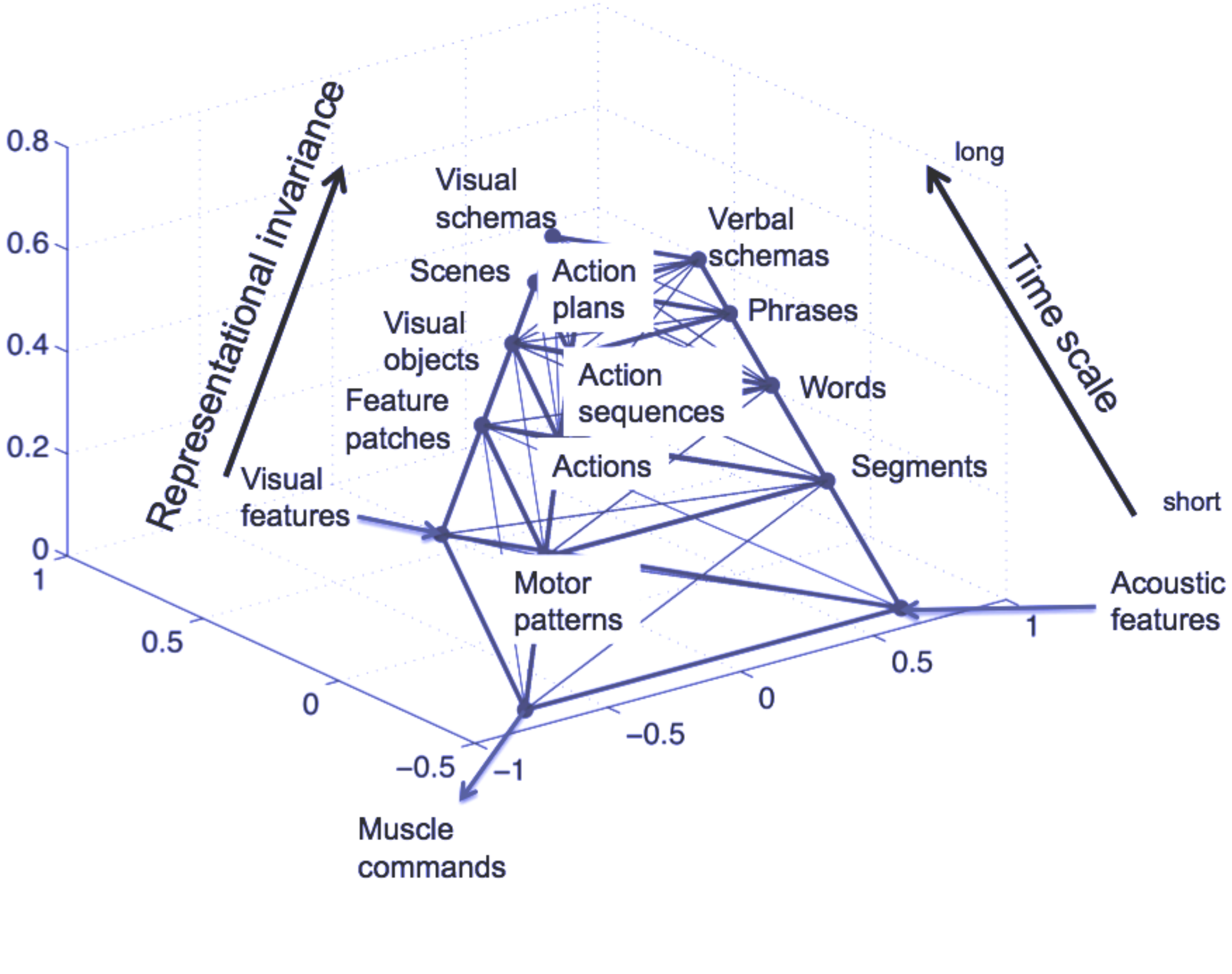
TECHNICAL WHITE PAPER:

AUTOMATIC LEARNING OF A TOPOLOGY OF ASSOCIATIONS FROM MULTIPLE DATA STREAMS

Okko Räsänen[1], Unto K. Laine[1] and Jukka P. Saarinen[2]

okko.rasanen@aalto.fi, unto.laine@aalto.fi, jukka.p.saarinen@nokia.com

[1]Department of Signal Processing and Acoustics, School of Electrical Engineering, Aalto University
[2]Nokia Research Center, Tampere, Finland



Keywords: unsupervised learning, associative learning, data mining, data analysis, predictive modeling, statistical learning, multimodal processing

Version 0.81

CONTENTS

SUMMARY

Huge amounts of data are constantly being produced and collected in the context of business analytics, industry processes, consumer products, robotics, scientific research, and, e.g., telecommunication networks. Making sense of massive amounts of parallel data streams and the contextual factors that contribute to the interpretation of the data is often extremely difficult. Automatic discovery of hidden patterns and the predictive dependencies between multiple data sources would therefore be extremely beneficial. An ability to represent original (e.g., sensory) data in terms of statistically significant patterns and their connections is also a necessary component in any cognitive computational system. Existing tools for associative or semantic data mining fail to address the discovery of important but a priori unknown patterns from sequential or spatially distributed data, but concentrate on the links between already known patterns such as words of the written language. Similarly, the majority of the pattern recognition applications concentrate on the classification of data into a finite number of pattern categories that are also known in advance. This limits the applicability of those methods to the domains where expert knowledge in the task is already available. In our research, we work towards an integrated cognitive architecture for unsupervised and hierarchical associative learning that can learn important patterns from multiple parallel data streams and how they are related to each other – we develop a system capable for automatic learning of a topology of associations (ALOTA). The goal is to perform intelligent data analysis with minimal human intervention and limited computational resources in any data domain with non-trivial temporally distributed patterns.

1. INTRODUCTION

Technological development has reached a point where countless devices such as phones, tablet-computers, medical systems, process industry machines or even network hubs and server clusters are able to sense their environment using a large variety of built-in sensors. Simultaneously, these devices can collect massive amounts of data regarding the internal operation of the system and, e.g., user actions on the device. Typical goals of the data collection include monitoring of anomalous situations, adaptation of the device behavior to the current internal or external context, or to simply achieve better understanding of the data through semi-automatic or manual analysis. Also, there is a constantly increasing interest towards autonomous (cognitive) machines that would succeed in their dedicated tasks by sensing, decision-making and making appropriate actions despite constantly changing or unpredictable external conditions that cannot be accounted for in the pre-programming of the system.

A major challenge in all the above applications is that the raw low-level data collected by the devices is typically too noisy and too variable to be used in high-level decision-making. It is not the instantaneous sensory readings, but how the instantaneous values combine over time or space to form larger patterns, that provides understanding of the data. Moreover, these patterns can depend on each other in similar manner that words that are connected to each other by the grammar and semantics of a language, forming higher-level patterns of patterns through compositional *hierarchies* (Pfleger, 2002). It is these *higher-level representations* that allow useful *generalizations* and thereby powerful *predictions* to be made on the basis of the input data (Hawkins, 2004; Haikonen, 2003).

Because relevant characteristics of the useful patterns are typically not known in advance, an intelligent system needs to be able to *learn* the patterns and their mutual dependencies in an unsupervised manner. This is also where the existing commercial data analysis solutions typically fail: they focus on modeling sets of individual data samples (e.g., SAS, SPSS etc.) or limit themselves to high-level data representations such as written text (Gavagai, Atigeo's xPatterns, Alphasense), being unable to capture structure of temporally distributed complex patterns of any generic data streams. When the need for patterning of low-level sensory data becomes explicit, the existing approach is to tailor a specialized solution to the given task by utilizing supervised pattern recognition techniques such as artificial neural networks or hidden-Markov models. This requires that the relevant pattern categories are known in advance in order to do the laborious manual preparation of training sets that contain exemplars of these patterns, effectively limiting these approaches to well-understood signal domains and excluding exploratory data analysis.

Another lately emerged approach is to use multilayered hierarchical artificial neural networks (ANNs) referred to as *Deep Learning* networks (see Bengio, 2009). Despite their recent success in many pattern recognition tasks, the ANN-based deep systems are not yet very well understood: they are difficult to optimize, their training to any specific sub-task requiring lots of time, manual effort, expertise, and iterative trial by error. Also, the logic behind the learned structures and the system decisions based on input data are difficult to analyze in deep networks, essentially making them black boxes with desired input/output characteristics. This makes their scalability to universal multimodal data analysis difficult as long as the theoretical considerations or practical rules of thumb in training of such networks are lacking. This is in contrast to our solution where we have explicit access to all processes in the learning system and thereby all decision made by the system can be traced back to the original data streams.

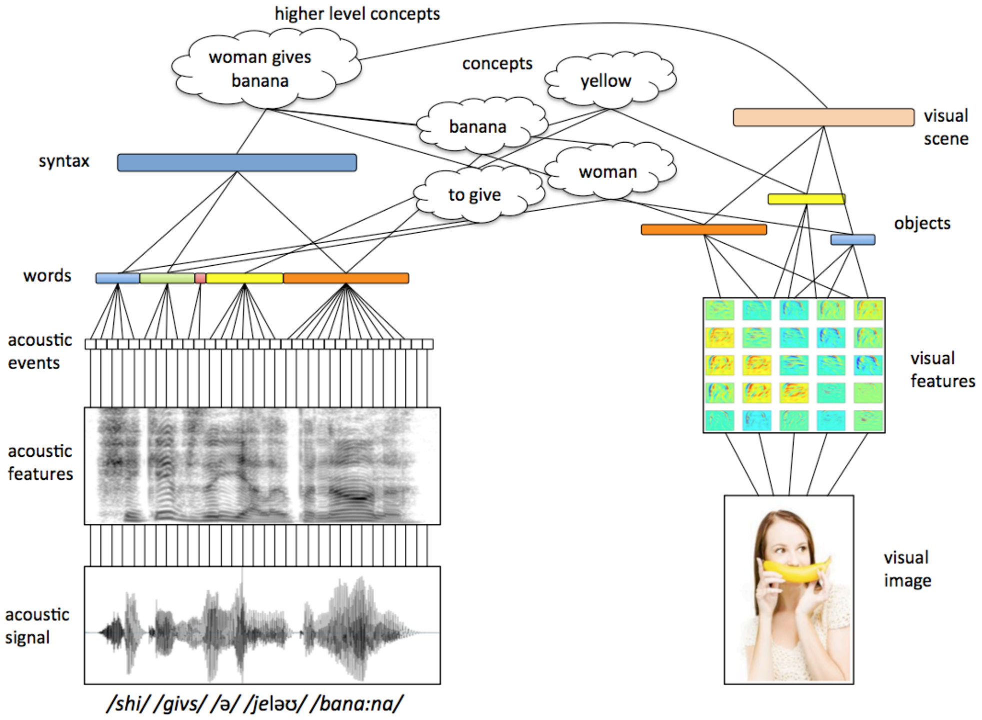


Figure 1: An example of compositional multimodal structure in speech understanding for utterance "*She gives a yellow banana*".

Finally, there is a central issue with the *meaning* of patterns. For any piece of data, be it a series of stock exchange values or a written word, the data itself carries no meaning. Instead, the meaning comes from way that the data is associated to some other entity in the mind of the agent making the interpretation (see Figure 1). In manual data analysis, it means that we use our subjective experience of the world to reflect on the importance of the patterns in the data. For autonomous computational systems, it means that the patterns in the data contain cues regarding desired functionality of the system. These cues can be based on correlations (things occur together) or causalities (things follow each other with a specific temporal order) between patterns. In cognitive systems, we call these links *associations*. In similar vein, *information* is data that is able to light up these associations when perceived. What is relevant is that *given a piece of data in one domain (e.g., hearing a spoken word), we receive information about the current state of the other domains and also about the possible future states of the domains*. Coupled to the understanding of how our own actions have consequences on the perceived environment, we are able to proactively manipulate our own behavior towards a desired state of being. This forms a so-called action-perception-loop.

Due to the nature of meaning as connectivity between patterns in different domains, *multimodal associative learning* becomes essential in autonomous understanding of any data. In fact, all traditional supervised learning schemes can be considered as a special case of associative learning where the learning simply takes place between the actual sensory signals and the manually prepared, artificial, label streams. In unsupervised associative learning, the meaning of the data emerges directly from the predictive associations between the patterns in the multiple input and output streams and there is no need for manual labelling. Value of the discovered dependencies is not dependent on human interpretation, but is an inherent outcome of the interplay between system's internal criteria of successful behavior and the environment that the system is subjected to through its input and output capabilities. Every action of an intelligent system is based on the predictive knowledge that the action leads to something useful or to new knowledge, and this knowledge can be incredibly difficult to pre-program for complex environments. Naturally, manual human interpretation of the discovered patterns is possible if explicit summaries of the learned dependencies are useful outside the autonomous behaviour of the learning system. Figure 2 illustrates how different sensory and motor modalities contribute to mental concepts such as objects, actions, or adjectives in the human mind.

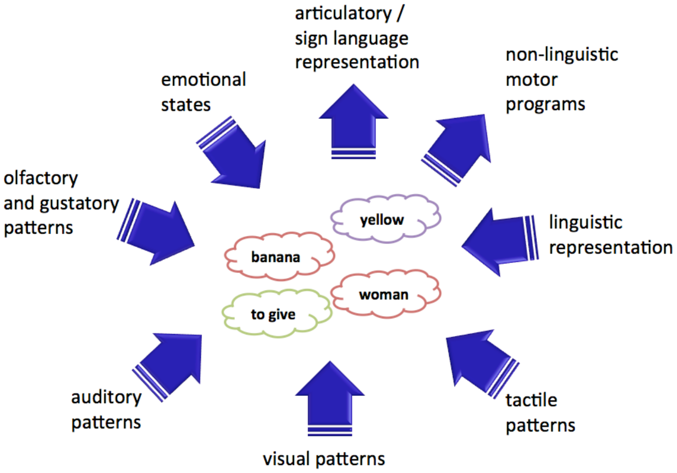


Figure 2: Meaningful concepts. Instead of being platonic ideas, meaningful concepts are multimodal associations with numerous components that have varying importance depending on the nature of the concept. These connections are learned instead of being inherited.

1.1 Learning multimodal concepts as Automatic Learning of a Topology of Associations (ALOTA)

From the perspective of mathematical theory, the general framework of modelling connectivity of patterns inside and across modalities can be formulated as a Topology of Associations (TOA). The goal of an intelligent learning system is then to derive such a topology, i.e., to perform **A**utomatic **L**earning **o**f **a** **T**opology **o**f **A**ssociations in the data (**ALOTA**).

In the basic topological framework, we define the following concepts: an element a_i from a vocabulary of elements $A \subseteq \mathbb{N}_1$ corresponding to input and output states of the system and/or patterns derived from them; an associative link $r(a_{i,},a_j) \rightarrow \mathbb{N}_1$ from a set of all associations R and the related association strength $P(a_i,a_j) \rightarrow \mathbb{R}$; and finally a topology τ_R over the set of all associations R. Technically, τ_R is a topology when it is a family of subsets of all associations R so that 1) empty set and R are included in τ_R, 2) any union of elements of τ_R are included in τ_R, and 3) any intersection of finitely many elements of τ_R is included in τ_R.

In ALOTA, the goal of the learning algorithm is to derive a topology τ_R of associations that maximizes the overall predictive capability of the system, i.e., maximizes association strengths $P(a_i,a_j)$ across all $i,j \subseteq \mathbb{N}_1$ hierarchically generated by associative links $r(a_{i,},a_j)$ when measured over all input and output modalities of the system. In other words, each association between two elements $r(a_{i,},a_j)$ generates a new element a_{new}, that can be then used as a component for further associations. Naturally, the associative links are not mutually exclusive, but in a fully connected R, all elements a_i are associated to all other elements a_j, $j \subseteq \mathbb{N}_1$. As each unique association maps to a new element a_{new} in the same element space, there are infinitely many elements and associations. In learning of a useful topology, only the associations that have statistical significance are learned ($P(a_i,a_j) > \delta$). Note that the elements a_i can originate from one or more data streams (modalities), and therefore the associative links are learned inside and across modalities.

1.2 Towards an integrated computational theory for ALOTA

In our research, we work towards an integrated computational theory and a practical implementation of unsupervised hierarchical learning of association topologies that could accomplish the following tasks:

1) Discovery of statistically significant patterns from sequential or multivariate data without a priori knowledge of the relevant pattern characteristics.

2) Discovery and quantification of statistical dependencies between patterns in multiple parallel data streams.

Also, in case of systems with the ability to act on the perceived environment

3) Automatic prediction of appropriate actions (outputs) based on the input data.

Moreover, we aim to achieve these goals with a computational implementation that

4) Performs in real time using real sensory data.

5) Can learn indefinite amount of data over time.

We will next review the basic advantages of ALOTA. The second section shortly reviews some neurobiological motivation and inspiration for our computational architecture, discussing Antonio Damasio's theory of how human memory is organized to support multimodal associations (Damasio, 1989; Meyer & Damasio, 2009). The third section gives an overview to the architectural principles in our work, and the fourth section provides demonstrations of computational experiments using the technology we have developed so far.

1.3 Benefits of unsupervised learning of association topologies

- ALOTA can be applied structure discovery in domains where a priori knowledge of the data patterns is not available or manual characterization of the patterns is difficult.

- Structure discovery is not limited to raw-data level, but the system performs hierarchical data granulation (abstractions) based both on 1) bottom-up statistical analysis and 2) predictive structure across multiple data streams, allowing the system to discover direct and indirect links between high-level concepts that are responsible for generating the data.

- ALOTA is a necessary component in cognitive computational systems that use sensing and data collection to infer appropriate actions in an autonomous manner. Structurally meaningful representation of the sensory domains forms the basis for action selection.

- Despite being unsupervised in nature, all processing in the system can be supplemented with human labelling of the data as an additional input stream, allowing the discovery of dependencies between data structures and a priori human knowledge.

2. NEUROBIOLOGICAL MOTIVATION: CONVERGENCE-DIVERGENCE ZONE (CDZ) ARCHITECTURE

The computational architecture for unsupervised hierarchical pattern discovery is inspired by the convergence-divergence zone (CDZ) architecture by Antonio Damasio (Damasio, 1989; Meyer & Damasio, 2009). The CDZ architecture provides a neurophysiologically valid description of how mammalian brain organizes sensory and motor patterns in its memory, how this organization enables associative learning between modalities, and how sensory and motor imagery are implemented in the human brain. The basic structure of the CDZ architecture is outlined in Figure 3.

The CDZ architecture is motivated by the finding that human perception is essentially multimodal and perceptual processing and imagery are characterized by time-locked joint neural activity across early sensory and motor cortices. For example, visual perception of lip movements drives our speech perception whereas mental rotation of visual objects is dependent on intact motor cortex. In similar manner, our tactual ability to interpret orientation of grated surfaces is dependent on the neural activity at the visual cortex and conflicting visual information or transcranial magnetic stimulation of the visual cortex interferes with the ability to feel different orientations. Multimodality not only increases the signal-to-noise ratio of perception due to the complementary information in different modalities, but also directly enables the construction of integrated representations of the external world from the series of separated percepts in different sensory organs.

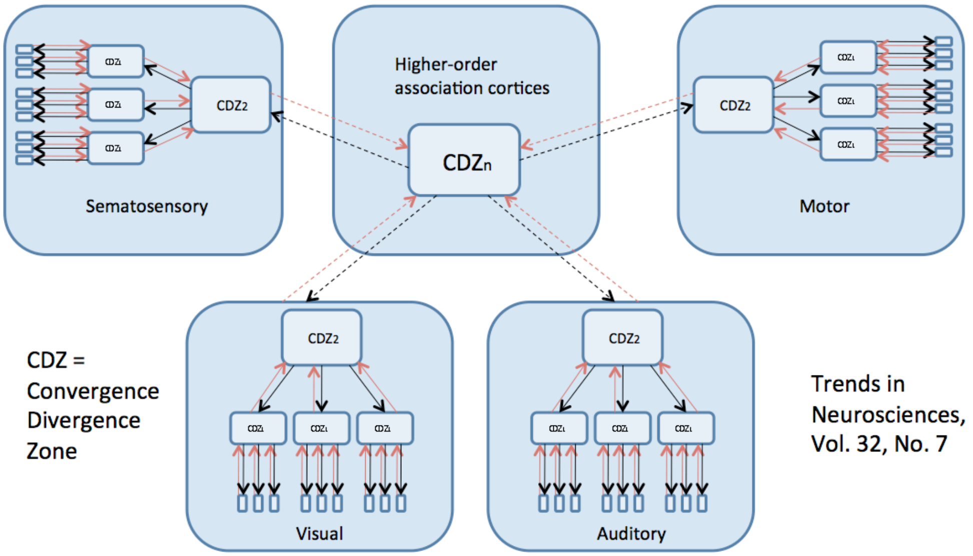


Figure 3: The CDZ architecture (adapted from Meyer & Damasio, 2009). All input and output modalities are reciprocally connected to each other through a hierarchical memory system. Representations at each intermediate layer are connected to integrative higher-level units through converging connections. Similarly, all units in the intermediate layers can reconstruct representations at the lower layer through the diverging connections.

The basic idea in CDZ is that each modality consists of feature detectors that are selective towards specific aspects of the incoming sensory stream. Signals such as parts of visual objects or individual acoustic features of spoken words embedded in the sensory stream activate a number of feature detectors (Figure 3). Each feature detector projects its activity to an integrative unit called convergence-divergence zone (CDZ) where the concurrent combinatorial activations of the low-level detectors are analyzed and combined to form higher-level patterns. For all but the most high-level association areas, each CDZ projects to a higher-level CDZ that integrates information across multiple CDZs. This allows a hierarchy of increasingly abstract representations, forming wholes from parts by learning what types of entities co-occur to build up larger entities. At the highest levels of the processing, the CDZs start to receive converging information from CDZs of other sensory and motor domains, leading to the emergence of multimodal concepts that link multiple aspects of the same entity into a unified representation.

The connections between CDZs are reciprocal so that one integrative unit in a CDZ can activate all lower-level feature detectors associated with the unit in a top-down manner, allowing reconstruction of the original sensory or motor patterns. Due to these reciprocal associative connections, partial feature representation in one modality can first lead to object completion in the same modality (e.g., visual object completion from partially occluded image), but also activate the object related representations in other domains (e.g., what are the typical sounds and haptic properties of the object).

In the CDZ architecture, the *original sensory data as such is not transferred* from the peripheral sensory systems to any kind of higher level memory units, but only the combinatorial arrangements between sensory features and links between higher-level links are learned and updated through experience. The memory does not reside in one specific point in the system, but the memory functions are served all over the system by the associative links between lower-level inputs. When memory recall is taking place, the original sensory activation is approximated by activating the top-down connections related to the concept in question, activating the CDZs and ultimately the proper feature detectors in the descending path. In other words, the CDZs record information of how to reconstruct multimodal sensory experiences from partial external or internal cues.

In ALOTA, the basic principles of memory organization follow the CDZ architecture. The essential idea is to process sensory data of each modality in an increasing hierarchy of abstraction, starting from low-level raw data and proceeding to features, typical feature combinations, and then typical temporal or spatial patterns of local features. These increasingly invariant representations of the modality specific patterns are then analyzed in the context of patterns in other modalities in order to learn multimodal associations. Also, the episodic memories can be stored as sequences of associations at the highest level of invariance, allowing the modeling of temporally sequenced combinatorial arrangement of events, objects, and actions perceived through the sensory and motor systems of the computational system. Our architecture also follows the generality principle of CDZ, namely that for everything else but the low-level sensors and actuators of the system, the processing principles are universal across modalities and signal representations are mediated using a universal coding scheme.

3. IMPLEMENTING A HIERARCHICAL ASSOCIATIVE MEMORY

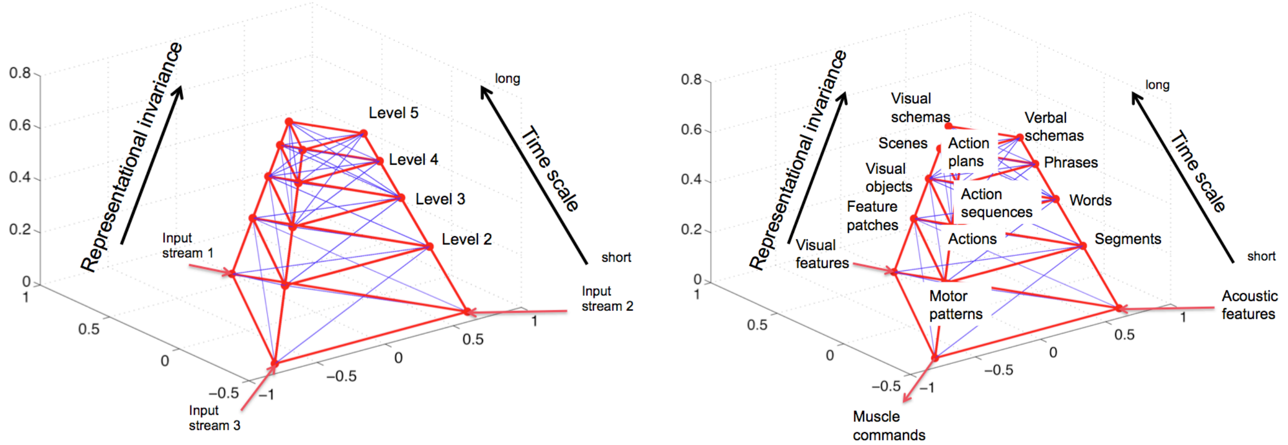


Figure 4: A schematic view of hierarchical associative learning with three input channels (left) and an example of representational hierarchy that could be learned for vision, auditory perception and motor actions (right). All modalities are connected to each other at the same level of hierarchy and across different layers of hierarchy. The horizontal connections allow associative learning between visual percepts of an entire object (e.g., *a lion*) and the acoustic patterns that go with the object (*a roar* and the word *"lion"*), whereas the vertical connections across modalities (blue thin lines) allow learning based on cross-modal supervision that give rise to more abstract and general, higher-level, patterns.

3.1 Basic specifications

A successful architecture for hierarchical associative learning must be able to 1) **sense** its sensory environment, and optionally, to **act** upon the environment, 2) to **learn** patterns from data and later **recognize** these patterns even from partial cues, 3) to **associate** patterns across time and across modalities. In addition, in all active applications such as intelligent devices the system should perform **on-line** and **scale up** for indefinite amount of data. In practice, the system cannot store every input seen in the past, but the size of the system memory must be fixed or grow in a sub-linear space with respect to time. Also, since sensory and motor representations required in different tasks vary in granularity and in detail, e.g., sets of patterns forming even larger patterns, it is most efficient to process patterns in a compositional **hierarchy** (see Figure 4) similarly to the CDZ architecture. As higher-level patterns are represented as combinations of lower level patterns, the system can flexibly encode a huge number of patterns using a finite number of computational units (see also Bengio, 2009).

Sensing and **acting** operations are mainly defined by the domain in which the system is used and by the available input and output modalities in that domain. For example, a full-scale cognitive robot could sense its environment through cameras, microphones, tactile sensors and accelerometers, and act upon the environment using its actuators. In analogy, a mobile phone can sense through microphones, accelerometers, GPS sensors and user actions on the operating system, whereas the output modalities can correspond to different software activities and displays (e.g., user interface behavior), wireless communications, or even speech produced by a built-in speech synthesizer. The detailed implementations of the input and output modalities are not of interest to the actual learning system, but they simply provide an interface between the external world and the learning system. Inside the system, all modalities are represented using a universal code. This coding is achieved by quantification of input and output signals into a finite number of discrete states on a nominal scale. Typically, a sensory specific intermediate feature representation can used between the raw signal and a quantized version of the signal in order to achieve compression of redundancies and to discard non-relevant noise (e.g., FFT for audio or Gabor patches for images).

Learning and recognition of patterns is the essence of the architecture. Patterning can be considered as an abstraction process where noisy and variable sensory time-series (realizations of patterns) are represented by occurrences of categorically perceived units, or patterns. Another viewpoint is that the sensory data is interpreted in terms of some contextual variable that links different physical patterns to each other. What is essential is that the system must be able to interpret the incoming data in a sufficiently invariant form that allows the generalization of learned dependencies between patterns to the whole group of functionally equivalent patterns. This is in contrast to the learning of dependencies between individual pattern realizations, which would be useless since the patterns will rarely if ever recur in the exactly same form. Recognition simply means that the system should be able to interpret new data in terms of the previously learned patterns whenever this type of generalization is beneficial and thereby to use the previously established associations to understand the current situation.

Associative learning of patterns refers to the discovery of mutual dependencies between patterns that can be used to construct an internal representation of how the external context is organized. To be precise, the pattern discovery process is also a form of associative learning between signal states in single- or multi-stream conditions. However, here we specifically refer to the learning of *predictive dependencies* between high-level patterns. These dependencies can either reside inside a single data stream (such as the grammar of a language or typical arrangements of physical objects), but more often they span across multiple streams, forming multimodal concepts (such as words acts as signs for some external referents). Ultimately, the whole idea of cognitive systems crystallizes to the ability to infer useful actions (motor patterns) from the sensed input (sensory patterns). In all but in the most trivial domains, the predictions of the appropriate actions in given circumstances must be based on the previously learned associations and the ability to choose the actions with highest expected rewards.

3.2 On the theory of patterns

The goal of learning meaningful patterns from sensory data without a priori knowledge of the actual patterns already poses two difficult questions: 1) *what is a pattern* and 2) *where does the meaning emerge from*? According to our view, these questions can be approached from two different perspectives (see also Figure 5).

According to the first (traditional) view, and assuming a finite state space representation for a sensory signal, a pattern can be considered as a probabilistic construct of elementary events (or observations) that are dependent on each other in time or space. The dependency does not have to be deterministic, but above chance level probability of observing two or more elementary events in a specific configuration can already be considered as a pattern. For example, an acoustic signal corresponding to a spoken word can be interpreted as a specific distribution of signal energy in time and frequency, analyzed up to a desired resolution. However, patterns discovered from a single data stream alone do not carry any meaning. According to the first view, the meaning of the pattern only emerges when the observed pattern is associated (*grounded*) to a jointly occurring contextual variable perceived through another modality or a variable representing an internal state of the system (e.g., active concepts in memory or current emotional state). For a spoken word such as "*a ball*", the contextual variable could correspond to the visual or haptic percept of a ball. In other words, the pattern as such can be defined without the grounding component, but the meaning emerges only through the grounding process.

The inherent problem with the first viewpoint is that even though the quantification of statistical dependencies in time and frequency is possible, it is not possible to derive an "optimal" and finite set of distinct patterns (or categories) for a data set without imposing some sort of a priori model for the data (this is true for maximum-likelihood, minimum description length and minimum entropy approaches). The goodness of a representation is always measured with respect to the task or context against which the patterns are reflected.

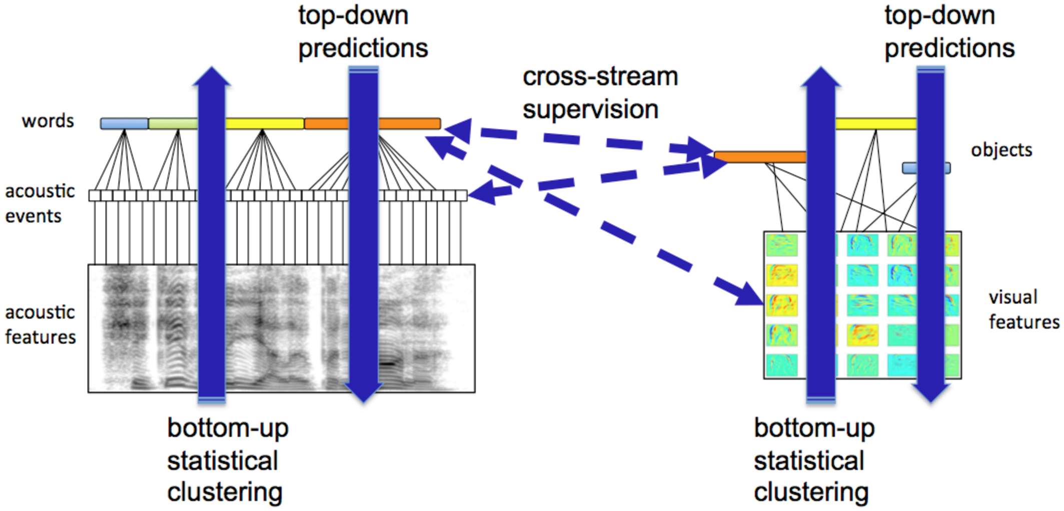


Figure 5: Bottom-up, top-down and cross-modal constraints in pattern discovery and recognition. Bottom-up learning is purely based on the statistical structure of the input signal, whereas top-down predictions may impose biases to the bottom-up processing through some prior knowledge or dynamic, task-related, factors. Importantly, information exchange between input and output modalities allows the system to learn dependencies across modalities, all modalities effectively supervising each other's learning processes.

The second viewpoint argues that the patterns and their meanings are inherently intertwined so that there is no other without the other. According to this view, any processing beyond the learning of low-level sensory receptive fields always takes place in the context of multiple temporally proximate perceptions and mental states (memory, emotions) of the perceiver, and that this context affects the way how incoming sensory information from each modality is interpreted. This automatically attaches a set of multimodal associations to each percept and the elementary sensory events become bound together not only by their mutual co-occurrences but also by their shared context. In this case, the learning of pattern categories is no longer a question of bottom-up statistical clustering, but the categories are actually a function of the context: the sensory inputs belonging to the same pattern category are those that have equivalent predictions of the state of the world in other modalities, or equivalently, the current context defines the boundaries of a pattern category. In a sense, the idea of non-chance level dependency of elementary events in the first viewpoint is expanded to allow these elementary events or states to occur across multiple input- and output modalities of the system.

The obvious challenge with the latter viewpoint is that the estimation of all cross-modal dependencies through, e.g., normal joint probabilities is not possible due to the high dimensionality of the problem. Also, the direct associations between low-level sensory events (e.g., spectral features and visual receptive fields) may not be meaningful, but the useful dependency structure only emerges when at least one of the signal representations is sufficiently invariant to act as "labeling" for the remaining modalities. Therefore both unsupervised and weakly supervised learning are needed in a successful architecture: the unsupervised learning process provides initial pattern representations based purely on the statistical dependencies in the input data. These initial patterns can then act as labels that supervise learning in other modalities, allowing learning and representation of patterns that are based on their multimodal links instead of bottom-up statistics.

3.3 Technological solutions for ALOTA

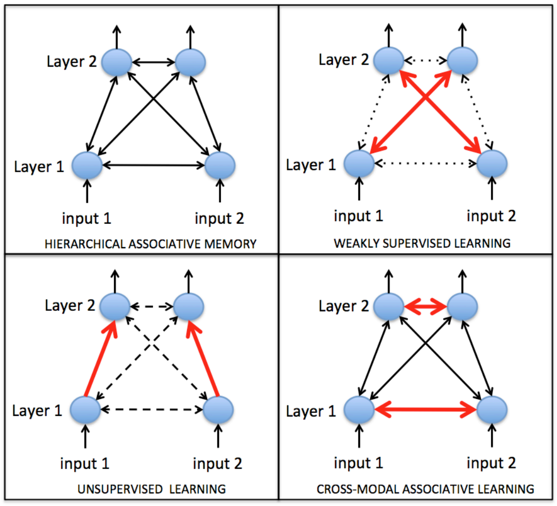


Figure 6: Types of learning processes required in hierarchical associative learning. The relationships between different modalities and layers of hierarchy are emphasized with red color for different learning processes.

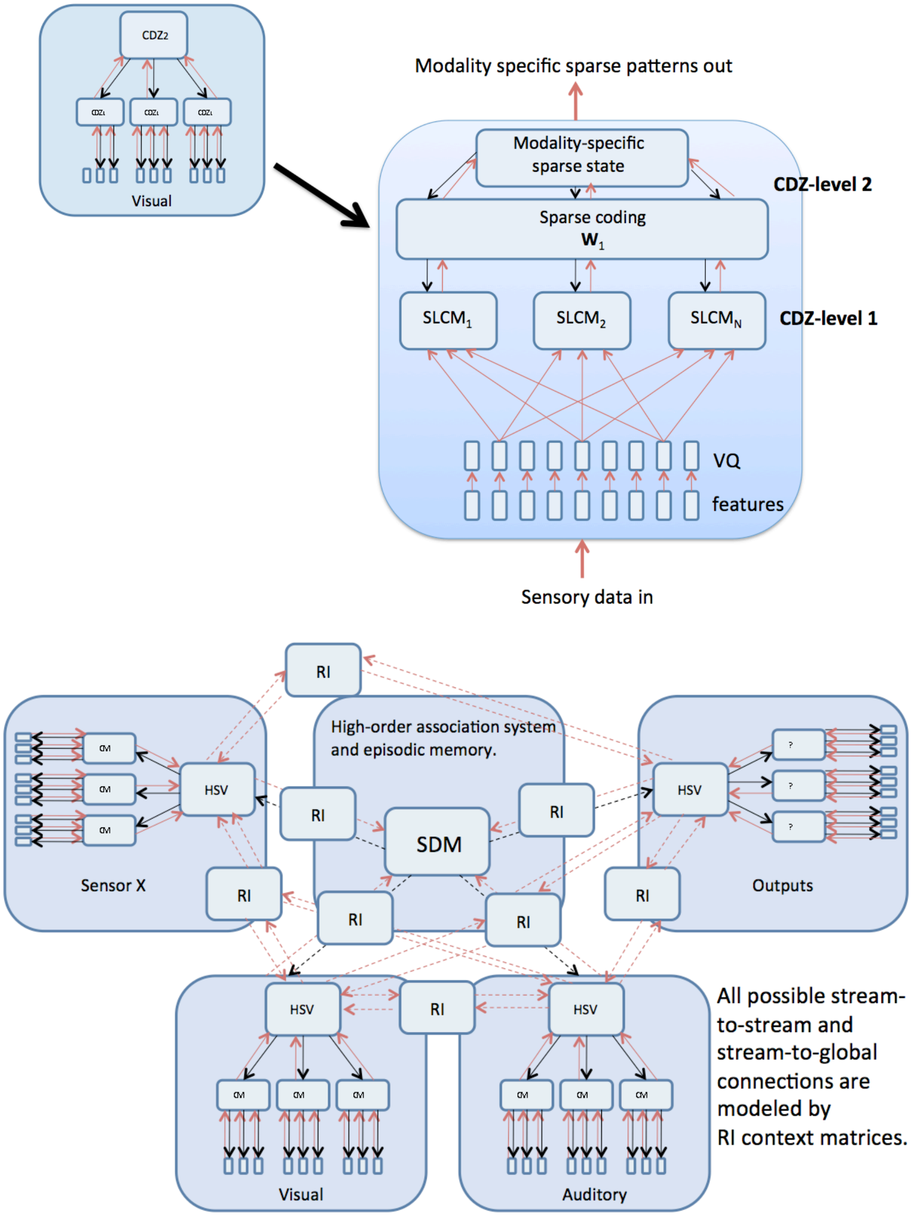
As motivated in the previous sections, the learning processes in ALOTA require both **unsupervised** and **weakly supervised learning** in order to perform pattern discovery and recognition from complex temporally distributed data, and to find compositional hierarchies of patterns (Figure 6). In addition, methods for robust **cross-modal associative learning** with different data qualities are needed, and all the different methods should integrate with each other. Finally, the ability to process and learn infinite amounts of training data within a finite memory is a desired property for the system.

In our studies, we have already developed basic understanding and methodology to address all these learning problems. In addition to using standard pre-processing; feature extraction, and vector quantization techniques, we have developed **Self-Learning Vector Quantization** (**SLVQ**; Räsänen et al., 2009) that allows incremental quantization of any multivariate time-series into a finite state space without a priori knowledge of the relevant number of state space partitions. In addition, the **Concept Matrix** (**CM**; Räsänen & Laine, 2012) algorithm and its purely unsupervised variant **Self-Learning Concept Matrix** (**SLCM**; Räsänen, 2011) are especially designed for modelling statistical structure in sequential data. Whereas CM accomplishes the weakly supervised learning problem and discriminative pattern recognition effectively with a minimal number of assumptions regarding the signal structure, the SLCM allows the discovery of recurring patterns from data without any supervising data streams, making it effective for bootstrapping of multimodal learning in cases where all available modalities correspond to low-level highly variable data streams. Both algorithms are computationally light and can be used in real-time implementations for multiple parallel data streams even in environments with very limited computational power. In addition, the **Hybrid Model Learner** (**HML**) –algorithm (Laine, 2011) allows the inference of hierarchical compositional structure (grammar) of sequential data using information theoretic criteria, allowing the learning of generative models for the data.

Since originally learned pattern category boundaries should not be hard-coded, but can depend dynamically on the current context and task of the system, we utilize the ideas from **Latent Semantic Analysis** (Landauer & Dumais, 1997) and **Random Indexing** (Kanerva et al., 2000; Sahlgren, 2005) to measure pattern synonymy in different contexts in different modalities in time-variant manner.

In order to solve scalability problems and to obtain generalizing high-level associative links that can be activated by partial cues, we can apply **Sparse Distributed Coding** as a universal code for the system, enabling learning in a fixed memory space (see Kanerva, 2009 for further details). This representation also directly enables the use of **Sparse Distributed Memory** (SDM; Kanerva, 1993) architecture for episodic learning at the topic of the abstraction hierarchy. Whereas the earlier problems with SDM-like memory systems have been that they are poor in accounting for temporally evolving structure (see Jockel, 2009), our tools for pattern discovery and recognition allow representation of time-varying data as more invariant abstract patterns for which SDM-like memory structures are designed.

Figure 7 shows a putative overview of a multimodal pattern discovery system along the lines of the CDZ framework. In this system, each modality is first processed separately to discover recurring pattern and code them in an invariant form. These patterns are then represented with sparse coding, and time-dependent random indexing is used to discover associative connections and synonymous patterns across different modalities. Finally, SDM is used as a heteroassociative memory for multimodal episodes, allowing storage and retrieval of holistic representations of previously experienced situations.



All possible stream-to-stream and stream-to-global connections are modeled by RI context matrices.

Figure 7: A putative computational implementation of ALOTA following the structural organization of the neural CDZ architecture. Top: a schematic view of one sensory channel with feature detectors and bottom-up organization of CDZs. Features of the sensory stream are first quantized into a finite state space, from which recurring pattern are discovered. These patterns are then coded with random sparse vectors, allowing a flexible compositional representation of the active patterns in the modality as a single point in the hyperdimensional space. Bottom: a schematic view of associative learning between modalities mediated by sparse coding. Combinations of patterns from different modalities can be effectively represented in the sparse hyperdimensional space. Synonymy of patterns and pattern combinations can be estimated dynamically using the principles of random indexing. Sparse distributed memory allows episodic coding of sequential representations that integrate information from all modalities.

4. TECHNOLOGY DEMONSTRATIONS

3.1 Unsupervised learning

3.1.1 Unsupervised learning of words from continuous speech

Word segmentation from continuous speech is a difficult task that is faced by human infants when they start to learn their native language. The difficulty is due to the fact that spoken words are rarely separated by pauses or any other universal cues that would signify word boundaries equally in all languages.

In our experiments (as reported in Räsänen, 2011), we applied our SLCM pattern discovery algorithm to word discovery from continuous speech. Instead of using any a priori linguistic knowledge of phones, syllables or words, our method analyzes the statistical dependencies between atomic acoustic events in a bottom-up manner. As a result, the method builds a collection of pattern recognizers that become selective to recurring structures in the data. In the case of continuous speech, the recognizers became selective towards specific spoken words or combinations of often co-occurring short words (Figure 8). As a result, we were the first ones to show that it is possible to find recurring word-like units from real speech in an unsupervised manner without storing the entire history of perceived signals to the system memory, but by simply storing statistical models of the encountered patterns in an incremental manner.

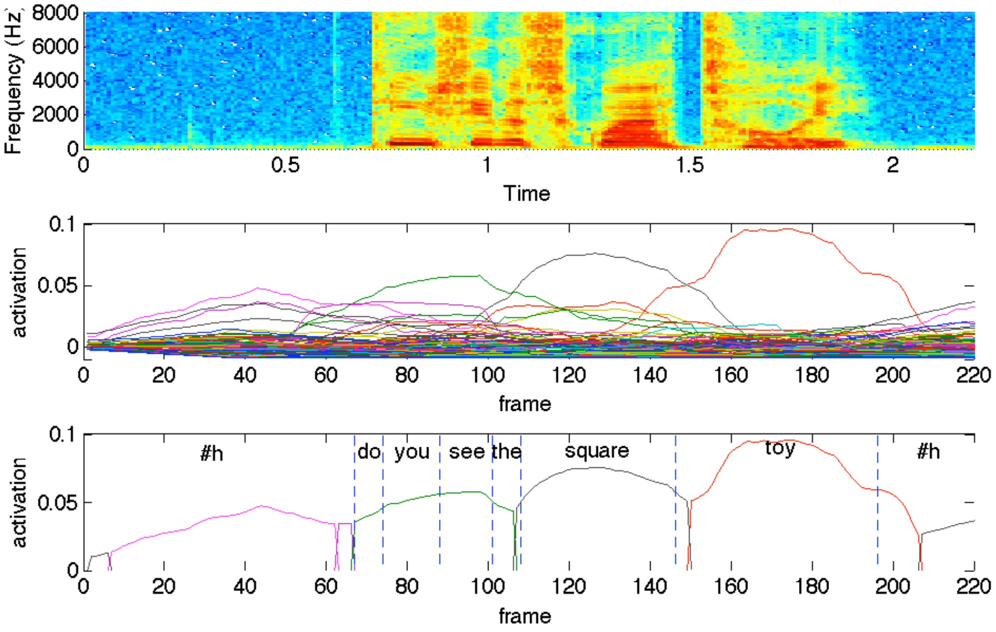


Figure 8: A recognition example for the utterance "Do you see the square toy?". Spectrogram of the utterance is shown at top and activation of all models are shown in the middle. At the bottom, only the winning model for each moment in time is chosen, leading to a segmentation of the input. Manually annotated segment boundaries (the references) are indicated by dashed lines.

3.2 Weakly-supervised and supervised learning

3.2.1 Audiovisual keyword recognition

Many learning tasks can be perceived as weakly supervised pattern discovery problems in the absence of explicit teaching. Weak supervision refers here to learning conditions where the training samples are not explicitly segmented and aligned to pinpoint their belongingness to a specific target category. Instead, a number of possible contextual variables (target classes) are presented in parallel with the input, but it is not known whether all these classes are present in the data and, if they are, where they are. The task of the learning algorithm is then to discover those patterns from the data that are relevant for each contextual variable.

In order to solve the task, our CM algorithm combines information from two input streams and finds co-occurrence relations between them. It learns recurring structures in similar contexts and recognizes them from new input. Contrary to the hidden Markov models (HMMs) that are the state-of-the-art in speech recognition and widely used in many other pattern recognition tasks, our approach does not make the Markov property assumption regarding independence of the subsequent states. This makes it capable of finding structures between non-adjacent events and robust against temporally local distortions.

We have evaluated the CM algorithm in the discovery and recognition of keywords from continuous speech when the spoken utterances are paired with visual (unaligned and unordered) labels simulating visual input attended by the learner. This simulates the word learning process of a human infant. The results show that the method is successful in acquiring high quality recognizers for all of the 50 keywords in the vocabulary without being explicitly taught any of the words (Figure 9). Also, we have evaluated the same system in spoken digit learning and recognition, leading to 96.11 % recognition rate with one-pass incremental training, and showing notable robustness against increasing levels of additive noise (see Räsänen & Laine, 2012 for more details). An example of digit recognition process is shown in Figure 10.

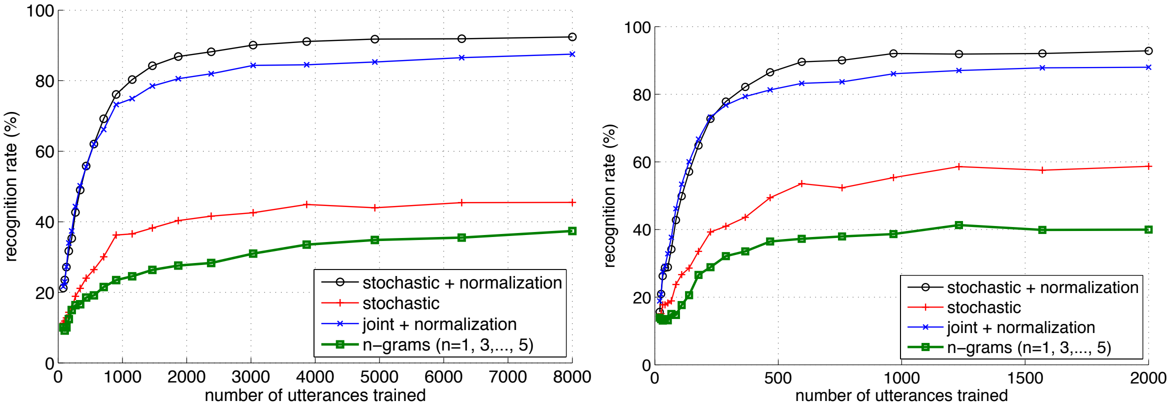


Figure 9: Recognition rates for 50–keyword vocabulary as a function of the amount of training data used to train the recognizers. Left: learning with data from 4 different talkers, right: data from only one talker.

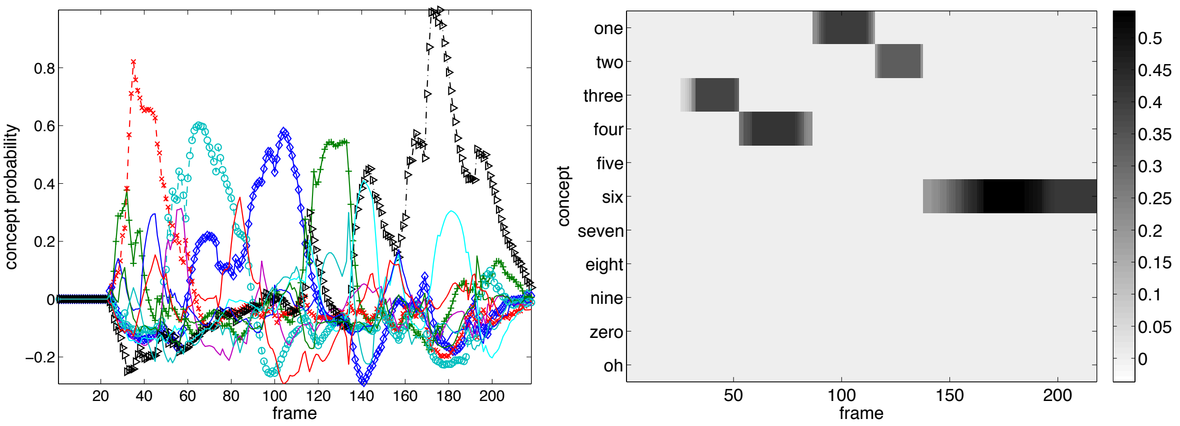


Figure 10: Left: cumulative activation curves of 11 digit recognizers in recognition of the utterance "three-four-one-two-six". Right: activation of the recognizers after temporal smoothing and inhibition.

3.2.2 Combination of input streams in supervised context recognition

The CM algorithm can be also used for easy decision-stage combination of multiple data streams for enhanced pattern classification (Räsänen et al., 2011). The probabilistic formulation behind CM guarantees that the inclusion of additional input streams to the classification stage does not bias the process towards erroneous classifications. Instead, if the additional data streams contain complementary information, the classification performance is increased.

Figure 11 shows an example of input stream combination. In this case, audio and accelerometer data measured from a mobile phone were used as input to the CM classifier, and the task was to find the most likely physical activity and environment of the mobile phone user. As can be observed, our CM classifier outperforms the other compared classifiers, namely discrete Hidden-Markov models (dHMM), minimum distance classifier (MDC), and k-nearest neighbors (kNN), in the combination. Also, the optimal combination of the input streams is achieved by simply taking the average of audio and acceleration classifier outputs. Finally, the CM classifier is computationally very light, and both training and classification can be easily done real-time in a standard smart phone.

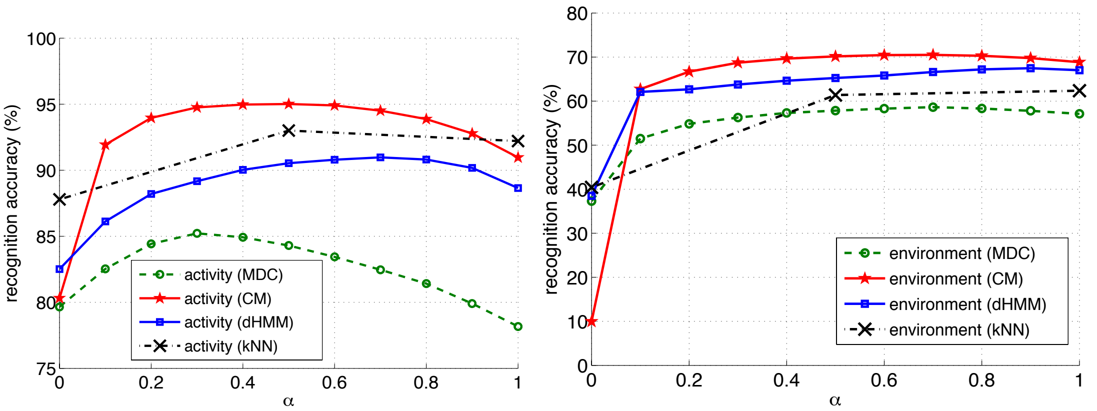


Figure 11: Mean recognition accuracies of physical contexts (left) and auditory environments (right) of mobile phone users using different classifiers. Results are shown as a function of weighting between acceleration and audio data (α = 0 corresponds to pure acceleration whereas α = 1 stands for pure audio). The CM algorithm performs outperforms other classifiers in combined classification, always providing optimal performance with α = 0.5 (unweighted) combination of the input streams.

3.3 Cross-modal associative learning

Normal bottom-up statistical pattern discovery can be used to discover recurring structures from a single data stream, allowing the representation of the data in terms of higher-level patterns that have a temporal or spatial extent (Figure 12). However, the patterns discovered in this manner have no intrinsic meaning, but the meaning emerges from the predictive associations between the patterns and some other states, or patterns, of the world. Therefore, cross-modal associative learning is required to learn these dependencies, enabling predictive modeling of multimodal data. In the general case, one does not know in advance which levels of representation (level 1 low-level features, level 2 patterns, level 3 or patterns of patterns) are most important in learning of the predictive dependencies, especially when different modalities can contain qualitatively different signals. Therefore it is beneficial to study the dependencies across modalities also across different levels of representational hierarchy (Figure 13).

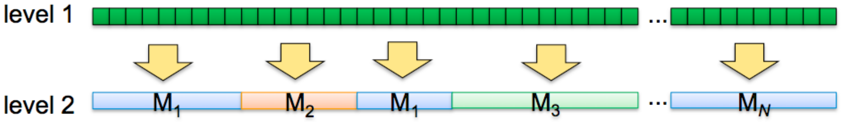


Figure 12: Unsupervised learning of recurring patterns in one data stream. The original frame-by-frame representation of the signal (level 1) can be replaced with a series of higher-level patterns (level 2)

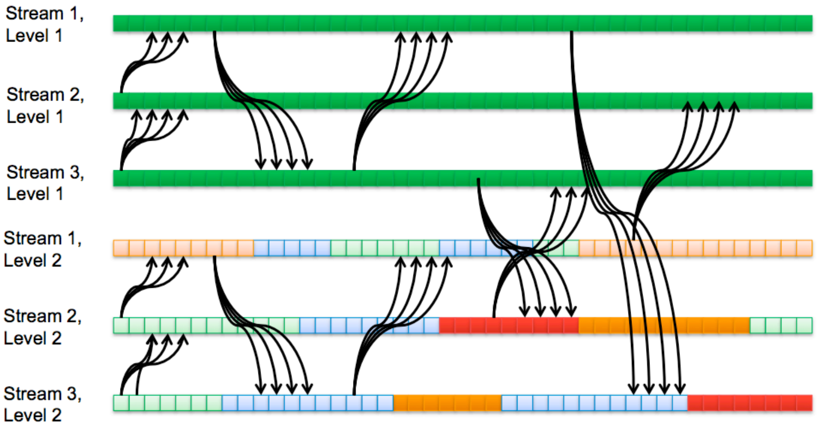


Figure 13: Unsupervised learning of level 2 patterns combined with cross-modal associative learning across multiple data streams and across levels of hierarchy.

3.3.1 Prediction of EEG-activity across electrodes

We have applied unsupervised pattern discovery and cross-modal associative prediction to the analysis of pre-term infant EEG signals. The goal was to estimate the activity in an EEG electrode given the signal in another electrode. In this process, the SLCM algorithm was first used discover patterns from sequences of EEG features (cf. Figure 12), and then the temporal dependencies between the patterns in different electrodes were modeled using our associative learning scheme. Given only 4.5 minutes of training data, the system was already able to predict EEG patterns in another electrode with 38% accuracy (4% chance level), whereas prediction at the feature level was only slightly above chance. This further supports the idea that hierarchical abstraction from sensory details is inevitable in order to learn stable connections between multiple input streams. Figure 14 shows an example of prediction output when activity of F4 electrode (in right hemisphere) is being predicted from F3 electrode (in left hemisphere).

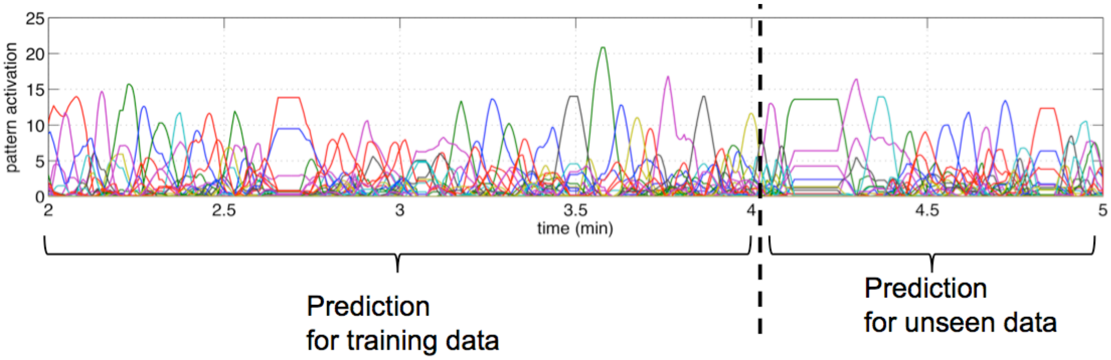


Figure 14: Prediction of EEG activity patterns in right hemisphere given the signal activity in left hemisphere.

3.3.2 Prediction of articulatory gestures from acoustic data

Another example of cross-modal associative learning is the interpretation of spoken language in terms of the articulatory gestures that were used to produce the perceived utterance (a so called *speech inversion problem*). In our experiments, we have created a learning agent that is equipped with auditory perceptual capabilities and the ability to produce speech with the help of an articulatory speech synthesizer. In the synthesizer, speech signals are created by modeling the locations and movements of different articulators (jaws, lips, tongue, etc.) in the vocal tract similarly to the speech production system in humans. Each speech sound, or phone, is defined in terms of the target positions of each articulator during the production of the sound. The parameters corresponding to the movement between these targets are then computed dynamically according to minimum-jerk principle. Finally, in order to allow learning, the system is equipped with the ability to discover recurring patterns from the auditory stream with the SLCM algorithm, and the ability to associate different representations (articulatory targets, articulatory parameters, audio features, and audio patterns) to each other using a so-called sparse distributed associative network. Figure 15 shows a schematic overview of the system.

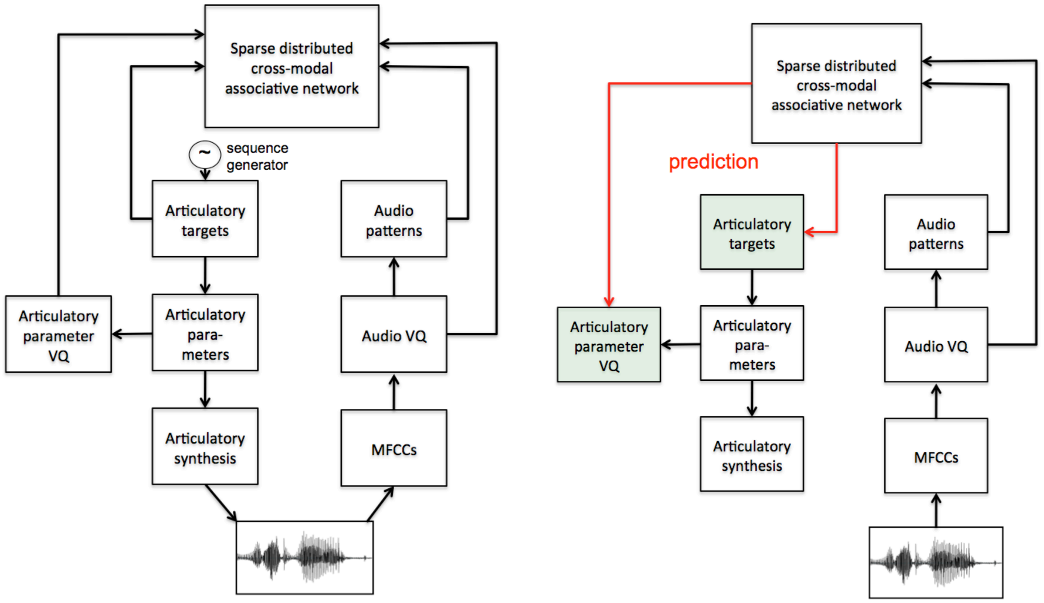


Figure 15: Learning scheme (left) and prediction scheme (right). During learning, randomly generated articulatory target sequences and corresponding articulatory parameters are associated to audio features and patterns discovered automatically from the sequences of audio features.

During the learning stage, the system spoke randomly generated phone sequences and the articulatory and acoustic consequences of these productions were analyzed by the associative model. During prediction stage, the system was only given an acoustic word form and its task was to find the most likely sequence of articulatory gestures from the acoustic features and the acoustic patterns extracted from these features. As a result, the system was able to recognize articulatory targets (phone categories) with an accuracy of 88.24 % correct recognitions. Figure 16 shows an example of articulatory target recognition from purely acoustic signal.

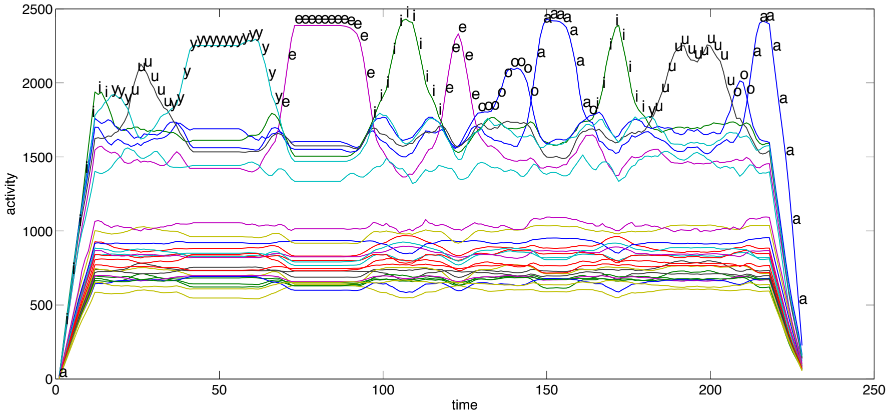


Figure 16: Example of articulatory target prediction from synthesized speech signal "*yuyyeeieoaiuuaa*". Correct targets can be seen as clear peaks in the activation curves.

One of the advantages in multimodal associative learning is also that it is possible to learn associative links between patterns that are never observed at the same time. This is called *indirect prediction*. For example, visual information may provide partial cues regarding the articulatory gestures of another speaker, but the visual information does not reveal the hidden phone targets of the speaker, just the superficial and reduced gestures. However, the listener may be able to map speech signals into articulatory parameters, and these articulatory parameters then into the hidden articulatory target based on listener's own articulatory experience. In terms of a computational implementation, the system must first estimate the likelihood of different articulatory parameters as a function of time, given a speech signal, and then estimate the likelihood of articulatory targets given the estimated articulatory parameters.

In order to test indirect prediction, a test setup was created where the learning agent was able to perceive articulatory targets, articulatory parameters (gestures), and the audio signals created by the articulations. However, the learner never received information about both articulatory targets and audio at the same time (Figure 17), making learning of direct associations between audio to articulatory targets impossible (Figure 18, left panel). Instead, the articulatory targets or audio signals were always paired with an intermediate representation of the articulatory parameters (Figure 17), enabling indirect mapping of audio to articulatory targets via this representation. After a period of learning, the system was successful in estimating first articulatory parameters from audio, and then estimating the articulatory targets from the articulatory parameters (Figure 18; middle and right panels, respectively).

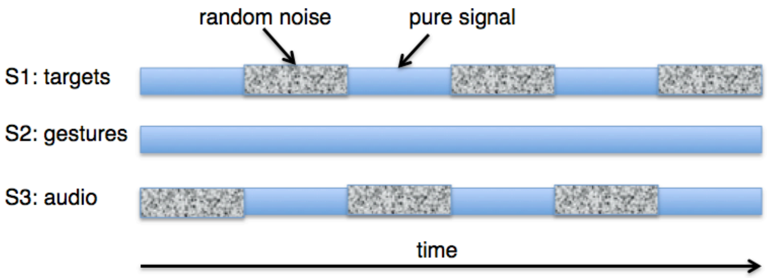


Figure 17: A schematic view of the interleaving of randomized sections of data in the experiment.

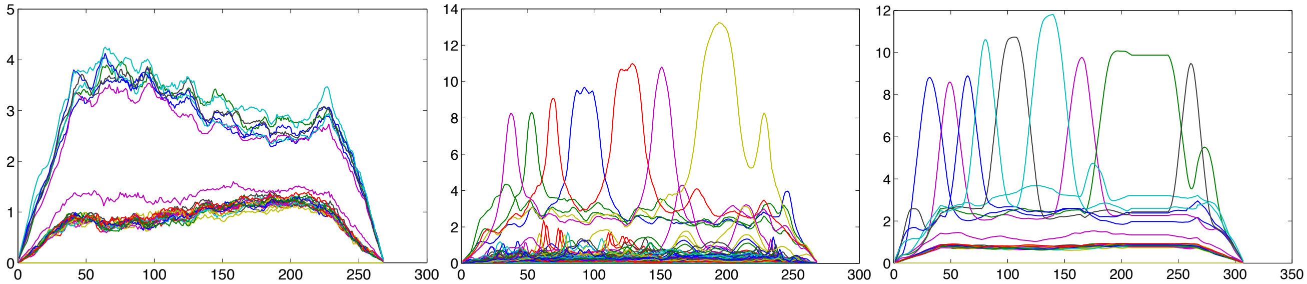


Figure 18: Articulatory targets estimated from audio (left), articulatory gestures estimated from audio (middle) and articulatory targets estimated from the obtained gesture estimate (right). Direct mapping from audio to targets is not possible since they never co-occur in the data. However, indirect prediction through gestures provides articulatory targets even when the gestures themselves are not perceived.

Finally, an autonomous system for cross-modal associative learning inherently utilizes predictive cues from all modalities that carry such information. This was demonstrated by running a speech inversion experiment in which the learner not only received acoustic speech signals, but also saw the lip and jaw movements of the speaker. It is well known that seeing the articulatory gestures of a speaker has a notable impact on speech understandability, increasing the signal-to-noise ratio of speech in noisy conditions. This is also what happened in our experiments (Figure 19), where the combination of audio and visual information led to increased phone recognition accuracy in adverse noise conditions. Note that the complementarity of auditory and visual data was not hard-coded to the system, nor was there any manually optimized weighting of the two modalities. The system simply combined the predictive information from the both streams according to the strengths of the predictive associations, automatically leading to a balanced representation of the hypothesized phone target.

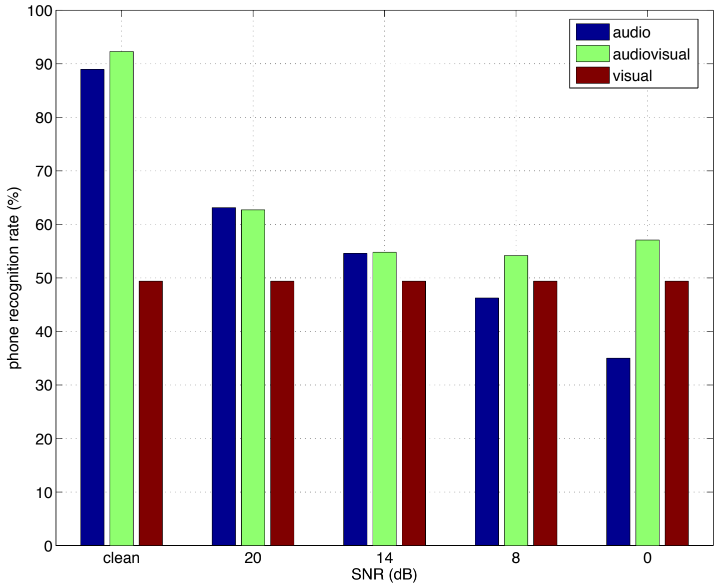


Figure 19: Phone recognition accuracy from CVCVCV structures with audio, visual, and audiovisual information. Results from varying levels of additive factory noise are shown.

3.4 Experiments with integrated architectures

3.4.1 Unsupervised learning of high-level contexts from low-level sensory data

One of the aims in context-aware computing is to infer higher-level abstract representations of the surrounding context from the sensory data that would provide useful information regarding the current use situation of the device (e.g., location such as *shop* or *home* or activity such as *walking*). Majority of the previous work in user context recognition has used supervised methods to train separate classifiers for different physical activities and auditory contexts of interest. The general finding of the studies is that the context recognition performance achieves relatively good levels when the training data has close correspondence to the actual testing conditions. When controlled in-lab data sets are evaluated in unconstrained situations, performance drops significantly. This calls for unsupervised methods that do not come with a priori assumptions regarding the relevant contexts but adapt themselves to the context patterns experienced by the system.

In our work (Räsänen, 2012), we have described a novel approach for unsupervised learning of high-level user contexts from any generic sensory data (Figure 20, left). The system combines unsupervised discovery of short-term sensory patterns to unsupervised acquisition of high-level context models in a hierarchical framework that is computationally feasible for platforms with low computational resources. The basic idea is to first discover statistically significant recurring structures in sensory streams and then to analyze the presence of these structures at a larger time-scale in order to find internally coherent segments of sensory activity. These segments are then clustered into context categories and on-line recognizers are trained for the categories using the discovered segments as the training data. We have shown that the system is able to 1) segment mobile sensory readings into segments corresponding to different physical and environmental contexts (Figure 20, right), and 2) to train selective recognizers for these contexts for on-line detection of the contexts directly from low-level sensory data (Table 1).

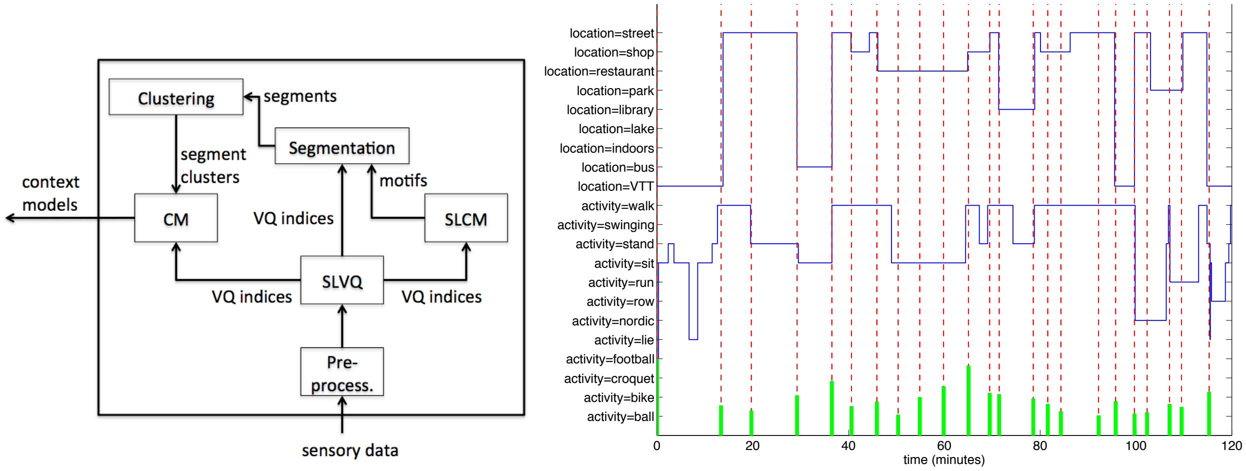


Figure 20: Left: a schematic view of the unsupervised context learning system. Long-term statistical analysis of pattern motifs and VQ indices is used to discover high-level context classes, for which on-line classifiers can be then trained. Right: unsupervised context segmentation based on audio data. Blue line denotes the current context and red dashed lines show detected activity segment boundaries.

Table 1: Contents of discovered context segments (left) and selectivity of learned on-line classifiers (right). Only 1-2 best matching classes are shown per token.

SEGMENTS			CLASSIFIER SELECTIVITY				
S	%	ID	M	%	ID	%	ID
1	95.8%	office	1	95.6%	street	4.4%	shop
2	93.2%	street	2	73.6%	shop	23.4%	street
3	99.7%	street	3	88.4%	street	8.5%	library
4	100.0%	bus	4	96.3%	street	3.7%	shop
5	96.0%	street	5	41.1%	shop	40.9%	street
6	70.5%	shop	6	100%	library		
7	95.2%	restaur.	7	93.2%	street	6.8%	office
8	100.0%	restaur.	8	97.0%	office	3.0%	street
9	100.0%	restaur.	9	50.8%	park	49.2%	street
10	96.8%	restaur.	10	89.7%	restaur.	10.3%	street
11	100.0%	shop	11	90.6%	park	9.4%	street
12	91.7%	street	12	100%	restaur.		
13	100.0%	library	13	100%	shop		
14	49.5%	shop	14	100%	office		
15	100.0%	shop	15	82.6%	bus	12.9%	street
16	75.1%	street	16	95.0%	shop	5.0%	restaur.
17	94.0%	street	17	89.5%	office	10.5%	street
18	100.0%	office	18	95.6%	street	4.5%	shop
19	98.8%	street	19	100%	restaur.		
20	84.4%	park					
21	100.0%	park					
22	86.5%	street					
23	100.0%	office					

4. POSSIBLE APPLICATION AREAS OF ALOTA

In general, we consider ALOTA as a step towards the concept of **machine understanding**: the ability of artificial computational systems to derive meaning from raw data. Machine understanding acts as a fundamental enabler for a new generation of artificial intelligence solutions, and therefore there are numerous application areas that can greatly benefit from it. Here, we simply give a number of examples to illuminate the variety of these opportunities. With the given state of the technology, some of the proposed application areas are closer to the practice than some others. However, we believe that all of them are achievable in near future with sufficient research effort and resourcing.

4.1 Adaptive UI

Understanding input data also means context awareness as long as the input data is somehow related to the current context of the computational device. In human-computer interaction (HCI), one long-term goal is to build computational devices that can adapt their behaviour to the current use context of the device. Since different contexts can have different use patterns or different desired input/output characteristics of the system, adapting the user interface to these conditions would allow easier and more fluid user experience with the device.

One of the central challenges in normal context recognition is that the different contexts of interest are difficult to anticipate for. The computational device, especially if mobile, may be used in an endless variety of sensory environments with an endless variety of use patterns, making supervised pre-training of context classifiers impractical for all but the most limited definitions of a context. With ALOTA, the system is able to derive useful abstractions of context from low-level sensory data without a priori assumptions regarding relevant context classes. In addition, the very same mechanism can be used to discover the different use patterns of the device and how the discovered high-level sensory contexts are related to these use patterns, allowing prediction of user needs based on sensory input and current state of the device. After that, it is simply up to the software and OS developers to decide how they like to utilize this predictive information in their programming.

4.2 Autonomous industry process control and medical diagnostics

The capability to learn the typical sensory patterns and their relationships across multiple sensors is useful in both process industry and in medical care. A system with ALOTA –like capabilities enables autonomous monitoring and control of complicated processes where a large number of parallel information sources are related to proper actions in different situations. The proper control behaviour in these situations can be learned by ALOTA if internal criteria for successful process behaviour are provided to the system. The autonomous learning can be complemented with learning from the actions of an expert in the given task, allowing automatic codification of implicit expert knowledge in many industrial and medical environments.

4.3 New generation speech recognition

One of the early inspirations for self-learning data analysis architectures comes from the field of speech recognition. The existing state-of-the-art methods, namely Hidden-Markov Model (HMM) based speech recognizers, clearly fall behind human speech perception performance outside highly controlled conditions. This is due to the fact that the existing speech recognizers simply do not understand speech or the context in which the speech takes place, but simply consist of statistical mapping of speech acoustics into text. These recognizers can only process input (e.g., acoustic characteristics of speech, vocabulary, tempo, grammar) that has been explicitly trained to the system in advance. This in contrast to human-like performance where communication takes place in a context and new language knowledge is accumulated on a daily basis. If a system could learn language similarly to a human infant by understanding how acoustic patterns make up words, and how these words connect to the objects, events, and actors in the surrounding environment, the system would naturally evolve to *understand* speech. If the system is also equipped with an ability to produce speech articulations, the system would be also able to learn how its own speech production can have an effect on other actors in the environment. This type of learning process is specifically what ALOTA is designed for: making sense of the sensory and motor environment in the absence of explicit teaching.

4.4 Robotics and AI

Cognitive machines, advanced robotics, artificial intelligence – all these concepts are related to the idea of an artificial system with the ability to sense its environment, make decisions regarding proper actions in the current situation, and to put these decisions into action. As motivated in the introduction, intelligent and generalized decision-making requires effective organization of the sensory inputs and the motor outputs. The system must be able to represent its previously learned experience at a sufficiently general level that it is applicable to unseen situations, but with sufficient detail so that the situations with different affordances and desired actions are properly recognized. Motor planning in complex environments also requires high-level abstract representation of the motor actions since there are infinite many combinations of possible percepts and motor commands to be learned from a finite experience. This all calls for ALOTA like organization of multimodal data streams, where the data is organized according to its intra- and cross-modal statistical connections, allowing prediction of proper actions from partially observed sensory state of the world.

4.5 Organizing the Big Data

The final application area that has been lately given increasing amounts of visibility is the so-called Big Data problem (see The Economist, 2010). Information technology society is producing and collecting massive amounts of data from all aspects of modern life, including, but not limited to, consumer behaviour, economics, logistics, internet and telecommunication, traffic, intelligent cities, weather, etc. A major problem is that there is simply too much data to be analyzed manually or even with data analysis tools that require indirect manual operation. Although many sophisticated data- and association-mining techniques already exist, they are typically dedicated to either time-invariant data (multivariate analysis), data on ordinal scale (time-series analysis), or pattern discovery from data streams with well-defined and stable elementary units (such as semantic analysis of text documents). Finally, the interpretation of the discovered patterns in a larger context is still left to the domain experts, requiring manual work. In contrast, the ALOTA is designed to discover statistically significant structures from data when they are distributed or interleaved in the data across time or space, when the elementary data points as such do not carry any meaning, and without assuming strong constraints to the nature of patterns that may exist in the data. Moreover, ALOTA discovers and thereby interprets the patterns in a context of other information sources, making meaning aspect of the discovered patterns an inherent part of the system.

5. CONCLUSIONS

Automatic analysis and understanding of multimodal data plays increasingly important role in the modern societies where massive amounts of data are being collected on a continuous basis. Discovering meaningful structure from the big data is essential for making use of the data, allowing computational devices to act autonomously or humans to understand the processes behind the data. In our work, we attempt to develop an architecture for unsupervised hierarchical associative learning (ALOTA) that could be used to solve the data analysis problem in any domain with continuous streams of poorly understood data. This document has reviewed the basic concepts and principles behind such a system, and we have also presented a number of technological demonstrations that have already been taken towards a fully functional integrated system. Meanwhile, the work is being continued.

RELATED LITERATURE

Bengio, Y., 2009. Learning Deep Architectures for AI. *Foundations and Trends in Machine Learning*, 2, 1-127.

Damasio, A., 1989. Time-locked multiregional retroactivation: a systems-level proposal for the neural substrates of recall and recognition. *Cognition*, 33, 25-62.

The Economist, 2010. Data, data everywhere. A special report on managing information, Feb. 27th, pp. 1-14.

Haikonen, P., 2003. The Cognitive Approach to Conscious Machines. Exeter, UK: Imprint Academic.

Hawkins, J., 2004. *On Intelligence*. Henry Holt & Company, New York, NY.

Kanerva, P., Kristoferson J., & Holst A., 2000. Random Indexing of Text Samples for Latent Semantic Analysis. *Proc. 22nd Annual Conference of the Cognitive Science Society*.

Kanerva, P., 2009. Hyperdimensional Computing: An introduction to computing in distributed representation with high-dimensional random vectors. *Cognitive Computation*, 1, 139-159.

Kanerva, P., 1993. Sparse Distributed Memory and Related Models. In Haussoun, M. (Ed.): *Associative Neural Memories: Theory and Implementation*, pp. 50-76, New York: Oxford University Press.

Jockel, S., 2009. *Prediction of Autobiographical Episodic Experiences using a Sparse Distributed Memory*. Doctoral Thesis, University of Hamburg, Department of Informatics.

Laine, U. K., 2011. Entropy-rate driven inference of stochastic grammars. *Proc. Interspeech'2011*, Florence, Italy, 2489-2492.

Landauer, T., & Dumais, S., 1997. A Solution to Plato's Problem: The Latent Semantic Analysis Theory of Acquisition, Induction, and Representation of Knowledge. *Psychological Review*, 104, 211-240.

Meyer, K. & Damasio, A., 2009. Convergence and divergence in a neural architecture for recognition and memory. *Trends in Neurosciences*, 32(7), 376-382.

Pfleger, K. R., 2002. *On-line learning of predictive compositional hierarchies*. Doctoral thesis, Stanford University, Department of Computer Science.

Räsänen, O., 2012. Hierarchical unsupervised discovery of user context from multivariate sensory data. *Proc. ICASSP'2012*, Kyoto, Japan, pp. 2105-2108.

Räsänen, O. & Laine, U., 2012. A method for noise-robust context-aware pattern discovery and recognition from categorical sequences. *Pattern Recognition*, 45, 606-616.

Räsänen, O., Leppänen, J., Laine, U., & Saarinen, J., 2011. Comparison of Classifiers in Audio and Acceleration Based Context Classification in Mobile Phones. *Proc. EUSIPCO'11*, Barcelona, Spain, pp. 946-950.

Räsänen, O., 2011. A computational model of word segmentation from continuous speech using transitional probabilities of atomic acoustic events", *Cognition*, 120, 149-176.

Räsänen, O., Laine, U., & Altosaar, T., 2009. Self-learning Vector Quantization for Pattern Discovery from Speech. Proc. *Interspeech'09*, Brighton, England, pp. 852-855.

Sahlgren, M., 2005. An introduction to random indexing. *Proc. Methods and Applications of Semantic Indexing Workshop, 7th Int. Conf. on Terminology and Knowledge Engineering*.