Project no. 034362

ACORNS

Acquisition of COmmunication and RecogNition Skills

Instrument: STREP
Thematic Priority: IST/FET

Periodic Activity Report

Period covered: from 1 December 2007   to 30 November 2008

Date of preparation: 23 December 2008

Start date of project: 1 December 2006   Duration: 36 months

Project coordinator name: Prof. Lou Boves
Project coordinator organisation name: Radboud University, Nijmegen
Revision [1]

Table of Contents

1. Executive summary

ACORNS, Acquisition of COmmunication and RecogNition Skills
www.acorns-project.org

Participants:

- Centre for Language and Speech Technology, Radboud University Nijmegen, The Netherlands (co-ordinator)
- Laboratory of Acoustics and Audio Signal Processing, Helsinki University of Technology, Finland
- Sound and Image Processing Laboratory, Royal Institute of Technology, Stockholm, Sweden
- Speech and Hearing Research Group, University of Sheffield, United Kingdom
- Center for Processing Speech and Images, Katholieke Universiteit Leuven, Belgium

Coordinator contact details: Dr. Lou Boves, CLST, P.O.Box 9103, 6500 HD Nijmegen, The Netherlands, l.boves@let.ru.nl; Tel: + 31243612902.

1.1 Summary description of project objectives

ACORNS aims at testing the viability of the memory-prediction theory of intelligent behaviour as a basis for modelling the acquisition of language and communication skills. The input for the model consists of audio signals in combination with representations of the context to which spoken utterances refer. The project work plan is structured in five technical work packages, four of which are devoted to several aspects of sensory processing, discovery of meaningful structures and associations in the input data and the representation of learned structures in memory. The fifth work package is dedicated to integrating results in a comprehensive system and conducting experiments that test the capability of the approach to account for the acquisition of language and communication skills.

The **front-end processing** module, under development in WP1, provides a representation of audio signals that can characterise and process all ecologically relevant sounds and model different sources independently, with a strong focus on features that are important for speech processing.

WP2 focuses on **pattern discovery**, by building computational models that detect recurring patterns in input signals. **Memory organisation and access** is the focus of WP3. We develop computational representations of the different types of memories that are implied in the memory-prediction theory and models of memory and processing resulting from research in psychology. WP4 investigates **Information discovery and integration,** with a focus on emergent representations that result from a combination of bottom-up and top-down signal processing.

For WP5, **Interaction and communication**, the aim is to integrate the results in a system that can simulate the acquisition of language and communication skills. In doing so, emphasis is put on the cognitive and biological plausibility of the procedures, processes and representations developed in the four preceding work packages. Three increasingly more complex stages, one for each year of the project, are defined. At the end of the first stage, our artificial infant had acquired the basic skills needed to understand that it is being addressed; in addition, it had learned some ten words. At the end of the second stage the artificial agent has learned some 50 words. In the third and last stage of language acquisition that we will simulate, we will investigate how previously acquired knowledge and skills can be harnessed to speed up further learning of language and communication skills. To demonstrate the learning skills we will show that the agent can learn new words when new concepts are introduced in the environment.

1.2 Work performed

While the results of the research in the first year of the project were very encouraging, inevitably we encountered a number of issues that required special attention during the second year.

Perhaps the most obvious one was that in the first year we had taken a shortcut with respect to the representation of the input in the visual (and tactile) channels. To bootstrap the experiments it was decided to represent the non-audio input in the form of crisp symbolic tags that identify the reference of the speech utterances. Yet, it had been clear all the time that this approach was not plausible from a cognitive or biological point of view. Therefore, substantial effort was spent in the second year to develop more plausible representations (cf. Deliverable 5.4.2).

A second major finding in the first year was that all existing theories and models of language acquisition and communication are so abstract and incomplete that all allow for several algorithmic implementations. On the one hand this is advantageous, because it made it possible to show the correspondences between the models developed in Psychology and the Memory-Prediction Theory (cf. Figs. 1 and 2). On the other hand, the need to investigate several algorithmic approaches made it impossible to integrate all results of the individual work packages in a single integrated system that simulates language acquisition. In year-2 substantial progress was made in relating independently developed memory and processing architectures (cf. Deliverable 3.2).

Even if alternative algorithms for discovering and representing structure and information in speech fit in the same abstract models of memory and processing, they may still require different software implementations of the architectures sketched in Figs. 1 and 2. For this reason it is not useful to try and develop a unique implementation of the memory architecture. Therefore, it was decided that ACORNS will take a two-pronged strategy: we will try to integrate as many findings as possible in an increasingly more powerful and plausible agent that simulates language acquisition, while there may be other results that do provide insight in processes but that cannot be integrated in an operational system. Moreover, we will construct parallel versions of the language acquisition agent, based on alternative strategies for information discovery and integration. These versions may share some modules, but evidently not all. One module that will be shared is the acoustic feature extraction under development in WP1.

Parallel implementations of the language acquisition agent can be compared in many different manners. One option would be to stage a competition between alternative implementations, with the eventual goal to select the 'best'. However, we feel that the state of the art in simulating language acquisition is not sufficiently advanced to allow for establishing clear-cut performance criteria with which 'competing' instantiations could be compared. For that reason, we decided to focus the comparison on the insights in language acquisition and language processing that can be gleaned from each of them. This also makes it possible to treat the partial approaches in a fair manner, since these too will contribute to advancing our understanding.

1.3 Results achieved

Representation of non-audio input

Guided by the results of the experiments in Year-1 and the goals set for Year-2 an extended corpus of speech utterances was recorded, annotated (as far as necessary) and made available to the partners for conducting experiments. The Year-2 corpus was recorded in Dutch, English and Finnish. Each corpus comprises recordings from 10 different speakers. Four of these are the same speakers as in the Year-1 corpus, enabling cognitively and biologically plausible experiments with more advanced language acquisition. Each of the four speakers 'acted out' 2000 utterances constructed to contain up to four key words. Six additional speakers produced a different subset of 600 utterances each.

To address the issue of too crisp and unique visual/semantic representations we have created a set of 64 visual/semantic features that can be used to represent the meaning or grounding of the utterances in the Year-1 and Year-2 corpora. For the Year-1 corpus, where each utterance refers to a single object, the features can be used to create ambiguity: each object can be referred to with multiple combinations of features. The use of features also makes it possible to represent similarities and differences between object classes: 'dog' and 'cat' can be given representations that are more similar than, for example, 'car' and

'mamma'. In addition, the use of features makes it possible to link multiple attributes to objects (e.g. a 'big green frog', where 'big' and 'green' are attributes of the object 'frog').
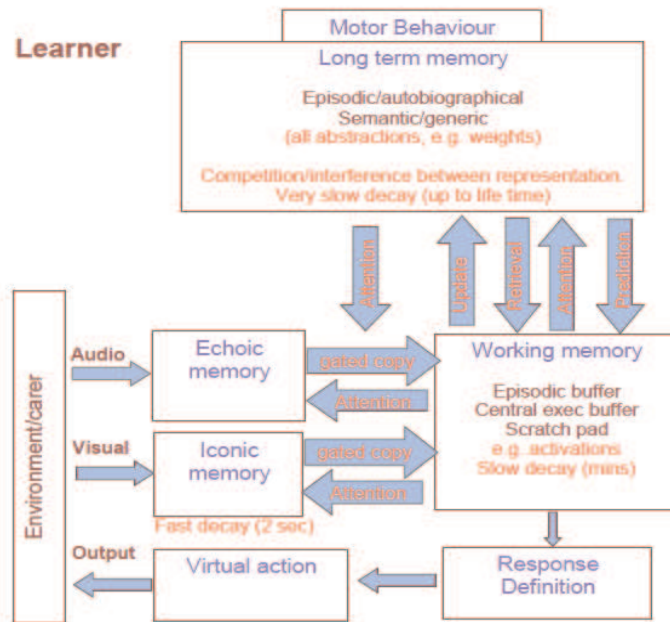
Figure 1 Hierarchical modular memory and processing architecture that reflects the results from research in Psychology on Memory, especially with regards to language processing

Pattern and Information Discovery and Integration

All partners worked to develop and elaborate methods for discovering patterns, structure and information in speech signals grounded by corresponding tags or semantic feature vectors. Virtually all conventional methods for mining data for meaningful patterns that are known from the literature assume that a massive amount of data is available at the start of the process and that all data can be accessed repeatedly. These assumptions are not compatible with biological and cognitive knowledge about how living agents learn. Therefore, much of the research in the second year was dedicated to making the structure discovery processes incremental, in the sense that each individual input utterance was heard only once. Of course, the presence of working and long-term memory in the architecture shown in Fig. 1 makes it possible to process input utterances multiple times, as long as a suitable representation remains accessible in the memory. We have succeeded in developing incremental representations of most pattern discovery algorithms under investigation (Non-negative Matrix Factorisation [NMF], DP-ngrams and Concept Matrices). This makes it possible to evaluate the learning behaviour of a specific pattern discovery algorithm by comparing the algorithm's interpretation of a novel stimulus with the ground truth.

The pattern discovery algorithms under investigation can be characterised in terms of the stage at which processing moves from sub-symbolic representations to representations that can be interpreted as meaningful symbols. We have come to the conclusion that approaches which rely on an early switch to symbolic processing (such as Computational Mechanics Modelling and multigran models) suffer from too large a symbol alphabet that must be considered in bottom-up processing of speech signals. Experiments with several conceptually different structure discovery methods showed that no existent method can deal with alphabets of more than about 100 symbols. Even with lower numbers of symbols existent methods fail if the symbols in the input stream are subject to uncertainty, which results in a very large number of different sequences. We have extended the theory underlying Computational Mechanics Modelling to enable it to cope with 'approximate causal states', i.e., symbol sequences that are similar to a degree that allows them to be considered as instantiations of a unique underlying sequence.

Algorithms (such as NMF and DP-ngrams) that postpone the switch to a later stage (when the number of potentially meaningful symbols is much smaller) appear to have a clear advantage in processing highly variable signals such as natural speech. .
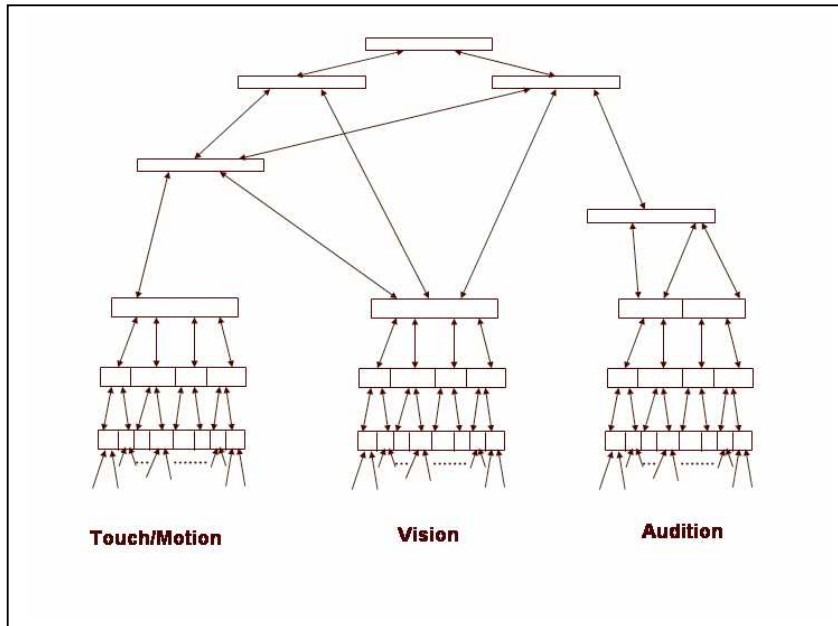
Figure 2 One possible view of the cortical hierarchy in the Memory-Prediction Framework. (After Hawkins, 2004)

.

Memory and processing architectures

The memory architecture designed in Year 1 was elaborated and made more concrete thanks to the research in Year 2. The updated model is shown in Fig. 1. We also made progress in establishing correspondences between the memory architecture based on the results of behavioural psychological research (basically, the architecture in Fig. 1) and the memory architecture suggested by the memory-prediction framework (sketched in Fig. 2). We analysed the seemingly incompatible architectures on the three levels suggested by Marr. It appears that both (classes of) models are so abstract and incompletely specified at the computational level that no existing theory can impose very strong constraints on the choices made at the algorithmic or the implementation levels.

From the results of the experiments with pattern discovery procedures it has become evident that multi-layer architectures have an advantage over single-layer ones. Also, it is difficult to imagine how a single-layer representation of language could be effective. For that reason we have investigated several multi-layer approaches, including multi-layered versions of NMF as well as Self Organising Maps and Restricted Boltzmann Machines. Interesting issues include the question to what extent 'old' representations remain accessible after newer, more powerful ones have been formed, and whether processes change over time or between levels of representation. These issues are central in all cognitive science research in language acquisition. The memory and processing architecture that we have built allows us to test all alternative hypotheses. The results obtained in year-2 suggest that none of the possibilities can be ruled out.

Interaction with the Scientific Advisory Committee (SAC)

ACORNS is supported by a multidisciplinary advisory committee. Towards the end of year-2, in time for guidance to have effect on the course of the research, we have convened meetings with the members of the SAC, to discuss the most promising directions and objectives for the last year. The results of these meetings have contributed substantially to the formulation of the concrete plans for year 3.

Publications

Up to now, ACORNS has resulted in 16 papers and 3 master theses, all of which can be accessed through the project's public website www.acorns-project.org

The public website also provides access to a paper that summarises the results of the first two years; the text of that paper was originally written to prepare the meetings with the SAC members.

2 Project objectives and major achievements during the reporting period

2.1 General Project Objectives

ACORNS aims to clarify the feasibility of the memory-prediction theory of intelligent behaviour as it applies to language acquisition and speech communication by developing and testing a computational model of language acquisition informed by the memory-prediction theory. The input for the model consists of audio signals in combination with symbolic representations of the environmental context, to provide grounding. Research in the first year has shown that there is not yet a software implementation of the memory-prediction theory that is powerful enough to build an agent that can simulate a process as complex and as badly understood as language acquisition. More conventional theories of language acquisition and processing suggest that the processes must be modular.

The project work plan is structured in five technical work packages, four of which are devoted to investigating specific aspects of the memory architecture and processing, while WP5 is responsible for the integration of the results and for conducting the experiments that will investigate to what extent the memory-prediction theory can account for the acquisition of language and communication skills. Here we briefly sketch the five work packages:

Front-end processing (WP1) results in a rich internal representation that is suitable to characterise and process essentially all ecologically relevant sounds and to model different sources independently.
Pattern discovery (WP2) investigates computational models that can detect recurring patterns in the input signals and that can be linked to memory representations.
Memory organisation and access (WP3) will focus on suitable computational representations of the different types of memories and the processing that takes place. An important aspect of memory processes is how representations of novel patterns can form and be stored.
Information discovery and integration (WP4) investigates how patterns in the input signals can be discovered, stored and accessed for the interpretation of novel input signals.

Interaction and communication (WP5) integrates the results of the WPs sketched above in a software platform that allows us to investigate aspects of language acquisition and processing in a setting that resembles human language acquisition: as the result of situated communication between a care giver and a learning agent. Care is taken to ensure that the learning algorithms and memory representations developed in WP1 – WP4 are plausible from a biological and cognitive point of view. For this purpose, it is emphasised that learning should be incremental, and that 'training' stimuli are offered to the learning agent only once.

In the first year of the project it has become clear that the borders between the work packages 2, 3 and 4 are fluid and permeable. Therefore, we have drawn up an updated pictorial representation of the structure of the project, which is shown here in Fig. 2.1.

2.1.1 Objectives for the reporting period

The focus of the research in the second year was on further development of techniques for structure discovery and information discovery. In the first year it appeared that all existing theories of language acquisition and processing (including the memory-prediction theory) are quite abstract, so that several different computational algorithms can be used to implement the memory and processing architectures that support language learning in interaction between infants and care givers. For that reason we decided to focus on a comparison of learning techniques, rather than on the construction of one complete system based on a single learning algorithm. Nevertheless, to be able to make fair comparisons it remains necessary to try and integrate all approaches in the common platform for conducting learning experiments.

The learning task in the second year was more complicated, in that the artificial agent must be able to handle a larger number (±50) of concepts (not only nouns, but also verbs and adjectives), to discover multiple semantic units in an utterance and to build internal representations that can be linked both to the acoustic

signals and the semantic/pragmatic value of an utterance and can be used as a stepping stone for learning additional words and concepts. If possible the learning process and the emergent internal representations should be linked to documented behavioural findings in human language acquisition, such as the loss of sensitivity to non-native phonetic contrasts and semantic confusions between similar objects.

WP0
Project management

WP5
Experiments

WP6
Dissemination

WP2
Signal Patterning

WP3
Memory Organization and Access

WP4
Information Discovery and Integration

WP1
Acoustic Features
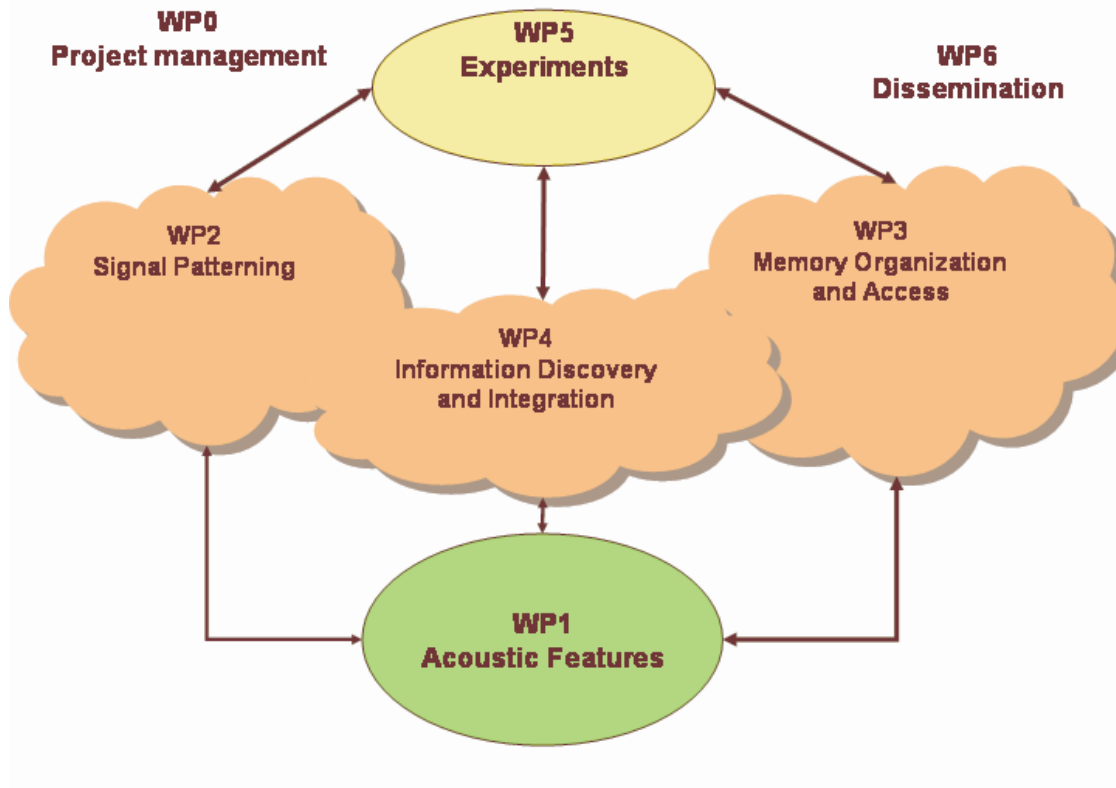
Figure 2-1 Updated representation of the structure of the ACORNS project.

2.2 Results

In this section we summarise the most important results of the research in the second year. More detailed accounts of the work performed and the results can be found in the deliverables for the second year, all of which have been produced according to schedule.

2.2.1 WP1

Substantial progress has been made in the development of acoustic features that can augment the standard spectral features that are conventionally used for speech processing with millisecond and deci-second features. In addition, progress has been made towards *selecting* the most powerful subset of a pool of more conventional acoustic features (mainly MFCCs) based on knowledge of the human auditory system.
Our work on millisecond features has resulted in an algorithm that automatically estimates the voicing onset of plosives. Our work towards defining deci-second features consists of a study on the efficacy of including measures of prosody (rhythm and pitch movement). We find that the measures are useful at the initial learning stage. Our work on feature selection led to an algorithm that selects features based on the ability of the features to describe the components of speech that are most clearly perceived. Experimental results confirm effectiveness of this generic strategy.

2.2.2 WP2

Promising results have been obtained from an incremental bottom-up learning method using segmental discrete model elements (DME) that was coupled with a new concept matrix (CM) approach for discovering

structure in continuous signals. The DME-CM tandem has been successfully applied to word learning. Also, the concept matrix approach was reformulated to incorporate transitional probabilities.

The importance of attentional focus to discovering word-like units in speech signals has been investigated. In this context a comparison of IDS/ADS was carried out to better understand how they may influence infant learning.

A new self-learning vector-quantization (SLVQ) method was developed that works incrementally and adapts to the properties of input data instead of forming a fixed number of clusters in a batch process.

Research into and further enhancement of clustering algorithms developed during the first year took place in order to address the challenges of incremental blind phone-like unit classification. This work also included the investigation of feature extraction methodologies for describing segmental phone-like units. This was deemed necessary due to imminent higher level processing needs, i.e., the formation of memory.

2.2.3 WP3

The relation between the memory-prediction theory and more conventional theories of memory and processing of sensory stimuli has been elaborated and further clarified. Specifically, attention has been paid to the implementation of selective attention mechanisms and the integration of working memory and semantic memory.

One specific model that has been developed that incorporates working memory is a recurrent self-organising architecture that is used to associate speech signals and semantic features to develop an emergent representation of speech. This model offers a hierarchical structure that incorporates certain features of the memory-prediction framework and as such can be extended to include more features.

Steps have been made toward the implementation of episodic and semantic memory in the memory-prediction framework. Episodic long-term memory has been implemented in the form of an extension of the instance based MINERVA2 approach. This model as been successful used to perform speech and keyword recognition.

Preliminary actives have been carried out towards the design of an actor-critic model of reinforcement learning for sensor-motor representation.

2.2.4 WP4

NMF-based learning was extended to enable incremental learning. Moreover, it is now possible to estimate where word-like units are present in a continuously spoken utterance.

Information integration in NMF was studied and it was shown that several information streams can be combined to achieve better accuracy. More specifically, we combined phone lattice information and quantized speech spectra as input streams.

NMF was applied to speech spectra to discover (without supervision) patches in the time-frequency plane which are relevant to speech and which showed noise robustness in a recognition experiment. This approach was extended to include a *second NMF layer* which is fed by the time-frequency-patch activations and learns words under weak supervision.

NMF was applied to learn a representation of grammar by *cascading* NMF layers. The first layer directly maps acoustic events to words. A second layer maps the word activation patterns at a 400 ms time scale to new word activations. This mapping now takes the context of words into account and models that some words are more likely to occur before or after others.

Three different activation verification mechanisms were proposed and studied and a top-down learning mechanism was tested. The results on a vocabulary of about 648 frequent words from the Wall Street Journal as well as from 90 words from the first and second year ACORNS database were disappointing. We attribute this effect to the strong context-dependency in our representation and in the data. For the time being a bottom-up approach seems a more appropriate route to follow.

The potential of an exemplar-based approach was evaluated. Preliminary results, obtained under constrained conditions, are promising.

2.2.5 WP5

The implementation of the caregiver has been upgraded so that it is now possible to respond to the learner with multiple reactions, which are especially relevant if the learner makes 'mistakes' in interpreting sensory

input. The learner is upgraded to enable dealing with vector-based features as representation for the information in the visual channel.

The internal and external loops, within which learning takes place, are improved with respect to the implementation of learning drive. The learning is now the result of the information provided by the caregiver via the external loop, and driven by the minimisation of a target function operating in the internal loop.

The integration of audio and visual information takes place at an early stage at the level of features. The learner combines the two streams of information, and is updated to generalise the use of the symbolic tags as used in the first year.

Experiments with strictly incremental NMF based learning show that the use of the more complex Year 2 ACORNS database yielded an accuracy of about 85 percent correct identification of concepts with a small number of training utterances, provided that trivial visual features are used. The Concept Matrix and the DP-ngram approaches are also able to cope with multiple concepts in an utterance, without prior knowledge of which words are present in a lexicon. NMF, DP-ngrams and Concept Matrices are able to discover that a new input utterance contains 'words' that have not been associated to visual input on the basis of previously processed stimuli.

2.3 Meetings with SAC Members

Because it appeared to be impossible to convene a meeting at which all SAC members could be present, we decided to organize two separate meetings, one in Nijmegen (12 November), and one in Stockholm (15 December 2008). All SAC members received a 29 page document that summarized and explained the major findings and results in the first half of the project and lists of questions and issues on which input of the SAC was solicited. The document is now available on the public ACORNS website. In addition, the SAC members had access to all published papers. In the meetings with the SAC members the consortium presented the work done and the plans for the remainder of the project's lifetime. The presentations were followed by intensive discussions. The most important comments and recommendations of the SAC members can be summarized as follows:

- Focus on understanding the processes, rather than improving some formal performance measure
  This recommendation came in two flavors. The SAC members with a background in cognitive informatics, machine learning and speech processing emphasized the need for better understanding of the differences and commonalities between the ACORNS approaches and the conventional HMM and ANN approaches to automatic speech recognition. The SAC members with a background in psycholinguistics emphasized the importance of relating the research to recent findings in the study of first language acquisition.
- Focus on scalability in terms of re-use of early representations to facilitate learning new 'words'
  All members of the SAC agreed that the scalability of the approaches to language acquisition and processing under development in ACORNS is a fundamental issue. However, there was some disagreement on what the aspect of scalability that the project should focus. A minority of the SAC members was of the opinion that we should address learning of adult-size vocabularies and adult-like language processing skills. The majority of the SAC members recommended that we should focus on elucidating how early representations can be re-used for learning new 'words' and how representations might change as the vocabulary grows.
- Do not 'waste' effort in building a 250-word corpus
  There was quite general agreement among the SAC members that the most interesting aspects of re-use of representations can be investigated with a 50-word corpus in three languages. All SAC members agreed that a 250-word corpus will be far too small to address the issues raised with respect to the evolution to adult-like skills and an adult-size vocabulary. It was pointed out that some 'technical' aspects of scaling can be addressed with existing medium-sized corpora such as TIMIT (Garofolo et al., 19990), Resource Management (Price et al, 1988), Wall Street Journal (Paul & Baker, 1992) and the Spoken Dutch Corpus (Oostdijk & Broeder, 2003).
- Make it clear how different approaches can be combined and/or compared
  All SAC members agreed that the parallel exploration of different approaches in ACORNS is one of the strong features of the project. However, they agreed that to reap the full benefits of the parallel approach it is necessary to clarify the relation between the approaches: to what extent can they be combined, and

in what ways can they be compared if they cannot be combined? Therefore, it was recommended that the consortium develop explicit measures to characterize the performance of the approaches, preferably in terms that are more informative than word error rate in a conventional test. Additional measures that were suggested include learning rate, re-use of early representations and cognitive plausibility. However, the SAC members agreed that defining and implementing such measures is far from trivial.

- Explain and justify the design of the ACORNS corpora on the basis of the planned experiments
  It appeared that the description and explanation of the ACORNS corpora in the summary document written in preparation for the meetings with the SAC members were not always sufficiently clear. Especially the way in which we try to simulate visual input and the arguments in favor of this choice caused some confusion.

The recommendations of the SAC members were discussed in a consortium meeting immediately following the SAC Meeting in Stockholm. Except perhaps for the recommendation that we should not create a 250-word corpus, the comments of the SAC members corroborate the basic objectives of the ACORNS project. We will come back to the recommendations of the SAC in section 2.5.

2.4 Addressing the recommendations from the first year review

The review of the work performed in the first year of the ACORNS project was positive in most respects. Still, a number of concerns were raised. In this section we address these concerns.

- Formalise and leverage outputs from Scientific Advisory Committee

  Section 2.3 reports on the interaction with the SAC members. In section 2.5 we summarize the impact of the SAC recommendations on the research plans for the third year of the project.

- Improve scientific coherence of whole project in relation to overall project vision. The most urgent requirement is for more feedback from WP5 to WPs 1 and 2, and from WP3 to WPs 2 and 4

  This concern has been addressed in several different ways. WP5 has taken the lead in the consortium internal discussions about the definition and prioritising of the experiments in the other WPs. Also, WP5 has taken the lead in defining the second instalment of the ACORNS corpora and the developments related to the visual features that have been pivotal in several experiments in the second year and that will shape most of the experiments in the third year.

  After intensive discussions about the relation between the memory architecture and processing in the consortium meeting in March 2008, a dedicated meeting was convened in Leuven where representatives of WP2, WP3, WP4 and WP5 discussed the issues at length and in depth. The results of that discussion were then consolidated in the consortium meeting in June 2008. The results of these activities are reflected in a new visual representation of the structure of ACORNS that was included in the summary document sent to the SAC members. They are also reflected in an improved and updated scheme of the ACORNS memory architecture and an explanation of the relations between the models of memory and processing resulting from psychological and psycholinguistic research on the one hand and the memory and processing models suggested by the memory-prediction theory. This is also explained in the document prepared for the SAC members.

- Stay in touch with cognitive informatics disciplines, e.g. by publishing in reputed journals and conferences in related fields (e.g. IEEE International conference on development and learning, International Conference on Epigenetic Robotics, CogSci, …)

  As could be expected, publication output of a project such as ACORNS will start fairly slowly and reach it peak during the third. In developing publication plans for the second and third year of the project (cf. section 2.5) the recommendations of the reviewers have been taken into account.

  In the second year a paper authored by ten Bosch, Van hamme, Boves and Moore was submitted to the journal Fundamenta Informaticae; it has been provisionally accepted. In addition, research done in ACORNS was presented in the 14th Annual Conference on Architectures and Mechanisms for Language Processing (AMLaP), 4-6 September 2008, Cambridge, UK. From the publication plans for

year 3 it can be seen that we will submit papers to journals in fields such as cognitive science, language learning, machine learning and speech processing.

- Study with serious care the ecological validity of the test databases (on the one hand concerning the (non-)systematic pairing of sounds and (un)-related tags, on the other hand concerning the use of abstract invariant symbolic tags)

    In the document prepared for the SAC members an attempt has been made to relate the research in ACORNS to the state of the art in language acquisition by referring the experiments to two overview papers (Saffran et al., 2006; Kaplan et al., 2008).

    The use of abstract symbolic tags as a crude approximation to visual features has been addressed by introducing visual/semantic features that allow experiments to scale gradually from unambiguous tags to very fuzzy representations of a visual scene.

- It is currently unclear how WP1 relates to WP2 and the rest of the project. Some of the work presented does not appear to be on the critical path of the project, and may be substitutable by readily available solutions. Resources employed do not appear fully justified at this stage.

    This issue is addressed in the progress report of WP1 (cf. section 3.1).

- Comments: Partners should review the project to identify cases of parallel development of similar topics, and set up a systematic and explicit methodology to operationally compare these parallel developments (e.g. concerning automatic discovery of word boundaries or automatic segmentation).

    This issue has been addressed in several meetings, and the results are reflected in several documents, perhaps most clearly in the document prepared for the SAC members.

2.5 Plans for the third year

Research in the third year of the project will focus on the processes involved in the creation and evolution of internal representations. In making concrete plans for the research in year 3 due attention was paid to the recommendations of the SAC members. In global terms, the research will address the following issues:

- Creating speaker dependent representations
- Generalizing speaker-dependent representations so that they become increasingly more speaker-independent
- Impact of the order in which new words are introduced: how do representations that already exist facilitate the creation of representations for new 'words'
- Impact of the degree of fuzziness of visual/semantic features on the rate of learning and the accuracy with which words can be recognized when spoken by as yet unknown speakers
- Developing ecologically relevant measures of language learning that allow comparing different approaches to modeling language acquisition

Concrete implementations of these global plans will be presented in the review meeting in January 2009. This amounts to a one month advance of Milestone M5.6 (Specification of third year experiments).

References

In addition to the papers published by the consortium members in the reporting period (cf. page 28-29 of this document) the text above draws on the following publications:

Garofolo, J. S., Lamel, L. F., Fisher, W. M., Fiscus, J. G., Pallett, D. S., Dahlgren, N. L. (1990) The DARPA TIMIT Acoustic-Phonetic Continuous Speech Corpus CDROM. *NTIS order number PB91-100354*.

Kaplan, F., Oudeyer, P.-Y., Bergen, B. (2008) Computational models in the debate over language learnability, *Infant and Child Development*, Vol. 17, pp. 55–80.

Paul, D.B. & Baker, J.M. (1992) The design for the wall street journal-based CSR corpus. *Proceedings of the workshop on Speech and Natural Language*, Harriman NY, pg. 357-362.

Price, P., Fisher, W.M., Bernstein, J., Pallett, D.S. (1988) The DARPA 1000-word resource manage-ment database for continuous speech recognition Proc. ICASSP-88, 651 – 654.

Saffran, J.R., Werker, J.F. & Werner, L.A. (2006) The Infant's Auditory World: Hearing, Speech and the Beginnings of Language. In: Damon, W., Lerner, R. M., Kuhn, D. & Siegler, R. S. (Eds.) *Handbook of Child Psychology, Volume 2: Cognition, Perception, and Language,* New York: Wiley, pp. 55-108.

Oostdijk, N.H.J. & Broeder, D. (2003) The Spoken Dutch Corpus and its exploitation environment In: *Proc. of the 4th International Workshop on Linguistically Interpreted Corpora (LINC-03).* 14 April, 2003. Budapest, Hungary

3 Workpackage progress of the period

3.0 Workpackage 0 Project Management

The Workpackage Management is divided into two Tasks: Scientific Management and Financial and Administrative Management.

The tasks for this reporting period were:
- Manage the project scientifically
- Prepare, conduct, and report on meetings
- Update the project website (www.acorns-project.org)
- Deliver Periodic Activity Report, and Management Report
- Take care of the financial issues in the project
- Organise Scientific Advisory Board meeting

Achievements:

- The project was well managed. All major milestones were met, all deliverables were on time and there is a good collaboration between the partners. Based on the comments made by the reviewers, special attention was paid to harmonize the work of the separate workpackages. The relationship between the workpackages was an issue that returned in all meetings of the consortium. Cross-fertilization and collaboration between the workpackages were stimulated. The results are reflected in the deliverables of the individual WPs.
- Four project meetings took place and four conference calls were held.
- The ACORNS website was updated every few months.
- The Activity, and management reports and audit certificates were delivered.
- The Scientific Advisory Committee meeting was organized and prepared. Since it turned out to be impossible to find one date on which all SAC members would be available, we decided to organize two SAC meetings, one in Nijmegen (12 November 2008), and one in Stockholm (15 December 2008).

Table 3.0.1: Deliverables List

Del. no.	Deliverable name	WP no.	Date due	Actual/ Forecast delivery date	Estimated indicative person-months *)	Used indicative person-months *)	Lead contract or
D0.2	Activity Report	0	M24	M24	3	1/2	RUN
D0.3	Dissemination and Use plan	0	M24	M24	2	1/3	RUN
D0.4	Management Report	0	M24	M24	1	1/3	RUN

*) if available

- List of milestones, including due date and actual/foreseen achievement date

For the milestones, see the deliverables.

3.1 WP1 Signal Representations

3.1.1 Workpackage objectives and starting point of work at beginning of reporting period

The primary objectives of work package 1 for the year-two reporting period were the completion of deliverable D1.2 and of milestone M1.3.

Deliverable D1.2 involves the completion of both software and a report on the definition of new features and a feature selection method. It is described in the Technical Annex as "*Modules for a) augmentation of standard spectral features with a stream of milli-second and deci-second features and evaluation on specific phone classification tasks and b) feature selection by sensitivity-analysis method (software and report)*". Item b) refers to the selection of features based on quantitative knowledge of the human auditory periphery in the form of sophisticated auditory models.

Milestone M1.3 was a midpoint check for progress towards deliverable D1.2. It is described in the Technical Annex as "*Validation of sensitivity analysis method and method based on augmentation with milli-second and deci-second features* ". Upon the completion of deliverable D1.2 milestone M1.3 is no longer of significance.

At the start of the reporting period WP1 had completed the conventional feature set and the auditory models for the feature selection by sensitivity-analysis. This is consistent with the planned outcome of the year-one reporting period.

3.1.2 Progress towards objectives – tasks worked on and achievements made with reference to the planned objectives, identify contractors involved

The objectives for the year-2 reporting period were completed according to plan. Milestone M1.3 was met and deliverable D1.2 was delivered on time.

The work towards defining milli-second features has taken the form of an algorithm that automatically estimates the voicing onset of time. The voicing onset time is the time between the onset of the burst and the onset of the voiced signal. This work was performed at K.U. Leuven and has resulting in a software module and a manuscript that is currently under review. The new feature will likely be included in the ACORNS feature set.

The work towards defining deci-econd features is a study on the efficacy of including measures of prosody (rhythm and pitch movement) in the feature set used in the ACORNS project. An improvement in performance was obtained for the learning stage. Once the performance of the recognition system has converged the addition of these deci-second features does not help. The work was performed by Sheffield and by K.U. Leuven and has resulted in a software module and a report that is included in deliverable D1.2.

The work toward selecting features based on a sensitivity analysis method has resulted in an algorithm that can select a subset of features from a larger set of features based on auditory model knowledge only. Tests showed that the resulting feature set performs well in recognition performance. The work has resulted in a software module and in a manuscript currently under review, with a second one in preparation. The work forms the first step towards defining (rather than selecting) features based on auditory model knowledge, which is the main goal of work package 1 in the next reporting period. Work towards this goal for the next reporting period has started. The work towards the selection and definition of auditory features was performed at KTH.

3.1.3 Deviations form the project work programme, and corrective actions taken/suggested

There are no deviations from the project work programme in workpackage 1 for the year-2 reporting period.

3.1.4 List of Deliverables

Table 3.1.1: Deliverables List for WP1

Del. no.	Deliverable name	WP no.	Date due	Actual/ Forecast delivery date	Estimated indicative person-months	Used indicative person-months	Lead
WP1							
D1.1	Modules for conventional feature set	1	M12	12	15	10	KTH
D1.2	Modules for a) augmentation of standard spectral features with a stream of milli-second and deci-second features and evaluation on specific phone classification tasks and b) feature selected by sensitivity-analysis method	1	M24				KTH
D1.3	Final Modules for features derived with sensitivity-analysis method criterion, with quantitative evaluation	1	M36				KTH

3.1.5 List of milestones

Table 3.1.2: Milestones List for WP1

Milestone no.	Milestone name	Workpackage no.	Date due	Actual/Forecast delivery date	Lead contractor
M1.1	Conventional feature set completed	1	M6	M9	KTH
M1.2	Auditory models for sensitivity analysis completed	1	M12	M12	KTH
M1.3	Validation of sensitivity analysis method and method based on augmentation with milli-second and deci-second features	1	M18	M18	KTH
M1.4	New features based on sensitivity analysis method	1	M30		KTH

3.2 WP2: Signal Patterning

3.2.1 Workpackage objectives and starting point of work at beginning of reporting period

The objectives for WP2's two tasks for the project's second year were as follows:

Task 1 - Pattern discovery using discrete model elements (DME)

- To complete the milestone which was deferred from M12 to M15 during the project's first year of operation pertaining to task T1.1, specifically the section that deals with **Auditory pre-processing with DME**s that includes milestone M2.1.1B.

- To complete the subtasks and milestones that were set out for the second year of WP2 Task 1, specifically:

  1. **T1.2 Enhanced PD for higher-level processing** where the aim is to adapt the pattern discovery (PD) module to provide both time and auditory domain segmental measures enabling the module's use in higher level processing. The expected completion date of milestone M2.1.2 "Enhanced PD for higher-level processing" was M18.

  2. **T1.3 Temporal Structures** where compact coding representations are defined for higher-level processes enabling the generation of higher-order representations. These higher-order representations are then to be used in clustering. This subtask was partially completed during the first year with the remainder due by M24.

  3. **T1.4 Auditory memory traces** where temporal structures in wider temporal windows are analysed by using chains of linked DMEs via a statistical model. Representation of auditory stimuli and memory by the use of *concept matrices* is also to be studied. Milestone M2.1.4 is to be reported by M24 and describes the theory and operation of concept matrices.

Task 2 - Pattern discovery with computational mechanics approaches

- To complete the deliverable and milestone that were deferred from the project's first year of operation pertaining to task T2.1, specifically the section that deals with the study of the *Applicability of CMM learning to ACORNS*.

- To complete the deliverable and milestone that were set out for the second year of WP2 task 2 operation dealing with **Learning different hierarchical units from acoustic patterns**. This includes milestone M2.2.2A whose completion date was extended from M15 to M21 at the end of Year 1.

3.2.2 Progress towards objectives – tasks worked on and achievements made with reference to the planned objectives, identify contractors involved

Task 1 - Pattern discovery using discrete model elements (DME)

During the second year's work period, WP2 task 1 concentrated on continuing the research undertaken during the first year related to pattern discovery and extending it to support the overall goals of the ACORNS project. Most of the work in WP2 has been carried out by TKK staff, with substantial contributions from KTH (Stockholm), Leuven and Sheffield.

- The potential use of auditory features was studied by comparing the effectiveness of using different spectral representations (MFCC vs. FFT) in segmentation and incremental clustering. The results

indicated that no significant improvement existed when using a perceptual scale. Details of this study are described in a project internal report published in March 2008 and available on the project's Wiki site: Räsänen, Laine, & Altosaar, "*Use of MFCC and FFT in segmentation and clustering.*" However, the MFCC representation developed by WP1 is being emplyed so that comparisons   can   be   carried   out   with   other   WPs   in   an   easier   manner.

- A study was carried out where incremental bottom-up learning using segmental discrete model elements (DME) was coupled with a new *concept matrix* approach. Word learning experiments using the Year 1 corpus were described and their results presented in March 2008 and published on the project's Wiki. Additionally, a conference paper was presented at the Interspeech'08 conference: Räsänen, Laine, & Altosaar, *"Computational language acquisition by statistical bottom-up processing"*, Proc. Interspeech 2008, pp. 1980-1983. A more generalized version developed from the concept matrix approach resulted in a patent application being filed during the autumn of 2008.

- In March 2008 the ACORNS project lead sugested that each WP look at how attentional focus to keywords and the IDS/ADS difference may influence infant learning. For this reason, several man-months of effort were used by WP2 to study the effects of selective attention in learning. An IDS/ADS comparison was also carried out that included an analysis regarding the possibility of the learner being able to differentiate between ADS and IDS as its input. The results of these studies were reported as follows: i) on the project's Wiki site, ii) in the form of a presentation at the project's quarterly meeting in June 2008, and iii) as a conference paper publication in August 2008: Räsänen, Altosaar, & Laine *"Comparison of prosodic features in Swedish and Finnish IDS/ADS speech",* Proc. of Nordic Prosody X.

- Reinforced learning in segmental transition probability tracking was studied by increasing the significance of correctly recognized keywords.

- During the late spring of 2008 a detailed ~30 page manuscript was submitted for publication in the *Speech Communication* journal. The article concerns the novel blind segmentation algorithm developed by WP2 during the first year of the ACORNS project as well as lays down new theory regarding the general problem of evaluation of automatic speech segmentation systems.

- The exploration and development of methods for detection of recurring patterns in segmental label sequences, provided by the developed incremental clustering algorithm, was performed regarding:
  o A generalizing reformulation of the concept-matrix approach for tracking transitional probabilities.
  o The use of non-linear filtering and dynamic programming approaches in cross-utterance, segmental and spectral distance based detection of recurring sub-word units. The preliminary conclusion of this study was that the utilization of segmental distance information seems to be an overly complex approach as long as there exist other unexplored possibilities based on DME approaches.
  o The detection of recurring units in segmental label sequences using approaches that are similar to the N-gram approaches were investigated.
  o A brief exploration dealing with the compression of recurring label sequences into prototype models that can be used is word detection scenarios was carried out.

- Considerable effort was expended on research into and enhancement of WP2's clustering algorithms in order to address the challenges of incremental blind phone-like unit classification. This work also included the investigation of feature extraction methodologies for describing segmental phone-like units. This was deemed necessary due to imminent higher level processing needs, i.e., the formation of memory.

- In collaboration with Leuven fixed frame vs. segmental based representations. WP2 and WP4 were compared; we are currently working on a joint publication regarding this theme.

- In collaboration with Sheffield a literature overview was carried out regarding acquisition of native phonetic categories in infancy. This overview was published in April 2008 and is available on the project's Wiki site: Räsänen & Aimetti, "*Perception of phonetic contrasts in infants and adults*".

- WP2 provided other ACORNS partners with segmentation and classification for several corpora on demand.

- A new version of the WP2 algorithm was implemented in MATLAB that provides automatic speech segmentation and segmental labelling using the latest available enhancements. This software was made available to the other partners on the project's Wiki site in early September 2008.

- A novel self-learning vector-quantization (SLVQ) method was developed by WP2. This new method works incrementally and adapts to the properties of input data instead of forming a fixed number of clusters in a batch process.

- WP2 has organised and coordinated the WP2 Year 2 deliverable that is in the form of an extended 60 page report entitled "*WP2: Methods for Enhanced Pattern Discovery in Speech Processing*." The report is a multi-site effort where individuals from Leuven (Joris Driesen), Sheffield (Guillaume Aimetti), and Stockholm (Gustav Henter) have participated in its authoring, in addition to WP2's own team.

The activities described above have all been largely motivated by fulfilling the responsibilities related to the second year WP2 deliverable as well as the targeted milestones specified in the ACORNS Technical Annex (TA). In the following section, we describe the relationship of the above activities to the specific WP2 Task 1 goals outlined in the TA.

Milestone M2.1.1B
For the completion of milestone M2.1.1B – that was rescheduled from M12 to M15 during the first year of the project – studies concerning *Auditory pre-processing with DMEs* were conducted and completed by M15. This work culminated with the publication of a project internal report entitled "*Use of MFCC and FFT in segmentation and clustering*" (Räsänen, Laine, & Altosaar, March, 2008) and was published on the project's Wiki.

Milestone M2.1.2
Adapting the pattern discovery (PD) module to provide both time and auditory domain segmental measures enabling its use for higher level processing (subtask T1.2 - Enhanced PD for higher-level processing) was completed as planned. An updated segmentation and incremental clustering software module was made available to the ACORNS group in September of 2008 on the project's Wiki site.

Milestone M2.1.3
Part of the work concerning *compact coding representations* was already completed in year 1. However, work in subtask T1.3 during Year 2 concentrated on refining as well as publishing the results obtained during the first year via submission and revision to a peer-reviewed journal. Finally, a new SLVQ method was developed to efficiently code representations that can be used in the memory architecture of ACORNS.

Milestone M2.1.4
By employing concept matrices to process output from SLVQ, WP2 was successfully able to form internal memory representations that combine time and frequency information of the auditory stimuli for mapping to multimodal concepts.

Task 2 - Pattern discovery with computational mechanics approaches
During the second year, WP2 Task 2 concentrated on continuing the research undertaken during the first year related to CMM and extending it to support the overall goals of the ACORNS project. Specifically, research conducted during Year 2 has covered the following areas: (In the following list, the contractor

involved for Task 2 activities has been TKK unless otherwise specified. The work on CMM theory and the modified CSSR algorithm has been performed by Gustav Henter located at KTH.)

- It was discovered that the CMM causal state representation of many processes is infinitely large, and thus not learnable with CSSR. Specifically, a proof was developed showing that many simple, finite HMMs that satisfy certain easy-to-check criteria must have an infinite number of causal states. This new insight coupled with a sound theoretical understanding explained the divergence problems of applying CSSR to noisy data that was observed during Year 1 work.

- Since the causal state description may be too complex to learn, a modified version of the CSSR algorithm was developed which includes a concept of *resolution*. This enables learning a finite approximation of the causal states of the original process; this is useful since the proof described above suggests that many causal states may be similar. The extension was performed in two parts:

  1. A concept of *resolution* was added to the statistical tests in the homogenization stage of the CSSR algorithm. This enables the modified algorithm to recover CSSR-learnable processes that have been disturbed by substitution noise.

  2. The CSSR algorithm was additionally modified at the determinization stage to allow for *partial determinization*. In many empirical cases this is the stage where explosive state growth occurs within the original CSSR code. The modification can contain the divergence also in cases where the data is not a finite state process with added substitution noise. Together, the improvements allow a trade-off between prescience and complexity, and may well be the first CSSR-based algorithm useful for practical real-world applications.

- Software capable of performing *CSSR* as well as *CSSR with resolution* was implemented in C++. This work took longer than expected due to unexpected subtleties in the original CSSR algorithm. The software with documentation was made available in mid-December 2008.

- Since the computational complexity can increase quickly with the number of symbols, CSSR may not be applicable at all hierarchical levels. The prime candidate for applying CSSR-based methods is currently the phone level where the alphabet of symbols is naturally limited to the order of ~40 phones.

The Task 2 activities described above have been motivated by fulfilling the responsibilities related to the first and second year WP2 deliverables as well as the targeted milestones specified in the ACORNS TA. In the following section, we describe the relationship of the above activities to the specific WP2 Task 2 goals outlined in the TA.

Milestone M2.2.1

For the completion of milestone M2.2.1 – that was rescheduled from M9 to M15 during the first year of the project – in-depth studies concerning the ***Applicability of CMM learning to ACORNS*** were continued from the first year. A presentation of the state of the research was presented at the project meeting in June 2008. The presentation entitled "*Making CSSR Work in ACORNS*" is available on the project's Wiki site (file: WP2_CMM_CSSR_Sheffield_08_v1_2.pdf). This report concluded that the CSSR algorithm with modifications would be able to deal with noise and could support important ACORNS architectural paradigms such as *reusability* and *hierarchy*.

By September of 2008 more work had been performed on re-implementing the original CSSR algorithm so that it could operate more effectively in noise. A report entitled "*The State of CSSR with Resolution*" was presented at the quarterly meeting in early September 2008 (Wiki file: WP2_CMM_CSSR_Helsinki_08_v1_1.pdf). New theoretical work in developing and implementing a suitable statistical test, the implementation of *tolerance* for non-deterministic transitions, as well as testing of these ideas using various data, was described.

CSSR software developed by Gustav Henter located at KTH was made available to the ACORNS project thus allowing its integration with the rest of the project to commence.

Milestone M2.2.2A
WP2 Task 2 was also scheduled to work on **Learning different hierarchical units from acoustic patterns**. This milestone whose completion time was extended originally from M15 to M21 was not reached yet since only preliminary tests have been performed on real data originating from speech.

3.2.3 Deviations from the project work programme, and corrective actions taken/suggested

Task 1 - Pattern discovery using discrete model elements (DME)
All planned milestones for WP2 Task 1's second year of operation were met. Currently, WP2 Task 1 is ahead of schedule since part of Year 3's subtask T1.5 (Self-directed search) was already competed during the first year. WP2 Task 1 will continue to focus on pattern discovery using the DME approach for the project's final year of operation.

Task 2 - Pattern discovery with computational mechanics approaches
The first year milestone M2.2.1 was reached much later than anticipated (M9 ➜ M24) due to significant problems and errors within the original CSSR algorithm that were not described nor made apparent in the literature. However, a more thorough understanding of these problems was gained that has allowed a theoretical proof to be developed that extends CMM/CSSR's applicability to noisy data, i.e., data that commonly exists in the ACORNS context. Due to this unforeseen delay, second year milestone M2.2.2A *Learning different hierarchical units from acoustic patterns* was not activated yet since M2.2.1 had not been reached in its entirety until the end of the second year. However, now with a better understanding of the shortcomings of the original CSSR algorithm, work on M2.2.2A can commence.

Since research up to now has determined that "CSSR may not be applicable at all hierarchical levels" (see section 3.2.2, Task 2, bullet point 4), it may be more realistic and rewarding to focus on applying CSSR to the phone level initially. Once this has been studied in depth, we should be in a better situation to indicate whether CMM/CSSR will allow itself to be applied to model other units practically, and furthermore, whether it is feasible to apply CMM/CSSR to learn syntactic structure. Since there are less than 12 months remaining in the project, it is doubtful whether the remaining two research milestones dealing with learning (M2.2.2A and M2.2.2B) can be studied in the full scope as was originally intended at the Technical Annex's time of writing. However, high interest in the new noise robust CSSR algorithm from other WP's, e.g., WP4, may help in testing its applicability more thoroughly and rapidly.

Taking into consideration the current state of progress, new target completion dates for the remaining milestones for Task 2 are being proposed as follows:

     M2.2.1:   Applicability of CMM Learning for ACORNS   M9 ➜ M24 (completed)
     M2.2.2A: Learning different hierarchical units from acoustic patterns   M15 ➜ M28 (started)
     M2.2.2B: Learning syntactic structure   M27 ➜ M32
     M2.2.3:   CMM learning in ACORNS.   M33 ➜ M36

3.2.4 List of Deliverables

Table 3.2.1: Deliverables List for WP2

Del. no.	Deliverable name	WP no.	Date due	Actual/Forecast delivery date	Estimated indicative person-months *)	Used indicative person-months *)	Lead contractor
D2.1	Task 1: PD/DME Modules. Task 2: Applicability of CMM Learning for ACORNS	2	M12 M9	M12 M24			TKK
D2.2	Task 1: Enhanced PD, Compact Coding, Linked DMEs Task 2: Learning linguistic units & syntactic structure with CMMs	2	M24 M24	M24 M36			TKK

*) if available

3.2.5 List of Milestones

Table 3.2.2: Milestones List for WP2

Milestone no.	Milestone name	WP no.	Date due	Actual/Forecast delivery date	Lead contractor
Task 1					
M2.1.1B	Auditory pre-processing with DMEs	2	M12 => M15	M15	TKK
M2.1.2	Enhanced PD for higher-level processing	2	M18	M18	TKK
M2.1.3	Temporal structures	2	M24	M24	TKK
M2.1.4	Auditory memory traces	2	M24	M24	TKK
M2.1.5	Self-directed search	2	M30	M30	TKK
Task 2					
M2.2.1	Applicability of CMM learning for ACORNS	2	M9	M24	TKK
M2.2.2A	Learning different hierarchical units from acoustic patterns	2	M15	M28	TKK
M2.2.2B	Learning syntactic structure	2	M27	M32	TKK
M2.2.3	CMM learning in ACORNS.	2	M33	M36	TKK

3.3 WP3 Memory Organisation and Access

3.3.1 Workpackage objectives and starting point of work at beginning of reporting period

The starting point for period 2 was a project internal report entitled 'Report focusing on the memory architecture requirements' which considered various approaches to modeling intelligent behavior for inclusion into the memory architecture and the requirements the make for the memory architecture.

The objectives of the Workpackage 3 are:

Creation of a synthetic memory structure that exhibits recognisable psychological behaviour as an emergent bi-product.

Design and implementation of mechanisms and computational models of working memory architecture and access.

Inclusion of attention within the overall memory architecture.

Investigation and implementation of episodic and semantic memory within the memory-prediction Framework

Linking memory-prediction framework with dual-purpose sensory-motor representations

3.3.2 Progress towards objectives – tasks worked on and achievements made with reference to the planned objectives, identify contractors involved

To reach the targets in Workpackage 3 described above, the following activities have taken place:

(i) We have developed a memory architecture based on evidence from psychological and emergent behaviour in which working memory and long-term memory models are incorporated .

(ii) Within the ACORNS architecture there as been the development of various semantic long-term memory models that perform selective attention mechanism. One attention model is used within keyword recognition (Deliverable Workpackage 3.2) Subsection and another acts as a gating mechanism to differentiate recognised speech (Deliverable Workpackage 3.2).

(iii) A mechanism as been developed to act as working memory, which produces activations patterns related to the auditory and semantic (visual) feature input, which is used to update weights structure that are stored in semantic and episodic memory. One specific model that has been developed that incorporates this working memory structure is a recurrent self-organising architecture that is used to associate speech signals and semantic features to develop an emergent representation of speech (Deliverable Workpackage 3.2). This model offers a hierarchical structure that incorporates certain features of the memory-prediction framework and as such can be extended to include more features.

(iv) Steps have been made toward the implementation of episodic and semantic memory within the memory-prediction Framework. In terms of the episodic long-term memory there has been the implementation of the extension of the instance based MINERVA2 approach so it is able to perform temporal based processing through a 'bag-of-frames' approach. This model as been successful used to perform speech and keyword recognition (Deliverable Workpackage 3.2). The use of an instance based model of episodic model will be considered further in period 3 of the project with the application of a new model known as TEMM (Temporal Episodic Memory Model). Within the overall ACORNS architecture semantic long-term memory as been used as stored weight structures in models for selective attention and the emergent speech

representation for instance. This interaction between the working memory and semantic long-term mechanism is a focus of ACORNS architecture and will be on considered further in period 3.

(v) Preliminary actives have been carried out towards the design of an actor-critic model of reinforcement learning for sensor-motor representation (Discussion Document AC1).

3.3.3 Deviations form the project work programme, and corrective actions taken/suggested

There as not been any clear deviation from the WP planned activities.

3.3.4 List of Deliverable

Table 3.3.1: Deliverables List for WP3

.

Del. no.	Deliverable name	Work package no.	Date due	Actual/Forecast delivery date	Estimated indicative person-months *)	Used indicative person-months *)	Lead contractor
D3.1	Report focussing on the memory architecture requirements	3	30/11/07	28/11/07	12	6	USFD
D3.2	Report focusing on the results of the initial ASR experiments comparing episodic and semantic memory	3	M24	26/11/08	28	28	USFD

*) if available

3.3.5 List of Milestones

Table 3.3.2: Milestones List for WP3

Milestone no.	Milestone name	Workpackage no.	Date due	Actual/Forecast delivery date	Lead contractor
3.3	Initial Release of the software simulations of working memory and attention Mechanisms	3	M18	30/06/2008	USFD
3.4	Initial Results of ASR experiments comparing episodic and semantic memory	3	M24	30/11/2008	USFD

Milestone 3.3 combines the (not numbered) milestones for the tasks 3.2, 3.3 and 3.5 mentioned on pg. 61 of the Technical Annex.

3.4 WP 4 Information discovery and integration

3.4.1 Workpackage objectives and starting point of work at beginning of reporting period

Workpackage objectives from the TA:
1.   To develop information discovery and integration mechanisms.
2.   To study how content addressable memory can be used for information representation and access.
3.   To investigate how to associate speech features and patterns with speech events and evidences.
4.   To integrate exemplar-based matching and high-dimension salient feature representation for access.

The second objective will be addressed in the final year (see also section 3.4.3) and we will not comment further on it here. Information discovery (objective 1) as well as objective 3 were already shown in the first year. In this reporting period, we made further progress towards objectives 1 and 4.

The following results were obtained in the first year and serve as the starting point for the second year of activity on WP 4:
1.   A framework for unsupervised and weakly supervised learning of speech objects based on NMF
2.   A method for pattern discovery in symbolic streams based on the *multigram* algorithm

3.4.2 Progress towards objectives – tasks worked on and achievements made with reference to the planned objectives, identify contractors involved

In the first year, we explored multigrams and NMF as information discovery algorithms. We concluded the research on multigrams, expanded the NMF research and report on progress on exemplar-based matching. NMF is viewed as an algorithm that discovers patterns and grounds them. Once the patterns are discovered, an *exemplar association algorithm* (to be developed) can attach the grounding information as meta data to the patterns. Hence a paradigm shift for learning and recognition could be implemented: initially, pattern discovery and recognition are both based on NMF, but as the pool of exemplars enriched with meta data becomes richer, exemplar-based recognition could take over. This is in line with human language acquisition, where initially phone transition statistics play an important role (modeled by NMF) (Safran *et al.,* 1996) and exemplar based learning and decoding takes over later. Hereto, we also envisage mechanisms in which the meta data for the exemplars can be refined based on success statistics. Hence, exemplar-based matching can be seen as either a competing model or as a method to complement the existing methods.

Multigrams

The multigram algorithm (Deligne and Bimbot, 1997) finds recurring patterns in symbolic streams. In the first year, we initiated an evaluation if it can be applied to word acquisition from spoken utterances that are accompanied by information from other modalities. In the original multigram algorithm, symbolic input is explained by a set of units, multigrams, that emit symbolic strings stochastically. In our (and their) setup, each multigram is modeled by a Hidden Markov Model (HMM). The set of multigrams, their topology and parameters need to be learned. We have extended the multigram learning algorithm to cope with ambiguity, i.e. the input is not a string of symbols any more, but a lattice of symbols that describes a large collection of possible input symbols. Secondly, we have designed a method to link the discovered patterns (multigrams) with the information in other modalities.

We have successfully acquired the 11 digits from phone lattices on a corpus of spoken digit strings and mapped the acquired word models to word identity tags (Driesen and Van hamme, 2008) and have shown that using lattice input indeed leads to better performance than the original multigram algorithm. Tests on artificially generated symbol sequences have shown that even if up to 35 % of the symbols are randomly substituted for another symbol, the multigram algorithm was successful in finding the underlying patterns that generated the stream. Though this was an encouraging result, we did find that discovering long patterns in noisy input does have limitations. For sequences of vector quantized speech spectra, we have found that initialization of the algorithm is critically sensitive. Because NMF does not have this down side, we decided

to *abandon* this method. If we succeed in building hierarchies such that the patterns in each layer are only a few symbols long, we may reconsider the multigram algorithm.

Non-negative Matrix Factorization

In NMF-based learning, counts data of the co-occurrence of acoustic events is stored in a matrix. Each column of this data matrix $\mathbf{V}$ contains the counts data (histogram) of an utterance. Learning is achieved by factorizing $\mathbf{V} \approx \mathbf{W}\,\mathbf{H}$ into a factor $\mathbf{W}$ that contains the learned internal representations along its columns and $\mathbf{H}$ whose columns contain the *activity* of each learned representation in an utterance. We extended the NMF-based learning as follows:

1. an incremental learning mechanism was proposed and evaluated (to be published). With our first-year batch method, when new utterances are presented to the learner, all utterances heard since birth (or as long as the memory reaches) needed to be reprocessed to find the updated internal word representations $\mathbf{W}$. With the incremental NMF update, only the current utterance in sensory store is used to update $\mathbf{W}$. In other words, the learning *state* is completely determined by $\mathbf{W}$, which resides in semantic long term memory. Of course, nothing would prevent us from simulating *rehearsal*, i.e. repeated offering of utterances that are stored in episodic long-term memory to strengthen their representation in $\mathbf{W}$.

2. a method for estimating the position (in time) of activated word representations was proposed and evaluated (Van hamme, 2008b). We reported last year that NMF-based recognition activates words from utterance-level data and hence only *detects* words in an utterance, without ordering them. We described a method to estimate the time at which these activated words occur within the utterance, and therefore allows to order the words. We further refined the resolution by sliding a window over the utterance and showing a recognition rate which is about 40% higher than that of a HMM trained on the same symbolic sequence but with supervision.

3. information integration in NMF was studied in (Van hamme, 2008a). We showed that several information streams can be combined to achieve better accuracy. More specifically, we combined phone lattice information and quantized speech spectra as input streams. With this, we also achieved the second part of the first objective.

4. NMF was applied to speech spectra to discover (without supervision) patches in the time-frequency plane which are relevant to speech and which showed noise robustness in a recognition experiment (submitted to ICASSP 2009). This is equivalent to learning receptive fields of speech. We later extended this work to include a *second NMF layer* which is fed by the time-frequency-patch activations and learns words under weak supervision. Hence we show we can *cascade* NMF-based learning layers.

5. NMF was applied to learn a representation of grammar (though a simple representation - see Deliverable D4.1) by *cascading* NMF layers. The first layer directly maps acoustic events to words (in a single layer, unlike the previous item, which would achieve the same mapping in two layers). As before, it uses co-occurrence statistics of acoustic events at a time scale of maximally 100 ms. A second layer maps the word activation patterns at a 400 ms time scale to new word activations. This mapping now takes the context of words into account and models that some words are more likely to occur before or after others. We showed that this second layer improves the recognition performance.

6. a feature selection mechanism for NMF was proposed (unpublished). This work is still in progress and a full report will be presented in the next year. The goal is to weigh the acoustic features according to their relevance for the classification task. An important application would be *a self-organizing layered NMF architecture* in which a layer at a higher level receives the same inputs as the lower layers plus the outputs of the lower layers. If a lower layer is doing a proper job, its outputs should be selected. Likewise, the higher layer should select the outputs of the best performing lower layers.

7. activation-verification mechanisms were proposed and studied (see Deliverable D4.1). We studied three mechanisms: (1) the activation level of a word needs to exceed a threshold, which leads to a detection formulation of utterance-level recognition. We provide DET-curves that show the trade-off between missed detections (false negatives) and false alarms (false positives), (2) the estimated word locations (see item 2) need to be consistent over time in a sliding window decoder, i.e. detected words are only accepted if they are placed in approximately the same position in subsequent analysis window positions, (3) the activated word sequence makes sense according to the language model, implemented as the top-level NMF layer of item 5. This top layer was trained on previous exposures to the language and hence verifies if the activated word sequence matches with the expectations.

8. a top-down learning mechanism was tested (Milestone M4.1.2). The idea was to first acquire a vocabulary $\mathbf{W}$ and then to factorize this vocabulary as $\mathbf{W} \approx \mathbf{X}\,\mathbf{P}$. In other words, each column of $\mathbf{W}$ (the co-occurrence statistics of a single word) is written as a linear combination of the columns of $\mathbf{X}$ (the co-occurrence statistics of sub word units) with weights in the corresponding column of $\mathbf{P}$. Given the phonemic structure of the vocabulary, $\mathbf{P}$ could reveal which phonemes the words are made of. Of course, there is no guarantee that we would discover phonemic structure like this, and the units could actually turn out to be allophones, syllables, morphemes, … The results on a vocabulary of about 648 frequent words from the Wall Street Journal as well as from 90 words from the first and second year ACORNS database were disappointing. Even for a large number of sub word units, the impact on accuracy is significant. For instance, even 150 subword units to model the 648 word vocabulary made the error rate rise from 8 % to 32 %. We attribute this effect to the strong context-dependency in our representation and in the data. The bottom-up approach of item 4 seems a more appropriate route to follow.

Exemplar-based matching

Exemplar-based matching differs fundamentally from NMF or HMM based approaches in that the reference data (the previous input, e.g. the train database) is not condensed into a model. Instead, those pieces of the reference data (preprocessed audio enriched with some meta-information) that are found to be relevant for decoding some new input data (e.g. a test sentence) are accessed directly from memory. This avoids the information loss typically seen when abstracting large amounts of data into compact models such as HMMs. To evaluate the potential of the exemplar-based approach, we first idealized the meta data, lexical and grammatical knowledge sources:

1. phone labels were provided by forced Viterbi-alignment with an existing HMM as an alternative to either the top-down (Milestone M4.1.2) or the bottom-up approach (see Deliverable D2.1) for segmentation
2. the transcription of all words in the vocabulary was provided instead of using discovered phone patterns such as what was achieved in Stouten *et al.* (2008).
3. the language model is a hand-crafted context-free grammar

The following steps were undertaken:

4. As a starting point we evaluated our existing example-based speech recognizer (De Wachter, 2007) on the ACORNS year 2 Dutch database. This resulted in a word error rate (WER) of 6.43%. An HMM system trained (supervised) on the same data yielded a WER of 0.23%. The two main problems in the template approach are (1) the bottom-up activations driven by the acoustics do not always match the top-down predictions by the lexicon and the language model, and (2) gaps (unexplained chunks of audio) occur in the input audio.
5. Introduce probability estimates. Fundamentally, example based approaches replace the parametric probabilistic models by non-parametric probability estimates. Rather than starting from bottom-up activation, the search is driven top-down with probabilities generated by nearest neighbors. The behavior of a simple k-NN (k nearest neighborhood) probability estimator was compared with the HMM approach when classifying frames, phones and words. This showed that the kNN (exemplar) approach outperformed the HMM on a frame by frame phone classification task and on continuous phone recognition. When traces of a fixed number of frames were used, the kNN approach extended its lead even further.
    Two variants for kNN based word recognition were explored. For each frame, the word identity is available from the meta data. A first experiment just integrated the frame by frame kNN-based word evidences. This approach does not impose any constraints on the order of the sounds in a word, nor does it require that all sounds are present. This approach yielded a rather high WER of 11% (at the time of writing this text). Describing a word as a sequence of phones using the same dictionary as the HMM system, solved the two main shortcomings and resulted in a 0.23% WER – identical to that of the state-of-the-art HMM system (though the errors were different)
6. Revise bottom-up activation mechanisms (a form of content-addressable memories and hence relevant to M4.1.3 – see also section 3.4.3). One of the main requirements for making the template based search computationally feasible is a method to quickly find the relevant frames or short traces from the reference data given some new input data. To that end, the roadmap algorithm was tested,

both on artificial data and on real speech data. The results are very promising. However, we still need to devise a method to create the roadmap structure for large amounts in an efficient manner.

7. In a masters thesis (De Tollenaere, 2008), spectral clustering was used as a technique for finding reusable acoustic units. Spectral clustering has the advantage that meta information such as likely segment boundaries or typical usage patterns (which word uses this type of audio) can be integrated in the clustering. Even without using such high-level information, De Tollenaere obtained very promising results. Spectral clustering could serve as the knowledge discovery and integration mechanism (referred to in the introduction of this section 3.4.2) in the exemplar-based approach.

8. Starting from the observations made in the previous attempts (top-down and bottom-up) to automatically derive reusable acoustic units, and based on the exemplar-based results obtained so far and on the work described in (De Tollenaere, 2008), a comprehensive plan for the next year was proposed. This encompasses a complete system, including a promising technique for finding reusable acoustic units.

More details about the progress made on exemplar-based matching can be found in (Demuynck, 2008).

3.4.3 Deviations form the project work programme, and corrective actions taken/suggested

Content-addressable-memories (CAMs) were proposed in the technical annex to address anticipated scaling problems on the assumption that searching in memory would be an issue. For the exemplar-based approach, it is indeed. We are addressing this problem with the roadmap algorithm (see item 6 under "exemplar-based matching" of section 3.4.2), so there seems to be little interest in studying the binary approximations involved in CAMs. For the NMF-based approach, we have learned in the project that building a successful hierarchy of self-learned representations is more important to achieve scalability. In other words, we are not confronted with a significant speed or search problem, so the binary approximations proposed in the technical annex do not seem to solve a problem of significance now. Hence, spending the effort on the hierarchy of representations seems to be more productive now. This modification will not impact the deliverables.

3.4.4 List of Deliverables

Table 3.4.1: Deliverables List for WP4

Del. no.	Deliverable name	WP no.	Date due	Actual/Forecast delivery date	Estimated indicative person-months *)	Used indicative person-months *)	Lead contractor
D4.1	Implementation and test of activation-verification mechanisms	WP 4	M12 = 30 Nov 2007	M24 = 30 Nov 2008			KUL
D4.2	Report on LSA representation and SVD dimension reduction	WP 4	M24 = 30 Nov 2008	M12 = 30 Nov 2007			KUL
D4.3	Report on exemplar-based and activation-based matching	WP 4	M36 = 30 Nov 2009	M36 = 30 Nov 2009			KUL

*) if available

3.4.5 List of Milestones

Table 3.4.2: Milestones List for WP4

Milestone no.	Milestone name	Workpackage no.	Date due	Actual/Forecast delivery date	Lead contractor
M4.1.1	Old description: Activation-verification mechanism implemented on 1st layer. New description: Implementation of activation mechanisms at the tag level.	WP 4	M12 = 30 Nov 2007	M12 = 30 Nov 2007	KUL
M4.1.2	Old description: Activation-verification mechanism implemented on 2nd layer. New description: Top-down learning of patterns and computation of activations	WP 4	M24 = 30 Nov 2008	M24 = 30 Nov 2008	KUL
M4.1.3	Old description: Activation-verification mechanism implemented on 3rd layer. New description: Activation/verification mechanisms in hierarchical CAMs	WP 4	M36 = 30 Nov 2009	M36 = 30 Nov 2009, relevance questioned – see section 3.4.3	KUL
M4.2.1	LSA representation and SVD dimension reduction	WP 4	M24 = 30 Nov 2008	M12 = 30 Nov 2007	KUL
M4.2.2	ASMs defined from WP1 and WP2 features and automatic segmentation	WP 4	M36 = 30 Nov 2009	M36 = 30 Nov 2009	KUL
M4.3.1	Time synchronous exemplar-based and activation based matching	WP 4	M24 = 30 Nov 2008	M24 = 30 Nov 2008	KUL
M4.3.2	Time-asynchronous matching and non-Euclidean distance	WP 4	M36 = 30 Nov 2009	M36 = 30 Nov 2009	KUL

References

J. Saffran, R. Aslin, and E. Newport, "Statistical learning by 8-month-old infants," Science, vol. 274, no. 5294, pp. 1926–1928, 1996.

S. Deligne and F. Bimbot, "Inference of variable-length linguistic and acoustic units by multigrams", *Speech Communication*, vol. 23, pp. 223–241, 1997.

Joris Driesen and Hugo Van hamme (2008). "Improving the Multigram Algorithm by using Lattices as Input", In Proc. *International Conference on Spoken Language Processing*, pp. 2086-2089, Brisbane, Australia, September 2008.

Hugo Van hamme (2008a). "Integration of Asynchronous Knowledge Sources in a Novel Speech Recognition Framework", In *Proc. ITRW on Speech Analysis and Processing for Knowledge Discovery*, Aalborg, Denmark, june 2008. 4 pages, ISBN 978-87-92328-00-7.

Hugo Van hamme (2008b). "HAC-models: a Novel Approach to Continuous Speech Recognition", In Proc. International *Conference on Spoken Language Processing*, pp. 2554-2557, Brisbane, Australia, September 2008.

Veronique Stouten, Kris Demuynck and Hugo Van hamme. "Discovering Phone Patterns in Spoken Utterances by Non-negative Matrix Factorisation", *IEEE Signal Processing Letters*, volume 15, pages 131-134, 2008.

Kris Demuynck. "Time synchronous exemplar-based and activation based matching", Technical Report PSI-SPCH-08-2, K.U.Leuven/ESAT, December 2008.

Mathias De Wachter. "Example Based Continuous Speech Recognition", PhD thesis, K.U.Leuven, ESAT, May 2007.

Joost De Tollenaere. "Zelfleren spraakherkenning: akoestische eenheden en woordmodellen", masters thesis, K.U.Leuven, ESAT, 2008.

3.5 WP5 Interaction and communication

3.5.1 Workpackage objectives and starting point of work at beginning of reporting period

The research in this WP is structured in four tasks.

Task 5.1 Creation of a platform for learning in the memory-prediction framework
The overall objective is to create the basic software environment that is needed to integrate the modules produced in WP1 – WP4 and to conduct experiments with language learning (these experiments are further specified in Task 5.4). The platform will come in two versions: one for off-line experiments, and one that can be used for demonstrations.
The starting point of the work in this reporting period is the platform in its status December 2007. In that platform, the leaner was able to process multimodal stimuli in the form of acoustic realizations (utterances) in combination with one invariant symbolic tag. The platform consisted of two interactive modules, the caregiver and the learner. The learning takes place in interaction between these modules. This platform was used for all Y1 experiments.

Task 5.2 Multimodal integration
This task is dedicated to the development of procedures and software for the integration of speech input and visual input for disambiguating spoken utterances and feedback that is equivalent to hugging.
The starting point of the work in this reporting period is the implementation of the integration module which is able to combine (fabricated, crisp) visual tags with (actual, realistic) audio information. The crisp tag was used by the learner to reinforce the creation of a new internal representation. This type of integration was used in all Y1 experiments.

Task 5.3 Architecture for interaction
Throughout the project we assume that our learning agent (Little Acorns) is endowed with an innate urge to communicate with people in her environment, especially her caregivers. We simulate this urge by designing the learner such that she will attempt to maximise the value of a function equivalent to 'caretaker attention' or 'caretaker appreciation'.
As the starting point of the work in this reporting period, the interaction was based on a simple learning loop between caregiver and learner, in combination with a simple (but conceptually defendable) learner design. This system's architecture has been in accordance with the assumption that learning results in a structure of layered perception-action loops. The learning agent is able to produce behaviour meant to attract the attention of its environment and show that it has noticed that it is being addressed by somebody. At the end of year 1, the computational model was in line with the basic structure of this architecture.

Task 5.4 Experiments with language learning
The overall objective in this task is to perform experiments, corresponding to three stages of language learning. At the start of this reporting period the learner showed to be able to perform basic simple word learning tasks on the Y1 database, containing utterances with about 10 target words per speaker per language. The accuracy of the learner varied dependent on the type of learning. Accuracy rates of 95 percent were obtained by an NMF-based learning approach. In general, the learning curves showed a dependency on the ordering in which stimuli are presented.

3.5.2 Progress towards objectives – tasks worked on and achievements made with reference to the planned objectives, identify contractors involved

All contractors have been actively involved in all activities in WP5.

The activities in WP5 during the second year span all tasks. There have been no swaps in this work package, and all subtasks (5.1-5.4) are addressed in parallel.

Task 5.1: Purposeful learning to communicate (c.q. platform)

In year 2 we have elaborated upon the learning drive in connection to learning loops. The computational model models this learning process within two connected loops: an interactive 'external' loop between learner and caregiver, and an 'internal' loop within the learner. In the first year of the project we have used a fairly simple target function of a number of observable variables, namely the number of correct and incorrect responses on an utterance-by-utterance basis. In the second year we have extended this function, by elaborating more on the connection between a high-level description of the learning goal ('need of the learner to receive care and protection and food') and a low-level description of mathematically explicit target function. As a result, the learner is now better able to react on the replies given by the carer.

Furthermore, with respect to platform enhancements in the second year, a new caregiver and learner have been devised. The caregiver can now respond with multiple types of replies (including corrective sentences such as 'no I mean FISH'); the new learner can deal with vector-based visual representations of the visual scene. The combination of new caregiver and new learner was applied for the NMF-based learning experiments as reported on the Y2 database. The experiments that were done so far on the Y2 database focus on specific aspects of the wide range of communication strategies that can now be modeled by the new platform (details are given in D5.4.2).

Task 5.2: Multimodal Integration

Realistic communication is of pivotal importance. In year 2, we therefore pursued the direction initiated in year 1, namely by specifically addressing the issue of cognitively plausible input representations. We extended the original approach with invariant symbolic tags that was used in the Y1-experiments to the ecologically much more plausible approach in which the visual channel is represented by a feature representation ('feature-based' approach). This provides cognitively much more plausible representations of the inputs in the visual channel, but at the same time inspired deep questions and practical issues concerning how to model a situation in which more than one concept is referred to in the audio channel. The method for representing visual inputs in a probabilistic manner is compatible with the basic tenets of the memory-prediction theory.

Task 5.3: Architecture

In year 2 (March-May), the memory architecture has been discussed, updated and consolidated. This has been prepared and carried out by a task-specific ACORNS task force.

The architecture is now based on a more explicit memory model, in which sensory store, short-term memory and long-term memory have a more specific function in the model, and in which the content of each memory is made more explicit.

The design of the learner as used in year 2 is based on the updated architecture. Therefore, there is a close connection with Task 5.1. The internal loop (a perception-action loop) within the learner allows the learner to handle repetitive and familiar situations 'from memory', i.e., without an explicit reasoning stage. The architecture of the learner also contains a module that models the evolving need of the learner 'to understand her environment'. This module actually drives the entire learning process. The module translates the need to understand the environment into the need to optimally parse the stimuli input in terms of what it knows (internal representations).

Task 5.4: Experiments

In year 2 the emphasis has been on learning 50 words and on building relations between acoustic patterns and simultaneously presented visual input. To that end, we recorded new data in three languages (Dutch, English and Finnish) with sentences that refer to increasingly more complex scenes (up to four significant concepts).

We extended the experiments done in year 1, by presenting utterances in which more than one concept is referred to, and by presenting corrective sentences of the kind 'no I mean FISH'. Furthermore, we exploited a number of learning techniques in parallel, including Non-Negative Matrix Factorisation, DP-ngrams, Multigrams, and Computational Mechanics Modelling, in order to functionally compare these techniques and to consider the interoperability options. The various experiments, based on different learning algorithms,

were able to show emerging representations in the form of episodes (DP-ngrams), emerging segmentations comparable to phone level (Concept Matrix-approach), and speaker-adaptive internal representations (NMF). All experiments use either tags or semantic features. A number of experiments focus on the exploration of hierarchical representations, while other experiments focus on a mono-layer versus multi-layer speech decoding step, or on the construction of sub-word units. More details are given in deliverable D5.4.2.

3.5.4 List of Deliverables

Table 3.5.1: Deliverables List for WP5

Del. no.	Deliverable name	WP no.	Date due	Actual/Forecast delivery date	Estimated indicative person-months *)	Used indicative person-months *)	Lead contractor
D5.4.1	System demonstrating the capacity for acquiring language and communication skills	WP5	M12 = 30 Nov 2007	M13 = 30 Nov 2007	26		RUN
D5.4.2	System capable of learning a 50 word vocabulary	WP5	M24 = 30 Nov 2008	M25 = 30 Nov 2008	23	23	RUN
D5.4.3	System capable of rapidly learning a large vocabulary	WP5	M36 = 30 Nov 2009		22		RUN

*) if available

3.5.5 List of Milestones

Table 3.5.2: Milestones List

Milestone no.	Milestone name	WP no.	Date due	Actual/Forecast delivery date	Lead contractor
M5.4	Specification of second year experiments	5	M15	M18	RUN
M5.5	Complete implementation of improved learning system	5	M22	M22	RUN

The milestones M5.4 and M5.5 have been met.. With respect to M5.4, a list of experiment has been specified (see Deliverable D5.4.2). With respect to M5.5, partners contributed to essential modules in the computational model of the learner. These modules have been used for the year-2 experiments (see Deliverable D5.4.2).

3.6 WP 6 dissemination and Use

There are five tasks in this workpackage.

The first task is related to the maintenance of a public website. The project website was regularly updated, see www.acorns-project.org. The public website provides access to all publications and public deliverable. In addition, it provides general information, contact information for the consortium and the consortium members, meeting dates, etc.

The second task relates to organising a workshop dedicated to topics of ACORNS. The first workshop was held at the end of the first year of ACORNS, instead at month 18 because additional funding by the European Science Foundation (ESF) for an exploratory workshop, and by the Dutch National Science Funding (NWO) for international activities offered the possibility to organise a high level workshop.
At this workshop it has been decided to write a proposal for an ESF Research Networking Programme (due 23 October 2008). The final workshop planned on 11th of September 2009, as a satellite event to the large-scale Interspeech Conference in Brighton, UK.

The third task relates to open source software. This task is planned for the final year.

The fourth task deals with the publications in ACORNS. Given the fact that only fundamental scientific research is done in this project, this tasks mainly relates to encourage writing of published papers. Ten papers, and one Master's thesis were published during the second year of ACORNS and more papers are planned for the final year.

One of the reviewers, Stephen Cox, wrote a small article about ACORNS (titled "Mighty Oaks from Little ACORNS grow") in the Summer 2008 issue of the IEEE Signal Processing Society Speech and Language Technical Committee (SLTC) e-Newsletter, http://www.ewh.ieee.org/soc/sps/stc/News/NL0807/index.htm.

The ESR Research Network proposal "Language models of Evolution, Acquisition, and Processing" (LEAP), which can be considered as a spin-off of ACORNS, because it is an outcome of the first workshop, was submitted to ESF in October 2008.

We have prepared an overview paper with contributions of all members of the project, summarizing the most important results of the first two year of the project. The first goal of this paper was to inform the members of the SAC, in preparation for the meetings. We will use this paper (that can be obtained from the ACORNS public website) as a basis for future publications addressing a wider audience. The text is also available as a new deliverable D6.5.

Publications
In the second year, ACORNS has resulted in ten conference papers, a book chapter, and one master thesis.

1. Hugo Van hamme "Integration of Asynchronous Knowledge Sources in a Novel Speech Recognition Framework", ISCA ITRW, *Speech Analysis and Processing for Knowledge Discovery*
2. Louis ten Bosch, Hugo Van hamme , Lou Boves "Unsupervised detection of words – questioning the relevance of segmentation", ISCA ITRW, *Speech Analysis and Processing for Knowledge Discovery*
3. Louis ten Bosch, Lou Boves "Language acquisition: the emergence of words from multimodal input", in Sojka, P., Horák, A., Kopecek, I & Pala, K. (Eds.) *Text, Speech and Dialogue, 11th Intern. Conference, TSD 2008*, Brno, pp. 261-268
4. Klein, M., Frank, S., van Jaarsveld, H., ten Bosch, L.F.M., & Boves, L. "Unsupervised learning of conceptual representations - a computational neural model", *Proc. 14th Annual Conference on Architectures and Mechanisms for Language Processing (AMLaP)*, 4-6 September 2008, Cambridge, UK

5. Okko Räsänen, Altosaar, T. & Laine U.K. (2008) Comparison of prosodic features in Swedish and Finnish IDS/ADS speech. *Proc. of Nordic Prosody X.*
6. Okko Räsänen, Unto K. Laine, Toomas Altosaar "Computational language acquisition by statistical bottom-up processing", *Proc. Interspeech 2008*, pp. 1980-1983
7. Joris Driesen, Hugo Van hamme "Improving the Multigram Algorithm by using Lattices as Input", *Proc. Interspeech 2008*, pp. 2086-2089
8. Hugo Van hamme "HAC-models: a Novel Approach to Continuous Speech Recognition", *Proc. Interspeech 2008*, pp. 2554-2557
9. Joost van Doremalen, Lou Boves "Spoken Digit Recognition using a Hierarchical Temporal Memory", *Proc. Interspeech 2008*, pp. 2566-2569
10. Louis ten Bosch, Hugo Van hamme, Lou Boves "A computational model of language acquisition: focus on word discovery", *Proc. Interspeech 2008*, pp. 2570-2573

Louis ten Bosch, Hugo Van hamme , Lou Boves "Discovery of words: Towards a computational model of language acquisition", in: France Mihelič and Janez Žibert (Eds.) *Speech Recogition: Technologies and Applications*, Vienna: I-Tech Education and Publishing KG, pp. 205 - 224

Additional papers are under review for journals and conferences.

Theses

Joost van Doremalen "Hierarchical Temporal Memory Networks for Spoken Digit Recognition", Radboud University Nijmegen, Dept. of Language & Speech, December 2007.
Joost De Tollenaere."Zelflerende spraakherkenning: akoestische eenheden en woordmodelllen" MSc thesis, K.U.Leuven, ESAT, 2008, (in Dutch)

All publications can be accessed through the public website maintained by the project.

Presentations about the ACORNS project

Trondheim, February 4-5, Unto Laine, Nordic Speech Meeting: ACORNS and other active projects of TKK Speech Technology Team

Leuven, 25 February 2008, Louis ten Bosch, Advanced MSc Programme in Artificial Intelligence ("Language Engineering Applications"): ACORNS - Acquisition of Communication and Recognition Skills: About a computational model for language learning

Merelbeke, 26 February 2008, Louis ten Bosch, Nuance lecture: ACORNS - Acquisition of Communication and Recognition Skills: How do we learn words?

Londen, 29 April 2008, Louis ten Bosch, University Collage London Colloquium: ACORNS - Acquisition of Communication and Recognition Skills: Design of a computational model for language learning

Nancy, 3 October 2008, Lou Boves & Els den Os, INRIA Colloquium: Acquisition of Communication and Recognition Skills Based on a Memory-Prediction Framework

Task five is devoted to spreading awareness beyond the scientific community. As a preparation for the SAC meetings, it was decided to write a general overview of results accomplished so far for the general public. This overview paper will also be used as a basis for new press releases, and for the ACORNS public website.

Table 3.6.1: Deliverables List

Del. no.	Deliverable name	WP no.	Date due	Actual/ Forecast delivery date	Estimated indicative person-months *)	Used indicative person-months	Lead contractor
D6.2.1	First Project Workshop	6	M18	M12	1	1	RUN
D6.4	Published papers	6	M24	M24			RUN
D6.5	Public Awareness	6	M36	M36			RUN

- List of milestones, including due date and actual/foreseen achievement date

For the milestones, see the deliverables.

4 Consortium Management

- **Consortium management tasks**

The management tasks for the second year were somewhat harder than for the first year, since attention had to be paid to better harmonize and integrate the work of the workpackages. All partners were willing to work on better integration, but it took some effort to accomplish this. We decided to have two and a half days project meetings (instead of 1 and a half days), so that there would be more room for in-depth discussion on specific topics.
The meetings were organised as planned, the minutes of the meetings and audio conferences were always sent in time.

In order to guarantee and optimise the quality of the deliverables we have assigned two or three senior members of the consortium –who were not involved in producing a specific deliverable- to review the drafts of the texts.

- **Contractors**

All partners have enthusiastic and dedicated teams working on the project. Only KTH was struggling to get the right persons in time. The senior staff members still spend relatively much time to the project. The project meetings were very useful and constructive meetings and always clear appointments were made for the next period.

No changes in responsibilities were necessary.

- **Project timetable and status**

WP		Month Task	13	14	15	16	17	18	19	20	21	22	23	24
WP0	Project Management	T0.1		Q		Q			Q			Q		D0.2.2 D0.3.2
		T0.2												D0.4.2
WP1	Signal Represent-ations	T1.1												D1.2
		T1.2						M1.3						
WP2	Signal Patterning	T2.1						M2.1.2						M2.1.3
		T2.2			M2.1.1B									M2.1.4 D2.2
WP3	Memory Organization and Access	T3.1												D3.2
		T3.2						M3.2						
		T3.3						M3.3.1						
		T3.4												M3.4
		T3.5						M3.5						
WP4	Information Discovery & Integration	T4.1												D4.1
		T4.2												M4.1.2
		T4.3												M4.3.1
WP5	Integration and Communic-ation	T5.1												
		T5.2										M5.5		D5.4.2
		T5.3												
		T5.4						M5.4						
WP6	Dissemination and Standardization	T6.1												
		T6.2												
		T6.3												
		T6.4												D6.4.2
		T6.5												

Updated Gantt Chart for Year 2

Overall, the work has proceeded according to the original plans, as updated at the end of the first year.

M2.2.2A, whose completion time was extended originally from M15 to M21, was not reached yet since only preliminary tests have been performed on real data originating from speech.

Milestone M5.4 (specification of the Year-2 experiments) was somewhat late, because it took more time to conclude the research and discussions on the use of visual/semantic features.

- **Co-ordination activities**

Four project meetings took place during the second year. It turned out that one and a half days meetings were no longer enough to discuss the work and to work together. Therefore an additional meeting was planned in Leuven on May 15. Representatives of almost all partners were present to prepare the updated memory and processing architecture, and to discuss the semantic features approach.

From the June meeting onwards, the meetings were planned for 2 and a half days. The following meetings took place:
1. Preparation of the Review meeting in Leuven, Belgium   23 January 2008
2. Project meeting in Berg en Dal, Netherlands   13-14 March 2008
3. Project meeting in Castleton, UK   16-18 June 2008
4. Project meeting in Espoo, Finland   3-5 September 2008

The SAC meeting with prof. Anne Cutler, prof. Paula Fikkert, and prof. Walter Daelemans took place on November 12th 2008. (The meeting with the other SAC meetings took place on 15 December 2008, at the very beginning of the third reporting period).

Audio conferences took place on:
1. 22 April 2008
2. 22 May 2008
3. 21 August 2008
4. 12 November 2008

In between the 'official' meetings and conference calls, a very lively exchange of e-mail traffic and data took place.

The PhD of Sheffield, Guy Aimetti, visited Leuven for a couple of days to work together with the Leuven people on prosody within the NMF-approach.

Annex 1: Plan for dissemination and Use

1 Exploitable knowledge and its Use

Table 5.1.1 Overview table of exploitable knowledge

Exploitable Knowledge (description)	**Exploitable product(s) or measure(s)**	**Sector(s)** of application	Timetable for commercial use	Patents or other IPR protection	**Owner** & Other Partner(s) involved
1. Procedure for blind bottom-up speech segmentation		1. Speech recognition 2. Industrial inspection; signature analysis	2010	patent application (FIN-20075696) filed	TKK
2. Software package for speech signal processing		1. speech recognition and speech coding	After 2010	n.a.	KTH
3. Structure detection by means of Non-Negative Matrix Factorisation		1. Speech recognition 2. Data mining	After 2010	n.a.	KU Leuven
4. Improved software implementation of the CSSR algorithm for Computational Mechanics Modelling		1. Data mining, 2. Structure discovery	2009	n.a.	KTH
5. Platform for conducting experiments with simulating language acquisition		Scientific research	2010	n.a.	RU and all partners

1. Procedure for blind bottom-up speech segmentation using Discrete Model Elements
   – Discrete Model Elements (DME) are a novel approach for finding local structure in continuously changing signals. Examples of such signals are speech, but also noise and vibration signals produced by machinery, natural systems, etc. The goal of bottom-up segmentation is to find points in time where the behaviour of the system generating the signals changes significantly, suggesting that the system is making a transition from one state to another.
   – Exploitation of the segmentation procedure will be pursued mainly by the originator, i.e., TKK. The other partners will assist TKK in contacting commercial companies.
   – Commercial exploitation will probably depend on finding commercial companies interested in developing the basic results obtained so far into an operational software module.
   – Actual deployment of the novel procedure will require additional research, among others to better understand the robustness of the procedure against additive and convolutional noise.
   – TKK, the originator of the novel procedure, has filed a patent application (FIN-20075696)

2. Software package for speech signal processing
–   The package contains tested software modules for conventional signal processing, primarily for use in the consortium, to guarantee that there are no differences between the results of processing identical input by different partners. The procedures can also be used outside the consortium.
–   Additional processing modules will be included to compute features that are especially salient from an auditory point of view
–   Features will be provided on millisecond, deci-second and centi-second time scales
    o   We see the major application of the software in scientific research, where there is a need for tested and verified procedures for basic speech signal processing routines.

3. Structure detection by means of Non-Negative Matrix Factorisation
–   Non-negative Matrix Factorisation (NMF) is a novel technique for discovering structure in matrices describing observations form physical processes, represented in terms of non-negative numbers (e.g. Energies, number of occurrences, etc.). We have developed NMF to detect structure in continuous speech, based on a representation that tracks the number of transitions between labels after vector quantisation.
–   The original NMF algorithms have been adapted to enable incremental decomposition, equivalent to incremental learning.
–   The work has been carried out mainly by KU Leuven.
    o   For the time being, we expect that the knowledge will mainly be used in the ACORNS project.

4. Causal State Splitting and Reconstruction algorithm for Computational Mechanics Modelling
–   New implementation of the algorithm, repairing small bugs in the publicly available implementation.
–   Extension of the algorithm to allow for approximate (instead of exact) causal states, to make the approach robust against natural variation in many kinds of data.

5. Platform for simulating language learning
–   The platform consists of a number of MATLAB scripts that implement the learner, the care giver and the interaction between learner and car giver.
–   Stimuli can be formatted with different amounts of information about the situational context in which speech utterances can be interpreted.
–   The actions of the care giver include the selection of new stimuli to offer to the learner, the interpretation of the learner's response and the decision on how to proceed.
–   The actions of the learner depend on the learning algorithm(s) selected by the experimenter.
–   The platform provides a choice of techniques for monitoring and interpreting the performance of the learner.

2 Dissemination of knowledge

Table 1.1 Overview table of past and future dissemination activities

Planned/actual Dates	Type	Type of audience	Countries addressed	Size of audience	Partner responsible /involved
15/01/2007	Press release	General public	Netherlands	16 Million	RU Nijmegen
26/11 – 28/11/2007	Workshop	Research	global	35	RU Nijmegen and USFD
11-9-2009	Workshop Interspeech-2009 Satellite event	Research	global	50	RU Nijmegen
Several dates	Publications; for details, see below	Scientific	global	15,000	all
01/02/2007	Project web-site	General Public, but mainly scientists	global	millions	RU Nijmegen
24/10/2008	Proposal for ESF Research Network	Scientific	Europe	thousands	RU Nijmegen

No specific dissemination activities were planned for the reporting period. However, a special session in the Interspeech conference in Brisbane was dominated by papers from the ACORNS project.

The project website (http://www.acorns-project.org) has been extended and maintained during the reporting period. The website is being kept up-to-date by the project coordinator.

The website will be maintained and updated for at least three years after the conclusion of the project. This will enable the project to provide access to publications that will see the light only after the formal end of the contract.

In October 2008 a proposal for an ESF Research Network "Language models of Evolution, Acquisition, and Processing" has been submitted, as a follow-up action to anchor the results of the research in ACORNS.

List of publications

Year-1

Papers

Lou Boves, Louis ten Bosch, Roger Moore "ACORNS -- towards computational modeling of communication and recognition skills", Proc. ICCI-2007.

Veronique Stouten, Kris Demuynck, Hugo Van hamme "Automatically Learning the Units of Speech by Non-negative Matrix Factorisation", Proc. Interspeech 2007.

Veronique Stouten, Kris Demuynck, Hugo Van hamme "Discovering Phone Patterns in Spoken Utterances by Non-Negative Matrix Factorization", IEEE Signal Processing Letters 2008

Louis ten Bosch, Bert Cranen "A computational model for unsupervised word discovery", Proc. Interspeech 2007.

Hugo Van hamme "Non-negative Matrix Factorization for Word Acquisition from Multimodal Information Including Speech", ESF Workshop, Leuven November 2007.

Theses:

Okko Räsänen "Speech Segmentation and Clustering Methods for a New Speech Recognition Architecture", MSc Thesis, Helsinki University of Technology, Espoo, November 5, 2007.

Alexander Bertrand "Zelflerende Spraakherkenning via Matrix-factorisatie", Katholieke Universiteit Leuven - Departement Elektrotechniek ESAT, 2007, [in Dutch].

Year-2

Papers:

Hugo Van hamme "Integration of Asynchronous Knowledge Sources in a Novel Speech Recognition Framework", ISCA ITRW, *Speech Analysis and Processing for Knowledge Discovery*

Louis ten Bosch, Hugo Van hamme , Lou Boves "Unsupervised detection of words – questioning the relevance of segmentation", ISCA ITRW, *Speech Analysis and Processing for Knowledge Discovery*

Louis ten Bosch, Lou Boves "Language acquisition: the emergence of words from multimodal input", in Sojka, P., Horák, A., Kopecek, I & Pala, K. (Eds.) *Text, Speech and Dialogue, 11th Intern. Conference, TSD 2008*, Brno, pp. 261-268

Klein, M., Frank, S., van Jaarsveld, H., ten Bosch, L.F.M., & Boves, L. "Unsupervised learning of conceptual representations - a computational neural model", *Proc. 14th Annual Conference on Architectures and Mechanisms for Language Processing (AMLaP)*, 4-6 September 2008, Cambridge, UK

Okko Räsänen, Altosaar, T. & Laine U.K. (2008) Comparison of prosodic features in Swedish and Finnish IDS/ADS speech. *Proc. of Nordic Prosody X.*

Okko Räsänen, Unto K. Laine, Toomas Altosaar "Computational language acquisition by statistical bottom-up processing", *Proc. Interspeech 2008*, pp. 1980-1983

Joris Driesen, Hugo Van hamme "Improving the Multigram Algorithm by using Lattices as Input", *Proc. Interspeech 2008*, pp. 2086-2089

Hugo Van hamme "HAC-models: a Novel Approach to Continuous Speech Recognition", *Proc. Interspeech 2008*, pp. 2554-2557

Joost van Doremalen, Lou Boves "Spoken Digit Recognition using a Hierarchical Temporal Memory", *Proc. Interspeech 2008*, pp. 2566-2569

Louis ten Bosch, Hugo Van hamme, Lou Boves "A computational model of language acquisition: focus on word discovery", *Proc. Interspeech 2008*, pp. 2570-2573

Book Chapter

Louis ten Bosch, Hugo Van hamme , Lou Boves "Discovery of words: Towards a computational model of language acquisition", in: France Mihelič and Janez Žibert (Eds.) *Speech Recogition: Technologies and Applications*, Vienna: I-Tech Education and Publishing KG, pp. 205 - 224

Theses

Joost van Doremalen "Hierarchical Temporal Memory Networks for Spoken Digit Recognition", Radboud University Nijmegen, Dept. of Language & Speech, December 2007.

Joost De Tollenaere."Zelflerende spraakherkenning: akoestische eenheden en woordmodelllen" MSc thesis, K.U.Leuven, ESAT, 2008, (in Dutch)

All publications can be accessed through the public website maintained by the project.

Presentations about the ACORNS project

Trondheim, February 4-5, Unto Laine, Nordic Speech Meeting: ACORNS and other active projects of TKK Speech Technology Team

Leuven, 25 February 2008, Louis ten Bosch, Advanced MSc Programme in Artificial Intelligence ("Language Engineering Applications"): ACORNS - Acquisition of Communication and Recognition Skills: About a computational model for language learning

Merelbeke, 26 February 2008, Louis ten Bosch, Nuance lecture: ACORNS - Acquisition of Communication and Recognition Skills: How do we learn words?

Londen, 29 April 2008, Louis ten Bosch, University Collage London Colloquium: ACORNS - Acquisition of Communication and Recognition Skills: Design of a computational model for language learning

Nancy, 3 October 2008, Lou Boves & Els den Os, INRIA Colloquium: Acquisition of Communication and Recognition Skills Based on a Memory-Prediction Framework

The PowerPoint files used in these presentations can be obtained from the project coordinators.

Planned presentations and publications

Louis ten Bosch, Hugo Van hamme , Lou Boves, Roger K. Moore "A computational model of language acquisition: the emergence of words", to be published in *Fundamenta Informaticae*

Guillaume Aimetti will have a presentation in the EACL Conference in 2009-01-07

Mark Elshaw and Roger Moore have submitted a paper to the Journal of Connection Science. Waiting for feedback.

In January 2009 Mark Elshaw, with other members of the ACORNS team who have contributed to the development of the Memory architecture and semantic features models, will submit a paper to a neural computation journal about work in WP3.

Mark Elshaw has been invited to give a talk at the Natural Computing Applications Forum in January 2009 (http://www.ncaf.org.uk/). He will present results on the recurrent SOM model, with contributions of Roger Moore and Michael Klein.

Mark Elshaw plans to produce two papers based on experiments in year-3 for computational neuroscience conferences in 2009.

Gustav Henter will submit a paper for the 25th Conference on Uncertainty in Artificial Intelligence (UAI 2009, URL http://www.cs.mcgill.ca/~uai2009/cfp.html). Arranged June 18-21, 2009 at McGill University, Montreal, Canada. The paper submission deadline is March 13.

Gustav Henter plans a journal publication about the extended CSSR algorithm (with additional proofs etc.); the target would be the Journal of Machine Learning Research

Gustav Henter considers the possibility of submitting a journal paper comparing CSSR with other related approaches. This would also be a publication within the framework of ACORNS, but possibly at a later time.

Joris Driesen and Hugo Van hamme will submit a journal paper about basics of NMF for word discovery; most likely to Computer, Speech and Language.

Joris Driesen, Hugo Van hamme, Louis ten Bosch will submit a journal paper about a cognitively motivated NMF-based learning.

Joris Driesen, Louis ten Bosch, Hugo Van hamme will submit a conference paper about incremental learning; possibilities include TSD, Interspeech and ICASSP.

Joris Driesen and Okko Räsänen plan to submit a conference paper regarding the segmental vs. fixed frame representations as a conference paper; possibilities include the ACL International Joint Conference on Natural Language Processing, TSD, or Interspeech.

Unto Laine, Okko Räsänen, Toomas Altosar will submit a paper on bottom-up speech segmentation; possible journals include JASA and Computer Speech and Language.

Okko Räsänen, Unto Laine, Toomas Altosar plan to submit a paper on self-learning vector quantization (SLVQ).

Unto Laine, Okko Räsänen, Toomas Altosar intend to submit a paper on pattern discovery using Concept Matrices.

Michael Klein, Louis ten Bosch, Boves will submit a paper entitled "Unsupervised Learning of Conceptual Representations" to the Cognitive Science Conference, Amsterdam, deadline 31 January 2009.

Michael plans to submit a paper entitled "Unsupervised Learning of Conceptual Representations" to the journal Connection Science; planned submission date is March 2009.

Michael Klein, in collaboration with Geoffrey Hinton, plans to submit a paper entitled "Learning Speech with Restricted Boltzmann Machines" to the journal Neural Computation or Cognitive Science; planned submission date is May 2009.

Kris Demuynck, Hugo Van hamme, Dirk Van Compernolle non-parametric densitities for speech recognition, to be submitted to Interspeech-2009 or TSD

Kris Demuynck, Hugo Van hamme intend to submit a paper entitled Phone discovery through spectral clustering to the journal Speech Communication

In addition, towards the end of the project we intend to submit overview papers to one or more journals such as IEEE Trans. on Autonomous Mental Development, Artificial Intelligence, Language Acquisition, Cognition, etc.

5.3 Publishable results

No public domain software was planned to be released in the reporting period.

We consider making available a novel implementation of the CSSR algorithm for Computational Mechanics Modelling that was developed in the reporting period available in Open Source format.