Issue N° 8 / April 2008
Editor : François Foglia, francois.foglia@idiap.ch

DIRAC
Detection and Identification of Rare Audiovisual Cues

Newsletter

Contents

COVER STORY
DIRAC topic as a focal point for the research at the Summer Workshop 1

FOCUS
- DIRAC Summer Workshop 2
- Preliminary AWEAR set-up 3
- Article published in «New Scientist» 3

INSIDE DIRAC
- Publications 4

News

DIRAC Review Meeting

8-9, April 2008 Martigny

Cover Story

DIRAC topic as a focal point for the research at the Summer Workshop at the Johns Hopkins University

Each summer, in 2007 in its 13th year, the intensive 8 week research workshop with the widespread worldwide impact on the speech and language community is held at the Johns Hopkins University in Baltimore, Maryland. These workshops focus on important and promising topics in speech and language engineering and lead to cross-fertilization of ideas between research groups as well as offering the opportunity to seed a variety of research projects that are continued well after the workshop is over. The workshops topics and the researchers with the winning proposals then attempts to put together an appropriate research teams from the worldwide pool of the available colleagues who are willing to spend up to two intense months at JHU. This effort is assisted by several (typically two) pre-workshop meetings, devoted to hammering out the details of the upcoming summer effort.

It is a big success for the DIRAC consortium that for the 2007 Workshop, one of the two awarded research topics was based on the

make significant contributions to the pool of trained specialists in the fields of speech and natural language processing by providing training to students, allowing researchers to learn from each other, and educating all workshop participants and other interested colleagues through guest lectures, participant seminars, and team research updates. Proposals for research topics are presented by individual researchers in a two-day workshop to the distinguished panel of leading experts from the Industry, Government, and Academia. The panel subsequently chooses several proposal that originated from DIRAC. The proposal aimed at identification of unexpected (out-of-vocabulary and out-of-language) words in machine recognition of speech http://www.clsp.jhu.edu/ws2007/groups/rmimsr/. The research team, organized and led by the DIRAC General Coordinator Prof. Hynek Hermansky from IDIAP Research Institute, Martigny, Switzerland, consisted of researchers from Georgia Institute of Technology, University of Singapore, Microsoft Research in Redmont, Washington, Brno University of Technology, Czech Republic, Magdeburg University,

to be continued on page 2

www.diracproject.org
DIRAC c/o IDIAP Research Institute, Centre du Parc, Av. des Prés-Beudins 20, P.O. Box 592, CH-1920 Martigny
info@diracproject.org - www.diracproject.org

Detection and identification of Rare Audio visual Cues (DIRAC) is an Integrated project funded by the EC's 6th Framework Program, managed by IDIAP (CH).

DIRAC
Detection and Identification of Rare Audiovisual Cues

Issue N° 8 / April 2008

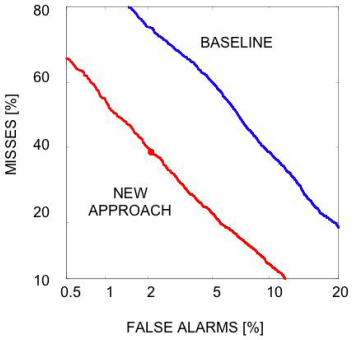
Editor : François Foglia, francois.foglia@idiap.ch

Newsletter

DIRAC topic as a focal point for the research at the Summer Workshop at the Johns Hopkins University *(continued from page 1)*

Germany (DIRAC Intern Mirko Hannemann), US Department of Defense, Trondheim University, Norway, Delhi University, India, University of Michigan, and Johns Hopkins University, Maryland. The group aimed at development of data-guided techniques that would yield unconstrained estimates of posterior probabilities of sub-word classes employed in the stochastic model solely from the acoustic evidence, i.e. without use of higher level language constraints. The goal was that these posterior probabilities then could be compared with the constrained estimates of posterior probabilities derived with the constraints implied by the underlying stochastic model in a state-of-the-art machine recognition system. Parts of the message where any significant mismatch between these two probability distributions would found could then be re-examined and corrective strategies applied. This would allow for development of systems that are able to indicate when they «do not know» and eventually may be able to «learn-as-you-go» in applications encountering new situations and new languages.

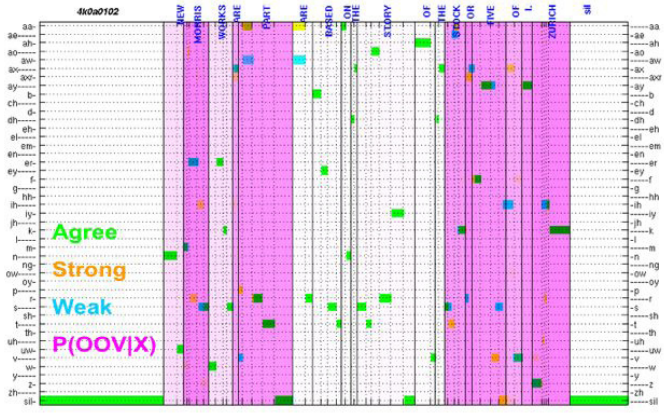
Prior to the workshop, researchers from Brno University of Technology modified large vocabulary continuous speech recognition system that was developed during their participation in another EC Integrated Projects AMI and AMIDA to yield the required posterior distributions and modified it to be able to also yield strings of phonemes without the use of higher-level language constraints, Microsoft Research provided their transducer-based decoder as well as their state-of-the art binary classifier, DoD made available their language-independent phoneme recognizer, Georgia Tech and University of Singapore brought in their work on phonetic attribute based phoneme recognition, and the whole group participated in preparation of the research data set that ended up to be the 5000 word Wall Street Journal (WSJ) database. During the workshop, all aspects of the system have been studied, with most effort devoted to 1) recognizing speech with a minimal use of the context, 2) comparing estimates from the “strongly-constrained” (i.e. both the acoustics ad the context constrained) and the “weakly-constrained” (mostly acoustics) recognition streams 3) the confidence measures, 4) phoneme recognition without use of any context. The data material consisted of WSJ data, down-sampled to 8 kHz, with about 20% of least frequent words left out from the lexicon, thus emulating the targeted unexpected words. In addition to the test data-set, a development set (used for a training of some of data-guided techniques) has been also created. A modified state-of-the-art LVCSR recognizer from the AMI Consortium has been used as the strongly-constrained recognizer. The weakly-constrained recognizer was the same LVCSR system modified for recognition of phonemes (rather than words). Both recognizers were trained on the independent telephone-quality data and not on the targeted WSJ data. Additionally, the strong constrains were also induced on the recognized phoneme string by a transducer-based system from Microsoft Research. A number of comparison techniques have been investigated in addition to many state-of-the-art confidence measures derived from both the strongly-constrained and the weakly-constrained recognition streams. In the final system, results from most of the investigated techniques have been fused using a state-of-the-art classifier from Microsoft Research. The progress has been evaluated by comparing the developed error-detection techniques to the state-of-the-art Cmax technique from the technical University Aachen. Results of the final comparison for the detected OOVs are shown in the Figure below.



Summary results for the detection of OOVs

More details of this effort can be found in
Christopher White, Geoffrey Zweig, Lukas Burget, Petr Schwarz, Hynek Hermansky, ‘CONFIDENCE ESTIMATION, OOV DETECTION AND LANGUAGE ID USING PHONE-TO-WORD TRANSDUCTION AND PHONE-LEVEL ALIGNMENTS’, Proceedings of IEEE Int. Conf. on Acoustics, Speech, and Signal Processing (ICASSP)», 2008

Lukas Burget, Petr Schwarz, Pavel Matejka, Mirko Hannemann, Ariya Rastrow, Christopher White, Sanjeev Khudanpur, Hynek Hermansky, Jan Cernocky, ‘COMBINATION OF STRONGLY AND WEAKLY CONSTRAINED RECOGNIZERS FOR RELIABLE DETECTION OF OOVS’, Proceedings of IEEE Int. Conf. on Acoustics, Speech, and Signal Processing (ICASSP)», 2008



True sentence (OOVs in red): Numerous works of art are based on story of Isaac.
Recognized sentence:New Morris works the part are based on the stock of five of I Zurich
Disagreeing posteriors from strongly-constrained recognizer in yellow, from weakly-constrained recognizer in blue and posteriors in agreement in both recognizers in green. Shades of magneta indicate estimated probability of OOV.

DIRAC and the speech recognition community

Recognizing the critical importance of being able to deal with unexpected items in machine information extraction, organizers of the bi-annual IEEE Workshop on Automatic Speech Recognition and Understanding (ASRU 2007) in Kyoto, Japan, decided that one of the panels should be devoted to the DIRAC topic of “Handling the unexpected acoustic data”. The lively discussion among the members of the panel and the audience of more that 200 leading researchers in machine recognition of speech and language from all parts of the world, touched on many problems associated with this issue. Even though no clear consensus was reached on how to handle this important problem, (slides of the panel presentations can be found at http://www.asru2007.org/agenda.php) it is obvious that the problem is prominently in the focus of the research community.

DIRAC c/o IDIAP Research Institute, Centre du Parc, Av. des Prés-beudins 20, P.O. Box 592, CH-1920 Martigny, info@diracproject.org - www.diracproject.org

DIRAC
Detection and Identification of Rare Audiovisual Cues

Issue N° 8 / April 2008

Editor : François Foglia, francois.foglia@idiap.ch

Newsletter

New scenes recorded using the preliminary AWEAR set-up

From 17 to 19 March 2008, the vision group from CTU (Prague) visited Oldenburg for the joint audio-visual recording of several new scenes using the preliminary AWEAR set-up. The recordings have taken place in a) an emptied seminar room and b) outside the university building on a pathway.

The outside recording stage, two 'passers-by' can be seen.

In the seminar room, several scenes contain dialogue between the AWEAR-user and a passer-by who seeks help from the AWEAR-user. The passer-by first tries his mother tongue and by request switches to english.

Other scenes contain an english dialogue between the AWEAR-user and somebody else, when all of a sudden a third person enters the room shouting «fire» and urgeing everyone to leave the room. In a second run, the same people interact, but this time the third person is looking for someone called «Meier» (in a far less hectic manner than with the fire, apparently). These scenes shall be used as benchmark scenes for the detection of rare events that should trigger an alarm («fire!»).

In the outside setting, the dialogue scenes have been recorded similarly to the seminar room recordings, although with significantly more visual and acoustic interference: Several uninvolved people pass the AWEAR-user, some chatting, some on their own, furthermore, the presence of sound sources outside the field of view yields more noise. Hence we expect these recordings to present a higher degree of difficulty than those recorded inside.

The scenes are suitable for all of the existing algorithms such as (visual) pedestrian detection, face detection in close-up shots, speech and language detection and (acoustic) direction of arrival estimation.

In between sessions, the preliminary results from the algorithms were inspected.

People involved were
(from CTU Prague): Michal Havlena, Akihiko Torii;
(from Oldenburg): Joern Anemueller, Joerg-Hendrik Bach, Hendrik Kayser, Thomas Rohdenburg

Article published in «New Scientist» issue 2640, 26 January 2008

Virtual reality to study the schizophrenic brain; VR simulations could be refined to diagnose schizophrenia more accurately

DOGS that moo, lawn mowers that sound like fax modems and red fluffy clouds in a clear blue sky are all components of a virtual reality (VR) game that could be used to explore the cognitive problems of people with schizophrenia.

At the Medicine Meets Virtual Reality conference in Long Beach, California, next week, computer scientist Daphna Weinshall of the Hebrew University of Jerusalem in Israel will describe a VR experiment in which volunteers wearing a head-mounted display navigated through simulated streets, shopping malls and a market. They were asked to flag up «incoherencies» such as objects that were the wrong colour or in the wrong position - a street sign at ground level, for instance - or making the wrong noise. While all the healthy volunteers spotted at least 87 per cent of the incoherencies, only six of the 43 volunteers with schizophrenia scored in this normal range.

As yet, the differences are not clear-cut enough for VR to replace established diagnostic methods, in which psychiatrists interview and observe patients to determine which of a checklist of symptoms they have. But the results suggest that VR simulations could be refined to produce diagnostic tests that provide a more detailed assessment of brain function, says Avi Peled, a psychiatrist at the Technion-Israel Institute of Technology in Haifa who co-developed the simulation.

Peled is also interested in using VR to study the details of what goes wrong in the brains of people with schizophrenia. For example, while it is already possible to probe cognitive ability using pen-and-paper tests, in VR more information can be presented using several senses simultaneously. «Why do we need this fancy technology? Because we can control vision and hearing in parallel,» says Peled.

Peter Yellowlees, a psychiatrist at the University of California at Davis who developed a simulation of schizophrenia symptoms in the virtual world Second Life, suggests imaging patients' brains while they play a VR game. That might reveal abnormalities in neural activity that underpin cognitive difficulties. «That is the winning combination,» says Peled. «It is surely the next step.»

Simulations could also allow doctors to test for cognitive problems that may affect treatment. Matthew Kurtz of Wesleyan University in Middletown, Connecticut, found that people with schizophrenia who were asked to take pills at a certain time in a VR apartment often took the wrong number at the wrong time (Schizophrenia Bulletin , DOI: 10.1093/schbul/sbl039).

BYLINE: Peter Aldhous

DIRAC
Detection and Identification of Rare Audiovisual Cues

Issue N° 8 / April 2008

Editor : François Foglia, francois.foglia@idiap.ch

Newsletter

DIRAC's Publications

(http://www.diracproject.org/publications/)

Journal papers

Chips in your head

F.W. Ohl and H. Scheich

Scientific American Mind, April/May, pp. 64-69

Damaged or diseased brains could soon get a boost from implanted prosthetics. This article summarizes recent experimental evidende how our understanding of neocortical dynamics during learning can support construction of a new type of interactive neuroprosthesis for sensory cortex.

The cognitive auditory cortex: Task-specificity of stimulus representations

H. Scheich, A. Brechmann, M. Brosch, E. Budinger and F.W. Ohl

Hear Res Volume 229, pp. 213-224

Auditory cortex (AC), like subcortical auditory nuclei, represents properties of auditory stimuli by spatiotemporal activation patterns across neurons. A tacit assumption of AC research has been that the multiplicity of functional maps in primary and secondary areas serves a refined continuation of subcortical stimulus processing, i.e. a parallel orderly analysis of distinct properties of a complex sound. This view, which was mainly derived from exposure to parametric sound variation, may not fully capture the essence of cortical processing. Neocortex, in spite of its parcellation into diverse sensory, motor, associative, and cognitive areas, exhibits a rather stereotyped local architecture.

The columnar arrangement of the neocortex and the quantitatively dominant connectivity with numerous other cortical areas are two of its key features. This suggests that cortex has a rather common function which lies beyond those usually leading to the distinction of functional areas. We propose that task-relatedness of the way, how any informtion can be represented in cortex, is one general consequence of the architecture and corticocortical connectivity. Specifically, this hypothesis predicts different spatiotemporal representations of auditory stimuli when concepts and strategies how these stimuli are analysed do change. We will describe, in an exemplary fashion, cortical patterns of local field potentials in gerbil, of unit spiking activity in monkey, and of fMRI signals in human AC during the execution of different tasks mainly in the realm of category formation of sounds. We demonstrate that the representations reflect context- and memory-related, conceptual and executional aspects of a task and that they can predict the behavioural outcome.

Conference papers

The distortion of reality perception in schizophrenia patients, as measured in Virtual Reality

A. Sorkin, D. Weinshall and A. Peled

Proc. 16th Annual Medicine Meets Virtual Reality Conference (MMVR), Long Beach CA

Virtual Reality is an interactive three-dimensional computer generated environment. Providing a complex and multi-modal environment, VR can be particularly useful for the study of complex cognitive functions and brain disorders. Here we used a VR world to measure the distortion in reality perception in schizophrenia patients. Methods: 43 schizophrenia patients and 29 healthy controls navigated in a VR environment and were asked to detect incoherencies, such as a cat barking or a tree with red leaves. Results: Whereas the healthy participants reliably detected incoherencies in the virtual experience, 88% of the patients failed in this task. The patients group had specific difficulty in the detection of audio-visual incoherencies; this was significantly correlated with the hallucinations score of the PANSS. Conclusions: By measuring the distortion in reality perception in schizophrenia patients, we demonstrated that Virtual Reality can serve as a powerful experimental tool to study complex cognitive processes.

Incremental Learning for Place Recognition in Dynamic Environments

J. Luo, A. Pronobis, B. Caputo and P. Jensfelt

IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS07)

This paper proposes a discriminative approach to template-based Vision-based place recognition is a desirable feature for an autonomous mobile system. In order to work in realistic scenarios visual recognition algorithms should be adaptive, i.e. should be able to learn from experience and adapt continuously to changes in the environment. This paper presents a discriminative incremental learning approach to place recognition. We use a recently introduced version of the incremental SVM, which allows to control the memory requirements as the system updates its internal representation. At the same time, it preserves the recognition performance of the batch algorithm. In order to assess the method, we acquired a database capturing the intrinsic variability of places over time. Extensive experiments show the power and the potential of the approach.

Robustness analysis of binaural hearing aid beamformer algorithms by means of objective perceptual quality measures

T.Rohdenburg and V. Hohmann and B. Kollmeier

Proc. IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA), pp. 315-318

In this contribution different microphone array-based noise reduction schemes for hearing aids are suggested and compared in terms of their performance, signal quality and robustness against model errors. The algorithms all have binaural output and are evaluated using objective perceptual quality measures. It has been shown earlier that these measures are able to predict subjective data that is relevant for the assessment of noise reduction algorithms. The quality measures showed clearly that fixed beamformers designed with head models were relatively robust against steering errors whereas for the adaptive beamformers tested in this study the robustness was limited and the benefit due to higher noise reduction depended on the noise scenario and the reliability of a direction of arrival estimation. Furthermore, binaural cue distortions introduced by the different binaural output strategies could be identified by the binaural speech intelligibility measure even in case monaural quality values were similar. Thus, this perceptual quality measure seems to be suitable to discover the benefit that the listener might have from the effect of spatial unmasking.

Direction of Arrival Estimation based on the Dual Delay Line Approach for Binaural Hearing Aid Microphone Arrays

S. Goetze and T. Rohdenburg and V. Hohmann and K.-D. Kammeyer and B. Kollmeier

Proc. Int. Symposium on Intelligent Signal Processing and Communication Systems (ISPACS)

Multi-channel beamformer algorithms are promising solutions for noise reduction in hearing aids as they exploit the spatial distribution of the interfering signals and therefore in general lead to less signal distortion than single channel algorithms. Beamformers need a priori information about the microphone array and the direction of arrival of the target speech source. For head-worn arrays it is usually assumed that the user physically steers the arrays' look direction toward the desired speech source. This may become unsatisfying for the hearing aid user for high directivity beamformers with a small main lobe and when the target signal source is moving. In this contribution an automatic steering (electronic control of the look direction) is applied based on the dual delay line approach after Liu et al. [1]. This approach is modified to be applicable for head-mounted hearing-aid arrays. We show that the original free-field approach does not work on a head-mounted array because of the inappropriate propagation model. If we apply the true HRTF or a spherical head propagation model, the estimate is reliable within ±8◦ degree mean estimation error for an input SNR of 10dB or higher. However, for lower SNR the method seems to be not robust enough.

DIRAC c/o IDIAP Research Institute, Centre du Parc, Av. des Prés-beudins 20, P.O. Box 592, CH-1920 Martigny,
info@diracproject.org - www.diracproject.org