Towards a Memory-Prediction Model of Speech Processing

Dates

26 – 28 November 2007, Faculty Club, Leuven

Location

Nijmegen

Applicants

Lou Boves & Louis ten Bosch
Dept. of Language & Speech,
Radboud University Nijmegen
Erasmusplein 1
6525 HT Nijmegen
The Netherlands

Roger Moore
Speech and Hearing Research Group
Department of Computer Science
University of Sheffield
Regent Court, 211 Portobello Street,
Sheffield, S1 4DP, UK

Keywords

Memory-Prediction Model; Automatic Speech Recognition; Probabilistic Pattern Recognition; Language Acquisition

Abstract

Automatic and Human Speech Recognition are modelled as probabilistic pattern recognition processes, where the basic patterns are pre-defined. However, this has had only a limited success. Therefore, we propose to develop a novel approach, based on the memory-prediction model, in which words and sub-word units are emergent properties. The novel approach will explain language acquisition and human speech recognition, and it will enable automatic speech recognition with human-like performance.

The case for an exploratory workshop

Exploratory, innovative character

Theory and modelling in automatic and human speech processing has been dominated by the assumption that processing is mainly symbolic: existing theories assume that speech is represented as a linear string of sounds. However, there is an increasing amount of evidence that this 'beads-on-a-string' representation is not adequate, neither for modelling speech recognition, nor for speech production. In automatic speech processing the symbolic representation is closely related to probabilistic pattern processing: the ubiquitous and very substantial variation in the physical appearance of the basic units (the beads) is modelled in the form of mixtures of (Gaussian) densities. However, the speech technology community now agrees that the performance obtainable with this approach will fall short of the level that is needed for many important applications (Moore, 2003).

In human speech processing models tend to be mainly verbal descriptions that are not easily amenable to computational implementation. This situation obviously does not facilitate the development of a commonly agreed theory and experimental methodology. Also, the lack of formal models of human speech processing has hampered the uptake of knowledge from human speech processing in models and systems for automatic speech processing.

Against the background of the state of the art sketched in the previous paragraph there is a clear need for novel and innovative approaches to describing and modelling speech. Preferably, such an approach should be solidly based on emerging knowledge and understanding o the human brain and its attendant cognitive processes. Such an 'embodied' approach would enable constant cross-fertilisation between the wide range of disciplines that investigate speech processing.

Fortunately, there is a number of recent developments that hold the promise of a fully embodied theory of speech processing that is amenable to efficient computational implementation, and that will allow to approach most –if not all- of the open problems in automatic and human speech processing in a common framework. These developments all revolve around episodic (rather than strictly symbolic) models of memory. One approach that seems to be especially interesting for the speech processing communities is the Memory-Prediction Model, proposed by Hawkins (2004) The Memory-Prediction Model (MPM) provides a basis for modelling language acquisition and robust speech processing in a single framework. MPM considers words and speech sounds not as pre-existing units that must be learned, but rather as emergent properties of speech signals that can be detected by a learning agent thanks to the fact that similar signals occur repeatedly in specific multimodal contexts.

For the time being, MPM as a model for language acquisition and speech processing is attractive, but also mainly speculative: there is virtually no speech research carried out in that framework. In the workshop we intend to bring together leading scientists from all relevant disciplines for an in-depth discussion of all aspects of the MPM, and its potential application to speech processing.

Potential impact on new developments

It is now more than ten years since Bourlard, Hermansky and Morgan (1996) called for fundamentally new and innovative approaches to automatic speech recognition, but so far mainly further developments of the existing approach have surfaced. One such extension is the use of Dynamic Bayesian Networks that allow to build more detailed models than the conventional Hidden Markov Model approach. However, the DBN approach still considers speech processing as a problem in probabilistic pattern processing.

At the human speech recognition side models based on episodic memories have been proposed (Goldinger, 1998), but as with symbolic models of human language processing, no implementation of an episodic model exists that could be used to process physical speech signals. Sme attempts have been made to carry over the idea of episodic representations to automatic speech recognition, but so far these have not seen the success that was hoped for (de Wachter, et al., 2003).

Given the fact that all disciplines related to automatic and human speech processing are in need of a comprehensive new theory and model, and that the MPM can provide the foundation for such a theory and model, we are confident that a high level workshop devoted to the possibility

of developing an MPM approach to speech processing will have an enormous impact in the field. We expect that the research to be initiated at the workshop will result in novel approaches to automatic speech recognition that will reach performance levels well beyond the ceiling approximated by current approaches. A comprehensive MPM approach of speech processing will also have an impact on future models and systems for speech synthesis, were the foundation has been laid –be it without any references to a theoretical background- in approaches based on unit selection concatenation. For the cognitive sciences in general, and all language-related disciplines in particular- the research to be spurned by the workshop will lay the foundation for comprehensive embodied theories of human language acquisition and processing. Given the fact that language is without doubt the most complex cognitive skill, an effective model of language processing will advance the cognitive sciences substantially.
We expect that the workshop will have several types of follow-ups. First and foremost, we are convinced that approaches to speech and language processing related to MPM will develop in the next decade. Therefore, we expect to see additional multidisciplinary and discipline specific workshops, special issues of journals and books that will refer to the workshop proposed here. We also believe that the workshop proposed here will result in a number of collaborative project proposals, some of which will bring together researchers from around the globe.

Scientific background and rationale
As sketched in the previous sections, the initiative for the workshop derives from a widely felt need for developing new approaches to speech and language processing. In concrete terms, we think that the workshop will succeed in defining a roadmap for experimental research that will result in operational computational models that go beyond the probabilistic pattern recognition approach to speech processing. The MPM approach enables building of neural representation of things t be recognised on several parallel levels. At the highest level complex concepts (e.g., frequently used phrases) can be represented in a holistic form that can be accessed by activating only a small part of the representation in a new input signal. At lower levels the constituent parts of the concept (e.g., the individual words) can be represented, and on a yet lower level the constituents that make up the words (speech sounds or phonemes) can be represented. A speech signal to be recognised will activate all representations quasi-simultaneously, thus providing an explanation for the fact that sometimes we only need half of a word to understand it completely. Multiple parallel representations also enhance robustness against competing signals.
Furthermore, the MPM representations can serve both recognition and generation, thus providing the much wanted uniform framework for all speech processing. Last but not least, an approach that considers recognition units fundamentally as emergent properties provides theory of language acquisition cast in exactly the same terms as the theory of language processing.
The fact that MPM offers a principled basis for modelling learning and perception in general, an operational model for speech processing will most certainly have an impact on our understanding of perception (and action) in other sensory modalities.

Interdisciplinarity
A cognitively and biologically plausible memory-prediction model of speech processing cannot possibly be constructed without contributions from all disciplines that make up the field indicated as the cognitive sciences. To inform research on neural representations and memory we need contributions from Theoretical Neuroscience, Neurophysiology and Neurobiology. To obtain appropriate models of audio signal processing we need expertise in Auditory Theory and Auditory Physiology. The need for expertise on Language Structure and Language Acquisition, as well as on Automatic Speech (and Language) Processing and Pattern Recognition is obvious. At the same time we will need expertise in Human and Machine Learning, to understand how language acquisition can be modelled.
Although there are multiple bi- and tripartite relations between the disciplines mentioned above, there is no tradition of all these disciplines meeting together, and there is as yet no established platform where such a gathering might happen spontaneously.

Need for European collaboration

As explained above, the problem that we want to address is extremely complex, and it involves a large number of disciplines that do not have a habit of working together in collaborative projects. In principle it is possible that one of the larger European countries houses all expertise that is needed for developing an embodied computational model of speech processing. However, it is far less likely that all of the leading groups in that country are interested in and have room for embarking on a high reward but also high risk enterprise. For this reason alone it will be necessary to bring together groups from several European countries, to build a maximally effective and efficient team.

The need for a novel approach to modelling speech processing is felt not just in all European countries, but rather all over the globe. The impact on the field of a national workshop in one of the large European countries would certainly be much smaller than what can be accomplished by a workshop with representatives of several different countries.

Novel and innovative models of speech processing are a concern that is not limited to Europe. On the contrary, all disciplines that make up the field of cognitive sciences have a strong global character. Therefore, it would definitely be counterproductive if the workshop were not open to non-European scientists. What is more, some crucial expertise is much more readily available in the USA than in Europe. Therefore, we propose to invite a small number of key researchers from the USA to contribute to the workshop.

Benefits and outcome, including future activities

Preliminary discussions with prospective participants in the workshop have already indicated that there is a pressing need for one or more books that summarise the state of the art in speech processing and lay the foundation for developing novel and innovative approaches. Therefore, we intend to prepare and organise the workshop in such a manner that at least one edited volume will be published within a year after the workshop.

We intend to use the workshop, the contacts during the preparation phase and during the phase in which the final publication is prepared to form a consortium that will apply for an Integrated Project in the Cognitive Systems area in the upcoming Seventh Framework Programme. Our involvement in the preparations of the Workplan have convinced us that there will be room for a large project aimed at developing novel architectures for speech processing (and multimodal perception). We intend to build a consortium that comprises a small number of crucial partners from the USA.

We expect that the workshop proposed here will be the starting point for a series of workshops, some more general, others more dedicated, but all addressing the topic of computational modelling of speech and language processing. We have already started exploring the possibility of organising a follow-up workshop as a satellite event connected to Interspeech-2009.

References

Bourlard, H., Hermansky, H., Morgan, N., 1996. Towards increasing speech recognition error rates. Speech Communication, 18, 205-231.

De Wachter, M., Demuynck, K., Van Compernolle, D. and Wambacq, P. (2003) Data Driven Ex-ample Based Continuous Speech Recognition. In Proc. European Conference on Speech Communication and Technology, pages 1133--1136, Geneva, Switzerland, September 2003.

Goldinger, S.D. (1998). Echoes of echoes? An episodic theory of lexical access. Psychological Review 105: 251-279

Hawkins, J. (204) On Intelligence. New York: Times Books.

Moore, R.K., 2003. A comparison of the data requirements of automatic speech recognition systems and human listeners. Proceedings of Eurospeech, Geneva, Switzerland, pp. 2581-2584.

Preliminary workshop programme

Scientific topics

The following topics will be addressed in invited tutorial presentations:

- Fundamental and operational aspects of cortical structure
- Fundamentals of the Memory-Prediction Model of Intelligence
- Fundamentals of Neural Signal Processing and Perception
- Fundamentals of Automatic structure learning from continuous multimodal input signals
- Fundamentals of Probabilistic Pattern Recognition
- Fundamentals of Exemplar-based (episodic) memory and recognition
- Fundamentals of Language Acquisition
- Role of computer simulation in future scientific research

The character, flavour and level of the presentations must be suitable for an audience that combines experts and non-experts, both of whom should benefit from the presentations. Therefore, we will instruct the presenters to organise their presentations in such a manner that they have a tutorial character to make them digestible for non-specialists, while at the same time not eschewing advanced topics that uncover relations that most experts might not readily see.

The workshop will cover three full days. The presentations that must lay the foundation for an innovative research programme will be given on the first two days. We aim for an intensive programme, which allows covering four major topics per day, in two hour sessions, separated by coffee, lunch and tea breaks. The evenings of the first two days can then be devoted to more general discussions.
The third day will be devoted to integrating the topics and issues and to produce a roadmap for future research.

Structure of discussion

Each of the presentations will be followed by a regular questions and discussions period.
The general after-dinner discussions and the discussions on the third day will be structured by the workshop organisers, in close collaboration with the presenters. To that end the presenters will be required to send comprehensive preparatory material to the organisers and introductory preparatory materials to the prospective participants at least one month before the start of the workshop. This will allow the organisers to draft a preliminary version of the eventual research roadmap, which can then be adapted as a result of the discussions.
All participants are obliged to study the introductory material before the start of the workshop, so that the presenters can safely assume a certain level of knowledge.

We are well aware that the format for the workshop proposed here is extremely intensive. However, there I general agreement in the field that such an intensive workshop is long overdue, and that too much time and energy has been spent over the last decade to workshops that were actually mini-conferences.