

STEVIN Midterm Review

International Assessment Panel Evaluation

The International Assessment Panel (IAP) of the STEVIN Programme consists of the following members:

- Prof. dr. Hans Uszkoreit (DFKI - Germany)
- Prof. dr. Gábor Prózszéky (Morphologic - Hungary)
- Prof. dr. Roger Moore (Sheffield University - UK)
- Dr. Nicoletta Calzolari (ILC - Italy)
- Mr. Paul Heisterkamp (Daimler - Germany)
- Dr. Gilles Adda (LIMSI - France)
- Prof. dr. Anne Abeillé (Université Paris 7 - France)
- Dr. Stelios Piperidis (ILSP - Greece).

The members of the IAP were asked to assess the midterm results and impact of the STEVIN programme as a whole and that of the individual projects. They discussed the results of the STEVIN programme with the researchers at the STEVIN midterm evaluation workshop that was held in Rotterdam on June 27 2008 and received the following documents:

- the STEVIN midterm Internal Evaluation written by the programme committee
In this document the STEVIN programme committee provided their assessment of the first phase of the STEVIN programme, by formulating answers to the nine questions specified by the HLT Board.
- the STEVIN midterm Fact File
This document contains background information concerning the STEVIN objectives and priorities, the STEVIN organisation, and information concerning the call procedures and results of the STEVIN calls (including statistics) and information concerning the STEVIN IPR policy. Also a list of supporting activities, networking grants and project publications is included.
- the STEVIN Project Progress Reports - May 2008
The report contains the progress reports of the STEVIN projects funded as a result of the three open calls for strategic research and HLT resources (data and tools), three calls for tender for specific HLT resources, three calls for demonstration projects and one call for educational projects.

At the end of the workshop the IAP met to discuss and formulate their findings. Unfortunately, Roger Moore was not able to attend the meeting in Rotterdam and Paul Heisterkamp had to leave early. Alice Dijkstra (NWO) was present as an informant.

The conclusions of the review panel are summarised in this report.

The Hague, October 2008

Evelien Krikhaar krikhaar@nwo.nl
(programme officer of the STEVIN midterm review)

Management Summary of the STEVIN Midterm Review International Assessment Panel

The IAP congratulates the Dutch and Flemish HLT community with the accomplishments STEVIN has reached so far. STEVIN's well structured and stratified approach, with open calls and call for tender each focusing on different aspects and levels of HLT needs, is well chosen and can serve as an example for other language communities. The set up as a joint Flemish-Dutch programme allows the direct co-operation of Dutch and Flemish research groups with excellent international track records. The co-operation between language and speech projects has also been growing, but is not optimal yet. Surprisingly good dynamics were observed in the collaboration between industry and academia, especially in small industrial enterprises, certainly due to the effective supporting activities initiated by STEVIN.

The IAP emphasizes the importance of the availability and maintenance of the developed resources and tools. The use of one Dutch HLT portal via the Dutch HLT Agency (TST-centrale) is highly appreciated and strongly encouraged.

The IPR issues, causing some delays in the starting phase of STEVIN, are now well served by the installed IPR Committee. However, IPR issues still remain a complex matter and they deserve continuous attention in the next phase of the programme. In this midterm review report the committee formulates some specific advices on dealing with IPR for the different products and results of STEVIN.

All scientific priorities and objectives are very well-covered by the STEVINI projects, apart from multimodal / multimedia projects and some semantic aspects, which are scarcely represented. The IAP acknowledges that it is not possible anymore to address these themes to their full extent within the current STEVIN Programme, but strongly suggests to consider these as important items on the research agenda for the near future.

It is considered as very important that the STEVIN programme is known outside the Dutch language community. The first STEVIN results have become visible, also internationally at conferences like LREC and InterSpeech. Now that, in addition to building resources, also research results are becoming available, publications in top level conferences and journals should be feasible. The high quality of recent results, also presented at the Midterm Review STEVIN conference in June 2008, certainly justifies an ambitious publication target, and the IAP sees this as a central aim for STEVIN in the next few years.

The IAP compliments the Programme Committee for their considerate and effective operation concerning some complicated procedural issues that arose during the first phase of STEVIN. Serious effort has been exerted to deal with possible conflicts of interests which exist in the well networked Dutch and Flemish language and speech technology community. Where present, they were handled appropriately and with integrity. Some good attempts are made for monitoring and evaluating scientific quality of projects inside the programme, although this remains a difficult subject. Participation in (international) evaluation exercises is encouraged in this respect but also publications in highly rated peer-reviewed journals.

In a final conclusion, the IAP assesses the first phase of the STEVIN programme as very successful, with excellent work done so far, which looks very promising for the results in the second phase. Moreover, STEVIN clearly provides an outstanding service to the Dutch language. Although still a lot of work needs to be done to allow the continued use of the Dutch language in ICT, funders and participants should be proud of their accomplishments so far.

Question 1

1. Do you think the STEVIN programme is sufficiently reaching its main aims and objectives described in the programme text? That is:

- **Realise an effective digital language infrastructure for Dutch, based on the BATAVO priorities (i.e., 'BasisTaal&spraakVOorzieningen' = BLARK 'Basic LANGUAGE Resources Kit');**
- **Carry out strategic research in the field of language and speech technology, especially in areas of high demand for specific applications and technologies;**
- **Advance the creation of networks and the consolidation of language and speech technology activities, educate new experts, and promote discussion and transfer of knowledge.**

Many of the main aims and objectives of the STEVIN programme already have been reached at the moment. Most of the high-priority goals were or will be addressed by the granted proposals, so that all targets can reasonably be expected to be reached at the end of the programme. The presented STEVIN projects cover indeed all strata in the stratified chain approach, from resources via applied research to real-world demonstrators.

The major part of an effective digital language infrastructure for Dutch will indeed be realised, based on BATAVO / BLARK priorities, as were formulated in the 2002 study by the Dutch and Flemish government in order to serve the goal of strengthening the economical and cultural position of the Dutch language in the modern ICT-based society and which were seen as essential for the development of language and speech technology for both the research and the industrial community.

The focus for realising an effective digital language infrastructure has successfully been set on 3 aspects: a) the creation of the relevant language resources, b) the availability, accessibility and maintenance of these resources, and c) the handling of the intellectual property rights (IPR). The open calls and the specific tenders successfully yielded projects for the creation of a semantic lexicon, an automatic speech recognition system and written Dutch corpus, of which the first two already are in their final phases. The start of the written corpus project (SONAR) evidently suffered from some delays in granting, for which however the reasons are adequately explained in the Internal Evaluation Report.

Concerning strategic research, excellent results arise from the various project reports as well as the scientific (poster) presentations at the STEVIN conference in June 2008. The IAP emphasizes the importance of the availability and maintenance of the resources and tools. The use of one Dutch HLT portal via the Dutch HLT Agency (TST-centrale) is strongly encouraged, and also the use of ELRA should be considered as a medium for availability and knowledge transfer in a broader European context. Re-usability and validation of resources is also an important theme for the near future of the programme.

IPR issues are certainly well served by the set up of an IPR committee, but apparently not all problems are completely solved at the moment, so these issues still need attention. More comments on this topic are formulated in the section under Question 3.

As regards the stimulation of network creation and the transfer of knowledge, an effective series of supporting activities has been developed, from demonstration and educational projects to academic conferences and brokerage events. The IAP stresses the importance of stimulating (PhD) students interests in HLT research and promoting HLT also in the educational academic programmes. Considering the typical characteristics of the Dutch language community – in size not large but with an excellent scientific tradition and reputation, especially in HLT - a relevant comparison can be made, for example, to the Scandinavian countries who are also seriously investing and performing research in this area. STEVIN is to be regarded as very successful, with the structured and well-planned approach as its strongest asset.

Question 2

2. Does the selection of the research projects reflect the stratified chain approach of the STEVIN programme, i.e. address all four levels in an appropriate way?:

- The basic language resources, explicitly limited to the data and their production;
- HLT research and development, ultimately resulting in available HLT components;
- Application development (HLT embedding);
- The user (`stimulating demand`)

Are there any critical gaps in the programme that should still be addressed?

STEVIN's well-structured and stratified approach, with open calls and calls for tender each focusing on different aspects and levels of Dutch HLT needs, is well chosen and can serve as an example for other language communities.

All four levels of the stratified chain approach of the STEVIN programme have been covered in a well-balanced way. The focus in amount of funding has obviously been on the creation of resources but strategic research, applied research and technology demonstration projects aimed at the user are also well represented. Moreover, funding is roughly equally distributed over speech and language technology projects.

Percentage of STEVIN funding per STEVIN priority		
• Speech technology resources	21,5%	
• Language technology resources	29,0%	
% STEVIN funding for basic resources		50,5%
• Speech technology research	14,5%	
• Language technology research	9,1%	
% STEVIN funding for basic research		23,6%
• Speech technology application-oriented research	7,4%	
• Language technology application-oriented research	8,2%	
% HLT Application-oriented research		15,6%
Speech technology demonstration projects	3,7%	
Language technology demonstration projects	6,6%	
% HLT Demonstration projects		10,3%
% STEVIN funding for speech technology		47,1%
% STEVIN funding for language technology		52,9%

The development of basic language resources has been successfully addressed, with five out of the six corpora projects already finished or in their final stages. Also the realisation of HLT strategic research and development is well underway, with two out of four projects already finished. The final call has resulted in a set of promising application-oriented research projects. The development of user-oriented applications is addressed by the demonstrator projects, which are discussed in more detail in the section under Question 5.

It is hard to detect some gaps in the programme that should still be addressed. Apart from the already in the Internal Evaluation mentioned absence of multimodal projects, the bases are very well covered by all projects. However, semantic discourse annotation as well as semantic web aspects are scarcely represented, which is rather unexpected given the rich tradition in good formal semantic research in The Netherlands. As regards the research on multimedia, the IAP acknowledges that this priority cannot be realised anymore within the STEVIN Programme, but they strongly recommend to consider it as an important part of the research agenda for the near future.

3. Is the (scientific) output of the STEVIN projects adequate in terms of quality and quantity? Which are the most outstanding results? Are the results made available in a non-discriminative way? Do you think that IPR policy is adequately dealt with in the STEVIN programme?

The IAP considers validation and evaluation within the STEVIN programme as very important aspects, but acknowledges that they are not easy to implement, since the required expertise usually comes from researchers who are themselves involved in the STEVIN projects. The IAP respects the serious attempts by the PC to install evaluation tools relying on circumstantial evidence, such as the number of publications in respectable peer-reviewed journals and the (re-) use of results (and especially resources and tools) by other projects. In addition, the IAP strongly encourages the participation of STEVIN in (international) evaluation exercises.

The most excellent results are clearly summarised by the PC in the Internal Evaluation Report. The high quality of recent results certainly justifies an ambitious publication target in high level conferences and journals, and the IAP sees this as a central aim for STEVIN in the next few years.

In the opinion of the IAP, all HLT corpora and tools and other parts of the Dutch BLARK should be made available via and actively distributed by the HLT portal of the Dutch HLT Archive. HLT research results should be actively shared in the research community via international peer-reviewed publications.

The IAP observes that the present construction where the Dutch Language Union, being a non-academic institution, claims all property rights to STEVIN results is different compared to the handling of IPR in other (inter)national funding schemes. They appreciate the premise that if all IPR is owned (or licensed) by the Dutch Language Union the (re)use of HLT results is enabled as widely as possible. However the IAP also feels that the present set-up may lead to unnecessary legal costs for the programme and administrative overhead for the researchers.

In the opinion of the IAP, the IPR should best be dealt with according to the following rules:

- IPR is transferred to the Dutch Language Union in case of foreground knowledge of HLT resources (especially corpora and other data collections) and to arrange user licenses for NTU for background knowledge (both for academic and commercial use) for these resources.
- Open Source arrangements are preferred for HLT tools as these will most often be constantly under revision. However, because for tools also background knowledge may be involved, this may not always be easy to realise.
- The IPR to STEVIN research *results* should be handled according to standard EU / FP7 regulations.

4. One aim of the programme is to stimulate collaboration between academia and industry in Flanders and The Netherlands. Is there enough evidence indicating that this aim is being achieved?

The realisation of collaboration between academia and industry is often difficult, due to the rather different - and in some cases even conflicting - aims and objectives. In this perspective, excellent co-operation between academia and industry has been observed. The IAP was surprised by the good dynamics, especially in small industrial enterprises, with short paths and close connections between the partners. Also the collaboration between Flanders and The Netherlands is very successfully developed. However, co-operation between language technology and speech technology still needs to be encouraged further.

The difficult situation around IPR issues in the research projects is probably one of the reasons why larger industrial enterprises did not want to collaborate in STEVIN projects and only small industrial partners are involved. It could be worthwhile to explore more opportunities in this respect in a larger, European context.

The statistics with respect to the distribution of STEVIN R&D funding over Dutch-Flemish recipients and the distribution of STEVIN R&D funding over academic and industrial recipients given in the table below are considered in line with the objectives of the STEVIN programme.

<i>Distribution STEVIN R&D funding over Dutch-Flemish recipients</i>			
Netherlands			
Universities	k€ 4.995		
Industry	k€ 1.245		
Total Netherlands		k€ 6.240	64%
Flanders			
Universities	k€ 3.133		
Industry	k€ 450		
Total Flanders		k€ 3.583	36%
Total STEVIN R&D funding		k€ 9.823	
<i>Distribution STEVIN R&D funding over academic and industrial recipients</i>			
Universities	k€ 8.128		83%
Industry	k€ 1.695		17%
Total STEVIN R&D funding		k€ 9.823	

5. An integral part of the STEVIN programme is its supporting programme which specifically aims at organising and/or financially supporting activities that stimulate network relations between the different types of actors that are involved in developing, implementing or embedding HLT technologies and encourage knowledge transfer between these actors. How do you judge the success of this part of the programme so far?

The IAP feels that the supporting programme of STEVIN is a very important and useful part of the STEVIN programme, which is considered to be very successful. Excellent work has been done in encouraging network relations between the complete spectrum of actors involved in developing, implementing and embedding HLT resources and technologies. This has been achieved by exploiting several activities, such as calls for joint strategic research projects and application-oriented projects involving industry and universities. In addition, STEVIN organised a series of major events, for instance brokerage events and the Dutch HLT platform event "Taal in Bedrijf" ("Language@Work"), that attracted a large number of researchers, developers, and users. STEVIN also supported various networking events, from scientific conferences (like CLIN and InterSpeech 2007) to more application-oriented meetings (like ICT-Delta).

Furthermore, STEVIN launched a call for proposals for demonstration projects where proven state-of-the-art language and speech technology was combined with existing applications and services, which should stimulate the demand for Dutch language and speech technology. It resulted in the realisation of 14 demonstrator projects, that enlarged the visibility of Dutch language and speech technology to, for example, governmental organisations, companies, educational organisations, researchers and the general public. The recent call for educational projects, aiming at informing senior high school children (age 17+) and bachelor students about HLT, can possibly also pay a good service to the promotion and incorporation of this topic in the academic curricula.

6. Are the STEVIN programme and its results visible to the (inter)national field of HLT research, the interested industry, and the interested public? Please provide some examples.

The first STEVIN results have become visible internationally, especially at LREC and some interesting ACL workshops, but in a rather modest way to the opinion of the IAP. Considering that the first phase of STEVIN was quite heavily focused on resources and data building, that is a logical choice. However, now that also research results are becoming available, the IAP strongly encourages the 'bold' publication of STEVIN's achievements, and that researchers aim at presentations on top level conferences and submit articles to top level peer-reviewed journals. The posters that were shown at the STEVIN Midterm conference show great potential for such high level submissions. Furthermore, it is important that the STEVIN programme is widely known, also outside the Dutch language community. Next to the external Dutch website of STEVIN, there should also be a project website in English that can be accessed by researchers from all over the world. The IAP suggests to present STEVIN and its projects at LANGTECH meetings and to compose an overview article of STEVIN's main aims and results and send it for instance to Springer Verlag, who is very interested in publications representing the cream of the "sonderforschungsbereich".

7. What is the (scientific) impact of the STEVIN programme with respect to Dutch and Flemish HLT?

Excellent stimulation of Dutch-Flemish co-operation and collaboration between academia and (mostly small) industry has resulted in a promising set of joint research and demonstration projects. With regard to the impact on larger industrial enterprises in the field of human language technology, it might be interesting to see how many of the STEVIN project partners link to KP6 and KP7 projects (with a more heavy industrial participation).

The impact of STEVIN in the academia, for instance in a rising number of PhDs, cannot be determined yet, as it is too early to measure this kind of impact at this moment. Certainly, the planned final evaluation (or "eindmeting") will provide a better perspective on this when such facts, as number of PhDs and other academic and educational impact figures, will be collected and compared to the situation at the start of the programme. The IAP considers the "nulmeting", although carried out rather late in the first phase of the programme, as a good tool for proper benchmarking and assessment of the scientific impact of STEVIN in the Final Evaluation, which will take place within two years.

8. Is the organisational structure of the programme, namely the International Assessment Panel, HLT Board, programme committee, and programme bureau appropriate? Does the programme offer sufficient guarantees on transparent, impartial and objective proposal evaluation, project monitoring and decision taking in general? Are potential conflicts of interest adequately dealt with on all levels of the programme?

The organisational structure of the programme is assessed as appropriate to serve the intended goals. The IAP appreciates the detailed and honest self evaluation formulated by the PC in the Internal Evaluation Report, concerning issues of transparency, evaluation, monitoring and decision taking. Conflicts of interest – which are inevitable in any relatively small research community – have been avoided and where present, they were handled appropriately and with integrity. The decision making structure that has been set up seems to be successful in dealing with this situation. The IAP appreciates it that their quality assessments and rankings were always accepted.

9. Were the installed procedures for granting, launching and monitoring the projects and for transferring and approving the project results adequate?

The IAP considers the procedures for granting, launching and monitoring the projects as appropriate. They are aware of the fact that a large part of this procedural system still had to be developed and refined in the first phase of the STEVIN programme, due to the rather new set-up and constellation of the programme.

The IAP wants to compliment the Programme Committee for their good analysis of problems and their considerate, effective and adaptive operation concerning some complicated procedural issues that arose during the first phase of STEVIN. In an international perspective, the delays in the start up of the projects, mentioned by the PC, are still considered relatively small.