The DISC Project

Niels Ole Bernsen and Laila Dybkjær
The Maersk Mc-Kinney Moller Institute for Production Technology
Odense University, Campusvej 55, 5230 Odense M, Denmark
emails: nob@mip.ou.dk, laila@mip.ou.dk
phone: (+45) 65 57 35 44 fax: (+45) 66 15 76 97

Based on a recent presentation at the SALT Workshop (Bernsen and Dybkjær 1997), this paper briefly presents the aims and assumptions of DISC, the Esprit Long-Term Research Concerted Action No. 24823 "Spoken Language Dialogue Systems and Components. Best practice in development and evaluation" which started on 1 June 1997. The DISC partners are: The Maersk Mc-Kinney Moller Institute for Production Technology (MIP), Odense University, Denmark (coordinator); Human-Machine Communication Department, Centre National de la Recherche Scientifique (CNRS-LIMSI), France; Institut für Maschinelle Sprachverarbeitung (IMS), Universität Stuttgart, Germany; Department of Speech, Music and Hearing, Kungliga Tekniska Högskolan (KTH), Sweden; Vocalis Ltd, UK; Daimler-Benz, Germany; Stichting Elsnet, The Netherlands.

The need for best practice in development and evaluation

No current scheme specialises software engineering best practice to the particular purposes of dialogue engineering, that is, to the development and evaluation of spoken language dialogue systems (SLDSs). The goals of dialogue engineering include optimisation of the user-friendliness of SLDSs which will ultimately determine their rank among emerging input/output technologies. DISC aims to develop a first detailed and integrated set of development and evaluation methods and procedures (guidelines, checklists, heuristics) for dialogue engineering best practice as well as a range of support concepts and software tools. The methodology produced by DISC will contribute towards establishing dialogue engineering as a sub-discipline of software engineering.

At this time there are no accepted standards or even widely understood benchmarks for assuring potential customers or users of SLDSs of the quality of systems. Neither are there any reliable methods for comparing the quality of two SLDSs before selecting one for deployment in the field. In an increasingly competitive marketplace, the ability to state that some system has been developed following a carefully designed and validated dialogue engineering methodology, along with the ability to report evaluation results in a standardised framework, is likely to give products developed in this way a competitive advantage. That in turn may stimulate take-up of the methodology by other organisations.

SLDS technology is taking off on a broad scale. For the year 2000, current estimates are that the global annual market for speech recognition alone will be $8 billion. According to the Ovum report on Voice Processing published last year, the global voice processing market in 1996 was $2.1 billion, and was expected to grow to $2.9 billion in 1997 and $3.75 billion in 1998. Even if, on a conservative estimate, only 1% could be described as SLDSs, that is still a very large number. The bulk of this business is taking place in the US but the opportunity to take a large and increasing share remains open to Europe.

Current commercial SLDSs are able to carry out routine tasks that were previously done by humans, thus generating significant savings in the companies or institutions that install them. Dur-

ing the last few years, interactive speech technology has begun significant deployment in real world applications in large vertical markets such as banking, finance and market research (Blyth and Piper 1994) as well as in telecommunications. An upcoming domain for advanced SLDSs is that of train information. Perhaps the most advanced SLDS in commercial use has been developed by Philips and is used by Swiss Rail. The system is based on the Philips Automatic Train Timetable Information System of which a demonstrator has been publicly available since February 1994, in Germany, on tel. +49 241 604020 (Aust et al. 1995, Aust and Oerder 1995). Similar train information systems are underway in the Netherlands, France and Italy. More advanced and flexible, large vocabulary SLDSs and systems integrating speech into multimodal systems are on their way from research laboratories to industrial exploitation and will have commercial significance by the end of DISC.

Publicly funded research has provided the major driving force for the technology advances exemplified by these systems. In the US, this has been coordinated by DARPA (latterly ARPA) through its competitive evaluations in large vocabulary speech recognition (Resource Management task) and spoken language understanding (ATIS task) (DARPA 1992; ARPA 1994, Young 1997). There has been a clearer focus on the special issues associated with spoken language dialogue in Europe than in the USA. Projects such as SUNDIAL (Peckham 1993, Peckham and Fraser 1994, Fraser and Thornton 1995, Peckham and Fraser forthcoming), the Danish project on Spoken Dialogue Systems (Dybkjær et al. 1995, Bernsen et al. forthcoming), MAIS, RAILTEL (Lamel et al. 1995) and VerbMobil (Wahlster 1993) have established a strong base of expertise in Europe in SLDSs. Results from several of those projects are currently being taken a step further towards commercialisation in the LE ARISE project.

Despite unquestionable progress, particularly in those parts of the SLDSs components field which have been delivering commercial applications for more than a decade, the design, development and evaluation of usable SLDSs is today as much of an art and a craft as it is an exact science with established standards and procedures of good engineering practice. The route from initial idea through analysis, requirements specification, design-and-evaluation cycles, prototype development, in-house and field testing to the final product and its evaluation is replete with unknowns and development steps that are undersupported in terms of procedures, concepts, theory, methods and software tools. Given the proven potential of SLDSs technologies, there is a need to take a significant step forward by creating a best practice methodology for SLDSs development and evaluation and start developing the concepts, methods and software tools required to integrate SLDSs development into mainstream software engineering. DISC aims to make central contributions to an SLDSs development and evaluation best practice methodology including novel concepts, methods and software tools.

DISC objectives, approach and envisioned results

As a long-term research Concerted Action aimed at making innovations responding to industrial needs, DISC aims to extend the state-of-the-art in dialogue engineering in four ways:

(1) by *generalising* current knowledge through combining state-of-the-art expertise to analyse a broad range of current SLDSs and components development and evaluation practices, thereby creating a detailed overview of current practice;

(2) by *maturing* promising novel concepts, methods and software tools existing in preliminary versions at partner sites, and bringing them to the industrial transfer stage, when possible;

(3) by *testing* on industrial and research cases, to the extent possible within the duration of the Action, a comprehensive scheme of dialogue engineering best practice; and

(4) by *systematising* results into a detailed, procedural dialogue engineering best practice methodology which takes a balanced view of competing approaches and technologies within current best practice, where such exist. The methodology should enable the user to specify the required behaviour (functionality, performance, ergonomics) and determine to what extent the system, its components and their interaction meet the stated requirements. Input from industry will be integrated in the methodology.

DISC will achieve its stated goals through three central work packages addressing current practice, best practice, and novel concepts, guidelines and software tools, respectively. Each of these work packages will focus on a set of aspects of SLDSs including speech recognition, speech generation, language understanding and generation, dialogue management, human factors and systems integration.

To ensure common approaches to each of the main results-building activities, and to ensure cross-aspect compatible results, each approach will have to include a set of agreed *evaluation criteria*. Thus, (a) the common approach for mapping out current practice includes criteria for the evaluation of current practice; (b) the common approach for testing best practice methods and procedures on industrial cases includes criteria for evaluating the transferability success of these methods or procedures; and (c) the common approach for iterative development and testing of novel concepts, methods and software tools includes criteria for deciding the feasibility of specific development and testing projects as well as for evaluating transferability success. The approaches and criteria (a)-(c) will form a basis of the quality assurance that can be issued with DISC-developed methods, procedures, concepts, and software tools.

All partners contribute to the Action access to products and running prototypes and their components as well as to prototypes under development. DISC will take advantage of existing practices, theories and tools, including results of the US ARPA exercise in comparative SLDS evaluation, as well as emerging results in the fields of de facto standards and guidelines for speech products, natural language components and evaluation from LRE EAGLES and experience from national initiatives in component evaluation methodologies, such as the German Morpholympics and the French GRACE project and other evaluation actions of the AUPELF group.

The envisioned industrial benefits of DISC will be:

- Progress towards the integration of SLDSs best practice into software engineering.
- Improved feasibility assurance of development projects (risk minimisation) and more exact feasibility assessment.
- Improved procedures, methods, concepts and software tools.
- Reduced development costs and time, improved maintenance and reusability.
- Improved product quality and increased flexibility and adaptability.
- Progress towards the establishment of dialogue engineering standards.
- Improved guarantees to end-users that a product has been developed following best software and cognitive engineering practice. Enabling end-users to objectively assess different systems and components technologies against one another and choose the right product according to quality, price and purpose.

The industries involved in DISC share the view that, for the emerging European SLDSs suppliers and end-user sector, a best practice dialogue engineering methodology consisting of detailed procedures and methods, concepts and software tools for development and evaluation will constitute an obvious competitive parameter.

References

ARPA. *Proceedings of the Speech and Natural Language Workshop.* San Mateo, CA: Morgan Kaufmann, 1994.

Aust, H., Oerder, M., Seide, F. and Steinbiss, V.: The Philips Automatic Train Timetable Information System. *Speech Communication* 17, 249-262, 1995.

Aust, H. and Oerder M.: Dialogue Control in Automatic Inquiry Systems. *Proceedings of the ESCA Workshop on Spoken Dialogue Systems,* 121-124, Aalborg, Denmark, 1995. Also in: *Proceedings of TWLT9,* 45-49, Enschede, The Netherlands.

Bernsen, N.O. and Dybkjær, L.: The DISC Concerted Action. In R. Gaizauskas (Ed.): *Proceedings of the SALT Club Workshop on Evaluation in Speech and Language Technology,* Sheffield, June 1997, 35-42.

Bernsen, N.O., Dybkjær, H. and Dybkjær, L.: *Designing Interactive Speech Systems. From First Ideas to User Testing.* To be published by Springer Verlag, 1997.

Blyth, B. and Piper, H.: Speech recognition: a new dimension in survey research. *Journal of the Market Research Society* 36(3), 1994, 183-203.

DARPA. *Proceedings of the Speech and Natural Language Workshop.* San Mateo, CA: Morgan Kaufmann, 1992.

Dybkjær, H., Dybkjær, L. and Bernsen, N.O.: Design, formalisation and evaluation of spoken language dialogue. *Proceedings of the TWLT9 Workshop*, Enschede, 1995, 67-82.

Fraser, N.M. and Thornton, J.H.S.: VOCALIST: a robust, portable spoken language dialogue system for telephone applications. *Eurospeech'95*, Madrid, 1995, 1947-50.

Lamel, L., Bennacef, S., Bonneau-Maynard, H., Rosset, S. and Gauvain, J.L.: Recent developments in spoken language systems for information retrieval, *Proceedings of the ESCA Workshop on Spoken Dialogue Systems*, Vigsø, Denmark, 1995, 17-20.

Peckham, J.: A new generation of spoken dialogue systems: results and lessons from the SUNDIAL project. *Eurospeech'93*, Berlin, 1993, 33-40.

Peckham, J. and Fraser, N.M.: Spoken language dialogue over the telephone. In H. Niemann, R. de Mori and G. Hanrieder (Eds.): *Progress and Prospects of Speech Research and Technology*. Sankt Augustin: Infix, 1994, 192-203.

Peckham, J. and Fraser, N.M.: *Speech Understanding and Dialogue*. Cambridge, MA: MIT Press, (forthcoming).

Wahlster, W.: Verbmobil - Translation of Face to Face Dialogues. Machine Translation Summary IV, Kobe, 1993.

Young, S.: Speech recognition evaluation: A review of the ARPA CSR Programme. In R. Gaizauskas (Ed.): *Proceedings of the SALT Club Workshop on Evaluation in Speech and Language Technology,* Sheffield, June 1997, 197-205.