A Model for Multimodal Humanlike Perception based on Modular Hierarchical Symbolic Information Processing, Knowledge Integration, and Learning

Rosemarie Velik
Vienna University of Technology
Gusshausstraße 27-29/384
1040 Vienna
+43 158801-38463
velik@ict.tuwien.ac.at

ABSTRACT
Automatic surveillance systems as well as autonomous robots are technical systems which would profit from the ability of humanlike perception for effective, efficient, and flexible operation. In this article, a model for humanlike perception is introduced based on hierarchical modular fusion of multi-sensory data, symbolic information processing, integration of knowledge and memory, and learning. The model is inspired by findings from neuroscience. Information from diverse sensors is transformed into symbolic representations and processed in parallel in a modular, hierarchical fashion. Higher-level symbolic information is gained by combination of lower-level symbols. Feedbacks from higher levels to lower levels are possible. Relations between symbols can be learned from examples. Stored knowledge influences the activation of symbols. The model and the underlying concepts are explained by means of a concrete example taken from building automation.

Keywords
Bionics, Humanlike Perception, Multisensory Integration, Symbolic Information Processing, Learning, Knowledge-based Systems, Building Automation.

1. INTRODUCTION
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, and scenarios in a real-world environment. Here, the number of possible occurring objects, events, and situations is almost infinite. As research from image processing and audio-data processing has shown, for a machine, recognition of real world situations is a task far from trivial. On the other hand, humans, even small children, can perceive such a "real-world environment" almost effortlessly. The challenging question is what gives humans the ability to perform these tasks and how to design machines that perform in a similarly efficient way.

In this article, a model for humanlike perception is introduced, which is inspired by findings from neuroscience. The proposed model is based on hierarchical modular fusion of multi-sensory data, symbolic information processing, integration of knowledge and memory, and learning. The model and the underlying concepts are explained by means of a concrete example taken from building automation.

2. RELATED WORK
The technical model for humanlike perception proposed in this article is based on modular fusion of multi-sensory data, symbolic information processing, integration of knowledge and memory, and learning. There have already been some attempts to model human perception, fuse multi-sensory data, process information symbolically, and to integrate different forms of knowledge. The approaches suggested in literature to solve these problems are disparate.

In [16], a mathematical model of the human perception process is presented. The proposed systems theoretical framework describes the principles of human perception as a concatenation of nonlinear vector mappings. [18] introduce a model for distributed perception systems for ubiquitous computing applications using a layered architecture. [4] 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. [7] outline a strategy and a control architecture to allow a mobile robot to navigate in an indoor environment on a planned path. The navigation system of the mobile robot integrates the position estimation obtained by a vision system with the position estimated by odometry, using a Kalman filter framework. Obstacle detection is performed by means of a set of ultrasonic sensors. [24] document the rationale and design of a multimodal interface to a pervasive/ubiquitous computing system

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.
Bionetics '07, December 10-13, 2007, Budapest, Hungary
Copyright 2007 ICST 978-963-9799-11-0.
.

that supports independent living by older people in their own homes.

[21] introduce a multimodal system for recognizing office activity that relies on a cascade of Hidden Markov Models (HMMs). The HMMs are used to diagnose states of user activity based on real-time streams of evidence from video, audio, and computer (keyboard and mouse) interactions. [22] attempts to develop a scientific and technological foundation for interactive environments. Therefore, machine perception techniques using acoustics, speech, computer vision, and mechanical sensors are employed. [5] 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 activity within a meeting or lecture.

[9] 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. [10] mention 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.

[2] 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. [25] 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. [8] 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. [6] propose a scheme where perception crucially involves comparison processes between incoming stimuli and expected perceptions built from previous perceptions.

Attempts to process sensor information symbolically have been made by [23] and [13] who suggest a layered architecture for this purpose. [15] attempt to achieve symbol grounding by adding a sensory concept to an abstract symbol.

3. NEUROSCIENTIFIC BACKGROUNDS

As the technical model for humanlike perception is inspired by neuroscientific insights about the perceptual system of the human brain, this section summarizes backgrounds which serve as archetype for the model. When elaborating such a technical model, it has to be considered that in neuroscience – up to now – there do not exist complete, unified concepts of the perceptual system of the brain. There are still many blind spots and controversial viewpoints, which make it difficult for an engineer to transform the neuroscientific model into a technical one.

3.1 Bottom-up and top-down processes in perception

Perception is the result of top-down processing and bottom-up processing working together. [12] Bottom-up processing, also referred to as data-based processing, is based on incoming data from the receptors of the human sense organs. Incoming data are always the starting point for perception. Without incoming data, there is no perception. Top-down processing, also labeled knowledge-based processing, is based on knowledge. This knowledge can be factual knowledge about objects, pre-experience, knowledge about the context in which an object occurs, or expectation.

3.2 Modular hierarchical architecture for information processing

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. Of most interest for the model proposed in section 4 are the cutaneous sensations, which are based on the stimulation of receptors in the skin responsible for tactile sensation, vibration sense, temperature sense, and pain sense. [11]

The perceptive 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. [17] A scheme of the structural organization of the human perceptive system is depicted in figure 1.

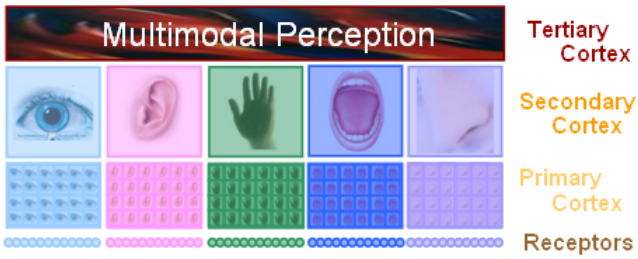

Fig. 1. Structural organization of the human perceptive system.

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 neurons of the primary cortical visual system only respond to the narrowly specialized properties of visual stimuli like the character of lines, shades of color, or the direction of movement. Neurons of the primary auditory cortex only respond to highly differentiated properties of acoustic stimuli. The primary areas are surrounded by systems of secondary cortical zones. In the secondary areas, information coming from the primary zones is processed. The cells of the secondary cortical zones have a much lower degree of specificity. Neurons of the secondary visual cortex respond to complex visual images like faces. Neurons of the secondary auditory context respond to complex acoustic signals like melodies. The tertiary zones, also referred to as zones of

overlapping, are responsible for enabling groups of several analyzers to work concertedly. Here, 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. The great majority of neurons of these zones are multimodal in character. An example for a processing result of this area would be the association of the visual image of a person to the auditory perception of the person talking.

From the lowest to the highest layer, information of increasing complexity is processed. The lowest level is fully developed at an early stage of development of the human brain whereas higher levels of information processing are not fully developed in a child's brain until the age of 7 years. Connections and associations of lower layers have to be learned and evolved first, before higher layers can evolve their full functionality.

3.3 Unimodal and multimodal binding of information

As just described, the perceptive system of the brain processes information in a modular hierarchical fashion. Simple features extracted from sensory information are combined to more complex unimodal representations and are then merged to a unified multimodal perception. The fundamental question, which is far from trivial, is how these combinations are performed. In neuroscience, this question is referred to as *binding problem*. Binding occurs in many different kinds of brain processes. To explain how coherent representations can be formed of information that is distributed throughout the brain, different binding mechanisms have been hypothesized which are not mutually exclusive. In literature, basically four different potential solutions to the binding problem are suggested and usually discussed in the context of visual perception: combination coding, population coding, temporal coding, and attention. [27], [20] The true solution of how binding is solved in the brain is still not known and might be a combination of some or all of these mechanisms.

3.4 Knowledge integration

Perception is facilitated by knowledge: 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 for perception is in fact memory. 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. [26] A fundamental question is, on which level knowledge interacts with sensory perception. The answers to this question are controversy.

3.5 Neural versus symbolic information processing

Information processing in the human brain is generally considered as being carried out by interacting neurons as well as chemicals like hormones and peptides. However, due to the complexity of mental processes, it is complicated – if not impossible – to project mental states to low-level explanations on the behavior of neurons, synapses, and chemicals.

To reduce complexity, information processing in the human brain can also be considered on the more abstract level of symbols. According to the theory of symbolic systems, the mind is a symbol system and cognition is symbol manipulation. [14] Examples for symbols are objects, characters, figures, sound, or colors used to represent abstract ideas and concepts.

4. TECHNICAL MODEL FOR HUMAN-LIKE PERCEPTION

In this section, a technical model for a humanlike perceptive system is presented. A humanlike perception system has a wide range of potential applications. Autonomous robots with such capabilities could navigate self-sufficiently in their environment, automatic surveillance systems could help to pare down personnel for monotonous observation tasks and increase safety of e.g. patients in a hospital. The application envisioned in this article is located in the field of building automation and automatic surveillance systems. For perceiving complex events and scenarios in buildings, different sensors are needed. The proposed system shall be capable of handling and processing such a flood of information. However, to clarify the underlying concepts, simplified examples are used. The principles of the model are first described generally, followed by a description by means of a concrete example: it shall be detected if a person is entering a room, leaving the room, walking around in the room, standing in the room, opening the door, or closing it.

First, the pursued strategy for sensor selection and information processing is outlined. Next, an architecture for modular, hierarchical processing of incoming sensor data is introduced followed by a description of how sensor data can be transformed into symbolic information. Fourth, a concept of how to integrate knowledge into symbolic information processing is drafted. Last, it is reflected about how to bind symbolic information and how learning can be integrated into this process.

4.1 Sensor selection for multi-sensory information processing

To perceive the environment, the human body has diverse sensory receptors. Accordingly, a technical system has to be equipped with diverse sensors to perceive events, situations, and scenarios in a building. The sensors to be chosen strongly depend on the desired situations to be recognized.

To perceive the different scenarios mentioned before, a room is equipped with the following sensor types: a motion detector, two tactile floor sensors, two light barriers, a door contact sensor, a video camera, and a microphone. The sensors are mounted at the positions depicted in figure 2 and have the property to have partly overlapping sensory fields of perception and to provide partly redundant information. The motion detector, tactile sensors, light barriers, and door contact sensor provide a binary sensor signal (*zero* or *one*). The motion detector detects movement in the room. The tactile floor sensors detect if an object is present in the room and whether it is present in the left or the right half of the room. The two light barriers detect whether an object passed the door and – by combining the information of both – in which direction it passed. The door contact sensor determines whether the door is open or closed. The video camera evaluates whether a person is present in the room. The microphone detects the noise that occurs when someone walks or when the door is opened or closed.

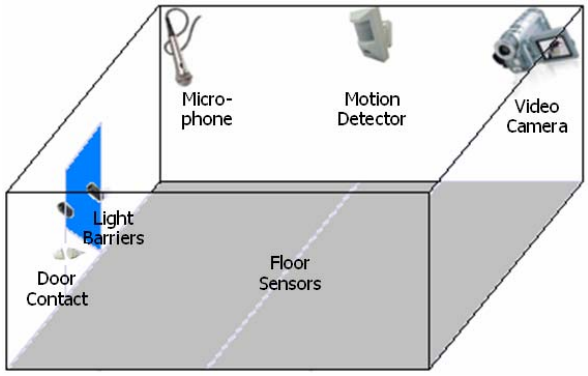

Fig. 2. Test bed: Room equipped with different sensors.

The information gathered from each sensor type is intentionally kept simple. The intended strategy is to utilize diverse sensors, extract information from these sensors, and gather information about complex coherencies by merging the information provided from different sensor types. By relying not only on one sensory modality but on different ones, the robustness and reliability of perception is increased.

4.2 Modular hierarchical information processing

Processing of vast amounts of sensory information requires a well thought out information processing structure. The information processing structure proposed in this paper is inspired by the modular hierarchical organization of the human brain as described in section 3.2. The sensory receptors have their analogy in the different sensor types. In a first stage – similar to the information processing performed in the primary cortices of the different sensory modalities – the sensory raw data are pre-processed to extract features suitable for a further processing. In a second stage, which corresponds to the information processing performed in the secondary cortices, the extracted features are combined to result in unimodal perceptions. As outlined for the cutaneous senses, each unimodal perceptual system can be further divided into subsystems. The processing of the first and second stage as well as its sub-stages is performed separately and in parallel for each sensory modality. In a third 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.

Figure 3 illustrates the modular hierarchical information processing structure for our example. Multimodal perception is achieved by combination of data from three unimodal perceptive systems – in analogy to its biological archetypes – referred to as visual perception, auditory perception, and cutaneous perception. The cutaneous perceptive system integrates information coming from four cutaneous sub-systems corresponding to the four different sensor types used.

4.3 Symbolic information processing

In section 3.5 it was mentioned that the human mind can be regarded as a symbolic processing system. This strategy of symbolization is applied to our model. For this purpose, sensory information is transformed into 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 termed *feature symbols*. In the next stage, the feature symbols of each sensory domain are either directly merged to *unimodal symbols* or 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*. All symbols can have properties, which comprise information that specifies the symbols in more detail. The concept of properties reduces the number of necessary different symbols to detect the defined possible situations.

The structure of the individual symbols used in our model has its paragon in the organization of biological neurons. Each symbol can receive input information from several other symbols which corresponds to the function of dendrites of neurons. The input information contains – among others – the activation grade of the symbol it origins from. The activation grades of all incoming symbols are summed up likewise in the cell body of a nerve cell. If this sum exceeds a certain threshold, the symbol passes a signal to other symbols it is connected to in analogy to the axon of a neuron. Like in the brain where many neurons are active at the same time, different symbols can also process information in parallel.

If a system is equipped with many different sensors and their values have to be processed, a lot of calculations have to be performed every instant. To pare down computational power, our system only reacts to changes of sensor values. New feature symbols are only generated and old feature symbols are only expired if sensor values are subject to changes. The same also holds true for all higher-level symbol types as they result from a combination of the diverse sensor values. At initial system startup, no feature symbol and no other symbol are active unimportant of the current values of the sensors.

The strategy of combining diverse lower-lever symbols to one higher-lever symbol corresponds to a bottom-up data processing principle. Through feedbacks, the generation of symbols can be influenced in a top-down manner. Additionally, symbols can be influenced top-down by knowledge that is stored in the system. The concept of knowledge integration and *memory symbols* will be described in more detail in section 4.4.

Applying the concept of symbolic information processing to our example, we get a modular hierarchical symbol structure as depicted in figure 4. As the explanation shall be kept simple, it is assumed that there is always only one person present that can trigger sensors. That way, only one of the six multimodal symbols will be generated at a certain time.

In a first processing stage, feature symbols – pictured as squares – are extracted from the sensory raw data. In case of the sensors of the cutaneous senses, the associations between the sensor values and the feature symbols are very simple due their binary sensor output values and the fact that the amount of sensors per sensor type is limited. However, if more sensors of each binary sensor type or analogous sensors were used, the relations between sensory raw data and feature symbols would become more complex. Conceivable feature symbols in the visual information processing flow are edges, lines, curves, colors, moving forms, etc. Feature symbols for auditory processing can be the frequency components or the loudness of a sound signal.

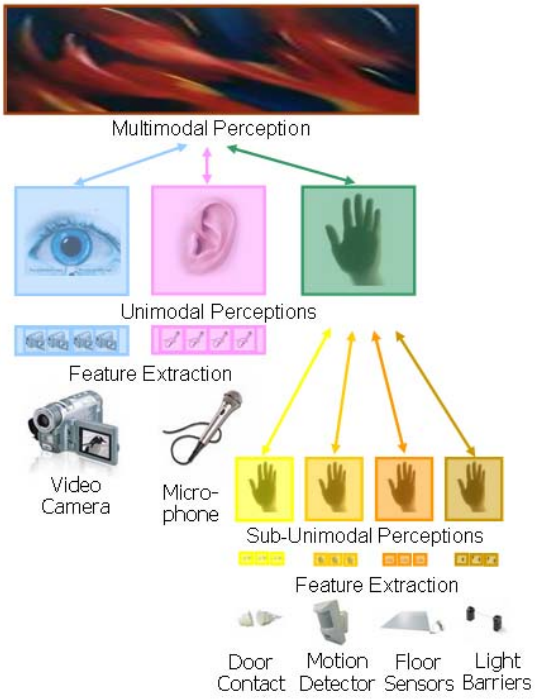

Fig. 3. Architecture for modular hierarchical information processing of multi-sensory data.

Out of the feature symbols, sub-unimodal symbols in case of the cutaneous senses and unimodal symbols in case of the visual and auditory sense are formed. 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 or the noise of an opening or closing door 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 or opening and closing doors 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 out of the raw data which correspond to our feature symbols.

The sub-unimodal and unimodal symbols of the cutaneous perceptive system describe the states and activities 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. Some of the sub-unimodal symbols contain properties which specify the symbol in more detail. The symbol "person present" has the property "location l", which indicates whether an object is present in the room and at what position (left or right half of the room) it is located. The symbol "object passes" has the property "direction d" which comprises the information from what direction the object passes the door. In other words, whether the object entered or left the room. The property "status s" of the symbol "door status" comprises the information whether the door is opened or closed. The sub-unimodal cutaneous symbols are combined to unified unimodal cutaneous symbols. What higher-lever symbols are generated by which combination of lower-lever symbols can be read out from figure 4. The symbols "object enters" and "object leaves" are both formed from the same symbols. The decision which of the two symbols is formed is only dependent on the property "direction d" of the symbol "object passes". The same holds true for the symbols "door is opened" and "door is closed" which depend on the property "status s" of the symbol "door status". Some of the unimodal cutaneous symbols also depend on the property "location l" of the symbol "person present".

Using this bottom-up data processing, a critical point that has to be considered is the fact that whenever one of the symbols "object enters", "object leaves", ""object moves", "door is opened", or "door is closed" is generated, there is also activated the symbols "object stands", because it results from a combination of a subset of the same sub-unimodal cutaneous symbols. Additionally, each activation of the symbol "object enters" or "object leaves" also triggers the symbol "object moves". To overcome the undesired activation of more than one symbol at a certain moment, inhibitory feedbacks are inducted. This structure is comparable to the neural feedbacks between different layers of neurons in the brain. In figure 4, these inhibitory feedback connections are depicted as dotted lines.

From the unimodal symbols of the visual, auditory, and cutaneous system, one of the six multimodal symbols is generated. The conditions under what circumstances which symbol is generated can be read out from figure 4. The generation of multimodal symbols is also influenced top-down by knowledge. The underlying concept therefore is described in section 4.4.

4.4 Knowledge integration

As outlined in section 3.4, perception does not only rely on sensory information but to a great extent also on knowledge. Integration of knowledge and awareness of what happened until now into the perceptive process can greatly facilitate perception. In a realistic situation, not only one out of six different scenarios has to be detected but one out of thousands. In such a case it may happen that a scenario cannot unambiguously be perceived from the current sensor values, because two or more scenarios might be triggered from 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. In our model of symbolic information processing, knowledge can influence the generation or non-generation of symbols.

One example for knowledge integration to our test scenarios would be to let the system know that a person can only walk around in the room, stand in the room, or leave the room if he entered before. This means that the symbols "person walks", "person stands", and "person leaves" can only be generated if the symbol "person enters" was generated 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" scenario if the door is already closed or open.

A utilization of information of that kind requires a sort of memory to store important information from past events. In our example,

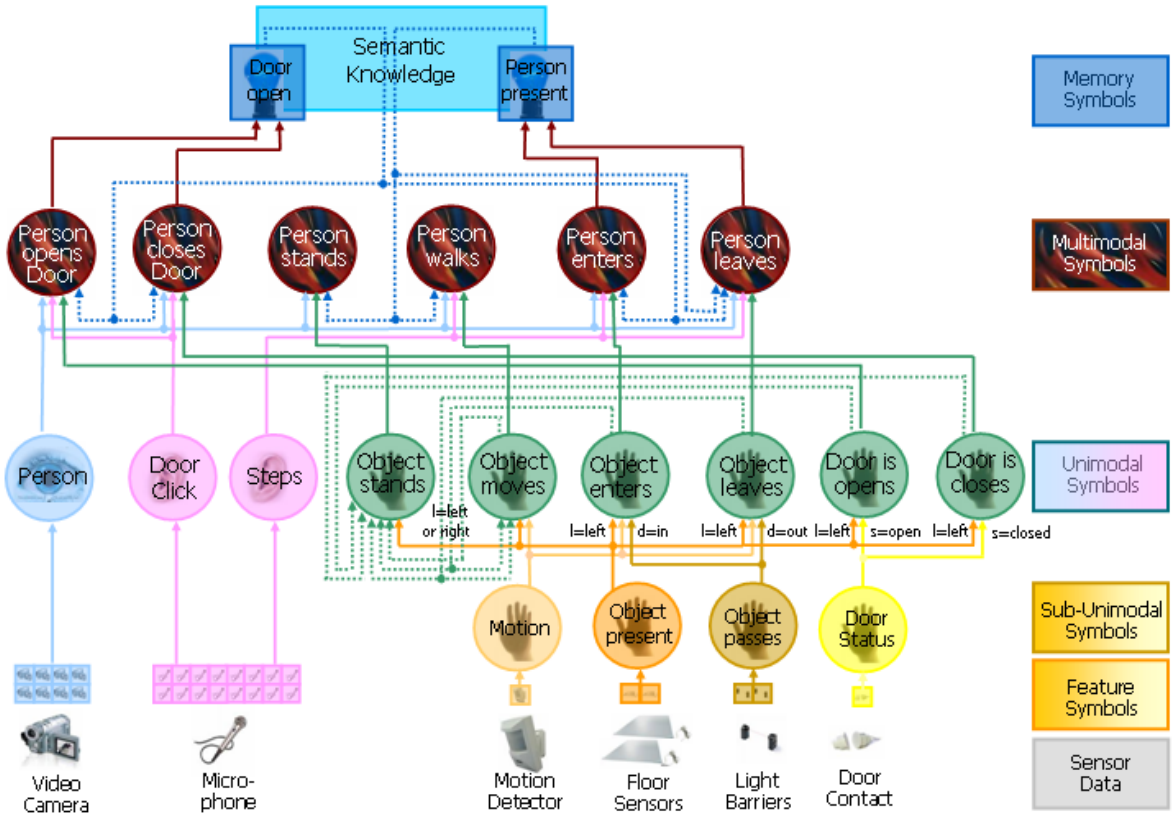

Fig. 4. Structure of hierarchical symbolic information processing.

up to the multimodal symbol layer, symbols are activated and deactivated when sensor values change. In these layers, there exists no memory that stores past states. Storage of events happened in the past is achieved in an additional layer with so called *memory symbols*.

A fundamental question is on what level knowledge should influence perception. It could either already influence unimodal perception or subsystems of unimodal perception or interact not until multimodal perception. The answers coming from neuroscience are controversy. The best way to find out is probably to try out the different possibilities and to evaluate what works most efficiently for a certain application. In our model, for a first implementation, knowledge only influences multimodal perception. The states of memory symbols are set or reset when certain multimodal symbols are activated. The symbol "person present" is set after the 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". The activation of certain multimodal symbols triggered bottom up by sensor values can be inhibited by top-down influence of *semantic knowledge* in form of rules in combination with stored states of memory symbols. In figure 4, inhibitory connections from the knowledge layer to the multimodal layer are depicted as dotted lines.

4.5 Symbol binding through learning

For our perceptive model, symbolic information processing was suggested. A fundamental question is how symbols can be formed from sensor values and how symbols can be formed by association of other symbols. This problem is adequate to the neuroscientific binding problem outlined in section 3.3.

For a flexible system it is desirable not to have to predefine everything but to learn coherences – preferable from examples. However, similar to the human brain, not everything can be learned during operation. Certain patterns must already be predefined at the system's initial start-up. The answers from neuroscience to what has to be predefined and what can be learned are controversy.

In our model, the symbols of the different hierarchy levels as well as their principal properties are predefined. Likewise it is fixed at initial start-up what feature symbols can be extracted from the sensory raw data. In contrast, there exists the possibility to learn from examples what combinations of feature symbols generate a sub-undimodal or unimodal symbol, which sub-unimodal symbols merge to unimodal symbols, and what combinations of unimodal symbols generate a multimodal symbol. As already mentioned, depending on the values of the properties of a symbol, this symbol can generate different higher-level symbols. It can also be learned which value of a property generates which higher-level symbol. It is also learnable what multimodal symbols trigger

which memory symbol. However, what inhibitory influence the activation of a knowledge symbol has on multimodal symbols is defined by explicit rules given by the system designer and stored as semantic knowledge.

The process of learning is divided into different phases. In section 3.2 it was outlined that higher cortical levels can only evolve if lower levels are already sophisticated. This strategy is applied to our model. In a first step it is learned what combinations of feature symbols form which sub-unimodal or unimodal symbols. Next, it is determined what sub-unimodal symbols form which unimodal symbols. At this level, certain unimodal symbols can have an inhibitory feedback on other unimodal symbols. These feedback connections are learned next. After having fully functioning unimodal levels, the relation between unimodal and multimodal symbols is learned. Finally, it is extracted out of examples by means of which multimodal symbols memory symbols are set and reset. The learning of symbol connections can be achieved by different methods. In a first implementation, quite simple statistical methods are used. However, also other techniques are conceivable. To set connections between feature symbols and sub-unimodal or unimodal symbols, neural networks could be applied. For higher levels, methods like fuzzy inductive reasoning (FIR) might also be possible. [3], [19]

5. IMPLEMENTATION

To test the model of humanlike perception just introduced, it was implemented and simulated in AnyLogic. The modeling language of AnyLogic has proven successful in the modeling of large and complex systems. [1] 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. Three further elements of AnyLogic that were acquired to simulate our model are ports, connections, and messages.

To simulate our model, active objects, ports, connections, and messages are used in the following way: The different symbols are modeled as active objects. Each active object has an input port and an output port. Through the output port, messages containing the actual activation grade of the symbol and the values of eventual properties are passed to other symbols to which this port is connected to. Through the input port, the symbol can receive information from other symbols which have a connection to it. Each of the incoming messages contains the activation grade of the symbol it was sent from. The activation grades are summed up. If this sum exceeds a certain threshold, the current symbol is activated and a message is passed to other symbols it is connected to via the output port. In the simulation, symbols that receive sensor values at the same time are computed in parallel.

As the system shall be able to learn connections between symbols from example, additional units have to be added that are only active during learning time and have the function to memorize the presented examples during the training phase until the correct connections are set. After the learning phase, they do no longer take influence on the system. However, due to shortage of space, the acquired concept for learning is not described in further detail in this article.

6. CONCLUSION AND OUTLOOK

Automatic surveillance systems as well as autonomous robots are technical systems which would profit from the ability of humanlike perception for effective, efficient, and flexible operation. In this article, a model for humanlike perception was introduced based on hierarchical modular fusion of multi-sensory data, symbolic information processing, integration of knowledge and memory, and learning. The model was inspired by findings from neuroscience. Information from diverse sensors is transformed into symbolic representations and processed in parallel in a modular, hierarchical fashion. Higher-level symbolic information is gained by combination of lower-level symbols. Feedbacks from higher levels to lower levels are possible. What symbols influence the activation of other symbols can be learned from examples. Stored knowledge of different forms influences the generation of symbols.

The model presented here is a first suggestion for a humanlike perceptive system. The following further investigations are planned:

Timing behavior of incoming information has only been considered marginally until now. The model also has to be extended to perceive scenarios that stretch over longer time periods and where the succession of events is important.

Until know, knowledge has been provided as explicit predefined rules. It would be interesting to extract these rules out of examples presented to the system.

Real human perception is also strongly influenced by emotions which have the function to evaluate the scenarios being perceived. Such an evaluation becomes especially important if the system needs to react adequately depending on the scenarios being perceived. An implementation of emotions into the model might be a step towards a perceptive system that is aware of the situation it perceives.

Until now, the model proposed has only been tested and verified on the PC with simulated parallel processing. To truly take advantage of the parallel distributed structure proposed, it would be interesting to implement the model into a chip which can perform real parallel processing.

7. REFERENCES

[1] *AnyLogic User's Manual*. Technologies Company Ltd, 2004.

[2] Bian, X., Abowd, G. D., and Rehg, J. M. Using Sound Source Localization to Monitor and Infer Activities in the Home. *GVU Technical Reports*, 2004.

[3] Cellier, F. E. and Greifeneder, J. *Continuous System Modeling*. Springer-Verlag New York, 1991.

[4] Costello, M. C. and Reichle, E. D. LSDNet: A Neural Network for Multisensory Perception. In *Proceedings of the Sixth International Conference on Cognitive Modeling*, 2004, p.341.

[5] Crowley, J. Situated Observation of Human Activity. In *Proceedings of the Computer Vision for Interactive and Intelligent Environment*, 2005, 97-108.

[6] Datteri, E., Teti, G., Laschi, C., Tamburrini, G., Dario, P., and Guglielmelli, E. Expected Perception: An Anticipation-Based Perception-Action Scheme in Robots. In *Proceedings*

of the International Conference on Intelligent Robots and Systems, 2003, 934-939.

[7] D'Orazio, T., Ianigro, M., Stella, E., Lovergine, F. P., and Distante, A. Mobile Robot Navigation by Multi-Sensory Integration. *IEEE International Conference on Robotics and Automation*, 1993, Vol. 2, 373-379.

[8] Ellis, D. P. W. *Prediction-driven Computational Auditory Scene Analysis*. PhD Thesis at the Massachusetts Institute of Technology, 1996.

[9] Ernst, M. O. and Bülthoff, H. H. Merging the Senses into a Robust Percept. *TRENDS in Cognitive Sciences*, 2004, Vol. 8, 162-169.

[10] Geisler, W. S. and Kersten, D. Illusions, Perception and Bayes. *Nature Neuroscience*, 2002, Vol. 5, 508-510.

[11] Goldstein, E. B. *Wahrnehmungspsychologie*, Chapter 14. Wadsworth Publishing, 2002.

[12] Goldstein, E. B. *Sensation and Perception*, Chapter 1. Spektrum Akademischer Verlag, 2007.

[13] Götzinger, S. O. *Processing and Symbolization of Ambient Sensor Data*. Master Thesis, Vienna University of Technology, 2006.

[14] Hanard, S. The Symbol Grounding Problem. *Physica D*, 1990, Vol. 42, 335-346.

[15] Joyce, D., Richards, L., Cangelosi, A., and Coventry, K. R. On the Foundations of Perceptual Symbol Systems: Specifying Embodied Representations via Connectionism. In Proceedings *of the Fifth International Conference on Cognitive Modeling*, 2003, 147-152.

[16] Kammermeier, P., Buss, M., and Schmidt, G. A Systems Theoretical Model for Human Perception in Multimodal Presence Systems. *IEEE/ASME Transactions of Mechatronics*, 2001, Vol. 6, 234-244.

[17] Luria, A.R. *The Working Brain: An Introduction to Neuropsychology*, Chapter 2. Basic Books, 1973.

[18] Michahelles, F., Antifakos, S., Schmidt, A., Schiele, B., and Beigl, M. Towards Distributed Awareness - An Artifact based Approach. In *Proceedings of the Sixth IEEE Workshop on Mobile Computing Systems and Applications*, 2004.

[19] Mirats Tur, J. M. *Qualitative Modelling of Complex Systems by Means of Fuzzy Inductive Reasoning. Variable Selection and Search Space Reduction*. Ph.D. Thesis, Universitat Politècnica de Catalunya, Barcelona, Spain, 2001.

[20] Muller, H. *Neural Binding of Space and Time: Spatial and Temporal Mechanisms of Feature-Object Binding*. Psychology Press, 2001.

[21] Oliver, N. and Horvitz, E. S-SEER: Selective Perception in a Multimodal Office Activity Recognition System. In *Proceedings of the First International Workshop of Machine Learning for Multimodal Interaction*, 2004, 122-135.

[22] Rhône-Alpes. *Project-Team PRIMA – Perception, Recognition and Integration for Interactive Environments*. Activity Report, 2005.

[23] Pratl, G., *Processing and Symbolization of Ambient Sensor Data*. Ph.D Thesis, Vienna University of Technology, 2006.

[24] Perry, M., Dowdall, A., Lines, L., and Hone, K. Multimodal and Ubiquitous Computing Systems: Supporting Independent-Living Older Users. *IEEE Transactions on Information Technology in Biomedicine*, Vol. 8, No. 3, 2004. 258-270.

[25] Sillanpää, J., Klapuri, A., Seppäne, J. and Virtanen, T. Recognition of Acoustic Noise Mixtures by Combined Bottom-up and Top-down Processing. In *Proceedings of the European Signal Processing Conference EUSIPCO*, 2000, Vol. 1, 335-338.

[26] Solms, M. and Turnbull, O. *The Brain and the Inner World: An Introduction to the Neuroscience of Subjective Experience*, Chapter 5. Other Press, New York, 2002.

[27] Zimmer, H. D., Mecklinger, A., and Lindenberger, U. *Handbook of Binding and Memory. Perspectives from Cognitive Neuroscience*. Oxford University Press, 2006.