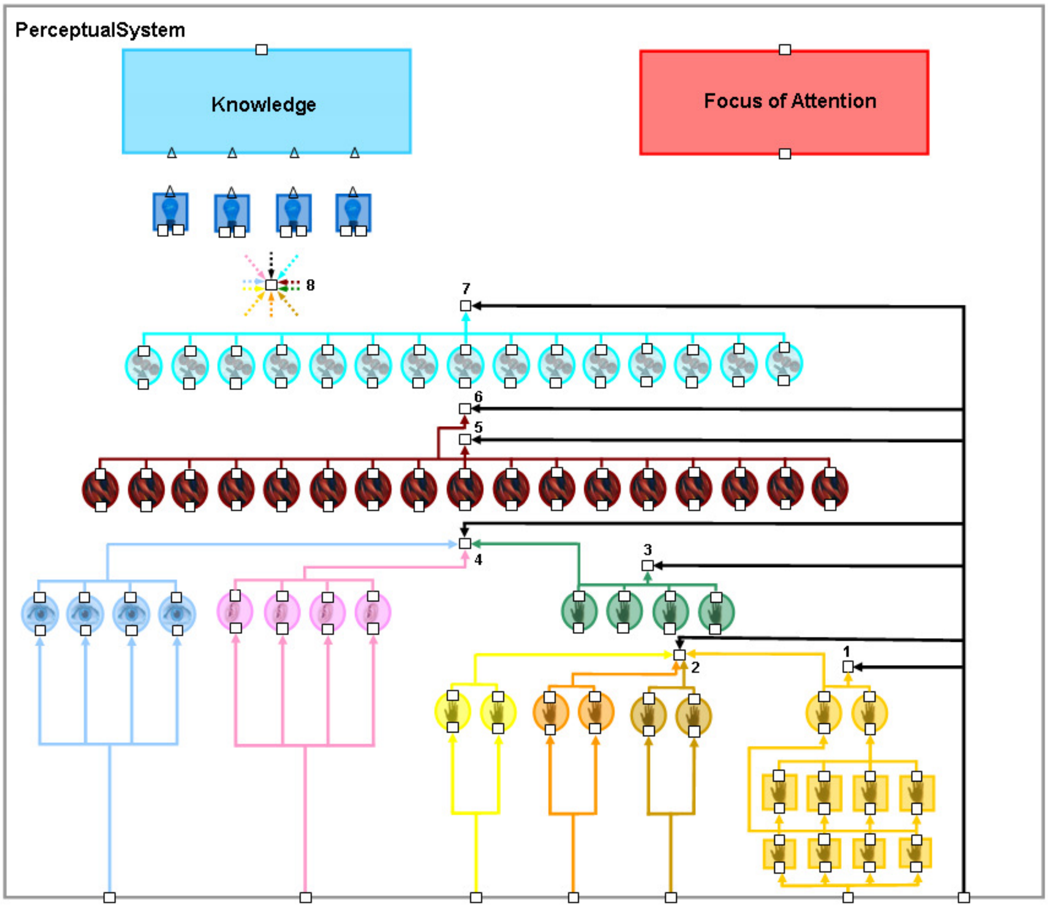


Figure 5.21: Connections from Neuro-symbols to Learning Ports

shall be able to learn connections between symbols from examples, additional variables, arrays of variables, and functions have to be defined for the perceptual system that are only active and necessary during learning. After the learning phase, the variables and functions do no longer take influence on the system. The variables and functions are responsible for memorizing the presented examples during the training in order to determine the most suitable connections and also to set adequate spatial areas within which a binding of information is possible. Additionally, the influence of other property values of neuro-symbols for symbol activation is specified. There also have to be determined the duration of time windows and the temporal succession of events. These processes are performed by the functions *memorizeExamples(msg)* and *calculateAndSetCorrelations()*. These functions exist for each learning port and differ for each port slightly in the tasks they have to perform. The general aim of the function *memorizeExamples(msg)* is to memorize a certain number of examples to allow the determination of correlations between examples representing the same situation.

The function *calculateAndSetCorrelations()* is the centrepiece of each learning phase and has the task to determine the correlations between the example data and to set correlations accordingly. To calculate correlations, different algorithms are conceivable. For the performed analyses, statistical methods are used for determining the most likely coherences (see below). After correlations have been determined, the connections between the symbols and further necessary data like properties, location data, and timing data are set. As described in section 4.5.3, there cannot only be determined correlations between symbols during the learning phases. It can also be determined if certain neuro-symbols are redundant and can therefore be deleted. The determination if a neuro-symbol is redundant is performed in the function *calculateAndSetCorrelations()* of the learning ports responsible for determining feedback connections. If there are identified redundant symbols, the connections to these symbols set in the former training phase can be cancelled. In a similar way, neuro-symbols can be added if needed. This is performed by the function *calculateAndSetCorrelations()* of the learning ports responsible for generating forward connections.

Learning Forward Correlations and Adding of Neuro-symbols

As already mentioned, for learning forward connections and correlations between neuro-symbols of different levels, there need to be presented examples to the system which have to be stored until correlations between them are calculated. The function *memorizeExamples(msg)* is responsible for storing these examples. There is stored the name of the received neuro-symbol and the x- and y-position where its perceptual image was perceived. If the neuro-symbol contains additional properties, the values of these properties are also stored. The function *memorizeExamples(msg)* is called whenever a message from a lower-level neuro-symbol is received at the corresponding learning port. In the training phases A, there is always presented and memorized the whole set of training examples corresponding to one particular perceptual image and then, the correlations between lower-level symbols and the higher-level symbol representing this perceptual image are calculated and set by calling the function *calculateAndSetCorrelations()*. The call of this function is invoked by messages sent from the control port.

In the implementation, the information of incoming neuro-symbols is stored in arrays. The function *calculateAndSetCorrelations()* searches through these arrays to calculate correlations. As already mentioned, for the calculation of the correlations, different algorithms are conceivable. The algorithm 5.1 used for a first implementation to calculate forward correlations – in the following labeled as algorithm for learning phase A – is described below.

Algorithm 5.1: Algorithm for Learning Phase A

```
For each modality
   Determine which neuro-symbols occur most often
If one neuro-symbol of one modality occurs in more than $c_1$ percent of all cases
   Set connection
If two neuro-symbols of one modality occur in more than $c_2$ percent of all cases
   Use two separate higher-level neuro-symbols and set connections
For each neuro-symbol type connected
   Calculate average x- and y-location
   Calculate average x- and y-location-deviation
   Determine property values
Calculate x- and y-inter-modality location deviations between neuro-symbols
If necessary
   Consider temporal character of data
```

After all examples belonging to one perceptual image have been stored, it is searched through the arrays and for each modality it is determined what neuro-symbols occur most often. If a particular lower-level symbol of one modality occurs in more than c₁ percent of all cases, a connection between this symbol and the higher-level symbol is set. As value for c₁, the rate 70 % proved to be suited for the test cases used during simulation (see section 6.1). The reason why a certain lower-level symbol does not need to occur in 100 % of all examples is to have the possibility to rule out faulty or not representative data as there might occur cases where sensors fail or where within the examples there are triggered or covered sensors from other events incidentally occurring at the same time. If two different lower-level neuro-symbols of the same modality occur always in more than c₂ percent of all examples, the perceptual image is represented by two separate neuro-symbols. Connections are set accordingly. For the used test cases, a value of c₂=35 % proved to be suitable. For the investigated test cases, these two distinctions turned out to be sufficient to calculate correlations. For more complex cases, the algorithm would have to be extended.

After the connections have been set, the arrays are searched through again. From each of the lower-level neuro-symbols which is of the type of the connected neuro-symbols, the location information is taken to calculate an average x- and y-position as well as an average x- and y-deviation according to the formulas 5.1 to 5.4, where the letter i represents the location information of the current example and the letter n the number of valid examples for one perceptual image[4]. This average x- and y-location as well as its average deviations represent the range within which a lower-level symbol can occur to contribute to the activation of the particular higher-level symbol.

$$AverageLocationX = \frac{\sum_{i=1}^{n} LocationX_i}{n} \tag{5.1}$$

$$AverageLocationY = \frac{\sum_{i=1}^{n} LocationY_i}{n} \tag{5.2}$$

$$AverageLocationDeviationX = \frac{\sum_{i=1}^{n} |AverageLocationX - LocationX_i|}{n} \tag{5.3}$$

$$AverageLocationDeviationY = \frac{\sum_{i=1}^{n} |AverageLocationY - LocationY_i|}{n} \tag{5.4}$$

After learning is finished, during system operation, lower-level neuro-symbols of the connected types lying within the range $AverageLocationX \pm 2 \cdot AverageLocationDeviationX$ and the range $AverageLocationY \pm 2 \cdot AverageLocationDeviationY$ can contribute to an activation of the higher-level neuro-symbol. This allows it to generalize to a certain extend over location data to handle also examples not seen during the training phase. For each neuro-symbol, the valid location ranges are stored as variables in the neuro-symbol. During operation, the check whether the symbol lies within the valid range is done in the *on receive* section of each neuro-symbol by the function *checkInputMessageType(msg)* (see section 5.4.1). If certain lower-level symbols comprise properties in addition to their location property, the values of the properties are also extracted and stored in variables of the higher-level symbol.

As outlined in section 4.4.3, besides the information what and within which spatial area lower-level neuro-symbols can principally contribute to the activation of a higher-level symbol, it also

[4] For calculating location data, other algorithms would also be conceivable. However, for the test cases outlined in section 6.1, the used algorithms deliver satisfactory results.

has to be considered how big the spatial deviation between lower-level neuro-symbols of different modalities belonging to one particular higher-level symbol can be. In the used simulation, this is calculated during the training phase A by determining one dominant modality and searching through the examples for the maximal location deviation between the symbol of this modality and the symbols of other modalities (see formulas 5.5 and 5.6).

$$InterModalityLocationDeviationX = \max_i(|domLocationX_i - LocationX_i|) \tag{5.5}$$

$$InterModalityLocationDeviationY = \max_i(|domLocationY_i - LocationY_i|) \tag{5.6}$$

In the simulations performed in chapter 6, the tactile floor sensor modality will be considered as most reliable and therefore dominant modality on the sub-unimodal level. On the unimodal level, tactile neuro-symbols directly take over the location values of the sub-unimodal symbols of this modality. For the used test cases, every higher-level neuro-symbol comprises information coming from this modality. If a certain symbol did not include information from this sensory modality, the modality with the next highest reliability would be taken as dominant modality for the calculation. During system operation, lower-level neuro-symbols can only be bound to a higher-level symbol if they lie within a distance of less than or equal $\pm 1.5 \cdot InterModalityLocationDeviationX$ and $\pm 1.5 \cdot InterModalityLocationDeviationY$.

In the sections 4.4.4 and 5.4.1, it was mentioned that in certain cases, neuro-symbols need to offer the possibility to handle data arriving over time. This is necessary if not all lower-level neuro-symbols being responsible for the activation of a higher-level symbol are active concurrently at least for a short instance of time. The learning of the size of time windows or the configuration of state charts for temporal successions based on examples is also performed in the training phase A of each level. To derive the length of the time window within which the incoming activations of neuro-symbols shall be considered, there just has to be determined the maximum time that occurred between the first neuro-symbol activated and the last neuro-symbol activated within the examples for a certain perceptual image. Taking this value and multiplying it with a factor of 2 results in the size of the time window for the corresponding neuro-symbol. As outlined in section 5.4.1, for handling temporal successions of incoming data, neuro-symbols contain a standard state chart, which is adapted to the special requirements of the corresponding perceptual image during the training phase A. Therefore, the examples with the lower-level symbols responsible for activating a certain symbol have to be stored in arrays for determining correlations between data in a next step. There are stored the succession of activated symbols, the location where they were perceived, their properties, and the time when they were activated. The storage is performed by the function *memorizeExamples(msg)*.

After all existing examples for one perceptual image have been seen, correlations for this image are determined and set. The determination what lower-level neuro-symbols are involved in the activation of a particular higher-level symbol as well as the adaptation of the state chart based on the succession of neuro-symbols is again performed by the function *calculateAndSetCorrelations()*. Based on the examples, it is learned what lower-level neuro-symbols are involved in the activation of a particular higher-level symbol and in what succession they take place, in what location range each of them occurs, and how long the time between two symbol activations is. This information is stored in the transition conditions between states. There can occur the case that different examples comprise activations of different multimodal symbols. The task of the learning algorithm is to determine what symbol activations all examples have in common and at what locations they occur around and to use only the common activations to configure the state chart. For determining the spatial area, the formulas 5.1 to 5.4 are used for each activated lower-level symbol. Similar as already described for neuro-symbols without state charts, all activated neuro-symbols of a certain type occurring within a spatial area of $AverageLocationX \pm 2 \cdot AverageLocationDeviationX$ and $AverageLocationY \pm 2 \cdot AverageLocationDeviationY$ are valid. This information is necessary to define rules for the forward transitions between states.

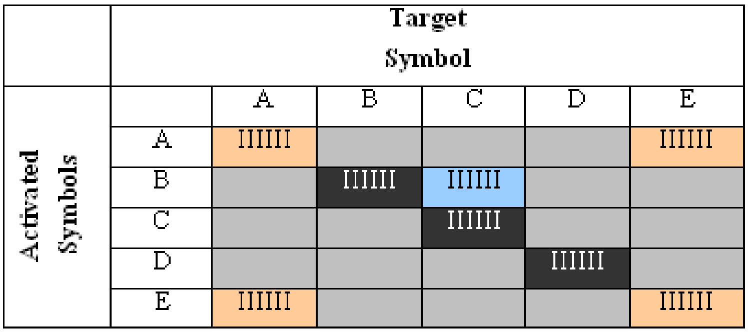
Additionally, it is determined what period of time typically passes between symbol activations of different symbols. This is done by taking the maximum time durations that occurred in the examples between two particular symbols (see formula 5.7).

$$ActivationTime = \max_i(T_i) \tag{5.7}$$

The values of the variable *ActivationTime* is important for setting timeouts responsible for transitions between a certain state and the initial state of the state chart. When a timeout occurs, a transition from the current state to the initial state is performed. This means that information about neuro-symbols activated until now is no longer considered for the perception of the scenario. A timeout occurs after a time period of $2 \cdot ActivationTime$. In section 6.1.2, it will be shown by means of a concrete example how a state chart can be configured.

Learning Feedback Connections and Correlations and Eliminating Neuro-symbols

As described in section 4.5, in the training phases A, forward connections between neuro-symbols of different levels are set, valid values for location properties as well as other properties are determined, and handling of data over time is considered. In the training phase B, based on these determined correlations, it is calculated what feedback connections are necessary and what symbols are redundant. This is performed by presenting the same examples as used in training phase A to the now connected neuro-symbols a second time and observing via the corresponding learning port if besides the desired neuro-symbol also other neuro-symbols are activated concurrently. However, unlike in training phase A, there are presented not always only the examples belonging to one particular perceptual image followed by a setting of the corresponding correlations, but all examples are presented to the system at once one after the other and for each example it is memorized what neuro-symbols were activated and what the target symbol was. Therefore, the function *memorizeExamples(msg)* is called whenever a message from a neuro-symbol connected to the learning port responsible for learning feedback connections is received. In the implementation, the information is stored in a multi-dimensional array. After all examples were presented and memorized, the function *calculateAndSetCorrelations()* is responsible for determining feedback connections and eliminating redundant symbols. The call of this function is invoked by messages sent from the control port. The function *calculateAndSetCorrelations()* searches through the array to calculate correlations. The mechanism for determining feedbacks and eliminating redundant symbols is best explained by representing target symbols and actually activated symbols of the presented examples in a tabular form (see figure 5.22). Ideally, for each presented example, only one neuro-symbol should be activated. For neuro-symbols for which this case is fulfilled, there only exist entries in the main diagonal of the table. This is fulfilled for the symbols B and D. If two diverse neuro-symbols are always activated concurrently no matter which of the two corresponding perceptual images was presented to the system, one of them is redundant. In the table, this case occurs for the symbols A and E. To eliminate a redundant symbol, all the existing connections from its ports to other ports are removed. If one neuro-symbol is always activated when another neuro-symbol is active but not vice versa, a feedback connection has to be set. In the table of figure 5.22, this is the case for symbol C, which needs a feedback connection to symbol B. In section 6.1.2, these facts will be illustrated by means of a concrete example.



		Target Symbol				
		A	B	C	D	E
Activated Symbols	A	IIIIII				IIIIII
	B		IIIIII	IIIIII		
	C			IIIIII		
	D				IIIIII	
	E	IIIIII				IIIIII

Figure 5.22: Determination of Feedback Connections and Redundant Symbols

What was not mentioned until now is that additionally to the setting of connections, it has to be determined within which spatial range the inhibition of a feedback shall be effective. Therefore, from the presented examples, it is memorized within what spatial area neuro-symbols were activated concurrently. From the memorized examples, there is then calculated an average x- and y-position as well as a deviation range (see formulas 5.8 to 5.11).

$$FeedbackAverageLocationX = \frac{\sum_{i=1}^{n} domLocationX_i}{n} \tag{5.8}$$

$$FeedbackAverageLocationY = \frac{\sum_{i=1}^{n} domLocationY_i}{n} \tag{5.9}$$

$$FeedbackAverageLocationDeviationX = \frac{\sum_{i=1}^{n} |FeedbackAverageLocationX - domLocationX_i|}{n} \tag{5.10}$$

$$FeedbackAverageLocationDeviationY = \frac{\sum_{i=1}^{n} |FeedbackAverageLocationY - domLocationY_i|}{n} \tag{5.11}$$

For the simulation, during system operation, feedbacks coming from neuro-symbols with location values lying outside the range $FeedbackAvarageLocationX \pm 2 \cdot FeedbackLocationDeviationX$ or $FeedbackAvarageLocationY \pm 2 \cdot FeedbackLocationDeviationY$ cannot inhibit the activation of the neuro-symbol receiving the feedback signal.

Learning Correlations between Neuro-symbols and Memory Symbols

Similar to learning of correlations between neuro-symbols, there can also be learned connections between neuro-symbols and memory symbols. Therefore, the learning principle for learning forward connections in training phase A can be taken over with the little difference that a memory symbols has two input ports and that there are now determined adequate connections to lower-level neuro-symbols for both input ports. Principally, there can exist connections from all neuro-symbols form the sub-unimodal layer upwards. In the test cases described in section 6, multimodal symbols will be the only neuro-symbols that set and reset memory symbols.

5.6 Design Methodology

In chapter 4, a model for human-like perception was introduced and explained. In the former sections of this chapter, it was described how this model can be implemented and simulated with the modeling language AnyLogic. This section now described how to apply the model and its implementation to a certain application and what design steps are therefore necessary.

The first step that has to be taken is to decide what objects, events, scenarios, and situations shall be perceived by the perceptual system to fulfill the requirements of the particular application. As described in chapter 4, such perceptual images are represented by neuro-symbols. Activated neuro-symbols indicate that the perceptual image they stand for has been perceived in the environment. Neuro-symbols up to the unimodal level are only accessible within the modules of the model. In contrast, multimodal symbols and scenario symbols additionally function as outputs of the system to indicate what was perceived by the system. Therefore, it first has to be defined what multimodal symbols and what scenario symbols shall exist. Neuro-symbols of these types can also have properties, which have to be defined.

In a next step, it is determined what sensory modalities and sub-modalities shall contribute to the perception process and by what unimodal and sub-unimodal neuro-symbols these perceptions shall be represented. It also has to be fixed if they shall comprise further properties additionally to their location property. Based on these decisions, the types, number, and position of sensors for the different modalities and sub-modalities are determined and what feature symbols shall be calculated from these sensor data.

Similar to neuro-symbols representing perceptual images, there also have to be defined memory symbols, which signal what influence these images have on later perceptions.

As already outlined in former sections, certain connections between model building blocks have to be predefined by the system engineer and others are learned based on examples. Connections and correlations that are already set and defined before learning starts are the connections and correlations between sensor data and the first neuro-symbolic level and in certain cases also the connections to the second neuro-symbolic level. Neuro-symbols that are not yet interconnected with each other are connected to several learning ports. Furthermore, the output port of the focus of attention module is already connected to the input ports of feature symbols. There also exist connections between memory symbols and the knowledge module as well as connections from the output port of the knowledge module to the inputs of neuro-symbols from the sub-unimodal level upwards. However, knowledge only interacts with neuro-symbols after learning has already been completed.

The next step in the design methodology is to select suitable examples for training the system. For each neuro-symbol for which correlations to other neuro-symbols shall be learned, there need to exist examples comprising information about what sensors (or lower-level symbols) are activated when the perceptual image represented by the neuro-symbol occurs. As already mentioned, learning takes place in different phases. There have to be learned correlations between lower levels first before higher-level connections can evolve. Neuro-symbolic learning starts at the sub-unimodal level and ends at the scenario symbol level. Similar like for neuro-symbols, there also have to be presented examples to the systems comprising information about under what circumstances certain memory symbols shall be set and reset. In this case, only a training phase A is needed. Training of memory symbols can only be performed if according connections between neuro-symbols have already been set. After all connections from neuro-symbols to memory symbols have been set, it has to be defined in what way top-down processes from knowledge shall influence the neuro-symbol activations. Therefore, rules have to be defined in the knowledge module.

Chapter 6

Simulation Results and Discussion

"The most exciting phrase to hear in science, the one that heralds new discoveries, is not 'Eureka!' (I found it!) but 'That's funny ..' "

[Isaac Asimov]

In chapter 4, a model for human-like perception was introduced. Chapter 5 gave a description how this model can be implemented in software for simulation and evaluation purposes. The aim of this chapter is to show simulation results for a number of test cases (see section 6.1) and to discuss important issues of the model based on these simulation results and insights gained during model development and implementation. The insights gained during this process also allow it to draw certain conclusions about the correctness, incorrectness, or incompleteness of neuroscientific and neuropsychological models (see section 6.2). Additionally, a comparison of the model to other existing models it is related to is made in section 6.3. There are discussed the differences to the ARS-PC model introduced in section 2.1.2 and the differences to neural networks and symbolic systems. Furthermore, it is pointed out in what manner the developed model could be classified as a model of sensor fusion or as a model for neuro-symbolic integration.

6.1 Simulation Results

To test and evaluate the model, a number of simulations were performed which are outlined in the following. In section 6.1.1, the used test environment and test cases are described. Section 6.1.2 shows by means of concrete examples how learning is performed in the neuro-symbolic network. Finally, section 6.1.3 illustrates the flow of information within the system for a number of cases after learning has already been completed.

6.1.1 Test Environment and Test Cases

For reasons outlined in section 5.1.2, the model is tested with simulated sensor data (or symbolic data) provided by the sensor value generator. To evaluate the model and allow a comprehensible illustration and discussion of the simulation results, a set of test cases is used that is extensive enough to test and explain all important design issues but limited in a way to avoid loosing track. Figure 6.1 shows the test bed, which is a room equipped with a stereo video camera, an array of microphones, motion detectors, tactile floor sensors, light barriers, and a door contact sensor. These sensor data are processed to perceive activities carried out by persons in the room. The room is of rectangular form and has a door and a window. In the figure, there is also depicted a Cartesian coordinate system, which will become important when discussing the location property of neuro-symbols. The reason why the origin of ordinates is not in the left lower corner but even further left is that there shall not only be considered activities within the room but also outside the room in an area close to the door[1].

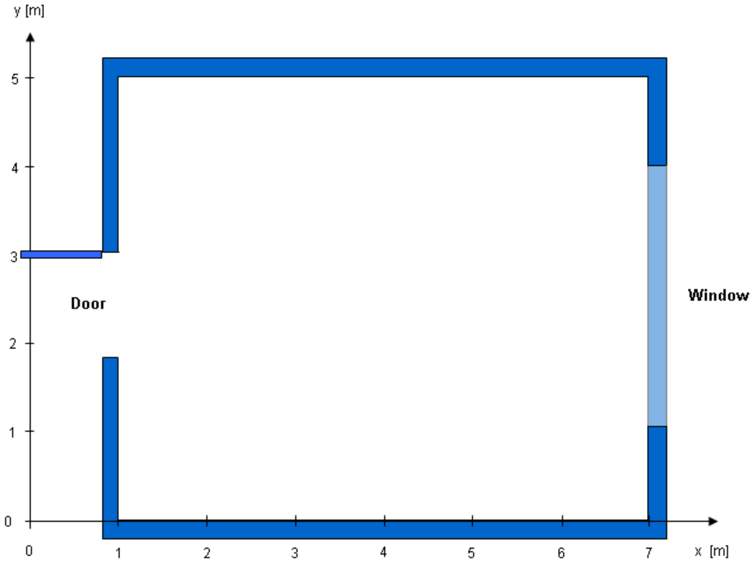


Figure 6.1: Test Environment for Illustrating Simulation Results

Figure 6.2 shows the "symbol alphabet" from the sub-unimodal level upwards for perceiving different activities of persons. Because determining of correlations between neuro-symbols from the sub-unimodal level upwards is subject to learning, the symbol hierarchy is unconnected at initial system startup. As outlined in section 5.1.2, there do not have to be processed sensory video and audio raw data, but there are already provided unimodal symbolic data from the sensor value generator. Therefore, for these two modalities, only unimodal symbols are used. From visual data, it can be detected that a person is somewhere present in the room indicated by the symbol "person", and from audio data, the noise of an opening or closing door or the sound of steps can be perceived represented by the symbols "door click" and "steps". These symbols can principally occur at each location in the room and also outside the room in an area around the door. In this section, for the tactile modalities, information processing is considered from the sub-unimodal level upwards. Therefore, no feature symbols and no sensors are presented in the symbol hierarchy of figure 6.2. For a description and discussion about lower-level information processing from sensor data to feature symbols and sub-unimodal symbols, have a look at section 6.2 and appendix A. On the sub-unimodal level of the tactile modality, there exist the symbols "motion", "object present", "object passes", and "door status". The first two symbols represent the presence and

[1] In the following, if not explicitly mentioned, location values will always be given in meters.

the motion of an object. As this information is derived from a number of sensors distributed over the room, these neuro-symbols can principally be perceived at every location in the room and outside the room around the door. It is assumed that the symbol “object present” derived from tactile floor sensors has a spatial resolution of 0.1 and the symbol “motion” resulting from motion detectors has a spatial resolution of 0.5. Therefore, the spatial resolution of corresponding neuro-symbols cannot be higher than these values. The symbols “object passes” and “door status” are derived from two light barriers and one door contact sensor, which are mounted at the door. Therefore, they are assigned to a fixed x- and y-position, which has the x-value 0.9 and the y-value 2.5. Additionally to their location property, the symbols “object passes” and “door status” have the properties “direction” and “status”, respectively, which indicate in which direction an object passed the door and whether the door was closed or opened. On the unimodal tactile level, the information of these four sub-unimodal tactile symbols is combined to get the symbols “object stands”, “object moves”, “object enters”, “object leaves”, “door is opened”, “door is closed”, and “object placed”. On the multimodal level, there are used the symbols “person opens door”, “person closes door”, “person stands”, “person walks”, “person enters”, and “person leaves”. Furthermore, there exist two symbols labeled as “symbol 1” and “symbol 2”, which will become important when discussing the topic of adding symbols during learning. In the test set, on the scenario symbol level, there only exists the symbol “person goes to window”. The memory symbol level comprises the symbols “door open” and “person present”.

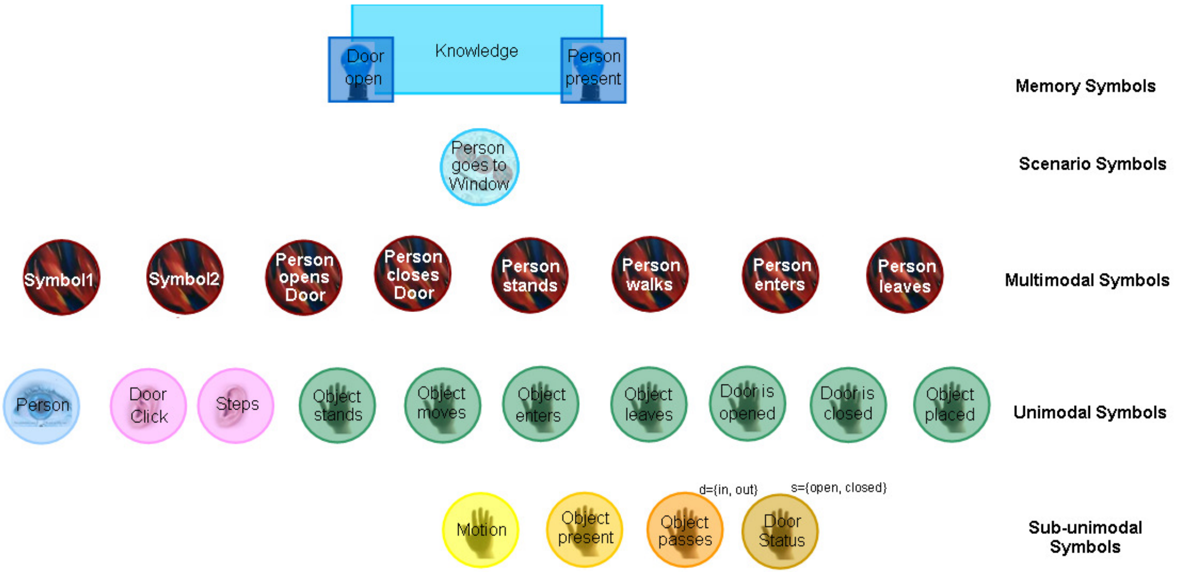


Figure 6.2: Symbol Alphabet for Specified Test Set

As already outlined, the finding of adequate connections as well as the determination of other correlations between symbols is subject to learning. This learning is performed in different phases. Figure 6.3 illustrates the final result of this learning. Location information and timing information is not depicted. Particularly interesting details of different learning phases will be discussed in the following. Afterwards, the flow of information from sensor data to symbolic perception will be illustrated by means of a number of selected examples. If not mentioned separately, for the simulations, the weight of connections from all modalities are set to 1, and the symbol activation thresholds have the value 0.85. Concerning the event-based information exchange, there is transmitted the value of the activation grade if it exceeds the threshold value and the value 0 if it is below this threshold. Feedbacks and messages from the knowledge module have an inhibitory influence on a neuro-symbol's activation.

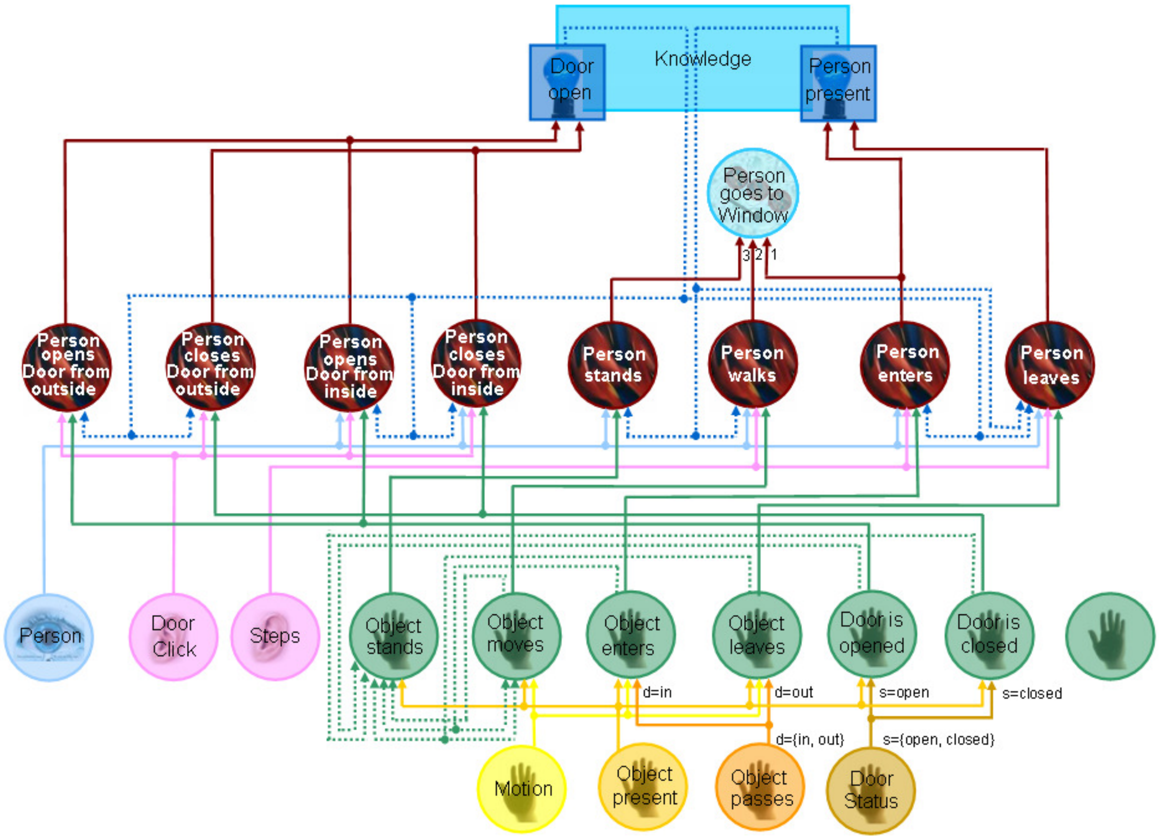


Figure 6.3: Symbol Hierarchy after Learning is Finished

6.1.2 Configuration, Learning, and Adaptability

As described in section 4.5, correlations between neuro-symbols of different layers and modalities and correlations between neuro-symbols and memory symbols can be determined from examples presented to the system. Figure 6.3 illustrates the connections resulting from such a learning process. In the following, the learning procedures and their results are illustrated for a number of cases of the chosen test set.

Learning of Correlations in Training Phase A

To illustrate the results of learning of forward connections and further correlations in the training phases A, there shall be discussed the case that a person walks around in the room, which is represented by the multimodal symbol "person walks". Additionally, to illustrate how to detect a necessary splitting of a certain perceptual image into two separate ones, the situation that a person opens the door represented by the symbol "person opens door" is used.

To determine the correlations between the symbol "person walks" and lower-level symbols, a number of training examples is necessary. Figure 6.4 and the table in figure 6.5 show the examples used for training to determine the correlations for the symbol “person walks”. Actually, they contain the neuro-symbols of the next lower level – the unimodal layer – activated when a person walks around in the room as well as the x- and y-location assigned to the symbols in the particular examples. For the concrete case of the perceptual image “person walks”, all lower-level symbols are active at the same time. Therefore, timing behavior of data does not have to be considered.

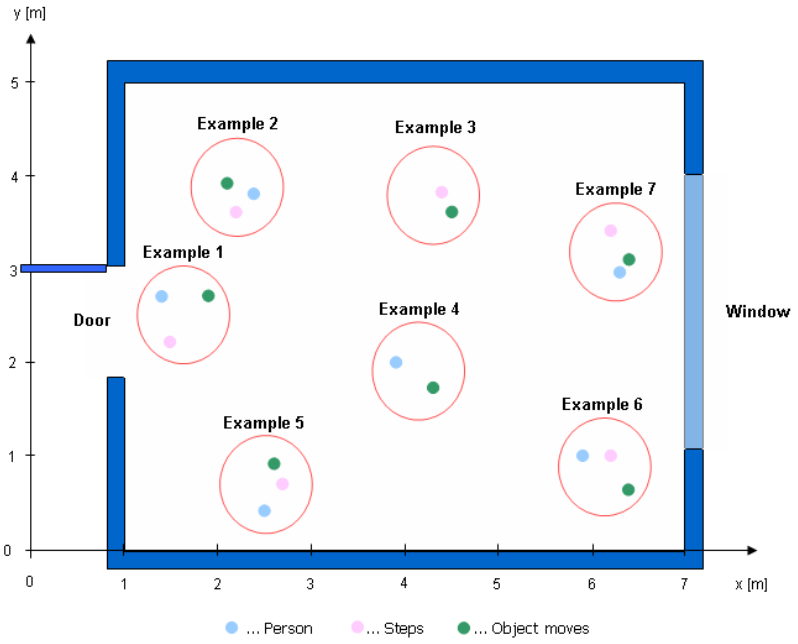
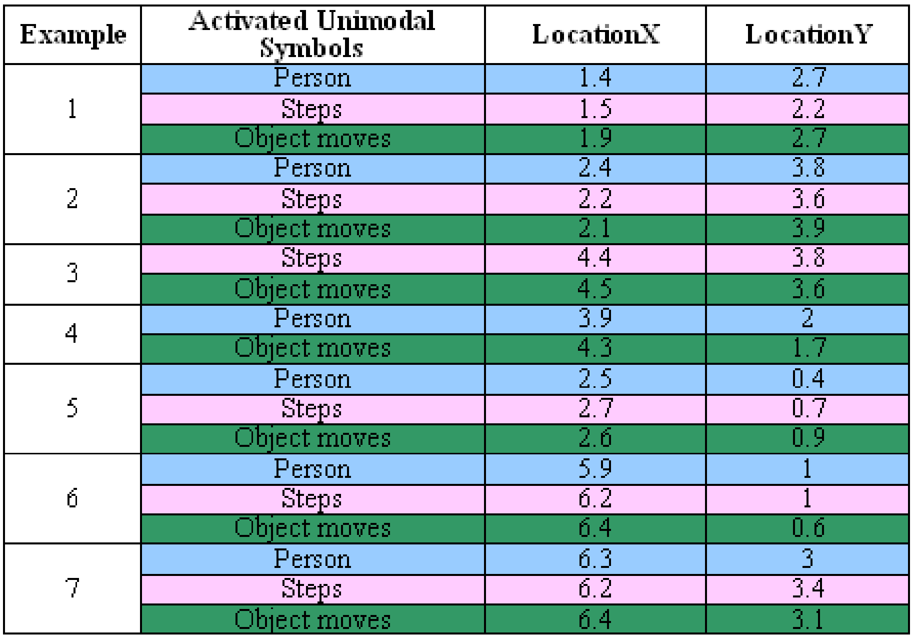


Figure 6.4: Graphical Visualization of Training Examples for the Multimodal Symbol “Person walks”

As can be seen from the examples, when a person walks around in the room, there are generally activated the unimodal symbols “person”, “steps”, and “object moves” concurrently. Looking at the table in figure 6.5, it attracts attention that in example 3, the symbol “person” and in example 4, the symbol “steps” is not active. Such cases might occur due to a failure in one modality or because the view of vision is covered by another object. Generally, if possible, corrupt or atypical data should be best avoided in the training data. However, as this is not always possible, the learning algorithm used should allow it to drop out a certain number of such atypical data. By applying the algorithm described in section 5.5, there are set forward connection from the visual unimodal symbol “person”, the auditory unimodal symbol “steps”, and the tactile unimodal symbol “object moves” to the multimodal symbol “person walks” (see figure 6.3).

After the connections have been set, for each connected symbol, certain location data are calculated based on the location information provided by the examples (see section 5.5, formulas 5.1 to 5.4). In the formulas 6.1 to 6.4, the calculation is illustrated for the symbol “person”. As can be seen from these formulas, the data from example 3 and 4 are excluded from the calculation process, because they are not representative for the particular perceptual image, which generally consists of three unimodal symbols instead of only two. The results for the symbols “steps” and “object moves” are shown in the table of figure 6.6. These values specify the valid range within which these lower-level symbols can principally occur to be bound to the higher-level symbol “person walks”. The valid location ranges are stored as variables in the neuro-symbol “person walks”. The check whether the symbol lies within the valid range is performed in the *on receive*



Example	Activated Unimodal Symbols	LocationX	LocationY
1	Person	1.4	2.7
	Steps	1.5	2.2
	Object moves	1.9	2.7
2	Person	2.4	3.8
	Steps	2.2	3.6
	Object moves	2.1	3.9
3	Steps	4.4	3.8
	Object moves	4.5	3.6
4	Person	3.9	2
	Object moves	4.3	1.7
5	Person	2.5	0.4
	Steps	2.7	0.7
	Object moves	2.6	0.9
6	Person	5.9	1
	Steps	6.2	1
	Object moves	6.4	0.6
7	Person	6.3	3
	Steps	6.2	3.4
	Object moves	6.4	3.1

Figure 6.5: Training Examples for the Multimodal Symbol “Person walks” in Tabular Form

section of the neuro-symbol “person walks” by the function *checkInputMessageType(msg)* (see section 5.4.1).

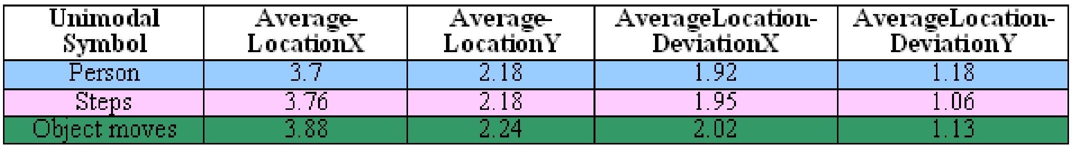
$$AverageLocationX = \frac{1.4 + 2.4 + 2.5 + 5.9 + 6.3}{5} = 3.7 \tag{6.1}$$

$$AverageLocationY = \frac{2.7 + 3.8 + 0.4 + 1 + 3}{5} = 2.18 \tag{6.2}$$

$$\begin{aligned} &AverageLocationDeviationX = \\ &\frac{|3.7 - 1.4| + |3.7 - 2.4| + |3.7 - 2.5| + |3.7 - 5.9| + |3.7 - 6.3|}{5} = 1.92 \end{aligned} \tag{6.3}$$

$$\begin{aligned} &AverageLocationDeviationY = \\ &\frac{|2.18 - 2.7| + |2.18 - 3.8| + |2.18 - 0.4| + |2.18 - 1| + |2.18 - 3|}{5} = 1.18 \end{aligned} \tag{6.4}$$

Additionally to the information where lower-level symbols can principally occur to be bound to a higher-level symbol, there also has to be considered the information how much their location values can deviate within each particular example. This information is calculated by the formulas 5.5 and 5.6. In the concrete case, the tactile modality is the dominant modality. How to calculate the inter-modality location deviation between the symbols “object moves” and “person” is shown in the formulas 6.5 and 6.6. The rest of the results is illustrated in the table of figure 6.7. As the



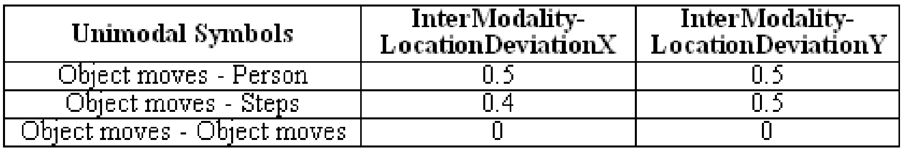
Unimodal Symbol	Average-LocationX	Average-LocationY	AverageLocation-DeviationX	AverageLocation-DeviationY
Person	3.7	2.18	1.92	1.18
Steps	3.76	2.18	1.95	1.06
Object moves	3.88	2.24	2.02	1.13

Figure 6.6: Calculation of Location Information of the Multimodal Symbol "Person walks"

symbol "object moves" is the symbol of the dominant modality, the deviations in the last row of this table are zero. In the implementation, again, these values are stored in variables in the symbol "person walks".

$$\begin{gathered} InterModalityLocationDeviationX = \\ \max(|1.9 - 1.4| + |2.1 - 2.4| + |2.6 - 2.5| + |6.4 - 5.9| + |6.4 - 6.3|) = 0.5 \end{gathered} \tag{6.5}$$

$$\begin{gathered} InterModalityLocationDeviationY = \\ \max(|2.7 - 2.7| + |3.9 - 3.8| + |0.9 - 0.4| + |0.6 - 1| + |3.1 - 3|) = 0.5 \end{gathered} \tag{6.6}$$

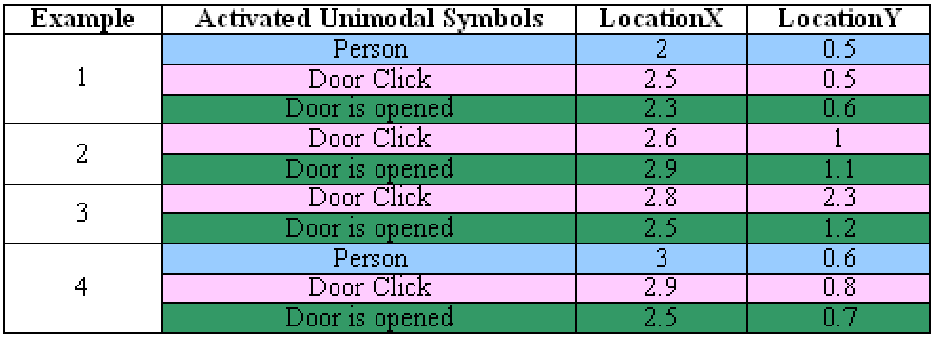


Unimodal Symbols	InterModality-LocationDeviationX	InterModality-LocationDeviationY
Object moves - Person	0.5	0.5
Object moves - Steps	0.4	0.5
Object moves - Object moves	0	0

Figure 6.7: Calculation of Inter-modality Location Information of the Multimodal Symbol "Person walks"

In algorithm 5.1 presented in section 5.5, for determining correlations between lower-level neuro-symbols and a higher-level neuro-symbol from examples, it was defined that there shall be used two neuro-symbols instead of only one if two different lower-level neuro-symbols of the same modality occur always in more than c_2=35 % of all examples. In the used test cases described in section 6.1.1, a splitting of symbols occurs for the symbol "person opens door" when using the examples given in the table of figure 6.8. A similar splitting will occur for the neuro-symbol "person closes door". Again, it is assumed that all symbols are active concurrently at least for one instant of time and that therefore a consideration of timing can be omitted.

As can be seen from the examples given in the table, in 50 % of the cases, the neuro symbols "door click", "door is opened", and "person" are activated when a person opens the door. In the other 50 %, only the symbols "door click" and "door is opened" are activated. This is plausible, because when a person opens the door from outside the room, the video camera cannot detect the visual symbol "person". Therefore, the situation that a person opens the door is split into the two cases "person opens door from inside" and "person opens door from outside", and connections are set according to the presented examples (see figure 6.3). The location information for these two symbols is calculated in a similar manner as described for the symbol "person walks".



Example	Activated Unimodal Symbols	LocationX	LocationY
1	Person	2	0.5
	Door Click	2.5	0.5
	Door is opened	2.3	0.6
2	Door Click	2.6	1
	Door is opened	2.9	1.1
3	Door Click	2.8	2.3
	Door is opened	2.5	1.2
4	Person	3	0.6
	Door Click	2.9	0.8
	Door is opened	2.5	0.7

Figure 6.8: Training Examples for the Multimodal Symbol "Person opens door" in Tabular Form

Learning of Correlations in Training Phase B

As described in section 4.5, in training phases A, forward connections between neuro-symbols of different levels are set and valid values for location properties as well as other properties are determined. Figure 6.9 shows the result of the training phase A of the unimodal tactile level for the used symbol configuration described in section 6.1.1.

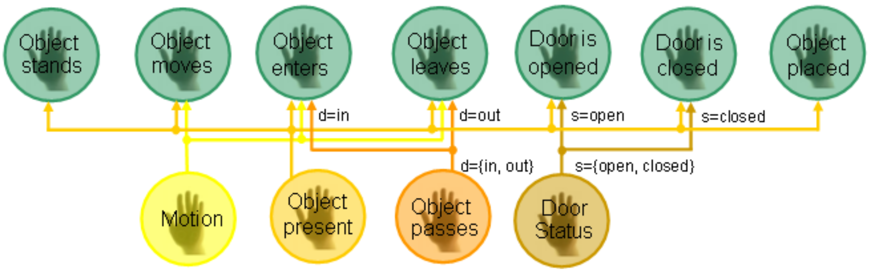
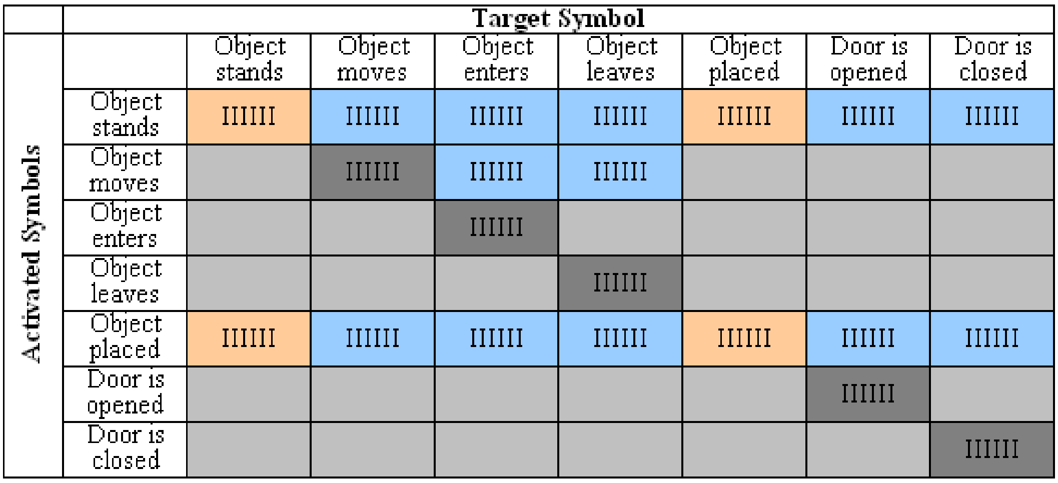


Figure 6.9: Connections between Sub-unimodal and Unimodal Tactile Layer after Training Phase A

In training phase B, based on these correlations, it is determined what feedback connections are necessary and if there exist redundant symbols. This is determined by presenting the same examples as already used in training phase A to the neuro-symbolic network a second time with the difference that now forward connections between neuro-symbols of the mentioned two layers are already set. For each neuro-symbol, a number of examples is used. For the simulation, for each unimodal tactile neuro-symbol, there exist six examples. In section 5.5, it was explained how to identify redundant symbols or symbols that need feedback connections from elements lying outside the main diagonal of a table. To get the entries of the table, in training phase B, it is observed what other neuro-symbols are always activated concurrently with the particular desired neuro-symbol.

The table of figure 6.10 shows the results of this observation. There are depicted the desired target symbols and the symbols actually activated when presenting the example data for a certain perceptual image. As can be seen from figure 6.9 and figure 6.10, one of the two symbols "object stands" and "object placed" marked orange in the table is completely redundant. Both symbols are always activated concurrently. Therefore, one of these two symbols can be omitted. For the simulation, the symbol "object placed" was selected to be the one to be eliminated. This happens by removing existing connections of its input and output port. For the symbols “object stands” and “object moves”, in the table if figure 6.10, there exist undesired symbol activations lying outside the main diagonal. To avoid such undesired activations, there have to be set feedback connections from the target symbols to these two neuro-symbols marked blue. Figure 6.3 shows the result of the process of eliminating redundant symbols and stetting feedback connections.



		Target Symbol						
		Object stands	Object moves	Object enters	Object leaves	Object placed	Door is opened	Door is closed
Activated Symbols	Object stands	IIIIII	IIIIII	IIIIII	IIIIII	IIIIII	IIIIII	IIIIII
	Object moves		IIIIII	IIIIII	IIIIII			
	Object enters			IIIIII				
	Object leaves				IIIIII			
	Object placed	IIIIII	IIIIII	IIIIII	IIIIII	IIIIII	IIIIII	IIIIII
	Door is opened						IIIIII	
	Door is closed							IIIIII

Figure 6.10: Determination of Redundant Symbols and Feedbacks between Unimodal Tactile Symbols

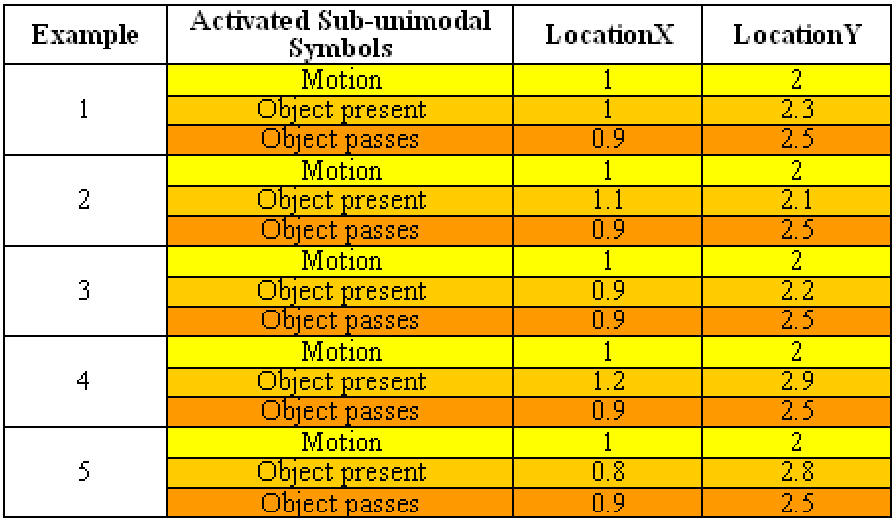
Besides determining needed feedback connections, it also has to be defined how close in position two symbols have to lie one next to the other for such an inhibitory feedback to be effective (see formulas 5.8 to 5.11). In the following, the results of these calculations are outlined for the tactile unimodal symbol “person enters”, which has inhibitory connections to the symbols “object stands” and “object moves”. The data used for the calculations are given in the table of figure 6.11. In the examples, the symbols “motion” and “object passes” are always detected at the same location, because the symbol “object passes” is determined by two light barriers mounted at a fixed position and the symbol “motion” is activated by data coming from motion detectors, which do not have a spatial resolution high enough to get different location values. The formulas 6.7 to 6.10 show the results of the calculations, which are valid for both the symbol “object stands” and “object moves”.

$$FeedbackAverageLocationX = \frac{1 + 1.1 + 0.9 + 1.2 + 0.8}{5} = 1 \tag{6.7}$$

$$FeedbackAverageLocationY = \frac{2.3 + 2.1 + 2.2 + 2.9 + 2.8}{5} = 2.46 \tag{6.8}$$

$$\begin{split} FeedbackAverageLocationDeviationX = \\ \frac{|1-1| + |1-1.1| + |1-0.9| + |1-1.2| + |1-0.8|}{5} = 0.12 \end{split} \tag{6.9}$$

$$\begin{split} FeedbackAverageLocationDeviationY = \\ \frac{|2.46-2.3| + |2.46-2.1| + |2.46-2.2| + |2.46-2.9| + |2.46-2.8|}{5} = 0.31 \end{split} \tag{6.10}$$



Example	Activated Sub-unimodal Symbols	LocationX	LocationY
1	Motion	1	2
	Object present	1	2.3
	Object passes	0.9	2.5
2	Motion	1	2
	Object present	1.1	2.1
	Object passes	0.9	2.5
3	Motion	1	2
	Object present	0.9	2.2
	Object passes	0.9	2.5
4	Motion	1	2
	Object present	1.2	2.9
	Object passes	0.9	2.5
5	Motion	1	2
	Object present	0.8	2.8
	Object passes	0.9	2.5

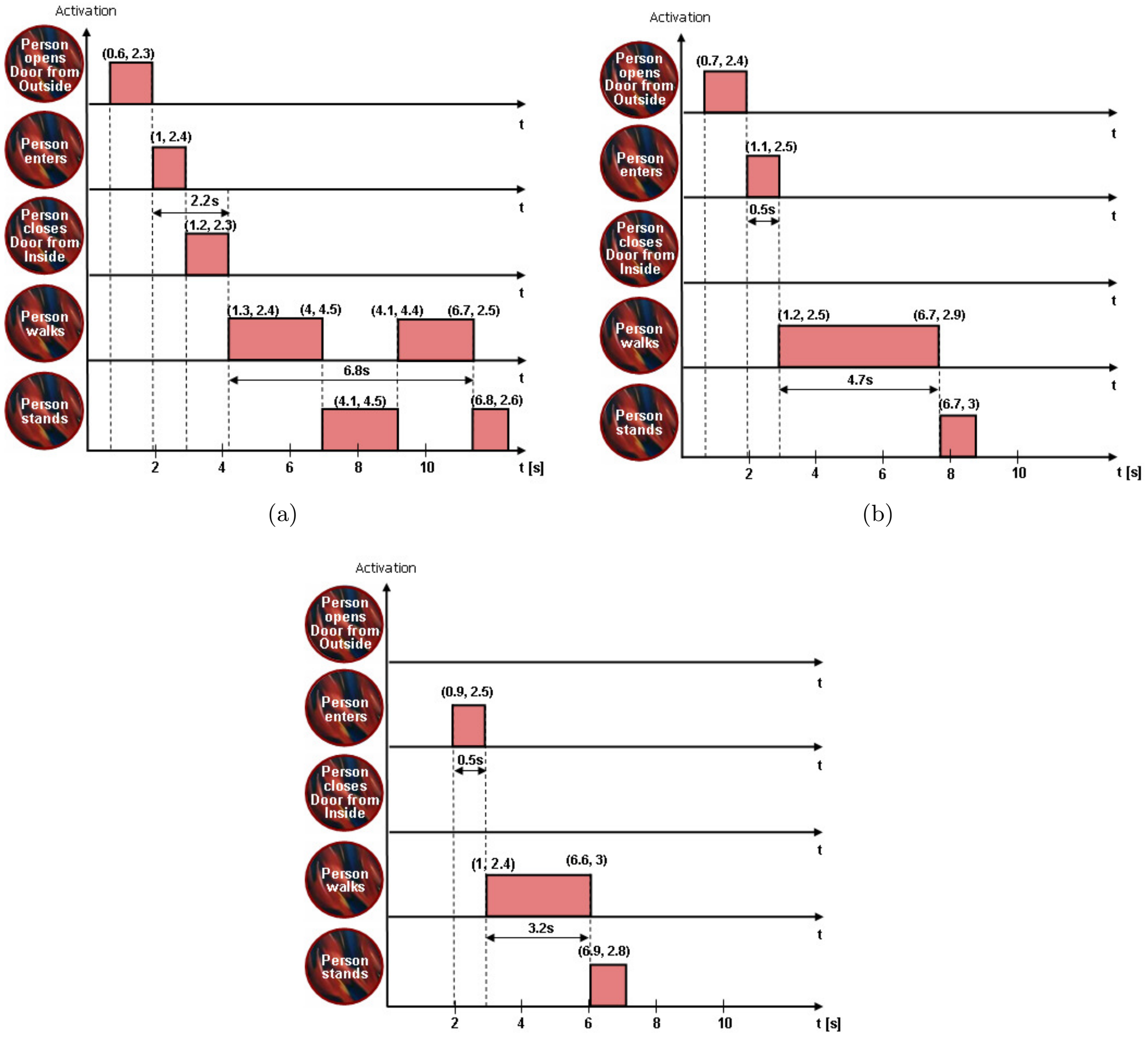
Figure 6.11: Training Examples for the Unimodal Symbol “Object enters”

Learning of Time Successions

As outlined in section 4.4.4, besides location information and other properties, the activation of sensors and symbols over time is of importance. In the model, data occurring within a certain time window as well as temporal successions of events can be handled. As described in section 5.5, for learning successions of events, state charts are used. How these state charts can be configured shall now be illustrated by means of a concrete example, which is the scenario that a person enters the room and goes to the window. This situation is represented by the scenario symbol “person goes to window” (see figure 6.2).

For the used test set, to configure the state chart, three different examples are presented to the system. The figures 6.12a to c show the examples. As can be seen, they can comprise the activation of different multimodal symbols. The numbers in brackets indicate at what x- and y-position the multimodal symbols corresponding to the scenario were perceived. The task of the learning algorithm is to determine what symbol activations all examples have in common and in what spatial area they occur. For the presented examples, the used algorithm will identify the succession of the symbols “person enters”, “person walks”, and “person stands” as common and connect them to the mentioned scenario symbol (see figure 6.3). Concerning their position, the symbol “person enters” will occur next to the door and the symbol “person stands” next to the window. The symbol “person walks” has a special meaning when considering symbol activations over time, because it changes the value of its location property due to the movement of the person. For the scenario “person goes to window”, only the start and the end position of this symbol are relevant. The person has to begin to walk near the door and has to go to the window within a certain time. In between, he/she can also stop and therefore activate the symbol “person stands”. Besides this, the symbol “person goes to window” is detected no matter if the person opened or closed the door.

Figure 6.13 illustrates how the state chart and its transition conditions look like after learning has taken place. Transitions not used are not depicted. To represent the scenario “person walks to window”, apart from the initial and the final state, four states are necessary. For this reason,



(a)

(b)

(c)

Figure 6.12: Training Examples for the Scenario Symbol “Person goes to Window”

state 5 and 6 of the state chart are not used. In transition T1, which stands for the activation of the symbol “person enters”, in the used test bed, no condition for the spatial area within which the symbol can occur needs to be defined, because there only exists one door through which a person can enter. Therefore, the symbol can only occur within the spatial area around this door. The situation would change if a test environment with a certain number of rooms were used. The location values of the transitions T2, T3, and T4 as well as the timing information of T8 and T9 are calculated as described in section 5.5 in the formulas 5.1 to 5.4. For T4 and T8, the calculations are shown explicitly in the formulas 6.11 to 6.15.

$$AverageLocationX = \frac{6.8 + 6.7 + 6.9}{3} = 6.8 \tag{6.11}$$

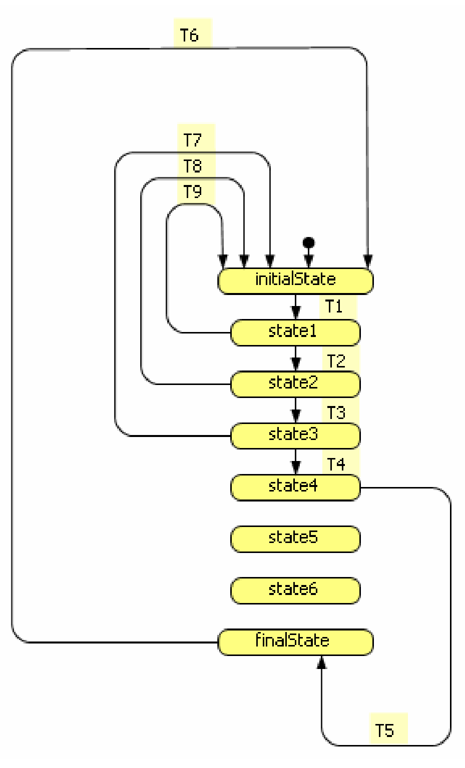
$$AverageLocationX = \frac{2.6 + 3 + 2.8}{3} = 2.8 \tag{6.12}$$

$$AverageLocationDeviationX = \frac{|6.8 - 6.8| + |6.8 - 6.7| + |6.8 - 6.9|}{3} = 0.07 \tag{6.13}$$

$$AverageLocationDeviationY = \frac{|2.8 - 2.6| + |2.8 - 3| + |2.8 - 2.8|}{3} = 0.13 \tag{6.14}$$

$$ActivationTime = \max(6.8, 4.7, 3.2) = 6.8 \tag{6.15}$$

In state 3, an activation of the symbol “person stands” causes a transition to state 4 when occuring within the range of $AverageLocationX \pm 2 \cdot AverageLocationDeviationX$ and $AverageLocationY \pm 2 \cdot AverageLocationDeviationY$. If within a time of $2 \cdot ActivationTime$ after the symbol “person walks” was perceived at a location close to the door, there is not activated a symbol “person stands” at a location close to the window, it is returned to the initial state of the state chart. For the transition T7, a default timeout value of 5 s is used. This is because T7 is the transition for indicating that a person walked to the position next to the window but did not stop there. This case cannot be covered by the positive examples of persons going to the window used for training. Therefore, this timeout value cannot be determined from examples. The transitions T5 and T6 are performed immediately after state 4 and the final state have been reached, respectively.



T1 … if (“Person enters” is active)

T2 … if ((“Person walks” is active) &
(0.95<LocationX<1.39) & (2.35<LocationY<2.51))

T3 … if ((“Person walks” is active) &
(6.59<LocationX<6.75) & (2.4<LocationY<3.2))

T4 … if ((“Person stands” is active) &
(6.66<LocationX<6.94) & (2.54<LocationY<3.06))

T5 … immediately

T6 … immediately

T7 … if (timerValue>5)

T8 … if (timerValue>13.6)

T9 … if (timerValue>4.4)

Figure 6.13: Configured State Chart of the Scenario Symbol “Person goes to Window” after Learning

6.1.3 Flow of Information and Symbol Activations during Operation

As described in chapter 4, in the proposed model of human-like machine perception, information is processed by neuro-symbols arranged in different hierarchical levels. Between neuro-symbols, there exist forward connections from lower-level symbols to symbols of the next higher level as well as feedback connections between neuro-symbols of a certain level and modality. Information processing between neuro-symbols is event-based. In figure 6.14, it is shown what neuro-symbols from the sub-unimodal level upwards are activated when a person enters the room. For the performed investigations, location information is not of importance and therefore omitted.

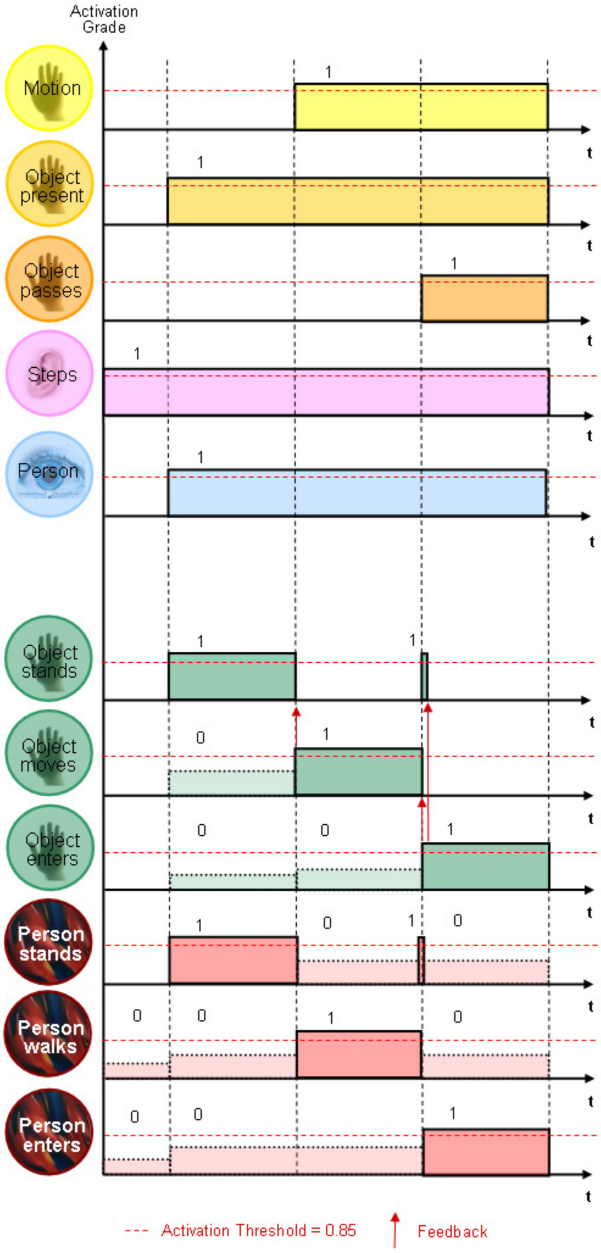


Figure 6.14: Symbol Activations when a Person Enters the Room

The first five signals of the diagram show the activation over time of the tactile sub-unimodal and the visual and auditory unimodal neuro-symbols, which are activated by data sent from the sensor value generator. The last six signals show what higher-level neuro-symbol activations result from these activations assuming connections between symbols as depicted in figure 6.3. The numbers above the time signals indicate the activation value that is sent from the corresponding neuro-symbol to other connected neuro-symbols. For symbol activations having a value below the threshold, the value zero is transmitted to other symbols via the output. In figure 6.14, these cases are depicted with dotted lines and in lighter color. Points in time where a feedback message is sent to another symbol are marked with a red arrow. As long as a feedback is effective, the activation grades of the symbols receiving the feedback signal are set to zero. Figure 6.14 shows that for the symbol “object stands”, there can occur a short activation peak when the symbol “object enters” is activated. This is due to the fact that the symbol “object moves” has an inhibitory influence on the symbol “object stands”. On the other hand, the symbol “object enters” has an inhibitory influence on the symbol “object moves” and on the symbol “object stands”. If now the inhibitory feedback message of the symbol “object enters” reaches the symbol “object moves” before the symbol “object stands”, the symbol “object moves” is deactivated and its inhibitory influence on the symbol “object stands” is rescinded before the feedback of the symbol “object enters” can get effective. This results in an activation of the symbol “object stands” for a very short time. In the simulation, where the time for transmitting and processing of data can be set to zero, the length of the peak can be assumed to be zero. However, when implementing the model into hardware, signal runtimes and time for data processing have to be considered, which will result in such peaks of a certain length. It has to be assured that these peaks do not influence succeeding information processing. Therefore, for setting and resetting memory symbols and delivering a valid value to the output of the system, neuro-symbols have to be active for a certain minimum time.

Besides forward connections and feedbacks, in the test set, knowledge can also have an inhibitory influence on neuro-symbol activations. In section 4.6, it was described that for storing effects and consequences of events happened, memory symbols are used. Via information stored in these memory symbols, knowledge interacts with neuro-symbols. Figure 6.15 illustrates a number of examples, which show when a memory symbol is set and reset and when the state of a memory symbol has influence on the activation grade of a neuro-symbol. For the intended investigations, location information and timing is irrelevant and therefore not considered. All depicted activations have the value one. Figure 6.15 shows that the memory symbol “person present”, which is set (or active) at the beginning, is reset (or deactivated) when the symbol “person leaves” occurs and reactivated when the perceptual image “person enters” is perceived. If the memory symbol “person present” is not set, the symbols “person stands” and “person walks” cannot be activated, even though according to the bottom-up information processing of sensory data, their activation grade would exceed the threshold value. In figure 6.15, these cases are depicted as dotted lighter colored rectangles labeled with the numbers 1 and 2. The memory symbol “door open”, originally set, is reset when the symbol “person closes door” is activated. After this event occurred, an activation of the symbol “person leaves” is inhibited by top-down influence of knowledge (see dotted light colored rectangle with labeling 3).

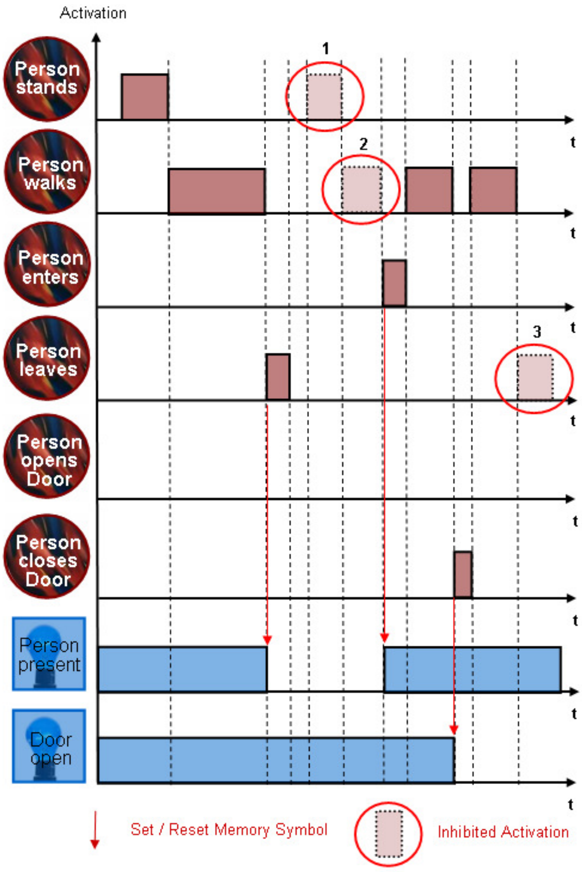


Figure 6.15: Examples for Interaction of Memory Symbols and Knowledge with Neuro-symbols

6.2 Discussion based on Lessons Learned from Implementation and Simulation Results

Section 6.1 gave an outline of simulation results for a given test set. Based on these simulation results as well as lessons learned during model development and implementation, this section comprises a discussion of important issues of the presented model. Additionally, from the gained insights, conclusions are drawn about the correctness, incorrectness, or incompleteness of neuroscientific and neuropsychological research findings and models of the brain.

Importance of Sensor Configuration and Feature Symbol Selection

The first issue that shall be discussed is the importance of the sensor configuration and the selection of feature symbols for further information processing. In traditional approaches, the structure and architecture of information processing are held responsible for the representational function of a cognitive system. However, in [Pes94], it is suggested that sensory systems play an equally important role in information processing and that the representational function of a cognitive system becomes only meaningful if it is physically coupled to the environment by a sensory (and motor) system. From the insights gained during model development, the theory that the configuration of the sensory system is important for information processing of perceptual images can be supported. First of all, of course, the types of used sensors are important for the effectiveness of the perception process. Additionally, the place where they are mounted as well as their number and spatial resolution play a decisive role. A high number of sensors and a higher spatial resolution potentially allow it to distinguish many different perceptual images. However, in such a case, the effort for information processing in the feature symbol level also gets higher and more computational resources are needed. As for many modalities the correlations between sensor data and neuro-symbols have to be predefined by the system designer, besides a higher hardware effort, this also requires a higher design effort and longer development times. If too many sensors are used, the effort and the needed resources increase without additional gain of perceptual discrimination abilities.

If only fewer sensors and therefore a smaller spatial resolution are available, many perceptual images that would be distinguishable by having available more sensors are assigned to the same neuro-symbol, because there cannot be detected a difference in sensor activations. The advantages of fewer sensors are simpler correlations between sensor data, feature symbols, and sub-unimodal symbols and therefore the necessity of fewer processing units.

For most efficient processing, a balance has to be found between the two extreme cases of many sensors and many different neuro-symbols causing much design effort and needing high computational resources and few sensors and few possibilities to distinguish perceptual images. From these observations, it can be concluded that the sensor configuration already plays an important role in efficient information processing. Additionally, an extraction of suitable features from sensor data is of great importance to allow efficient processing. Generally, concerning the number and spatial resolution of sensors, it is not mandatory to have the same spatial resolution at all places. Drawing a conclusion back to neuroscience, this means that the type, number, and arrangement of sensory receptors of our body already play a crucial role for the efficiency and manifoldness of our perception.

One Single Structure for Storage and Processing of Information

In classical computer architectures like suggested by von Neumann, storage and processing of information is performed by different modules. In contrast to this technical approach, [JLN05] hypothesize that in the brain, storage is mediated by the same brain structures that process perceptual information. The neuro-symbolic network developed in this thesis shows such an architecture being responsible for storing and processing perceptual information at the same time. Neuro-symbols in neuro-symbolic networks are both memory cells and processing units. Therewith it was shown that such a structure is principally feasible. Additionally, it turned out that such a structure is very efficient for information processing as there are needed no explicit memory access and no explicit comparative operations for comparing input data and stored data. Having shown that an architecture incorporating storage and processing of information is feasible and is expected to work very efficiently, the hypothesis that in the perceptual system of the brain the structure for storage of information is also responsible for processing of information gains additional support. Taking over this principle for computational systems, this could lead to more efficient controllers.

Adaptability of Design

In section 4.5, it was described how flexibility and adaptability of the design can be achieved by supervised learning from examples. During testing, it turned out that the efficiency of learning strongly depends on the quality of the example data. The more representative the examples, the better are the achieved results. If the examples contain too many atypical or faulty data, there might be learned undesired connections and correlations. If necessary, the used learning algorithm needs to have the capability to rule out or to detect and eliminate faulty or atypical data. Additionally, similar as for neural networks, the algorithm needs to offer a certain ability to generalize over presented examples to be able to handle also unseen cases.

Concerning the ability to add neuro-symbols in different layers, the algorithm needs a certain ability to generalize to not create new neuro-symbols for every single example representing one and the same object, event, or situation. For the function of such a splitting of neuro-symbols, it is advisable to let the designer provide information in what cases an addition of neuro-symbols shall be performed and to allow the designer also to circumvent a suggested splitting. Concerning the capability of the system to eliminate redundant symbols, the learning algorithm should not generalize too grossly over data to not eliminate too many designated symbols. Generally, it makes sense to have different generalization abilities for different parameters. It turned out that in most cases, it makes sense to generalize to a certain extend over location data, because it would require too much effort to generate examples of one and the same event at all different locations where it can occur and train the system with these data. In the same manner, it makes sense to "generalize over time" as the duration of certain events or situations can vary.

Redundancy, Fault Tolerance, and Conflict Resolution

Human perception relies on information coming from different sensory receptors. In a similar manner, the introduced model for human-like perception acquires information from different sensor types. However, the usage of multiple sensor types brings some problems along that have to be solved by an intelligent system design. Data belonging to one event can be redundant, incomplete, contradicting, or of different reliability. In the following, it is discussed how the model can handle redundant data, how fault tolerance can be achieved, and how conflicts caused by ambiguous sensor data can be resolved.

In many cases, multisensory data originating from one and the same event include redundancy. For example, when using video cameras, motion detectors, and tactile floor sensors, all three sensor types can detect movement of objects. The data from the different sensors have to be merged and it has to be avoided that – instead of one single moving object – three moving objects are detected. The model presented in chapter 4 allows it to merge redundant data quite effectively by its modular hierarchical neuro-symbol structure. The connections between the neuro-symbols of different levels as well as their properties contain the information which sensor data are likely to belong to the same object or event. Connections in lower levels have to be predefined, connections in higher layers are derived from examples. Location information plays a very important role in the process of integrating diverse sensory data.

Redundancy of data can be exploited to achieve fault tolerance if a certain modality fails partly or completely. There can occur the cases that a modality delivers no data, or that it delivers faulty data. The case that a certain modality just does not contribute data is the case easier to handle. The case that faulty data are transmitted is critical, because these faulty data could be coincidentally assigned to neuro-symbols representing perceptual images they actually do not belong to. First, the case shall be discussed that sensory modalities fail and just do not contribute sensory data to the perception process:

Neuro-symbols, which receive input information from different information sources, are potentially more robust to failures, because they can exploit a certain redundancy in data. If for example a video camera, motion detectors, and tactile floor sensors are used to detect movement of objects, a moving object could be perceived even if for example the visual modality failed. For allowing fault tolerance, it is important to set the threshold of neuro-symbols to an appropriate value. If the threshold is too high, the symbol will only be activated if all modalities serving as input data provide a contribution. If it is too low, there will be activated symbols that do not correspond to the perceptual image currently perceived. The value of the threshold best to be chosen depends on the probability for the occurrence of failures. If the probability for a failure is low, a high threshold value leads to the best result. If the number of errors occurring is high, a lower threshold value is more recommendable. Additionally, it might have an influence on performance what information correlating with the activation grade of a neuro-symbol is transmitted to other neuro-symbols. The efficiency of perception in case of faulty data is by far better if the system "knows" what modality failed. If the system does not know about the failure, in the worst case, it can occur that there are activated neuro-symbols not corresponding at all to the object, event, scenario, or situation currently going on in the environment. This particularly occurs if neuro-symbols of one level are activated by largely the same lower-level neuro-symbols and therefore comprise little redundancy. Therefore, if possible, the system should be "aware" of occurrences of malfunctions of certain modalities. Comparing this concept to the human brain, humans are generally also aware that they have e.g., momentarily closed their eyes and can therefore not acquire visual information for perception. If the malfunction is known, by influence of knowledge, perception can be facilitated. The question is how the system can notice the malfunction of a modality. Possibilities would be to build in self tests into the system or to integrate mechanisms into the system that detect that certain neuro-symbols are triggered remarkably often or seldom and that therefore a sensor malfunction is probable.

In the last paragraphs, it was described how perception of the system can be preserved if one sensory modality fails and does not contribute information to the perception process. Now, it shall be investigated how the system behaves if one modality fails, and the sensors do not simply deliver no data but, instead, delivers faulty data. If the system is "aware" of this malfunction, it can exclude the faulty data and the problem handling works like described in the last paragraphs where one modality falls out completely. If the system is not "aware" of the failure, for one single event, different sensor modalities can provide contradicting data. If there is enough redundancy in the sensor data of the different modalities, this does not pose a great problem. For example, if motion of an object is detected by a video camera, motion detectors, and floor sensors, if one modality fails, there are still two modalities left working correctly that can overrule the faulty data. The situation gets more complicated if the faulty modality coincidentally activates neuro-symbols that can contribute to the activation of other neuro-symbols. In such a case, also higher-level symbols could be activated that do not represent the real events and situations going on in the environment. The number of undesired higher-level neuro-symbol activations can greatly be decreased exploiting location information about data. Additionally, top-down influences from knowledge can facilitate perception in such situations.

Binding and the Importance of Location Information

In section 3.4, the so-called binding problem, which is regarded as one of the most puzzling problems ever faced in neuroscience, and its potential solutions were introduced. Binding has to be performed across space and time. However, each of the currently existing theories suffers from certain weak points. Chapter 4 comprised a suggestion how to combine and supplement the different proposed mechanisms of binding in order to get a feasible bionic system being capable of parallel distributed information processing of sensory data. In lower levels of information processing, combination coding proved to be a reasonable solution. In higher levels, a combination of principles inspired from population coding and temporal coding turned out to be suitable. Additionally, top-down mechanisms coming from memory and knowledge showed their utility in the binding process.

Furthermore, as outlined in section 4.4.3, besides the utilization of information occurring at the same instant or interval of time for binding, location information of sensory data is crucial for the binding process. First of all, location information is of urgent importance if different objects, events, and situations are going on concurrently. Location information allows it to correctly assign activated lower-level neuro-symbols to higher-level symbols. Location information is also very useful to detect faulty data and to avoid undesired binding of such data. To consider location information, the used learning algorithms need a certain ability to generalize over location data. The higher the resolution of the location data, the better the generalization capabilities of the algorithm have to be. If perception is overloaded by too many events happening concurrently, focus of attention can help to correctly assign sensory and neuro-symbolic information. This mechanism also acquires location information by restricting the spatial area within which information is processed. As can be shown from the description just given, location information seems to be necessary at all hierarchical levels of perception for binding of perceptual information and might therefore be the key to solve the binding problem in perception.

In neuroscientific basic literature, it is generally reported that in the visual system, there exist two distinct, separate pathways for object recognition and spatial object location being directed towards two different brain areas. However, as insights gained during model development in this thesis showed, the hypothesis of completely distinct streams of information about object type and object location is relatively unlikely. In this case, location information could not be used for the binding process and there would arise the problem how to merge this separated information again in later processing stages. Therefore, it should rather be sympathized with the hypothesis of [GHT96] according to which it would be surprising if the brain did not make use of spatial information freely available at least partly for solving the binding problem.

Focus of Attention versus Parallel Processing

In section 4.4.5, different mechanisms were introduced to handle objects, events, and situations occurring concurrently. There principally exist the possibilities of parallel symbol representations, of the usage of group activity symbols, and the mechanism of focus of attention, each of them having certain advantages and weak points.

The method that leads to the best and fastest perception results is the method of parallel symbol representation, because with this method, events can actually be processed in parallel. However, when considering the design effort, there has to exist a certain number of neuro-symbols for each perceptual image from the sub-unimodal layer upwards. This requires lots of basic information processing units and therefore high computational resources. Additionally, mechanisms are necessary to avoid that different neuro-symbols representing the same type of perceptual image are activated concurrently by this perceptual image occurring in the environment. Therefore, this method is only recommendable for perceptual images occurring concurrently very frequently and/or that are triggered from sensor values and low-level symbols being active only for a short time.

For other cases, the usage of focus of attention is generally preferable, because it offers the possibility to get along with a smaller number of neuro-symbols. However, one problem can occur when using focus of attention instead of parallel symbol representation. As focus of attention processes information sequentially from the sub-unimodal level upwards, it can happen that information being represented by activated feature-symbols only for a short instance of time cannot be processed if the focus of attention is currently directed to another spatial area. This information is then lost. From the neuroscientific point of view, there cannot yet be given an answer by what mechanisms the focus of attention is steered. For test purposes, in the simulation, this information was provided to the system by an external source[2]. A contribution to determine the current area of focus of attention could be made by using group activity symbols. The activation of such symbols could steer the focus of attention to certain spatial areas if a more detailed investigation of certain situations is desired.

If no parallel symbol representations are used and too many activities are going on in the environment to represent them all by different neuro-symbols, group activity symbols can be used. Information represented by group activity symbols is of less detail, which can be an advantage but also a disadvantage depending on what shall actually be perceived. A further restriction is that lower-level group activity symbols can only be combined with other symbols of such a group activity character.

Neuro-symbolic Level for Interaction with Knowledge and Memory

In section 4.6, mechanisms were introduced to allow knowledge and memory about past events to influence the activation of neuro-symbols. It was described that neuroscience and neuropsychology do not give a decisive answer on what level interaction between knowledge and perception takes place in the brain. Therefore, it was declared that this interaction can principally take place in layers between the sub-unimodal level and the scenario symbol level. During model implementation and simulation, it turned out that an interaction at higher hierarchical neuro-symbolic levels is generally more efficient than an interaction at lower levels. The effect of suppressing or increasing the activation of neuro-symbols by knowledge already at lower levels is in many cases too crude, because a certain lower-level symbol can be combined to many different higher-level symbols with quite different meaning. An interaction at higher hierarchical levels is more specific. Generally, single neuro-symbols of lower layers comprise less specific information, which is also often less intuitively interpretable than higher-level information. Therefore, they offer less possibility for a meaningful interaction with knowledge and memory. Additionally, neuro-symbols of lower levels are based on data from fewer sensory modalities and are therefore often less redundant and more likely to be subject to false activations. For the performed simulations, the multimodal layer and the scenario layer generally seemed to be the best levels for interaction. They are also the only levels that can deliver information used as output information of the system. Using only these two levels for interaction with knowledge and memory reduce the implementation effort for learning mechanisms to determine what neuro-symbols shall set and reset memory symbols and the number of necessary connections between the modules.

[2] e.g., by an algorithm searching through space or by the system engineer

Drawing a conclusion from the bionic model back to neuroscience, this would mean that – besides for the very lowest levels – an interaction between abstract knowledge and perception is principally conceivable in all layers. However, it might be more likely to occur in higher cortical layers.

Stability of Perception

In section 4.2.3, it was mentioned that information in neuro-symbolic networks is processed bottom-up from sensor data, that neuro-symbols are influenced top-down from knowledge and focus of attention, and that there can exist feedbacks within neuro-symbolic levels. Information exchange is carried out event-based by passing messages between units (see section 4.3.1). As discussed in section 6.1, due to the existence of top-down information flow and feedbacks, this can lead to a circulation of messages between neuron-symbols several times before a stable value is reached. During this "message circulation phase", there can temporarily be activated certain neuro-symbols, which do not correspond to the actual perceptual image being currently present in the environment.

For a computer simulation, this does not pose a problem, because in the simulation, the time for transmitting and processing messages between neuro-symbols can be assumed to be zero. Concerning the activations of neuro-symbols in the multimodal level and the scenario level, which also serve as system output, (and for lower-level neuro-symbols if they set or reset memory symbols), only the last values of the neuro symbols that were delivered at a certain instant of time are taken into account. However, if the model is implemented into hardware, there have to be considered certain signal runtimes and processing times, because the correct activation values for neuro-symbols will only be available after the message circulation phase is terminated.

The occurrence of circulating messages was the reason why in the implementation, for communication between memory symbols and the knowledge module, interface variables and state charts with certain transition conditions were used instead of ports and messages like for other parts of the model. Without this measure, it can happen that messages are circulated within the model ad infinitum without reaching a stable value. When using only event-based information processing, whenever a memory symbol receives a message, it processes this information and sends a message to the knowledge module no matter if the data in the message have actually changed or not. This module calculates information and sends information back to diverse neuro-symbols. The neuro-symbols again process information and send it upwards to higher neuro-symbols and again to memory symbols and finally to the knowledge module, which transmits information to neuro-symbols and so on. By using interface variables and state charts, a calculation of top-down influences from knowledge to memory symbols is not performed whenever a message is received but only when the state of a memory symbol changes. There are only sent messages from the knowledge module to neuro-symbols, which are influenced by these changes. This measure acts as a kind of filter mechanism and allows it to reduce the number of circulating messages and the number of periods needed for circulation until a final stable value is reached.

The occurrence of circulating messages in the model caused by feedbacks and top-down information flow is an interesting result. Although in the brain, information does not seem to be necessarily exchanged event-based, there can be drawn certain conclusion about information transmission in the brain. The brain is made up of approximately 100 billions of neurons being interconnected and exchanging information every instant of time. Within these neural structures, information does not flow only in one direction, but there exist plenty of feedbacks from higher cortical layers to lower ones. It takes a certain amount of time to transmit information via neurons. Therefore, in the brain, there might occur comparable “circulating signals”.

Taking a look at the perceptual system of the brain concerning signal circulations, information coming from sensors will be processed bottom-up. In later processing stages, higher-level information can influence and change neural activations in lower layers, which again influence the higher levels and so on, which can theoretically proceed ad infinitum. Additionally, every instant of time, the activations can be influenced from parts of the brain lying outside the perceptual system of the brain. Feedbacks and top-down influences might not only be negative (inhibitory) but also positive (excitatory). This arises the question how the brain can ever get to stable signal representations and therefore to a stable, unified perception of the world. In the presented model, higher-level functions were responsible for filtering and inhibiting the activation of neuro-symbolic processing units. In the brain, there might also have to exist certain higher-level filter or inhibition mechanisms to suppress the activation of certain processing units and to reduce the amount of transferred information.

6.3 Comparison with Existing Models

The model proposed in this thesis introduces a new principle of information processing called neuro-symbolic information processing to interpret data coming from sensors. The model is designated for applications in the field of machine perception – particularly for monitoring and surveillance systems in buildings. The aim of this section is to demarcate the model to already existing related models proposed from different research domains. First, there are outlined the main differences to the model developed in the project ARS-PC (see section 2.1.2). Next, there is made a comparison to neural networks and symbolic systems. Finally, it is attempted to bring the model in line with the research fields of neuro-symbolic integration and sensor fusion.

6.3.1 Demarcation to Prior Work

As mentioned in section 2.1.2, the developed model of human-like machine perception described in this thesis was influenced from prior work in the project ARS-PC. In the context of this project, there was developed a model for bottom-up information processing of sensor data in three layers in order to perceive different scenarios. The information within these three layers is processed in terms of symbols. Correlations between lower-level symbols and higher-level symbols have to be predefined by the system engineer in form of rules. However, although the model claims to apply neuroscientific, neuropsychological, and neuro-psychoanalytical concepts, besides the fact that information is processed hierarchically in different layers and that information is processed in terms of symbols, no such concepts are used for the implemented model.

In contrast to this prior model, this thesis attempted to actually concentrate on neuroscientific and neuropsychological research findings and to use them to design a technically functioning and implementable model. The model introduced neuro-symbols as basic processing units, and therefore allows it to combine advantages of both neural and symbolic processing, which are considered as two disparate approaches to explain information processing in the human brain. Additionally, there is defined a strict modular hierarchical structure how information from different sensory modalities has to be processed, which is derived from the structural organization of the perceptual system of the human brain. Accordingly, neuro-symbols are combined to neuro-symbolic networks. The functions of neuro-symbols in different modalities and hierarchical levels correspond to the functions of processing units in the different parts of the perceptual system of the brain. Location information plays a crucial role in binding of information. The lowest neuro-symbolic level is topographic in structure whereas higher-level neuro-symbols respond to perceptual images independently of their location. In contrast to the former model where information was only processed bottom-up beginning with sensor data, in this model, there exist feedbacks between neuro-symbols as well as an information flow, which is directed top-down. Furthermore, in comparison to its precursor, the model offers the possibility to learn correlations between neuro-symbols from examples. Additionally, there were introduced mechanisms like memory, knowledge, and focus of attention to facilitate perception when sensor data are ambiguous and when many events happen in parallel.

6.3.2 Neuro-symbolic Networks versus Symbolic Systems and Neural Networks

Symbolic Systems

In section 2.3.2, a short overview about symbolic systems was given. What neuro-symbolic networks have in common with symbolic systems is the fact that knowledge – in the designated application *perceptual knowledge* – is represented symbolically. The figures in chapter 4 illustrating neuro-symbol hierarchies might convey the impression that neuro-symbolic networks resemble semantic networks. However, there exist many differences between neuro-symbolic networks and symbolic systems. This allows it to circumvent certain problems that arise when using pure symbolic approaches.

With symbolic systems, knowledge can be represented explicitly. However, in section 2.3.2, it was mentioned that sensory processes are generally considered as implicit knowledge, which is hard to model by symbolic AI frameworks. In contrast, the model presented in this work is designed to fit the needs of sensory processes.

A problem of symbolic system, which is generally referred to as the frame problem, is to let the system know what are relevant features that should be tracked. In neuro-symbolic networks, this problem is evaded as each neuro-symbol only receives input information that is relevant for its processing and during learning there can be set time intervals within which arriving information shall be considered. Additionally, a mechanism called focus of attention is designated which allows it to consider only information within a relevant spatial range.

Neuro-symbolic networks also circumvent another problem of symbolic systems described in section 2.3.2, which was referred to as the symbol grounding problem. By deriving neuro-symbolic data from sensor data, neuro-symbols become grounded.

Symbolic systems are often said to be unable to generalize and to be fault tolerant. In contrast to most symbolic systems, neuro-symbolic networks have the potential to learn correlations and relations between neuro-symbols from examples. Learning algorithms used in neuro-symbolic networks allow it to configure the network in a way to give it the ability to generalize to unseen examples. Additionally, by using partly redundant data from different sensory sources and integrating knowledge into the perception process, fault tolerance is obtained. Neuro-symbols can also handle data that involve time, which is critical or even not feasible in many methods used for symbolic systems.

An important characteristic of neuro-symbolic networks is that the neuro-symbolic structure that represents perceptual knowledge is at the same time also the structure that processes the data. This means that neuro-symbols are memory cells and processing units at the same time, which allows fast and effective information processing.

Unlike symbolic systems that have sequential and centrally controlled algorithms, neuro-symbolic structures allow distributed parallel processing. This is a second reason why they guarantee high performance.

There exists one part in the model of human-like perception where for the current state of development, the use of pure symbolic systems would make sense. This part is the knowledge module (see figure 5.3). In section 4.6, it was mentioned that the abstract knowledge contained in this module is currently represented by rules. For coding these rules, symbolic systems would be suitable. However, the focus of future work will be on finding mechanisms to represent this knowledge in a similarly efficient and neuroscientifically related way like perceptual images within the neuro-symbolic structure. Unfortunately, until now, neuroscience and neuropsychology do not provide much information to this topic utilizable for a bionic model.

Neural Networks

Section 2.3.1 gave a brief overview about the research field of neural networks. There exist certain similarities between neural networks and the proposed concept of neuro-symbolic networks. Similar as in neural networks, the basic processing units of neuro-symbolic networks sum up input information about the activation grade of connected entities and are themselves activated if this value exceeds a certain threshold. Like in neural networks, basic processing units are combined to perform complex tasks. In both cases, connections between units can be weighted. However, besides these similarities, there also exist many differences between neural networks and neuro-symbolic networks. In contrast to neural networks, where information is represented in a distributed and generally not interpretable form via weights of connections, every single neuro-symbol has a certain interpretable meaning as each neuro-symbol represents a certain perceptual image. When a neuro-symbol is activated, this indicates that the perceptual image it stands for has been perceived in the environment. Unlike neurons, neuro-symbols can contain properties, which specify a perceptual image in more detail. A special property of neuro-symbols is their location property, which represents information about the location where a perceptual image assigned to a particular neuro-symbol occurs in the environment. This allows it to correctly merge information in case of different events triggering diverse sensors concurrently and offers a mechanism for fault detection.

For artificial neural networks, only the structure and function of a single nerve cell serves as biological archetype. For connecting particular neurons, there are generally not applied concepts taken from neuroscience. In contrast, in neuro-symbolic networks, the structural organization of the perceptual system of the human brain serves as archetype for their architecture.

A further difference is the function of weights of connections in neural networks and neuro-symbolic networks. Whereas weights in neural networks are altered by a learning algorithm to achieve a mapping of input values to output values, in neuro-symbolic networks, weights of connections are used when different sensor modalities deliver information of different reliability.

Another difference between neural networks and neuro-symbolic networks in the way how learning is performed. In contrast to neural networks, in neuro-symbolic networks, learning is performed in several stages for different modalities and hierarchical levels. Lower-level correlations have to be determined before higher-level correlations can evolve. For each hierarchical level, two training phases exist. The first is responsible for calculating forward connections and the second is responsible for calculating and setting feedback connections. In these learning phases, there can also be determined further parameters like location and timing data and values of properties. Additionally, the possibility exists to adapt the neuro-symbolic architecture by adding or eliminating neuro-symbols.

The next difference is the way how information is exchanged between neuro-symbols in contrast to neural networks. For calculating output values from certain input values, in common neural networks, all the necessary information has to be present at the input of the network at one instant of time. To process time signals, there can be used time delays. However, the purpose of the time delays is just to make available the whole needed signal information at the input layer of the network at the same time. In neuro-symbolic networks, neuro-symbols comprise mechanisms, which make it possible to process information arriving asynchronously within a certain time window or in a certain succession. Besides this, information exchange between neuro-symbols is event-based. This means that information is only processed if a new input message is received. This method allows it to reduce the communication and information processing effort.

A big advantage of neuro-symbolic networks compared to neural networks is their interpretable structure as each processing unit is assigned to a certain perceptual image. In neural networks, output information is represented by numeric values, which have to be mapped to a desired meaning. This fact also makes it difficult to connect neural networks to ensemble and modular multi-net systems [Sha98]. In neuro-symbolic networks, output information is represented symbolically. This symbolic output information can be used as input for further neuro-symbolic networks.

6.3.3 Neuro-symbolic Networks for Neuro-symbolic Integration and Sensor Fusion

In section 2.3.3, an overview about existing research approaches in the field of neuro-symbolic integration was given. In this thesis, a model for neuro-symbolic information processing was introduced. As in this model, principles from neural as well as from symbolic information processing are used, it can be regarded as a particular method of neuro-symbolic integration. In section 2.3.3, a classification scheme for categorizing existing neuro-symbolic integration approaches was introduced. It shall now briefly be discussed how the proposed model fits into this classification scheme.

In the classification scheme depicted in figure 2.2, neuro-symbolic systems were divided into unified approaches and hybrid approaches, which are further divided into neuronal symbolic and connectionist symbolic processing and into translational and functional hybrids. When trying to fit the proposed model of neuro-symbolic information processing into this scheme, it best fits into the category of functional hybrids. The characteristic of functional hybrids is that they comprise complete symbolic and connectionist components, which is also the case for neuro-symbolic networks. Functional hybrids can be further classified into loosely coupled architectures, tightly coupled architectures, and fully integrated architectures. Neuro-symbolic networks can be assigned to the category of fully integrated architectures, which have the characteristic to show no discernible external difference between symbolic and neural modules.

In section 2.3.3, for unified approaches, it was outlined that connectionist symbol processing can be further divided into localist and distributed architectures. Localist architectures contain one distinct node for representing each concept. Distributed architectures comprise a set of non-exclusive, overlapping nodes to represent each concept. Considering neuro-symbolic networks, they would be classified as localist architecture.

In the sections 2.3.1 and 2.3.2, it was mentioned that both neural networks and symbolic systems suffer from certain weak points. In section 2.3.3, it was mentioned that the weak points of neural networks and symbolic systems are complementary and that with hybrid approaches, they could be overcome. By using neuro-symbolic networks, certain advantages of neural networks and symbolic systems can be combined. Potential advantages in contrast to pure connectionism or symbolic approaches were already outlined in section 6.3.2.

Section 2.2 comprised a brief description of the research field of sensor fusion. Similar like in sensor fusion, the model introduced in this thesis aims to combine sensor data from diverse sensory sources (and sometimes also other sources) to achieve a better perception of the environment. Therefore, the presented model can also be considered as a model for sensor fusion. As it was outlined in section 2.2, existing models for sensor fusion strongly depend on the application, and up to now, there does not exist a generally excepted model. When discussing biological sensor fusion, it was pointed out that sensor fusion in the perceptual system of the human brain is of far superior quality than sensor fusion achieved with existing mathematical or algorithmic methods. This makes it particularly useful to study the biological principles of sensor fusion. Therefore, the presented model for merging and interpreting sensor data based on studies about the perceptual system of the brain could lead to a more efficient and more generally applicable method in sensor fusion.

Chapter 7

Conclusion and Outlook

"Science is a wonderful thing if one does not have to earn one's living at it."

[Albert Einstein]

The aim of this work was to present a bionic model for human-like machine perception. In section 1.1, it was outlined that up to now, machine perception can by far not compete with the perceptual capabilities of humans. This was the motivation to use the perceptual system of the human brain as archetype for developing a model for machine perception. The foundation for model development was laid by an extensive study of neuroscientific and neuropsychological backgrounds about the human brain and particularly about its perceptual system. In section 7.1, the most important aspects of the developed model are summarized briefly, it is outlined how the characteristics and requirements of human perception described in section 1.1 are fulfilled, and the key results of the work are recapitulated. In section 7.2, based on lessons learned from the performed investigations, interesting directions for further research are identified. Section 7.3 gives an outlook about expected and recommended developments in the research field the thesis is embedded in.

7.1 Model Recapitulation and Key Results

Model Overview

In this thesis, a model was introduced, which aims to emulate the perceptual system of the human brain to build next generation machine perception systems. Therefore, neuroscientific and neuropsychological research findings about the organization and function of the perceptual system of the brain served as archetype. Having available systems capable of a human-like perception of their environment would be valuable for many applications like the surveillance of public and private buildings for safety and security reasons and the increase of comfort of the occupants, the automatic observation of the activities and state of health of persons in retirement homes and hospitals to detect critical situations, for autonomous robots, and interactive environments. The first designated application for the model is in the field of building automation for monitoring systems. By automating such processes, personnel for monotonous observation tasks could be pared down.

For this purpose, buildings need to be equipped with a large number of diverse sensors. The challenge that has to be taken is to merge and interpret the information coming from these various sensory sources. To get a unified perception of what is going on in the environment, information provided by these sensors has to be "bound" within sensory modalities, across diverse modalities, and across space and time. Current solutions are barely capable of handling this task. Therefore, a new information processing principle was introduced called neuro-symbolic information processing. According to this method, sensory data are processed by so-called neuro-symbolic networks to result in "awareness" of what is going on in the environment. The basic processing units of neuro-symbolic networks are neuro-symbols. The inspiration for the utilization of neuro-symbols came from the fact that humans think in terms of symbols, which emerge from information processed by neurons. In the model, neuro-symbols represent perceptual images of different grades of complexity. Neuro-symbols can have so-called properties that specify them in more detail and consider location and timing information of input data. To perform complex tasks, neuro-symbols need to be connected in order to exchange information and to interact. Therefore, neuro-symbols are arranged in a modular hierarchical manner, which was derived from the structural organization of the perceptual system of the human brain. Similar like between neurons in the brain, within neuro-symbolic networks, there exist forward and feedback connections between particular neuro-symbols. Additionally, mechanisms called focus of attention, memory, and knowledge – which are also concepts taken over from neuroscience and neuropsychology – influence the perceptive process in a top-down manner and help to devote processing power to relevant features and to resolve ambiguous sensory information. In the proposed model, correlations between neuro-symbols do not all have to be predefined but can be acquired from examples in a supervised manner. Similar like in the brain, only the lowest-level connections do already have to be fixed at startup, because it is not feasible to "get everything from nothing". Based on these prerequisites, higher-level correlations can then be learned successively.

Fulfillment of Requirements

In section 1.1, a number of characteristics and requirements for human perception were outlined, which were the starting point for model development. Furthermore, it was claimed that the developed design shall allow it to integrate already existing workable video and audio processing methods, and that the model has to be actually realizable in a technical system. How these characteristics and requirements are fulfilled by the introduced model are briefly summarizes in the following:

Diverse Sensory Modalities and Parallel Distributed Information Processing: As outlined in section 1.1, to perceive the environment, the brain needs to integrate a huge amount of information coming from various sources being processed in parallel and in a distributed fashion. The challenge was to develop a technical model with an architecture that allows a similar distributed parallel processing and to combine the separate processing results to one unified perception. In the proposed model, to solve the problem of processing data from various sources, a layered modular hierarchical information processing architecture was suggested which is inspired by the modular hierarchical structure of the perceptual system of the brain. Sensors of different types are grouped to different sensory sub-modalities and modalities and are finally merged to a unified multimodal perception. Information of different modalities or sub-modalities is processed separately and in parallel. A merging always takes place at the next higher layer. By the usage of neuro-symbols, which are activated when the sum of their input signals exceeds a certain threshold, and by structuring them to neuro-symbolic networks, complementary, redundant, and also contradicting and inconclusive data can be handled. In case of contradicting and inconclusive data, influence from knowledge and memory can help to solve ambiguity. For handling events happening concurrently, there were introduced the concepts of parallel symbol representations, of group activity symbol representation, and of focus of attention.

Neural and Symbolic Information Processing: In the human brain, perceptual information from different modalities is processed by interacting neurons. However, humans think in terms of symbols. Therefore, in the model, so-called neuro-symbols were introduced which are structured to neuro-symbolic networks. In a first processing stage, neuro-symbolic information is derived from sensory raw data. In the following steps, neuro-symbolic information of a more and more abstractive level is processed. Higher-level neuro-symbolic information can be interpreted by humans intuitively.

Information Integration across Time and Asynchronous Information Processing: To perceive object, events, and situations in an environment, it is necessary to consider information arriving asynchronously, information arriving within a certain time interval, and to conceive certain successions of events. To handle such data, in section 4.4.4, standard neuro-symbols where extended in a way that they are capable of considering data arriving within a certain time window or certain temporal successions of data.

Learning and Adaptation: To allow flexibility and adaptability to different situations, a system needs to have the ability to learn certain correlations within data. In the model, there can be learned correlations between neuro-symbols representing perceptual images in a supervised learning process. Therefore, a number of examples comprising sensor data being triggered when certain perceptual images occur is used. The learning process is divided into several learning phases, starting with lower levels and continuing with higher hierarchical levels.

Influence from Focus of Attention: In real world environments, it can happen that at a certain moment, more perceptual information is available than can be effectively processed. In particular cases, instead of trying to process all objects simultaneously, processing needs to be limited to a certain area of space at a time by focus of attention. Therefore, a concept was introduced to integrate such a focus of attention into the developed bionic model. This focus limits the spatial area within which information is processed. Interaction takes place at the neuro-symbolic feature symbol level, which is topographic in structure. A method for considering only events happening within a certain time interval is directly implemented into every single neuro-symbol.

Influence from Knowledge: For a perceptual system that can perceive lots of different objects, events, scenarios, and situations, the case can occur that sensor data are ambiguous. This means that very similar sensor patterns can correspond to different object, events, scenarios, and situations. In such a case, knowledge can help to interpret ambiguous sensory signals. In the model, memory symbols and an interface between perception and knowledge were introduced to increase or decrease the activation of neuro-symbols representing perceptual images.

Hybrid System Design: In section 1.1, additionally to the fulfillment of the requirements and characteristics of human perception, it was claimed that the developed model shall be able to integrate existing methods and algorithms for audio and video data processing to reduce the design effort. This is enabled by the modular hierarchical structure of the model which allows it to substitute certain neuro-symbolic modalities by other solutions. The usage of existing workable solutions is also permitted for other sensory modalities or sub-modalities if they provide for a particular case an advantage compared to the suggested neuro-symbolic information processing principle.

Model Feasibility: Besides the pure suggestion of a model for human-like machine perception, it was demanded that this model is also actually technically realizable. The proof of technical feasibility of the proposed model was given in chapter 5 and chapter 6, where the implementation of the model and its simulation results were outlined.

Key Results

The suggested model was implemented and simulated with the modeling language AnyLogic. Test data came from simulated sensory and symbolic values being triggered when certain activities are going on in a virtual building. As illustrated in section 6.1, the model proved to be successful for the specified test environment and it is expected to achieve satisfying results also for larger system configurations and a greater number of perceptual images to be detected. In section 6.2, based on lessons learned from model development, implementation, and simulation, important issues of the bionic model were discussed including the necessity of generalization for learning from examples and the achievement of fault tolerance and conflict resolution by redundancy of sensor data.

Furthermore, besides pure technical performance considerations, the insights gained during model development, implementation, and simulation allowed it to draw certain conclusions about the correctness, incorrectness, or incompleteness of neuroscientific and neuropsychological research findings. The most important results concerning this issue are briefly summarized in the following. For a more detailed description please review section 6.2.

- The first point to mention is that unlike assumed in traditional approaches, the “configuration” of sensory receptors in the body and the lowest levels of information processing turned out to be already of crucial importance for the efficiency and manifoldness of perception.
- A second outcome is that it is very likely that the neural units in the perceptual system of the brain are both memory cells and processing units of perceptual information at the same time. It is presumable that for the rest of the brain, a similar concept applies. It was argued that compared to classical computer architectures like suggested by von Neumann, such structures incorporating both storage and processing of information can work very efficiently as there do not have to be performed explicit memory access and comparative operations.
- One question not yet being conclusively answered from neuroscience and neuropsychology is on what level abstract knowledge about the world interacts with perceptions based on sensory information. Concerning this question, it was shown that besides on the very lowest levels of perception only providing information about simple features derived from sensor data, an interaction is principally conceivable and possible at all levels. However, it turned out that an interaction at the higher levels might be more efficient, because the perceptual information being available at lower levels is often quite crude and can belong to many different higher-level perceptual images with quite different meaning.

- Another point that was discussed was how to combine and supplement different proposed solutions to the binding problem – which is considered as one of the key questions to brain function – to get from distributed parallel sensor data processing to a unified perception of the environment. In lower levels of processing, it was suggested to use combination coding for binding of sensor data whereas in higher levels, a combination of principles inspired from population coding and temporal coding turned out to be suitable. Additionally, top-down mechanisms coming from memory, knowledge, and focus of attention showed their utility in the binding process. A further result that could be derived is that in perception, for solving the binding problem, besides the utilization of data occurring at the same instant of time or within a certain time interval, location information is of crucial importance. Furthermore, it is outlined that the textbook line that there exist two distinct, separate pathways for object recognition and spatial object location in the visual system being directed towards two different brain areas should be reconsidered. The hypothesis of completely distinct streams of information about object type and object location is relatively unlikely, because in this case, location information could not be used for the binding process, and there would arise the problem how to merge the separate information again in later processing stages.

- Last but not least, it was discovered that when using structures being made up of a huge number of particular processing units where information is not only processed bottom-up, but where also feedbacks and top-down interactions are possible, the stability of the system is raised to question. In such a case, signals could theoretically circulate ad infinitum within the different layers and units and not allow the system to ever get to a stable state. This arises the question how the perceptual system of the brain (or the brain as a whole), which without doubt comprises such feedbacks and top-down influences, can ever be stable and lead to a conclusive perception and experience of the world. For this purpose, certain higher-level filter or inhibition mechanisms might be necessary.

7.2 Recommendations and Hints for Future Research

During model development, implementation, and simulation, there were identified a number of topics that are worth to be subject of further research work to round out the model and make it even more efficient and applicable to a broader range of problem domains. This potential of improvement is outlined in the following. It is started with issues being realistic to be achieved and resolved in near future and being rather tasks for engineers. Afterwards, it is succeeded with topics, which – due to their complexity – rather have to be considered as long term goals and will require interdisciplinary efforts from neuroscientists, neuropsychologists, and engineers.

Hardware Realization: Until now, the introduced model has only been tested and verified on the PC by simulating parallel processing. However, the time required for simulation increases with the number of neuro-symbols, which limits the possible expansion of the model. To truly take advantage of the parallel distributed structure of the model, it would be interesting to implement the model into a chip, which would allow it to perform real parallel processing.

Neuro-symbolic Network Toolbox: The proposed model provides a powerful and flexible tool for information processing of sensor data to perceive objects, events, scenarios, and situations in an environment. By making certain adaptations, neuro-symbolic information processing could also be used for other applications. What will be subject to further work is to develop a toolbox for neuro-symbolic networks comparable to the existing toolboxes for neural networks, which will allow fast and comfortable development and testing. With this toolbox, the method of neuro-symbolic information processing shall be made attractive for a broader group of users.

Learning Spatial Configuration of Sensors: As outlined in section 4.4.3, location information is important for binding perceptual data. Location information is derived from the location of triggered sensors. In the introduced model, it was assumed that the positions where individual sensors are mounted as well as their detection ranges are known to the system at initial system startup. This approach seems to be in accordance with the circumstances in the human body and brain. Sensory receptors of the body are connected via nerve cells to the primary cortices of the perceptual system, which are topographic in structure. It is very likely that these connections already exist at birth and are therefore defined by genes. However, for a technical system, the exact assembling and configuration of sensors in the environment require a certain effort. Therefore, it would be desirable to have available a method to learn the positions where sensors are mounted only during operation. This could be performed by having sensors of different modalities with overlapping detection ranges. If an event happens in the environment, diverse sensors covering the same spatial detection range are activated concurrently. From this concurrent activation, there could be drawn conclusions about the position of the sensors.

Emerging a Dominant Modality and Detection of Failures: In section 3.3.2, it was described that there generally exists one sensory modality that dominates over the other modalities, because it delivers generally the most reliable information. Vision has traditionally been considered as the dominant modality. The classical view is that visual dominance is an inherent physiological advantage of visual over other sensory perceptions in the brain. In accordance to this, in the proposed model, it was also predefined what sensor modality shall be dominant. However, an alternative hypothesis suggests that visual dominance is not inherited but emerges during development, because visual information proves to be most accurate and reliable in the majority of cases. Following this suggestion, it would be interesting to find a mechanism to determine only during system operation what sensory modality of a certain sensor configuration delivers the most reliable information. Concerning the reliability of sensory information, a second issue would be of interest. In section 6.2, it was discussed how the system behaves when a sensory modality fails partly or completely. As pointed out, in case of failure, the performance of perception is significantly higher when the system "knows" about the failure of the modality. Therefore, it would be desirable to have available effective mechanisms to detect the malfunction of the modality. For instance, besides the integration of low-level self tests into the systems, there could also be integrated mechanisms that detect when certain neuro-symbols are triggered remarkably often or seldom, which is an indication for sensor malfunction.

Unsupervised Learning within Neuro-symbolic Networks: As outlined in section 4.5, for learning correlations between neuro-symbols, supervised learning methods are used. However, in the human brain, especially in the lower levels, which evolve at early developmental stages, it is very unlikely that there can be made use of supervised learning. Here, neural structures either have to be predefined or the brain needs to offer the possibility to learn correlations between data in an unsupervised manner. For learning, data of the same sensory sources changing over time as well as data from different sensory sources being assigned to the same spatial position in the environment might be the key to extract such correlations. Identifying and emulating these mechanisms of unsupervised learning, which might be the same through all hierarchical levels, could allow it to make a great leap forward in understanding the brain as well as in the field of cognitive science and artificial intelligence and would bring us closer to the aim of developing technical systems with real cognitive capabilities.

Knowledge Acquisition: In section 3.3, it was pointed out that knowledge and memory of what happened in the past is very important for perception. The complexity of bottom-up sensor data processing as well as its error-proneness can be drastically reduced by using knowledge. In the model, until know, knowledge and its influence on neuro-symbol activations have been represented by explicit predefined rules. This approach is obvious, because abstract knowledge – especially semantic knowledge – can easily be described in an encyclopedic form of third-person information. Nevertheless, it would be interesting and desirable to extract a certain amount of these rules in a learning process and to represent knowledge not in a rule-based form but to find another suitable representation being inspired from how knowledge is "coded" in the brain. However, brain research does not yet provide detailed information about these issues. For acquiring knowledge, there might have to exist reinforcement mechanisms or concepts like emotions. Coding of knowledge might follow similar principles as discovered for the perceptual system of the brain, where units seem to be memory elements and processing elements at the same time. Cooperation between neuropsychologist and engineers could lead to new insights not being achievable by these two disciplines when researching independently from each other.

Steering the Beam of Attention: Section 3.4 described how focus of attention can help to bind information when different events happen in parallel. In section 4.4.5, it was proposed to constraint the spatial area within which information is processed by a focus of attention. The interaction between focus of attention and neuro-symbols was suggested to take place on the feature symbol level. However, neuroscience and neuropsychology do not yet provide a decisive answer by which mechanisms the beam of attention is steered and directed to certain areas. Therefore, this part was left out in model development by defining that the steering of the focus of attention is performed by an external source. Nevertheless, this issue is a topic that is recommendable for future research work. Where focus of attention is directed to is for sure influenced to a certain extend by perception itself. However, there are most certainly also involved mental processes not being part of the perceptual system of the brain. Emotions, knowledge, expectation, and even action might play a role.

Experience of Time: In the model, within neuro-symbols, mechanisms were introduced to allow it to consider input signals arriving within a time window and certain successions of events. It was outlined how time windows and successions of events could be modeled by neural structures including feedbacks. Such circuits are conceivable to exist at the lower cortical levels of the brain. However, at the highest hierarchical levels responsible for experiencing long time periods, such mechanisms might no longer be applicable. It would be an interesting topic to find out how such a time experience can emerge from neural or neuro-symbolic structures and whether the usage of such mechanisms could bring advantages also for technical systems.

Neural Code: As outlined in section 3.1, neurons exchange information by spike trains. However, the neural code by which this information is transmitted is still not well understood. In chapter 6, it was described that structures including not only bottom-up information flow but also feedbacks and top-down directed connections can be subject to "signal circulations". It is most obvious that the brain is a system to which this description applies. Under these considerations, it would be conceivable that the transmission of information in neurons via spike trains is no general characteristic of neural tissue but a result of these circulating signals. However, therefore, feedbacks would have to be effective down to the lowest layers of neural processing. Nevertheless, in any case, circulating signals could have an influence on the characteristics of these spike trains. It would be interesting to investigate and analyze the characteristics of spike trains considering these circumstances.

Stability, Epilepsy, and Consciousness: As outlined in section 6.2, signal circulations caused by feedbacks and top-down mechanisms can challenge the stability of systems. In the brain, which answers to this description, stability can generally be achieved. However, stability is not self-evident as can be seen from patients with epileptic fits. The World Health Organization estimates that there exist 40 to 50 million people with epilepsy throughout the world. It seems quite plausible that such epileptic seizures are the visible effect of stability problems within the brain. The tempting question that results from this is what mechanisms are responsible for keeping the brain stable. In the suggested bionic model, undesired multiple signal circulations were avoided by integrating certain filter mechanisms into the knowledge module, which is the topmost module of the introduced perceptual model . Drawing conclusion from the bionic model to its biological archetype, which is of course a critical step, this would mean that the highest cortical levels of the brain are responsible for guaranteeing stability by filtering and suppressing lower-level neural activations. Consciousness, which allows us to get a unified stable experience of ourselves and our world [ST02], might be what remains after these filtering and suppressing mechanisms. Proving or negating this theorem is a further interesting issue recommended for future research. In any case, when emulating the information processing structure of the brain, further analyses of stability need to be performed.

7.3 Intelligent Machine, quo vadis?

"Within a generation ... the problem of creating 'artificial intelligence' will substantially be solved." This is the famous declaration of Marvin Minsky made in 1967. About 30 years later, in 1996, a young researcher named Push Singh, who happened to be working under Minsky, published a paper called *"Why AI Failed"* [Nut08]. What can be learned from this wrong forecast is that making predictions in artificial intelligence (AI) and related research fields is far from easy. Therefore, in the following, there shall rather be made a recommendation what way should be pursued in future instead of making prognoses.

First of all, to avoid mistakes already made in the past, it is important to understand the history that this research field went through until now: About 50 years ago, in the fifties, with the advent of the computer, the research field of artificial intelligence emerged with the aim to build machines being equally intelligent as humans are. The first days of AI were marked by strong optimism, and it was believed that computers would soon think, reason, and behave in a similar manner as humans do. However, at the end of the sixties, it got clear that making computers think – even on a childlike level – is an extremely complex problem. Therefore, researchers started to focus on by far less complex problems like planning algorithms, pattern recognition, expert systems, reacting to situations by using certain rules, etc., each of them being dedicated and applicable only to very narrow and specific problem domains. Until today, there does not exist any technical system that can even nearly compete with the capacity and the capabilities of the human mind. Within the last years, it had to be admitted that such reduced approaches often focused on in current AI projects can never lead to technical systems with skills and capabilities comparable to humans' mental abilities. Therefore, like at the beginning of artificial intelligence research, again, findings about how natural intelligence works have to be the basis for developing concepts for technical approaches trying to achieve intelligence.

In this thesis, a neuroscientifically and neuropsychologically inspired approach for human-like machine perception was presented based on insights about the internal structural organizations and information processing principles of the perceptual system of the human brain. Nevertheless, in future applications, it will not only be desirable to make machines perceive their environment in a human-like fashion, but to also to let them react on perceived situations in a more human-like way, which means to analyze the situations and to choose the most advantageous actions from a range of possibilities during a decision making process. Like taking the perceptual system of the human brain as archetype for modeling perception, concerning the action control system, it is recommendable to take the mechanisms of the human brain responsible for action generation as archetype for model development. For advanced systems incorporating perception as well as action, it has to be considered that taken actions cause changes in the environment and therefore again have influence on perception. For this reason, there have to exist connections between the system of perception and the system for decision making and action control.

Emulating the structural organization and the mechanisms and principles of information processing of the perceptual system and the action system of the human brain will most certainly lead to more effective and efficient autonomous technical systems. However, it is doubtable that this alone will suffice to achieve real human-like capabilities and behavior. As outlined in section 3.2, state of the art neuropsychological research findings about the brain describe it to consist of three principal functional units: (1.) the unit for receiving, analyzing, and storing information arriving from the outside world, which is responsible for perception, (2.) the unit of programming, regulation, and verification of mental activity, which is responsible for action, (3.) the unit for regulating tone and waking of mental states, which is closely connected to metabolic processes. These three functional units do not work in isolation. Complex forms of mental activity take place through the combined working of all three brain units. According to Aleksandr Luria [Lur73], one of the leading neuropsychologists of his time whose research findings are still considered as being state of the art, *"an insight into the nature of the cerebral mechanisms of mental activity can only be obtained by studying the interaction of these three units"*. Therefore, for achieving machines really performing, perceiving, and behaving in a human-like manner, a model of all three principal functional units as well as a definition of their interaction will be necessary. Additionally, it has to be noticed that the human brain and mind and the body are tightly interwoven and one influences the development and abilities of the other. Perception requires information from sensors located in the body. Actions need the human body to be carried out. Information about internal body states being processed in the unit for regulating tone and waking of mental states influences perception as well as action. Therefore, it might not be sufficient to emulate the information processing principles of the brain, but also its information exchange within the body might have to be considered.

To develop such models of the brain, today's knowledge of brain research will not suffice. A deeper understanding about the function principles and mechanisms taking place in the brain needs to be gained. The key to future brain research and AI research is most certainly interdisciplinarity. Different disciplines concerned with the investigation of the brain, which today provide diverse and even contradicting theories, will have to join forces and work out a unitary, unambiguous model. Engineers trying to transform these research findings into technically feasible systems will make very important contributions, because by this process, lots of weak points, contradictions, blind spots, and errors of existing brain theories will be discovered. The incorporation of different research disciplines might lead to synergy effects allowing it to make a leap forward in the understanding of the human brain and mind. However, as long as this level is not reached, and AI continues focusing on pure technical and algorithmic solutions, we will still have to admit the famous sentence of the Austrian computer pioneer Prof. Heinz Zemanek: *"If the computer is intelligent, then I am something different."*

Appendices

Appendix A

Low-Level Information Processing of Data from Tactile Floor Sensors

In the section 4.2.6, there were pointed out characteristics of neuro-symbolic information processing in the lower hierarchical layers from sensor data up to the sub-unimodal layer. These characteristics are principally the same for all used sensory modalities. To make the information processing concepts even better understandable, in this appendix, they will be outlined in detail for one sensory modality. For the explanation, the modality of tactile floor sensors is used. The result of information processing from sensor data up to the sub-unimodal layer shall be the detection of objects of different sizes at different positions, being static or moving with different velocities and in different directions. Section A.1 outlines the used sensor configuration. Section A.2 illustrates how to get from sensor data to feature symbols, and section A.3 shows the transition to the sub-unimodal level.

A.1 Sensor Configuration

As already outlined, for explaining characteristics of low-level information processing, the modality of tactile floor sensors is used. Therefore, a room is equipped with a number of quadratic tactile floor sensors covering the whole floor (see figure A.1). Each floor sensor provides a binary signal (zero or one) depending on whether it is activated by an object or not.

A.2 Deriving Feature Symbols from Sensor Data

From the sensor data provided by the tactile floor sensors, different information can be derived. Among others, it can be determined where an object is present and what size, shape, and orientation the base of this object has. It can also be identified if the object stands still or moves. If an object is moving, it activates different floor sensors over time. In case of movement, the location and direction of the movement and its velocity can be calculated. Such information derived from sensor data is represented by feature symbols. The feature symbol level comprises a certain number of different feature symbol layers. Different feature symbol types are responsible for detecting different features derived from sensory raw data or lower-level feature symbols. In correspondence to their biological archetype, there are used different feature symbols to represent

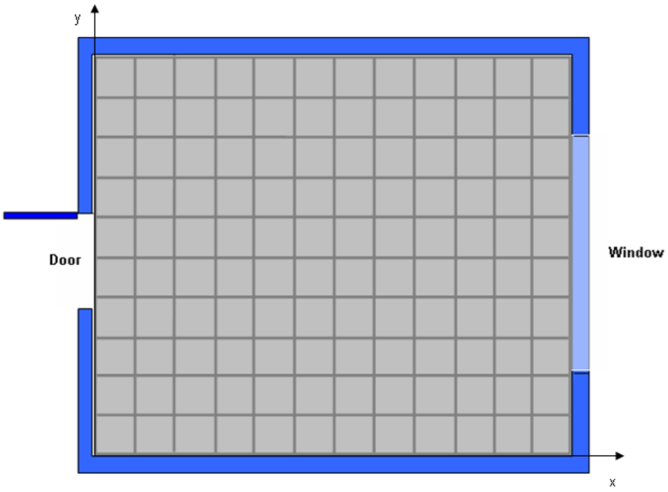


Figure A.1: Room Equipped with Tactile Floor Sensors

objects of different sizes and shapes and for moving objects also different directions and velocities of movement. There can exist feature symbols reacting to small round non-moving object, others reacting to small square object, again other ones reacting to medium-sized quadratic objects, and so on. Again, different types of feature symbols are triggered for moving objects of a certain velocity and direction of movement. One feature symbol of a certain type only reacts to a specific feature at a certain position. Another feature symbol of the same type reacts if the same feature is detected at another position. To illustrate the principle how to get from sensor data to feature symbols, in the following, it shall be described how to perceive non-moving and moving objects that trigger tactile floor sensors. The objects have different sizes and can move in different directions with different velocities.

A.2.1 Detecting Non-moving Objects

In the following, it is described how feature symbols can be extracted from sensory raw data correlating to the size of non-moving objects standing on the floor. For detecting objects of different sizes, different feature symbol types are necessary. For the examples given in the following, there are always used objects with a cylindrical form, which stand on the circular base. However, the configuration could be extended to perceive also objects of other forms by introducing additional feature symbol layers.

Concerning the feature symbols for detecting objects of different sizes, there is made a distinction between three different classes of objects. This distinction is made according to the maximal number of tactile floor sensors that are activated by an object either in x- or y-direction. Each class is represented by a certain type of feature symbol:

- **Small Non-moving Object:** Objects belonging to this class trigger one or two sensors in x- or y-direction (see figure A.2). Objects are depicted as red circles. Activated sensors are always colored pink. The diameter of objects is smaller than the length of two tactile floor sensors.

- **Medium-sized Non-moving Object:** Objects of this class activate three to five sensors in x- or y-direction. Examples therefore are given in figure A.3. The diameter of these objects is bigger than the length of one tactile floor sensor and smaller than the length of five tactile floor sensors.
- **Big Non-moving Object:** Objects of this class are objects that activate five to nine floor sensors in x- or y-direction. The diameter of objects is bigger than the length of three tactile floor sensors and smaller than the length of nine tactile floor sensors.

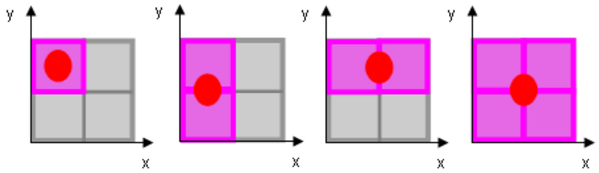


Figure A.2: Sensors Triggered by Small Objects

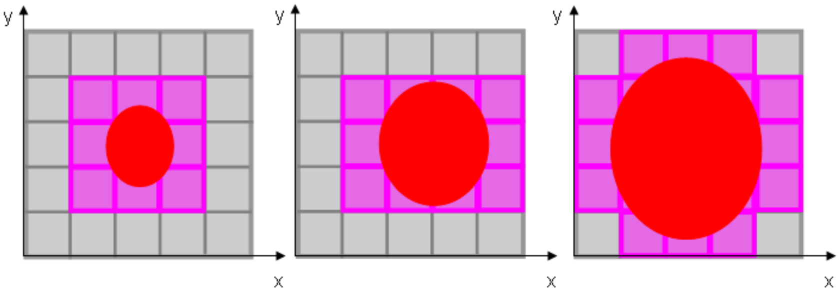


Figure A.3: Sensors Triggered by Medium-sized Objects

The distinction of these three sizes is exemplary to illustrate the underlying concept. It can be extended easily to detect bigger objects. A particularity concerning the dimension of the objects is that the three defined classes have no sharp borders. For categorizing an object into one of the three classes, not the actual size of the object is of importance but the maximum number of sensors it triggers in x- or y-direction. Therefore, not only the size but also the position of the object influences the categorization. This results from the limited spatial resolution of the sensors.

To detect objects of three different sizes, three layers of feature symbols are needed, which are structured one above the other (see figure A.4). The enumeration of the feature symbol levels with odd numbers (layer 1, 3, and 5) will get clear when adding layers to detect moving objects, which will be labeled with even numbers (see section A.2.2). In the first layer, information coming from the sensors is processed. This layer is responsible for detecting small objects. The third layer receives information from the first layer and detects whether middle-sized objects are present. The fifth layer receives information from the third layer and detects big objects. The number of layers could be increased to detect even bigger objects. However, to explain the underlying principle, three layers are sufficient.

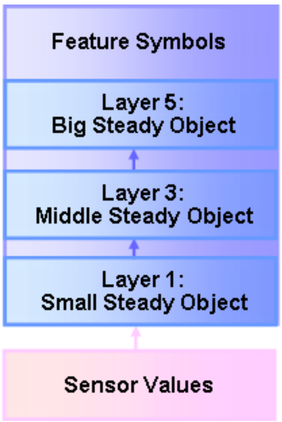


Figure A.4: Feature Symbol Layers for Detecting Steady Objects of Different Sizes

Each of the three layers is further divided into four sub-layers. Figure A.5 illustrates this fact for layer 1. The principle for detecting steady objects of a certain size will now be explained beginning with the description of layer 1, which is responsible for detecting small, non-moving objects.

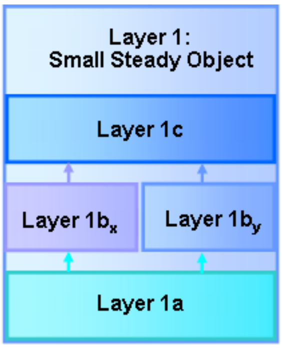


Figure A.5: Sub-layers of Layer 1

Detecting Small Objects

To detect where in the room a small, non-moving object is present, layer 1 is divided into four sub-layers labeled as *layer 1a*, *layer 1b$_x$*, *layer 1b$_y$*, and *layer 1c*. The layer 1a receives information directly from the sensors. The layers 1b$_x$ and 1b$_y$ receive information from layer 1a. Layer 1c receives information from the layers 1b$_x$ and 1b$_y$.

As mentioned in section 4.2.6, feature symbols are topographic in structure. Therefore, for each perceptual image (small object, medium-sized object, big object) there exist a lot of equal parallel feature symbols differing only in their position. This fact is best illustrated graphically. Figure A.6 shows the dependency between tactile floor sensors and the corresponding feature symbols of layer 1. Tactile floor sensors are depicted as grey cubes. Feature symbols of different sub-layers are illustrated as cubes of different colors. In the picture, the pink squares indicate where an object activates sensors and therefore also feature symbols. The picture makes clear that it depends on the location of the object what feature symbol of a certain level is activated.

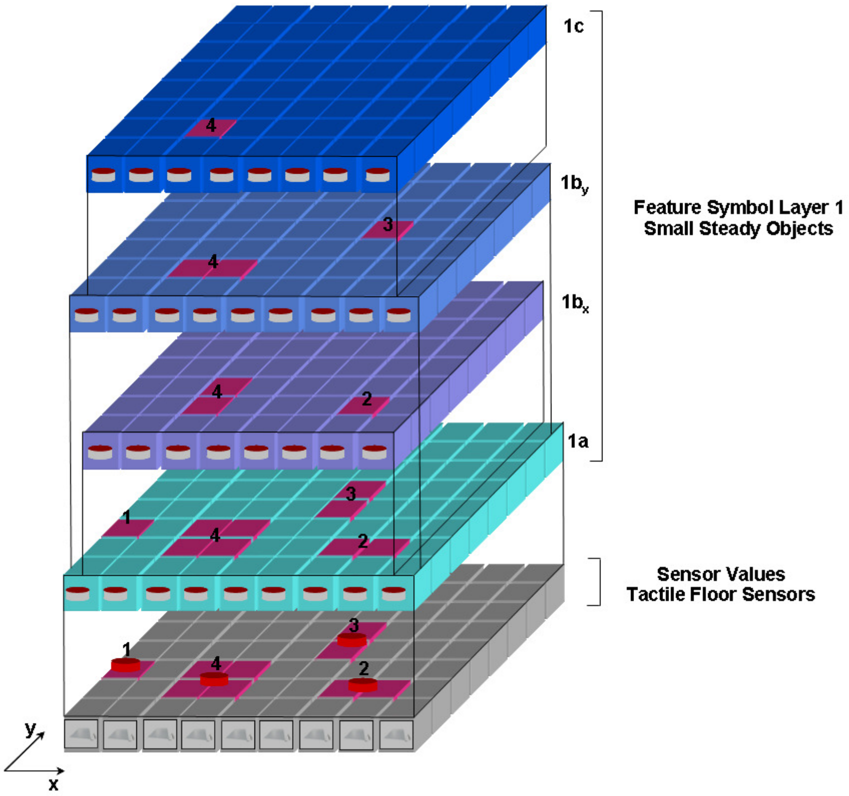
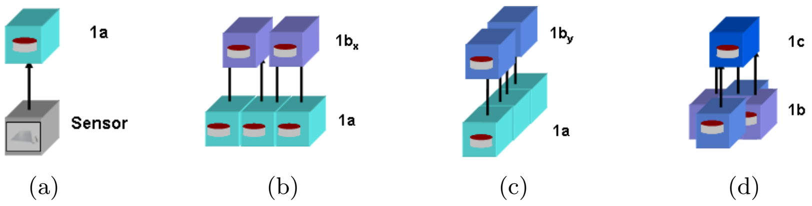


Figure A.6: Topographic, Hierarchical Arrangement of Feature Symbols in Layer 1

Connections between the sensor level and the feature symbol level 1a are one to one connections. Each sensor is assigned to one single feature symbol of this level (see figure A.7a). In the case that an object triggers only one single tactile floor sensor, only one feature symbol in layer 1a is activated (see figure A.6, object 1).



(a) (b) (c) (d)

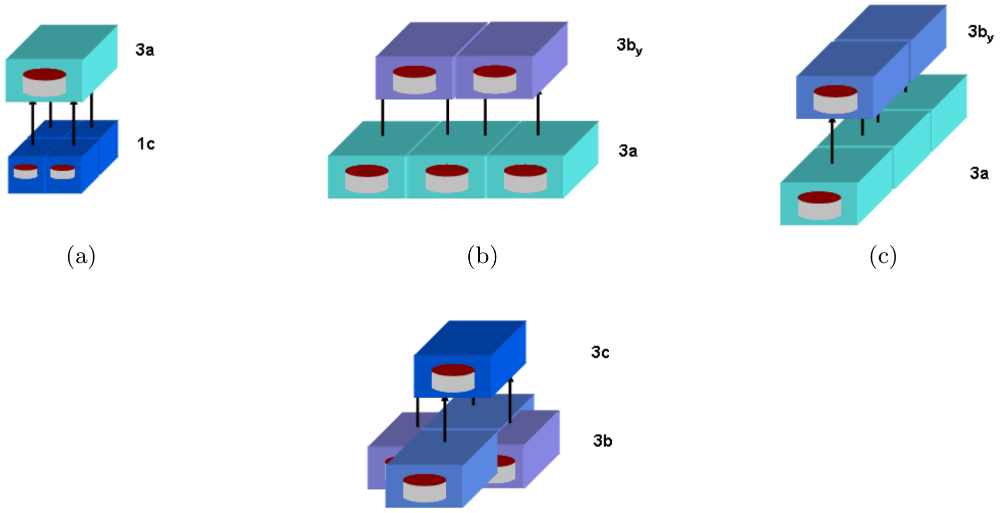
Figure A.7: Correlations between Sensors and Feature Symbols of Different Sub-levels of Layer 1

The situation gets a little bit more complicated if the object activates two tactile floor sensors. As can be seen from figure A.6 (object 2 and 3), in this case, two feature symbols of layer 1a are activated although there is only one object present. To resolve this problem, the sub-layers $1b_x$ and $1b_y$ are used. In comparison to layer 1a, layer $1b_x$ is shifted in x-direction by half the length of one tactile floor sensor. A feature symbol of the level $1b_x$ receives information from always two feature symbols of the former layer 1a (see figure A.7b). In this layer, a feature symbol is activated when the two feature symbols in layer 1a it receives information from are active. By using a mechanism considering signal runtimes and processing times, only the activation of this higher level feature symbol is processed further in later processing stages (see section A.3). Similar to layer $1b_x$, there has to exist a layer $1b_y$, which is shifted by half the length of a tactile floor sensor in y-direction. Like layer $1b_x$, layer $1b_y$ processes information coming from always two feature symbols of layer 1a (see figure A.7c). A feature symbol of this layer is activated if the two symbols of layer 1a it receives information from are active.

To cover the case that one object activates four tactile floor sensors (see figure A.6, object 4), again another layer has to be added. In this layer, which is labeled as 1c, each feature symbol receives information from always two feature symbols of layer $1b_x$ and $1b_y$ (see figure A.7d). A feature symbol is activated if all four feature symbols from layer $1b_x$ and $1b_y$ it receives information from are active. Again, mechanisms considering signal runtimes and processing times are responsible for resolving conflicts caused by concurrent symbol activations in different sub-layers corresponding to the same object (see section A.3).

Detecting Medium-sized Objects

Layer 3 of figure A.4 is responsible for detecting medium-sized objects. Similar like in layer 1, layer 3 consists of four sub-layers one built above the other. Except for layer 3a, which has the special property to allow semi-activation, the sub-layers function equally as in layer 1. The concept of semi-activation is explained further below. Each symbol of layer 3a receives information from four feature symbols of layer 1c (see figure A.8a). Symbols of layer $3b_x$ and $3b_y$ receive information from always two features symbols of layer 3a (see figures A.8b and A.8c). Each symbol of layer 3c receives information from always two symbols of layer $3b_x$ and $3b_y$ (see figure A.8d).



(a) (b) (c)

(d)

Figure A.8: Correlations between Feature Symbols of Layer 1c and Sub-levels of Layer 3

To activate a symbol of layer 3a, at least two of the four feature symbols of layer 1c it receives information from have to be active. For the activation of a feature symbol of layer $3b_x$ and $3b_y$, both lower-layer symbols they are connected to have to be active or semi-active (see below). A symbol of layer 3c is activated if all four corresponding symbols of the layers $3b_x$ and $3b_y$ are active. Similar like in layer 1, mechanisms exploiting runtimes and processing times avoid the activation of multiple symbols by one single object in the sub-unimodal layer (see section A.3).

In figure A.9, the arrangement of feature symbols up to the level 3c is given. The figure also shows the symbol activations for a medium-sized object for a concrete example.

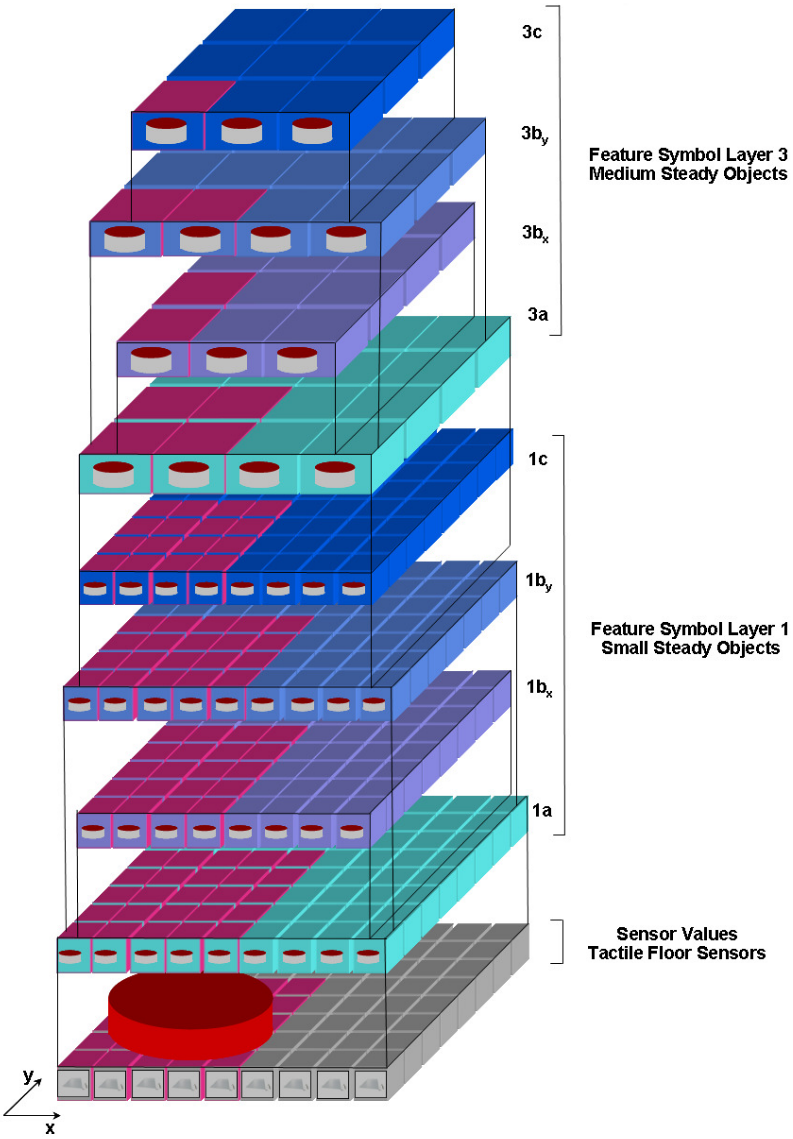


Figure A.9: Hierarchical Arrangement of Feature Symbols in Layer 1 and 3

In contrast to layer 1a, layer 3a (and also layer 5a) have the possibility of semi-activation of symbols. Unlike in layer 1a where each feature symbol receives information from always only one sensor, in layer 3a and 5a, a feature symbol is connected to always four neuro-symbols of layer 1c and 3c, respectively. In the following, the principle of semi-activation is explained for layer 3 (see figure A.10). For layer 5, the same principle applies. To activate a symbol of layer 3a, at least two of four symbols it receives information from have to be active. For the case that only one symbol of layer 1c is active, the symbol of layer 3a is semi-activated (indicated as yellow square). Semi-activation of a symbol means that it can activate a symbol of a higher layer when this higher level symbol additionally receives information from at least one other active or semi-active lower-level symbol. An example for activating a symbol of layer $3b_x$ by semi-activated symbols of layer 3a is illustrated in figure A.10.

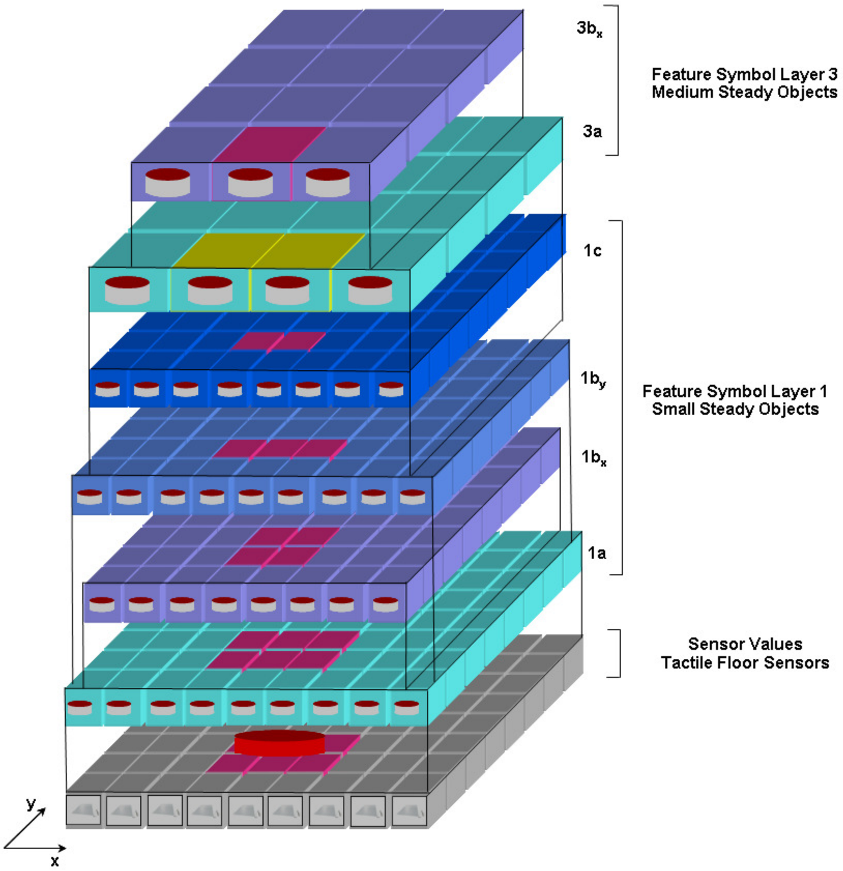


Figure A.10: Principle of Semi-activation in Layer 3a

Detecting Big Objects

The detection of big objects is performed by layer 5. Similar to layer 3 and layer 1, this layer consists of four sub-layers, which follow the same principles as outlined for layer 3. Figure A.11 illustrates the sum of feature layers and sub-layers necessary for detecting small, middle-sized, and big steady objects.

As can be seen from figure A.11, due to the shift in x- and y-direction, for bigger objects, there do not exist feature symbols at the borders. However, as for realistic applications this area is small in comparison to the whole spatial area equipped with floor sensors, this fact can generally be neglected.

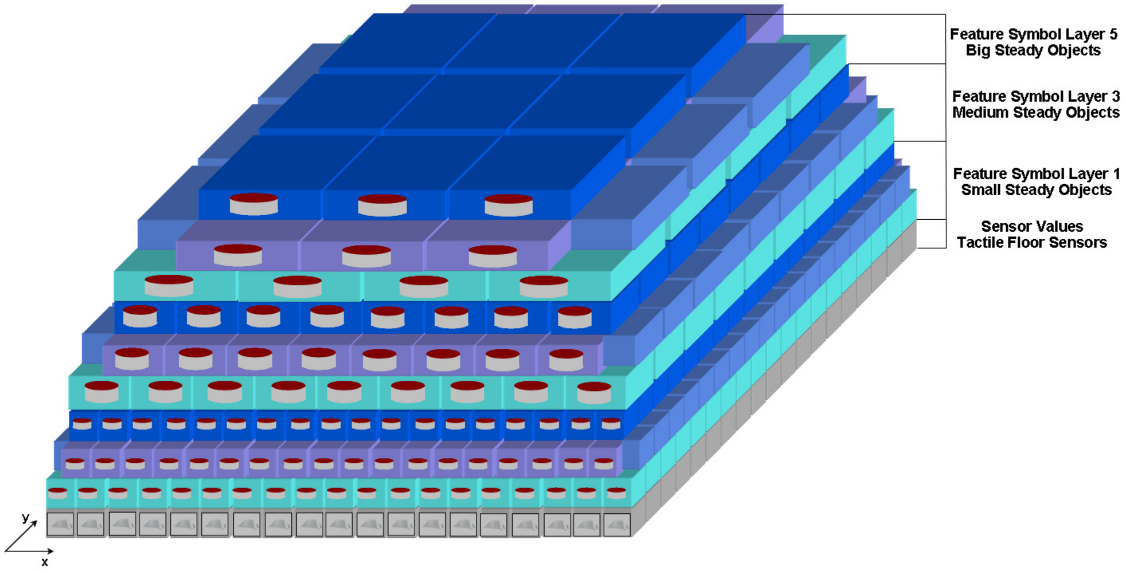


Figure A.11: Feature Symbol Layers for Detecting Steady Objects

A.2.2 Detecting Moving Objects

As mentioned at the beginning of this chapter, there shall not only be detected steady objects from sensor data but also moving objects. Until now, it is not known how motion of objects is perceived in the primary visual cortex of the brain [RP05]. In the following, a concept is introduced how moving objects (moving in different directions and with different velocities) can be detected by feature symbols. Therefore, additional layers have to be added to the feature symbol level of the tactile floor sensor sub-modality (see figure A.12). These layers are referred to as layer 2, 4, and 6 and are responsible for detecting the movement of small, medium-sized, and big objects. Similar to the layers 1, 3, and 5, they consist of a certain number of sub-layers. These sub-layers are responsible for detecting objects moving in different directions. The sub-layers are further divided to detect objects moving with different velocities. However, unlike in the layers 1, 3, and 5, these sub-layers always process information from the layers 1, 3, or 5 and do not use information coming from their own sub-layers.

Detecting Small Moving Objects

In the following, it is described how small moving objects can be detected. As an object can move in different directions, different feature symbols are needed to cover the different cases. There are distinguished the cases of movement in positive and negative x- and y-directions as well as diagonal movements (see figure A.13).

For detecting objects moving in these eight directions, eight parallel sub-layers are needed, each of them being responsible for one particular direction (see figure A.14). According to the direction of movement they can detect, the layers are labeled as $2x_{+}$, $2x_{-}$, $2y_{+}$, $2y_{-}$, $2x_{+}y_{+}$, $2x_{-}y_{-}$, $2x_{-}y_{+}$, and $2x_{+}y_{-}$.

The principle for detecting objects moving in a certain direction shall be explained by means of the direction x_{+}. Detecting objects moving in other directions works analogously. When detecting

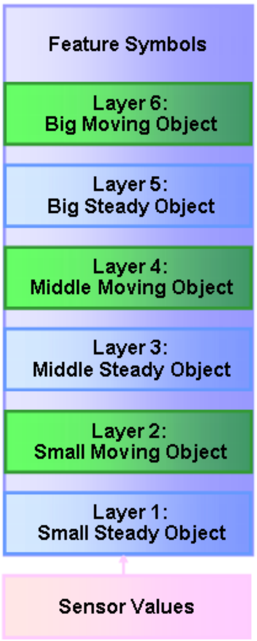


Figure A.12: Feature Symbol Layers for Detecting Steady and Moving Objects of Different Sizes

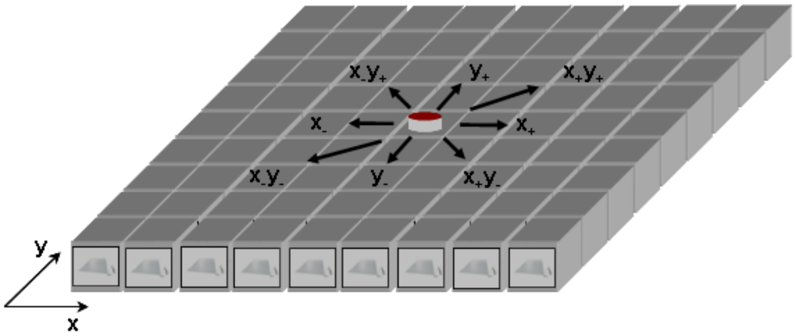


Figure A.13: Possible Directions of Movement

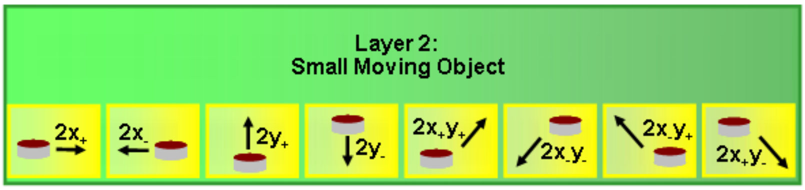


Figure A.14: Parallel Layers for Detecting Movement in Different Directions

a moving object, additionally to its direction, there also has to be considered its velocity. For different velocity ranges, different feature symbols are needed. To outline the used concept, three different velocities are considered labeled as fast (v_1), middle (v_2) and slow (v_3). For each of these velocities, a two-layered feature symbol level is needed processing information from the layers $1b_x$ or 1c. Therefore, these layers are labeled as $2x_{+}v_{1}b_{x}$, $2x_{+}v_{1}c$, $2x_{+}v_{2}b_{x}$, $2x_{+}v_{2}c$, $2x_{+}v_{3}b_{x}$, and $2x_{+}v_{3}c$ (see figure A.15).

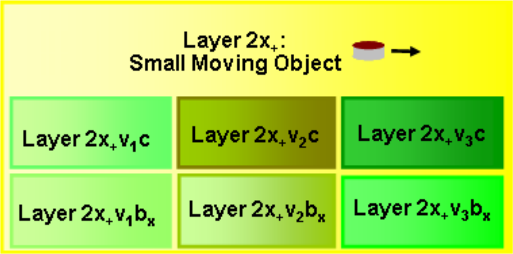


Figure A.15: Sub-layers for Detecting Movements of Different Velocities in Layer $2x_+$

Each symbol of these layers receives information from two feature symbols of the layer $1b_x$ and 1c, respectively (see figures A.16a to f). The signal coming from the left feature symbol is temporally delayed via a delay element with a certain delay time (T_1, T_2, and T_3). By this measure, movements of objects within a certain velocity range can be perceived. The delay element guarantees that – for a certain velocity range – the signal of the left lower-level feature symbol, which is activated before the right symbol during object movement in x_+-direction, arrives not until the right symbol is active. That way, the activation threshold value is exceeded and the higher-level neuro-symbol is activated.

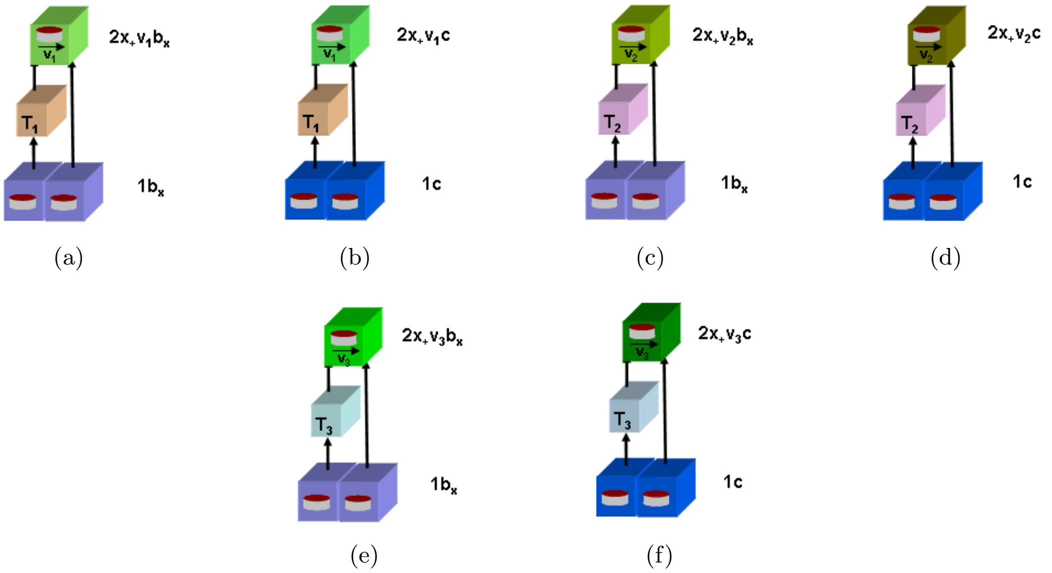


Figure A.16: Correlations between Feature Symbols of Layer 1 and Layer 2 for Detecting Movement in Direction x_+

In figure A.17, an example shows what neuro-symbols of different layers are activated when an object moves in direction x_+ with a velocity of about v_1. Again, mechanism considering signal runtimes are responsible for resolving conflicts caused by concurrent symbol activations in different sub-layers (see section A.3).

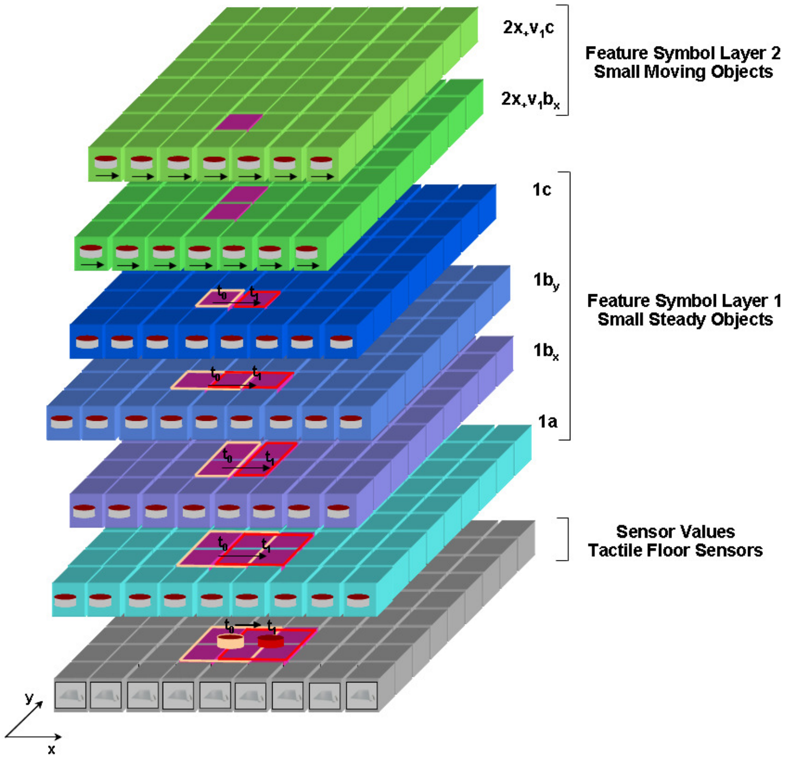


Figure A.17: Hierarchical Arrangement of Feature Symbols in Layer 1 and $2x_+v_1$

Similar to layer $2x_+$, the layers $2x_-$, $2y_+$, and $2y_-$ are also divided into sub-layers receiving information from the layers 1b and 1c, respectively (see figures A.18a to c).

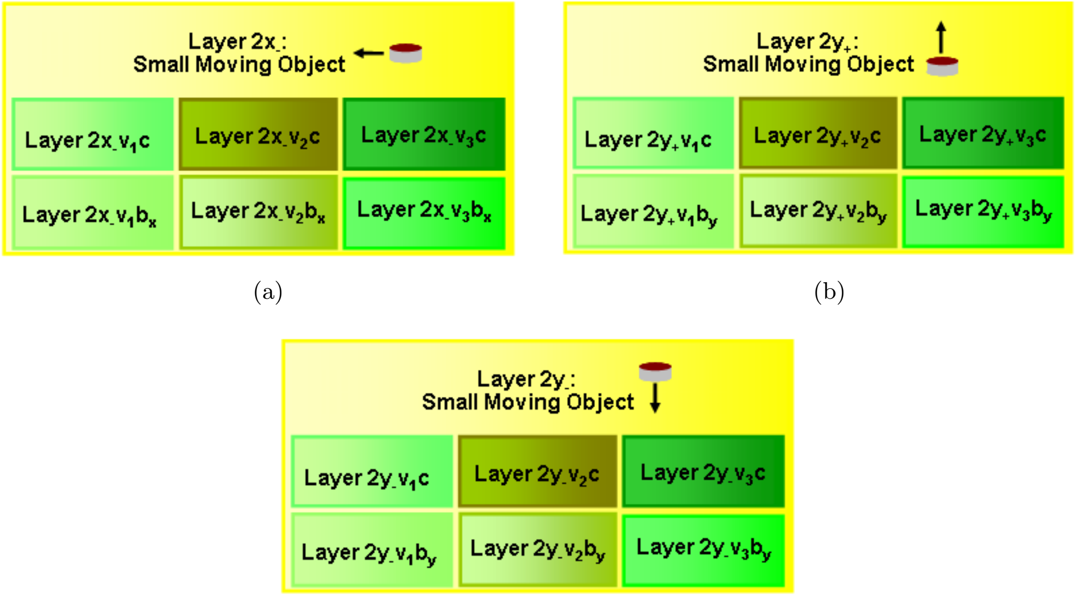
The figures A.19a to f illustrate the connections between feature symbols for detecting movement with a velocity of about v_1.

For detecting movement in diagonal directions, which is performed by the layers $2x_+y_+$, $2x_-y_-$, $2x_-y_+$, and $2x_+y_-$, the arrangement of layers changes slightly. These layers only receive information from layer 1c (see figures A.20a to d and figures A.21a to d).

The figures A.21a to d illustrate the connections between feature symbols for detecting movement in diagonal directions with a velocity of about v_1. What has to be considered for diagonal movements is the fact that the moving objects must have a certain minimum diameter, which is equal to the length l_{tfs} of one tactile floor sensor multiplied by the factor $\sqrt{2}/2$. For objects with a diameter smaller than this size, there will be alternately detected a movement in x- and y-direction (see figure A.22).

Detecting Medium-sized and Big Moving Objects

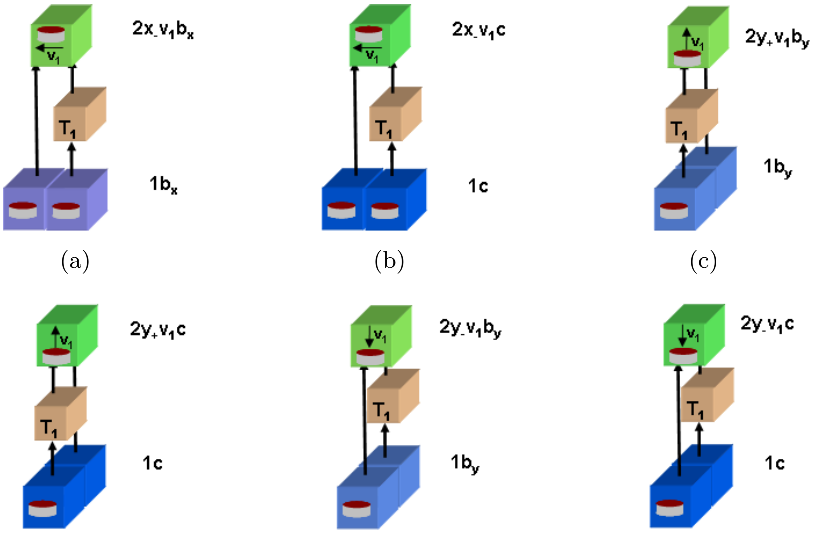
For detecting movement of medium-sized and big objects, the same principles as just described for layer 2 are applied to the layers 4 and 6. Layer 4 receives information from the layers 3b and 3c, layer 6 from the layers 5b and 5c.



(a) (b)

(c)

Figure A.18: Sub-layers for Detecting Movement of Different Velocities in the Layers $2x_-$, $2y_+$, and $2y_-$



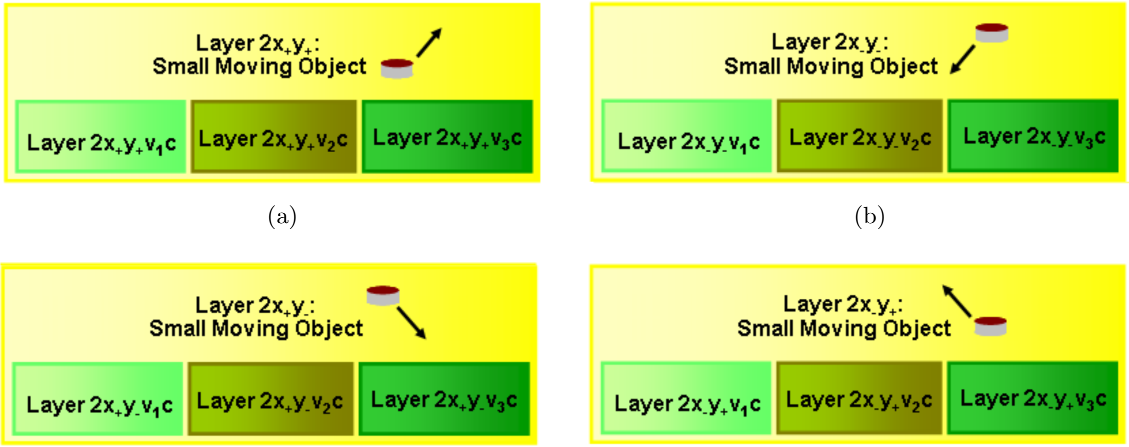
(a) (b) (c)

(d) (e) (f)

Figure A.19: Correlations between Feature Symbols of Layer 1 and Layer 2 for Detecting Movement in Direction x_-, y_+, and y_- for the Velocity v_1

A.3 Deriving Sub-unimodal Symbols from Feature Symbols

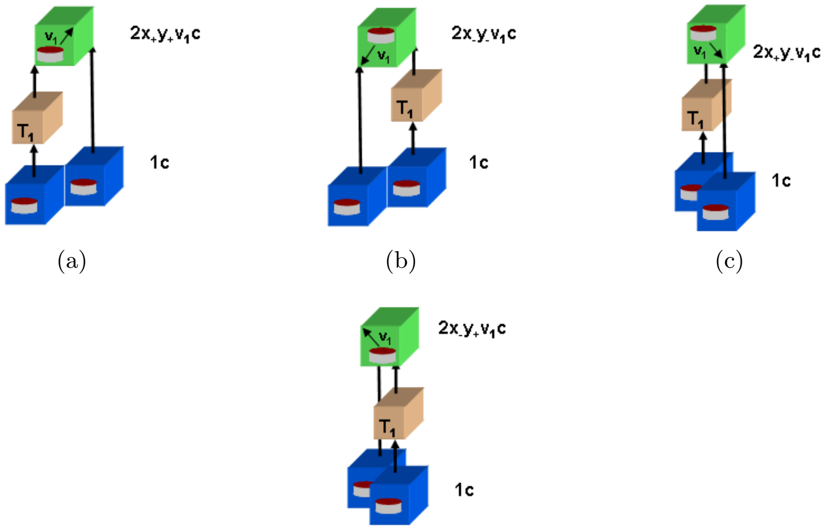
In section A.2, it was described how to get feature symbols out of sensor data coming from tactile floor sensors. This section explains how to get from feature symbols having a strong location dependency to sub-unimodal symbols, which contain location information only as a



(a) (b)

(c) (d)

Figure A.20: Sub-layers for Detecting Movements of Different Velocities in Layer 2x+y+, 2x-y-, 2x+y-, and 2x-y+



(a) (b) (c)

(d)

Figure A.21: Correlations between Feature Symbols of Layer 1 and Layer 2 for Detecting Movement in Direction x+y+, x-y-, x+y-, and x-y+ for the Velocity v_1

property. The transition from the feature level to the sub-unimodal level is the place where location dependent information is transformed into information, which is mostly independent of the location it originates from and only comprises the location information as a property. In the following, it is explained how the transition between location dependent feature symbols and location independent sub-unimodal symbols looks like for the modality of tactile floor sensors. In the used example, on the sub-unimodal level, there exist the neuro-symbols “small steady object”, “middle steady object”, “big steady object”, “small moving object”, “middle moving object”, and “big moving object”. The direction and velocity of movement of objects are coded

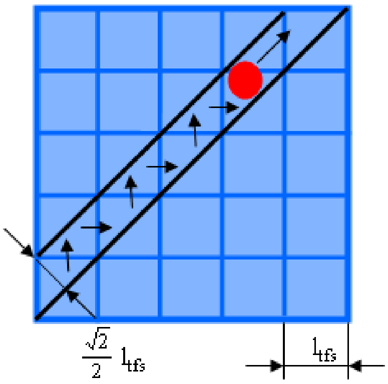


Figure A.22: Minimum Size of Objects Moving along a Diagonal

as properties of the corresponding sub-unimodal symbols.

As outlined in section A.2, to detect static and moving objects of different sizes, a number of feature symbol levels and sub-levels are necessary arranged in a hierarchical manner. E.g., by a small object, which triggers a number of tactile floor sensors, neuro-symbols in different sub-layers can be activated. Therefore, there has to exist a mechanism that avoids it to erroneously perceive more than one object if only one object is present. This problem can be handled by exploiting signal delays caused by signal runtimes and signal processing times. The used principle is illustrated in the figures A.23a to d. It is assumed that the time for processing within a feature sub-layer takes longer than the transmission of signals to the sub-unimodal layer.

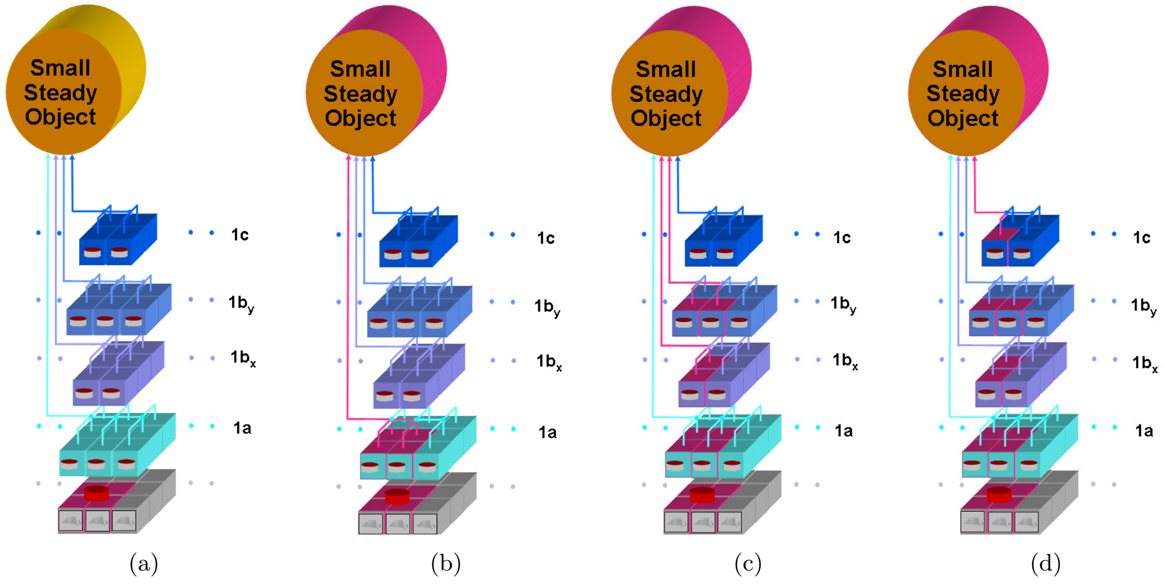


Figure A.23: Transition from Feature Level to Sub-unimodal Level

Figure A.23a shows the case that four tactile floor sensors are activated by one small object. This results in an activation of four feature symbols of layer 1a. These four activated feature symbols send a message to the sub-unimodal symbol "small steady object", which is activated subsequently (see figure A.23b). By using different signal runtimes for neighboring connections, it can be achieved that the four messages do not reach the sub-unimodal symbol concurrently but one after the other. Concerning the value of the location property of the sub-unimodal symbol, the location value of the feature symbol arriving last overwrites prior location values. Parallel to the sub-unimodal symbol “small steady object”, the information about symbol activations is also transmitted to the feature symbol layers $1b_x$ and $1b_y$. After the information is processed, this results in an activation of neuro-symbols of these layers. Again, neuro-symbols of layer $1b_x$ and $1b_y$ transmit information to the sub-unimodal layer (see figure A.23c). By this information, the location value of the symbol is again overwritten. Concurrently, the information of layer $1b_x$ and $1b_y$ is transmitted to layer 1c. In this layer, only one single feature symbol is activated and from this symbol, information about its activation as well as its location is transmitted to the sub-unimodal symbol “small steady object”. As this information arrives at the symbol after the messages coming from lower levels, it overwrites them (see figure A.23d). By exploiting these delays caused by signal runtimes and processing times, a distinct assignment between the feature symbol level and the sub-unimodal symbol level can be achieved.

The detection of steady objects of other sizes as well as moving objects works similarly. However, one question has been let open until now: If for example an object of middle size is present in the room, according to the description given until now, there would be activated the sub-unimodal symbols “middle steady object” as well as “small steady object”, because in the feature symbol hierarchy, there are activated feature symbols of both levels. However, in reality, only one object of middle size is present. To avoid the activation of the symbol “small steady object”, there has to exist a feedback connection to this symbol from the symbol “middle steady object”. The utility of feedback connection has already been outlined in section 4.2.5. Via the feedback, whenever the symbol “middle steady object” is activated, it inhibits the activation of the symbol “small steady object” presumed that their location values lie close together. In figure A.24, it is illustrated what inhibitory feedback connections have to exist between the sub-unimodal symbols of the tactile floor sensor modality. Necessary feedback connections between sub-unimodal neuro-symbols can be learned from examples (see section 4.5.2).

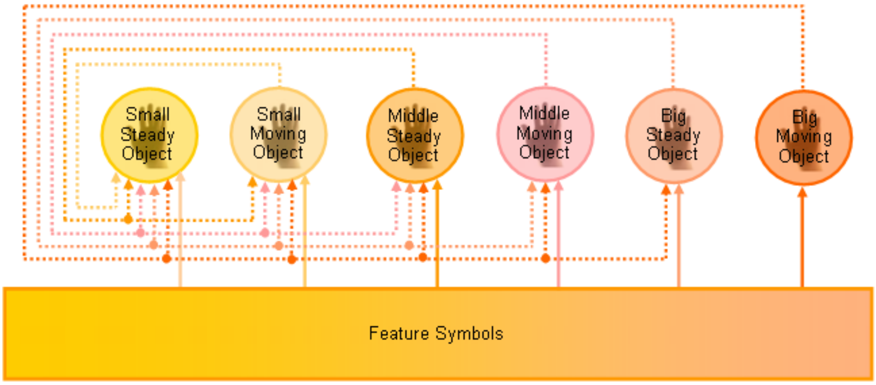


Figure A.24: Feedbacks between Sub-unimodal Symbols

B Literature

[Any04] Anylogic User's Manual. Technologies Company Ltd, 2004.

[Bar72] H. B. Barlow. Single Units and Sensation: A Neuron Doctrine for Perceptual Psychology. *Perception*, 1:371–394, 1972.

[BAR04] Xuehai Bian, Gregory Abowd, and James Rehg. Using Sound Source Localization to Monitor and Infer Activities in the Home. GVU Center Technical Report GIT-GVU-04-20, Georgia Institute of Technology, 2004.

[Bau05] Frank Bauer. *Visual Attention and Temporal Binding.* PhD thesis, Ludwig-Maximilians-Universität München, November 2005.

[BKVH08] D. Bruckner, J. Kasbi, R. Velik, and W. Herzner. High-level Hierarchical Semantic Processing Framework for Smart Sensor Networks. In *Proceedings of the Conference on Human System Interactions, to be published*, 2008.

[BLPV07] Wolfgang Burgstaller, Roland Lang, Patrizia Pörscht, and Rosemarie Velik. Technical Model of Basic and Complex Emotions. In *5th International Conference on Industrial Informatics*, pages 1033–1038, 2007.

[BLS06] J. Beyerer, F. Puente León, and K.-D. Sommer, editors. *Informationfusion in der Mess- und Sensortechnik.* Universitätsverlag Karlsruhe, 2006.

[BRG96] Eloi Bossé, Jean Roy, and Dominic Grenier. Data Fusion Concepts Applied to a Suite of Dissimilar Sensors. In *Canadian Conference on Electrical and Computer Engineering*, volume 2, pages 692–695, 1996.

[Bru07] Dietmar Bruckner. *Probabilistic Models in Building Automation: Recognizing Scenarios with Statistical Methods.* PhD thesis, Vienna University of Technology, 2007.

[BSR06] Dietmar Bruckner, Brian Sallans, and Gerhard Russ. Probabilistic Construction of Semantic Symbols in Building Automation Systems. In *IEEE International Conference of Industrial Informatics*, pages 1–6, 2006.

[Bur07] Wolfgang Burgstaller. *Interpretation of Situations in Buildings.* PhD thesis, Vienna University of Technology, 2007.

[Caw98] Alison Cawsey. *The Essence of Artificial Intelligence.* Prentice Hall Europe, 1998.

[CR04] Matthew C. Costello and Eric D. Reichle. LSDNet: A Neural Network for Multisensory Perception. In *Sixth International Conference on Cognitive Modeling*, pages 341–341, 2004.

[Cro05] James L. Crowley. Situated Observation of Human Activity. *Computer Vision for Interactive and Intelligent Environment*, pages 97–108, 2005.

[CS97] Patricia S. Churchland and Terrence J. Sejnowski. *Grundlagen zur Neuroinformatik und Neurobiologie – The Computational Brain in deutscher Sprache.* Vieweg Verlag, 1997.

[CSS04] Gemma Calvert, Charles Spencer, and Barry E. Stein, editors. *The Handbook of Multisensory Processes.* MIT Press, 2004.

[CW01] Marvin M. Chun and Jeremy M. Wolfe. *Blackwell's Handbook of Perception*, chapter 9, pages 272–310. Oxford, 2001.

[DA99] Anind K. Dey and Gregory D. Abowd. Towards a Better Understanding of Context and Context-Awareness. In *1st International Symposium on Handheld and Ubiquitous Computing*, pages 304–307, 1999.

[Dam94] Antonio Damasio. *Descartes' Error – Emotion, Reason, and the Human Brain.* Penguin Books, 1994.

[Dav97] J. Davis. Biological Sensor Fusion inspires Novel System Design. In *Joint Service Combat Identification Systems Conference*, 1997.

[DB05] Howard Demuth and Mark Beale. *Neural Network Toolbox – For Use with MATLAB.* The MathWorks, User's Guide, 2005.

[dGBG01] Artur S. d'Avila Garcez, Krysia Broda, and Dov M. Gabbay. Symbolic Knowledge Extraction from Trained Neural Networks: A Sound Approach. *Artificial Intelligence*, 125(1-2):155–207, 2001.

[dGGB02] Artur S. d'Avila Garcez, Dov M. Gabbay, and Krysia B. Broda. *Neural-Symbolic Learning System: Foundations and Applications.* Springer-Verlag New York, 2002.

[DGLV08] T. Deutsch, A. Gruber, R. Lang, and R. Velik. Episodic Memory for Autonomous Agents. In *Proceedings of the Conference on Human System Interactions, to be published*, 2008.

[DLP$^+$06] Tobias Deutsch, Roland Lang, Gerhard Pratl, Elisabeth Brainin, and Samy Teicher. Applying Psychoanalytic and Neuro-scientific Models to Automation. In *2nd IET International Conference on Intelligent Environments*, volume 1, pages 111–118, 2006.

[DRT$^+$01] Dietmar Dietrich, Gerhard Russ, Clara Tamarit, Gerald Koller, Wolfgang Ponweiser, and Markus Vincze. Modellierung des technischen Wahrnehmungsbewusstseins für den Bereich Home Automation. *Elektrotechnik und Informationstechnik*, 118:545–555, 2001.

[DS00] Dietmar Dietrich and Thilo Sauter. Evolution Potentials for Fieldbus Systems. In *3rd IEEE International Workshop on Factory Communication System*, pages 343–350, 2000.

[DTL$^+$03] Edoardo Datteri, Giancarlo Teti, Cecilia Laschi, Guglielmo Tamburrini, Paolo Dario, and Eugenio Guglielmelli. Expected Perception: An Anticipation-Based Perception-Action Scheme in Robots. In *International Conference on Intelligent Robots and Systems*, pages 934–939, 2003.

[EB04] Marc O. Ernst and Heinrich H. Bülthoff. Merging the Senses into a Robust Percept. *TRENDS in Cognitive Sciences*, 8(4):162–169, April 2004.

[Eic06] Howard Eichenbaum. Memory Binding in Hippocampal Relational Networks. In Hubert Zimmer, Axel Mecklinger, and Ulman Lindenberger, editors, *Handbook of Binding and Memory – Perspectives from Cognitive Neuroscience*, pages 25–51. Oxford University Press, 2006.

[Ell96] Daniel P. W. Ellis. *Prediction-driven Computational Auditory Scene Analysis.* PhD thesis, Massachusetts Institute of Technology, 1996.

[Elm02] Wilfried Elmenreich. *Sensor Fusion in Time-Triggered Systems.* PhD thesis, Vienna Univertiy of Technology, 2002.

[Elm07] Wilfried Elmenreich. A Review on System Architectures for Sensor Fusion Applications. In *Software Technologies for Embedded and Ubiquitous Systems*, pages 547–559. Springer Berlin / Heidelberg, 2007.

[Fre91] Sigmund Freud. *Zur Auffassung der Aphasien.* Fischer Taschenbuch, 1891.

[Fre15] Sigmund Freud. Triebe und Triebschicksale. In *Studienausgabe Band 3: Psychologie des Unbewussten*. Fischer Taschenbuch Verlag, 1915.

[Fre23] Sigmund Freud. Das Ich und das Es. In *Studienausgabe Band 3: Psychologie des Unbewussten*. Fischer Taschenbuch Verlag, 1923.

[Fre96] Robert M. French. Review of "The Engine of Reason, the Seat of the Soul". *In Minds and Machines*, 6(3):416–421, 1996.

[Gab04] Andreas Christian Gabriel. *Traveling gamma-Waves: New Insights into Coupling Processes in Visual Texture Coding*. PhD thesis, Philipps-University Marburg, August 2004.

[GHT96] H. D. R. Golledge, C. C. Hilgetag, and M. J. Tovee. Information Processing: A Solution to the Binding Problem? *Current Biology*, 6(9):1092–1095, September 1996.

[GJ07] Dileep George and Bobby Jaros. The HTM Learning Algorithms. Numenta, 2007.

[GK02] Wilson S. Geisler and Daniel Kersten. Illusions, Perception and Bayes. *Nature Neuroscience*, 5(6):508–510, June 2002.

[GM99] Geoffrey M. Ghose and John Maunsell. Specialized Representations in Visual Cortex: A Role for Binding? *Neuron*, 24:79–85, September 1999.

[Gol02] E. Bruce Goldstein. *Wahrnehmungspsychologie*. Spektrum Akademischer Verlag, 2002.

[Gol07] E. Bruce Goldstein. *Sensation and Perception*. Thomson Wadsworth, 2007.

[Gra99] Charles M. Gray. The Temporal Correlation Hypothesis of Visual Feature Integration: Still Alive and Well. *Neuron*, 24:31–47, September 1999.

[Gre97] Richard L. Gregory. Knowledge in Perception and Illusion. *Phil. Trans. R. Soc. Lond. B*, 352:1121–1128, 1997.

[Göt06] Siegfried Otto Götzinger. Scenario Recognition based on a Bionic Model for Multi-Level Symbolization. Master's thesis, Vienna University of Technology, 2006.

[Har90] S. Harnad. The Symbol Grounding Problem. *Physica D*, 42:335–346, 1990.

[Har08] Harald Hareter. *Worst Case Szenarien Simulator für die Gebäudeautomation*. PhD thesis, Vienna University of Technology, to be published, 2008.

[Hau98] Matthias Haun. *Simulation Neuronaler Netze – Eine praxisorientierte Einführung*. Expert Verlag, 1998.

[Hen05] John M. Henderson. Introduction to Real-word Scene Perception. *Visual Cognition*, 12(6):849–851, 2005.

[HG06] Jeff Hawkins and Dileep George. Hierarchical Temporal Memory – Concepts, Theory, and Terminology. Numenta, 2006.

[HH92] R.L. Harvey and K.G. Heinemann. Biological Vision Modles for Sensor Fusion. In *First IEEE Conference on Control Applications*, pages 392–397, 1992.

[HH98] Andrew Hollingworth and John M. Henderson. Does Consistent Scene Context Facilitate Object Perception? *Journal of Experimental Psychology*, 127(4):398–415, 1998.

[HH99] John M. Henderson and Andrew Hollingworth. High-level Scene Perception. *Annual Reviews Psychology 1999*, 50:243–271, 1999.

[Hil97] Mélanie Hilario. *An Overview of Strategies for Neurosymbolic Integration*, chapter 2, pages 13–35. Lawrence Erlbaum Associates, 1997.

[HM07] Bernhard Hommel and Bruce Milliken. Taking the Brain Serious: Introduction to the Special Issue on Integration in and across Perception and Action. *Psychological Research*, 71:1–3, 2007.

[Hom04] Bernhard Hommel. Event Files: Feature Binding in and across Perception and

Action. *TRENDS in Cognitive Sciences*, 8(11):494–500, November 2004.

[HPB05] Harald Hareter, Gerhard Pratl, and Dietmar Bruckner. A Simulation and Visualization System for Sensor and Actuator Data Generation. In *6th IFAC International Conference on Fieldbus Systems and their Applications*, pages 56–63, 2005.

[HU94] Vasant Honavar and Leonard Uhr. Symbolic Artificial Intelligence, Connectionist Networks, and Beyond, August 1994. Iowa State University of Science and Technology, Department of Computer Science.

[Huy99] Christian R. Huyck. Combining Symbolic and Connectionist Models in Artificial Intelligence, 1999.

[HW62] David H. Hubel and Torsten N. Wiesel. Receptive Fields, Binocular Interaction and Functional Architecture in the Cats Visual Cortex. *J. Physiol.*, 160:106–154, 1962.

[IAMN03] Teijiro Isokawa, Jun Adumi, Nobuyuki Matsui, and Haruhiko Nishimura. A Coupled Oscillatory Neural Network Model for Binding Problem. In *SICE Annual Conference*, volume 3, pages 2774–2777, 2003.

[JLN05] John Jonides, Steven C. Lacey, and Derek Evan Nee. Processes of working memory in mind and brain. *Current Directions in Psychological Science*, 14(1):2–5, 2005.

[Jov97] Emil Jovanov. A Model for Consciousness: An Engineering Approach. In *Brain and Consciousness, ECPD Symposium*, pages 291–295, 1997.

[JRCC03] D. Joyce, L. Richards, A. Cangelosi, and K.R. Coventry. On the Foundations of Perceptual Symbol Systems: Specifying Embodied Representations via Connectionism. In F. Detje, D. Dorner, and H. Schaub, editors, *Fifth International Conference on Cognitive Modeling*, pages 147–152, 2003.

[KBS01] Peter Kammermeier, Martin Buss, and Günther Schmidt. A Systems Theoretical Model for Human Perception in Multimodal Presence Systems. *IEEE/ASME Transactions on Mechatronics*, 6(3):234–244, 2001.

[KC01] Andrew J. King and Gemma A. Calvert. Multisensory Integration: Perceptual Grouping by Eye and Ear. *Current Biology*, 11(8):R322–R325, 2001.

[KKvW$^+$06] Markku Känsäkoski, Marika Kurkinen, Niklas von Weymarn, Pentti Niemelä, Tero Eerikäinen, Seppo Turunen, Sirkka Aho, and Pirkko Suhonen. *Process Analytical Technology (PAT) Needs and Applications in the Bioprocess Industry.* VTT, 2006.

[KZK97] Moshe Kam, Xiaoxun Zhu, and Paul Kalata. Sensor Fusion for Mobile Robot Navigation. In *IEEE*, volume 85, pages 108–119, 1997.

[LBP$^+$07] Roland Lang, Dietmar Bruckner, Gerhard Pratl, Rosemarie Velik, and Tobias Deutsch. Scenario Recognition in Modern Building Automation. In *7th IFAC International Conference on Fieldbuses and Networks in Industrial and Embedded Systems*, pages 305–312, 2007.

[LK89] Ren C. Luo and Michael G. Kay. Multisensor Integration and Fusion in Intelligent Systems. *IEEE Transactions on Systems, Man, and Cybernetics*, 19:901–931, 1989.

[LP73] J. Laplanche and J. B. Pontalis. *Das Vokabular der Psychoanalyse.* Suhrkamp, 1973.

[Lur73] Alexander R. Luria. *The Working Brain – An Introduction in Neuropsychology.* Basic Books, 1973.

[LYS02] Ren C. Luo, Chih-Chen Yih, and Kuo Lan Su. Multisensor Fusion and Integration: Approaches, Applications, and Future Research Directions. *IEEE Sensours Journal*, 2:107–119, 2002.

[LZD$^+$08] R. Lang, H. Zeilinger, T. Deutsch, R. Velik, and B. Mueller. Perceptive Learning – A Psychoanalytical Learning Framework for Autonomous Agents. In *Proceedings of the Conference on Human System Interactions, to be published*, 2008.

[Mas04] George A. Mashour. The Cognitive Binding Problem: From Kant to Quantum

Neurodynamics. *NeuroQuantology*, 1:29–38, 2004.

[MCM96] Phil Mars, J. R. Chen, and P. Mars. *Learning Algorithms – Theory and Applications in Signal Processing, Control and Communications.* CRC Press, 1996.

[Min06] Marvin Minsky. Die Maschine muss fühlen lernen. Interview in Technology Review, July 2006.

[MJS07] Wolfgang Maass, Prashant Joshi, and Eduardo D. Sontag. Computational Aspects of Feedback in Neural Circuits. *Computational Biology*, 3(1):1–20, 2007.

[Mur96] R. Murphy. Biological and Cognitive Foundations of Intelligent Sensor Fusion. *IEEE Transactions on Systems, Man and Cybernetics, Part A*, 26:42–51, 1996.

[Nev82] Ramakant Nevatia. *Machine Perception.* Prentice-Hall, Inc., 1982.

[New04] Fiona N. Newell. Cross-Modal Object Recognition. In *The Handbook of Multisensory Processes.* MIT Press, 2004.

[NGCV04] Tudor Nicosevici, Rafael Garcia, Marc Carreras, and Miquel Villanueva. A Review of Sensor Fusion Techniques for Underwater Vehicle Navigation. In *OCEANS 04*, volume 3, pages 1600–1605, 2004.

[Nut08] Tom Nuttall. Artificial Intelligence Tragedies. The Prospect Magazine, February 2008. http://blog.prospectblogs.com/2008/02/12/artificial-intelligence-tragedies/.

[Pai07] Jocelyn Paine. AI Lecture 1. Online Script, 2007. http://www.j-paine.org/students/lectures/lect1/node1.html.

[Pal08] Brigitte Palensky. *Introducing Neuro-Psychoanalysis towards the Design of Cognitive and Affective Automation Systems.* PhD thesis, Vienna University of Technology, 2008.

[Pan98] Jaak Panksepp. *Affective Neuroscience: The Foundations of Human and Animal Emotions.* Oxford University Press, 1998.

[PDHP07] Gerhard Pratl, Dietmar Dietrich, Gerhard Hancke, and Walter Penzhorn. A New Model for Autonomous, Networked Control Systems. *IEEE Transactions on Industrial Informatics*, 3:21–32, 2007.

[Pes94] Markus F. Peschl. Embodiment of Knowledge in the Sensory System and its Contribution to Sensorimotor Integration. *IEEE*, 0-8186-6482-7:444–447, 1994.

[PF07] Gerhard Pratl and Laurentiu Frangu. Smart Nodes for Semantic Analysis of Visual and Aural Data. In *5th International Conference on Industrial Informatics*, pages 1001–1006, 2007.

[PLC07] Peter Palensky, Brigitte Lorenz, and Andrea Clarici. Cognitive and Affective Automation: Machines Using the Psychoanalytic Model of the Human Mind. In *Engineering and Neuro-Psychoanalysis Forum*, 2007.

[PLD05] Gerhard Pratl, Brigitte Lorenz, and Dietmar Dietrich. The Artificial Recognition System (ARS): New Concepts for Building Automation. In *6th IFAC International Conference on Fieldbus Systems and their Applications*, pages 48–55, 2005.

[PP05] Gerhard Pratl and Peter Palensky. Project ARS – The Next Step towards an Intelligent Environment. In *International Workshop on Intelligent Environments*, pages 55–62, 2005.

[Pra06] Gerhard Pratl. *Processing and Symbolization of Ambient Sensor Data.* PhD thesis, Vienna University of Technology, 2006.

[PWM03] Leonid I. Perlovsky, Bertus Weijers, and Chris W. Mutz. Cognitive Foundations for Model-based Sensor Fusion. *Proceedings of SPIE*, 5096:494–501, August 2003.

[RD99] John H. Reynolds and Robert Desimone. The Role of Neural Mechanisms of Attention in Solving the Binding Problem. *Neuron*, 24:19–29, September 1999.

[Ric07] Andreas Richtsfeld. Szenarienerkennung durch symbolische Datenverarbeitung mit

Fuzzy-Logic. Master's thesis, Technischen Universität Wien, 2007.

[RL07] Heinrich Ruser and Fernando Puente León. Informationfusion – Eine Übersicht. *Technisches Messen (Oldenbourg Verlag)*, 74(3):93–102, 2007.

[RN07] Eyal Reingold and Johnathan Nightingale. Artificial Intelligence Tutorial Review. Online Script, 20007. http://www.psych.utoronto.ca/users/reingold/courses/ai/ai.html.

[Rob03] L. C. Robertson. Binding, Spatial Attention, and Perceptual Awareness. *Nature Reviews: Neuroscience*, 4(2):93–102, February 2003.

[Roj96] Raúl Rojas. *Theorie der neuronalen Netze – Eine systematische Einführung.* Springer Lehrbuch, 1996.

[Ros99] Adina L. Roskies. The Binding Problem. *Neuron*, 24:7–9, September 1999.

[RP99] Maximilian Riesenhuber and Tomaso Poggio. Are Cortical Models Really Bound by the "Binding Problem"? *Neuron*, 24:87–93, September 1999.

[RP02] Maximilian Riesenhuber and Tomaso Poggio. Neural Mechanisms of Object Recognition. *Current Opinion in Neurobiology*, 12:162–168, 2002.

[RP05] Khaleel A. Razak and Sarah L. Pallas. Neural Mechanisms of Stimulus Velocity Tuning in the Superior Colliculus. *Neurophysiol.*, 94:3573–3589, 2005.

[Rös07] Charlotte Rösener. *Adaptive Behavior Arbitration for Mobile Service Robots in Building Automation.* PhD thesis, Vienna University of Technology, 2007.

[Rus03] Gerhard Russ. *Situation-dependent Behavior in Building Automation.* PhD thesis, Vienna University of Technology, 2003.

[SA97] Ron Sun and Frederic Alexandre. *Connectionist-symbolic Integration.* Lawrence Erlbaum Associates, 1997.

[SB04] Richard J. Stevenson and Robert A. Boakes. Sweet and Sour Smells: Learned Synesthesia Between the Senses of Taste and Smell. In *The Handbook of Multisensory Processes.* MIT Press, 2004.

[Sch97] Andreas Scherer. *Neuronale Netze – Grundlagen und Anwendungen.* Vieweg Verlag, 1997.

[Sch01] Jan Scholz. The Binding Problem. Theoretical Neuroscience, Dezember 2001. University of Osnabrück.

[SGP01] Simon R. Schultz, Huw D.R. Golledge, and Stefano Panzeri. Synchronisation, Binding and the Role of Correlated Firing in Fast Information Transmission. In 2036, editor, *Emergent Neural Computational Architectures Based on Neuroscience*, pages 212–226. Springer Berlin / Heidelberg, 2001.

[SH99] Gijsbert Stoet and Bernhard Hommel. Action Planning and Temporal Binding of Response Codes. *Experimental Psychology: Human Perception and Performance*, 25(6):1625–1640, 1999.

[SH02] Gijsbert Stoet and Bernhard Hommel. Interaction between Feature Binding in Perception and Action. In Wolfgang Prinz and Bernhard Hommel, editors, *Common Mechanisms in Perception and Action*, volume 19, pages 538–552. Oxford University Press, 2002.

[Sha98] Amanda J.C. Sharkey. *Combining Artificial Neural Nets Ensemble and Modular Multi-Net Systems.* Springer-Verlag, June 1998.

[Sin01] Wolf Singer. Consciousness and the Binding Problem. *Annals of the New York Academy of Sciences*, 929:123–146, 2001.

[Sin03] Wolf Singer. Synchronization, Binding and Expectancy. In *The Handbook of Brain Theory and Neural Networks*, pages 1136–1143. MIT Press, 2003.

[SKS04] Ladan Shams, Yukiyasu Kamitani, and Shinsuke Shimojo. Modulations of Visual

Perception by Sound. In *The Handbook of Multisensory Processes*. MIT Press, 2004.

[SKSV00] Jukka Sillanpää, Anssi Klapuri, Jarno Seppänen, and Tuomas Virtanen. Recognition of Acoustic Noise Mixtures by Combined Bottom-up and Top-down Processing. In *10th European Conference on Signal Processing (EUSIPCO2000)*, 2000.

[SM99] Michael N. Shadlen and J. Anthony Movshon. Synchrony Unbound: A Critical Evaluation of the Temporal Binding Hypothesis. *Neuron*, 24:67–77, September 1999.

[SRT00] Stefan Soucek, Gerhard Russ, and Clara Tamarit. The Smart Kitchen Project – An Application of Fieldbus Technology to Domotics. In *2nd International Workshop on Networked Appliances*, pages 1–9, 2000.

[ST02] Mark Solms and Oliver Turnbull. *The Brain and the Inner World – An Introduction to the Neuroscience of Subjective Experience.* Other Press New York, 2002.

[Sta04] Stanford Encyclopedia of Philosophy. Online Encyclopedia, 2004. http://plato.stanford.edu/entries/time-experience/.

[Sta07] Stanford Encyclopedia of Philosophy. Online Encyclopedia, 2007. http://plato.stanford.edu/entries/cognitive-science/.

[SZ96] Charles F. Stevens and Anthony Zador. Information through a Spiking Neuron. In David S. Touretzky, Michael C. Mozer, and Michael E. Hasselmo, editors, *Advances in Neural Information Processing Systems*, volume 8, pages 7–81. The MIT Press, 1996.

[TBM97] D. A. Thurman, D. M. Brann, and C. M. Mitchell. An Architecture to Support Incremental Automation of Complex Systems. In *IEEE International Conference on Systems, Man, and Cybernetics*, 1997.

[TDDR01] Clara Tamarit, Dietmar Dietrich, Keith Dimond, and Gerhard Russ. A Definition and a Model of a Perceptive Awareness System (PAS). In *IFAC International Conference on Fielbus Systems and their Applications FeT*, pages 1–7, 2001.

[TF03] Clara Tamarit-Fuertes. *Automation System Perception – First Step towards Perceptive Awareness.* PhD thesis, Vienna University of Technology, 2003.

[TFR02] Clara Tamarit-Fuertes and Gerhard Russ. Unification of Perception Sources for Perceptive Awareness Automatic Systems. In *Africon 02*, pages 1–4, 2002.

[TG80] Anne M. Treisman and Gertrude Gelade. A Feature-Integration Theory of Attention. *Cognitive Psychology*, 12:97–136, 1980.

[Tod04] James T. Todd. The Visual Perception of 3D Shape. *TRENDS in Cognitive Sciences*, 8(3):115–121, 2004.

[Tre96] Anne Treisman. The Binding Problem. *Current Opinion in Neurobiology*, 6:171–178, 1996.

[Tre99] Anne Treisman. Solutions to the Binding Problem: Progress through Controversy and Convergence. *Neuron*, 24:105–110, September 1999.

[Tul83] Endel Tulving. *Elements of Episodic Memory.* Oxford University Press New York, 1983.

[TvdM96] Jochen Triesch and Christoph von der Malsburg. Binding – A Proposed Experiment and a Model. In *Proceedings of the ICANN 96*, pages 685–690. Springer Verlag, 1996.

[van98] Joris van Dam. *Environment Modelling for Mobile Robots: Neural Learning for Sensor Fusion.* PhD thesis, University of Amsterdam, 1998.

[Vel06] Rosemarie Velik. Neuronale Muster- und Objecterkennung durch Analyse im Zeit- und Frequenzbereich. Master's thesis, Vienna University of Technology, 2006.

[Vel07] Rosemarie Velik. A Model for Multimodal Humanlike Perception based on Modular Hierarchical Symbolic Information Processing, Knowledge Integration, and Learning. In *2nd International Conference on Bio-Inspired Models of Network, Informa-*

tion, and Computing Systems, pages 1–8, 2007.

[VLBD08] R. Velik, R. Lang, D. Bruckner, and T. Deutsch. Emulating the Perceptual System of the Brain for the Purpose of Sensor Fusion. In *Proceedings of the Conference on Human System Interactions, to be published*, 2008.

[von81] Christoph von der Malsburg. The Correlation Theory of Brain Function. Technical report, Biophysical Chemistry, MIP, 1981.

[von95] Christoph von der Malsburg. Binding in Models of Perception and Brain Function. *Curr Opin Neurobiol*, 5(4):520–526, August 1995.

[von99] Christoph von der Malsburg. The What and Why of Binding: The Modeler's Perspective. *Neuron*, 24:95–104, September 1999.

[VPL07] Rosemarie Velik, Gerhard Pratl, and Roland Lang. A Multi-sensory, Symbolic, Knowledge-based Model for Humanlike Perception. In *7th IFAC International Conference on Fieldbuses and Networks in Industrial and Embedded Systems*, pages 273–278, 2007.

[WC99] Jeremy M. Wolfe and Kyle R. Cave. The Psychophysical Evidence for a Binding Problem. *Neuron*, 24:11–17, September 1999.

[Wer98] Stefan Wermter. Hybrid Neural Symbolic Integration. In *Workshop as Part of International Conference on Neural Information Processing Systems, Breckenridge, Colorado*, December 1998.

[WK05] Ilana B. Witten and Eric I. Knudsen. Why Seeing is Believing: Merging Auditory and Visual Worlds. *Neuron*, 48:489–496, November 2005.

[WS00] Stefan Wermter and Ron Sun. *Hybrid Neural Systems.* Springer, 2000.

[Zad98] Anthony Zador. Impact of Synaptic Unreliability on the Information Transmitted by Spiking Neurons. *Neurophysiology*, 79:1219–1229, 1998.

CURRICULUM VITAE

PERSONAL INFORMATION

Name:	Rosemarie VELIK
Birth:	08.09.1981 Klagenfurt, Austria
Martial Status:	single
Nationality:	Austria

EDUCATION

July 2006 – now	University Assistant, Vienna University of Technology, Institute of Computer Technology
Aug. 2006 – April 2008	PhD Study of **Technical Sciences**, Vienna University of Technology PhD Thesis: "A Bionic Model for Human-like Machine Perception", Institute of Computer Technology, Prof. Dr. Dietmar Dietrich, Prof. Dr. Peter Palensky, Prof. Dr. Wolfgang Kastner
Sept. 2007	Language Course in Spain
May 2006 – July 2006	Language Course in Italy
Oct. 2001 – April 2006	Baccalaureate Study **Electrical Engineering and Information Technology** and Master Study **Automation Technology,** Vienna University of Technology, (both passed with excellence) Diploma Thesis: "Neural Pattern and Object Recognition by Analyses in the Time and Frequency Domain", Institute of Automation and Control Technology, Prof. Dr. Bernard Favre-Bulle
Sept. 1996 – June 2001	Technical High School for **Computer Engineering** and **Internet Engineering**, Klagenfurt, (passed with excellence)

PROFESSIONAL EXPERIENCE

since March 2008	Teaching "**Networked Systems**"
since Oct. 2007	Teaching "**Information Technology**"
since March 2007	Teaching "**Artificial Intelligence and Cognitive Sciences**"

since Dec. 2006	AKG Representative for Women Equal Treatment at the Vienna University of Technology, Faculty of Electrical Engineering
since Oct. 2006	Teaching “**Digital Integrated Circuits**”, “**Computer Technology**”, and “**Process Control Technique**”
since July 2006	Project ARS (Artificial Recognition System)
since July 2006	University Assistant at the Vienna University of Technology, Institute of Computer Technology, Research Area: **Artificial Intelligence**, **Neural Networks**, **Cognitive Sciences**, **Cognitive Computation**, **Cognitive Automation**, **Machine Perception**, **Neuro-symbolic Networks**
June 2004 – Jan. 2005	Austrian Research Center, A-2443 Seibersdorf, Bachelor Project about the Effects of Electro-magnetic Fields on the Body
May 2000 – May 2001	First Components, A-9500 Villach, Matura Project about Control of High-speed Welding Processes for Packaging Machines
July – Aug. 2000	Siemens, A-1100 Vienna, Traineeship
July 1998	Kelag, A-9020 Klagenfurt, Traineeship

PRIZES AND AWARDS

2006	Würdigungspreis of the Austrian Federal Ministry for Education, the Arts and Culture
2005	Leistungsstipendium of the Faculty of Electrical Engineering and Information Technology
2004	Leistungsstipendium of the Faculty of Electrical Engineering and Information Technology
2003	Leistungsstipendium of the Faculty of Electrical Engineering and Information Technology
2002	Leistungsstipendium of the Faculty of Electrical Engineering and Information Technology
2001	Jugend Innovativ
2000	Känguru der Mathematik

LIST OF PUBLICATIONS

- R. Velik, R. Lang, D. Bruckner, T. Deutsch: Emulating the Perceptual System of the Brain for the Purpose of Sensor Fusion. To be published in: Proceedings of the Conference on Human System Interactions, Krakow, 2008.
- T. Deutsch, A. Gruber, R. Lang, R. Velik: Episodic Memory for Autonomous Agents. To be published in: Proceedings of the Conference on Human System Interactions, Krakow, 2008.
- D. Bruckner, J. Kasbi, R. Velik, W. Herzner: High-level hierarchical Semantic Processing Framework for Smart Sensor Networks. To be published in: Proceedings of the Conference on Human System Interactions, Krakow, 2008.
- R. Lang, H. Zeilinger, T. Deutsch, R. Velik, B. Mueller. Perceptive Learning – A Psychoanalytical Learning Framework for Autonomous Agents. To be published in: Proceedings of the Conference on Human System Interactions, Krakow, 2008.
- R. Velik: A Bionic Model for Human-like Machine Perception, PhD Thesis, Vienna University of Technology, Institute of Computer Technology, 2008.
- R. Velik: A Model for Multimodal Humanlike Perception based on Modular Hierarchical Symbolic Information Processing, Knowledge Integration, and Learning. In: Proceedings of the 2nd

International Conference on Bio-Inspired Models of Network, Information, and Computing Systems (BIONETICS 2007), p. 8, Budapest, 2007.

- R. Lang, D. Bruckner, G. Pratl, R. Velik, T. Deutsch: Scenario Recognition in Modern Building Automation. In: Proceedings of the 7th IFAC International Conference on Fieldbuses & Networks in Industrial & Embedded Systems (FeT 2007), 2007, pp. 305–312.
- R. Velik, G. Pratl, R. Lang: A Multi-sensory, Symbolic, Knowledge-based Model for Human-like Perception. In: Proceedings of the 7th IFAC International Conference on Fieldbuses & Networks in Industrial & Embedded Systems (FeT 2007), 2007, pp. 273–278.
- W. Burgstaller, R. Lang, P. Pörscht, R. Velik: Technical Model for Basic and Complex Emotions. In: INDIN 2007 Conference Proceedings, 2007, pp. 1033–1038.
- R. Velik: Neuronal Pattern and Object Recognition by Analyses in the Time and Frequency Range, Diploma Thesis, Vienna University of Technology, Institute of Automation and Control Technology, 2006.

ADDITIONAL QUALIFICATIONS

2002 – 2007	Attendance to Courses in the Field of Biomedical Technology, Biophysics, Medicine, and Biology

LANGUAGE SKILLS

German	native language
English	business fluent
Italian	fluent
Spanish	fluent