Technology News

The Speechdat Project for Spoken Language Databases

The Commission of the European Community (CEC) is funding a collaborative programme of work to produce databases in various languages to meet defined specifications. Marconi Communications (formerly GPT) is one of the main partners†. Also involved in the project, as a subcontractor, is Marconi Research Centre (MRC).

The languages covered within Speechdat II are the main European languages, which include British English*.

Aims of the Project

The project was initiated to fulfil the demand for speech resources (that is, databases) to assist in the development of speech recognition technologies and speaker verification. Many European companies are active in the field of voice-driven teleservices and delivering the necessary speech technology. The aim is to produce speech databases recorded over the telephone network that meet the same requirements specification.

† Other partners include Siemens, British Telecom, Centro Studi E Laboratori Telecomunicazioni, Institut Dalle Molle d'Intelligence Artificielle Perceptive, Knowledge, Lernout & Hauspie, Matra, Philips, Portugal Telecom, Speech Processing Expertise Centre (SPEX), Vocalis (the contractors). Also involved in the project, as associates or subcontractors are Aalborg University, University of Munich, University of Maribor, Swiss Telecom PTT, University of Patras, Tampere University of Technology, Kungl Tekniska Hogskolan, INESC, Universitat Politecnica de Catalunya and Telenor.

* Other languages are Belgian French, Danish, Dutch, Finnish, Finnish Swedish, Flemish, French, German, Greek, Italian, Luxembourgish French, Luxembourgish German, Norwegian, Portuguese, Slovenian, Spanish, Swedish, Swiss French, Swiss German, and Welsh.

Such databases are needed to allow the development of multilingual voice servers by European industries, with potential for export.

The British English databases include a Fixed Network Database (FDB) of 4000 speakers and a Speaker Validation Database (SDB) of 120 speakers × 20 calls each, recorded by GMRC, and a Mobile Network Database (MDB) of 2000 speakers recorded by BT. FDBs are being created for all languages stated above. The SDBs and MDBs are being created for selected languages only. The aim of the MDB is to improve channel adaptation and noise reduction techniques, whilst the SDB is used in speaker verification algorithms.

Currently, many automated teleservices rely on isolated word recognition – that is, words can be spoken in isolation only after a system prompt. Speech recognition systems are being developed that can perform **continuous** speech recognition (that is, no pause is required between words). When keywords are uttered within a phrase the recognizer will be able to extract the relevant command and discard the rest. The overall aim is to make speech recognizers more user-friendly, so that, for example, users will be able to interrupt recorded prompts at any time and use natural, spontaneous speech when interacting with automated systems.

Speech Collection

The two complementary databases were recorded by MRC using the same recording platform. Speakers for both were recruited by a Market Research company in accordance with specified requirements. The databases were validated and are accompanied by design documentation, stating any deviations from the specifications.

In order to elicit utterances that conform to the requirements specification, speakers were provided with prompt-sheets and the recording platform gave a verbal prompt for each item – that is:

'please say item 2 after the tone'.

Forty-six items were recorded for each speaker for the FDB (of which 40 were mandatory) and 22 (of which 21 were mandatory) for the SDB. The FDB contains spontaneous utterances that were prompted by a verbal question such as:

'do you think it will rain tomorrow?'

The mandatory items included:

- phonetically-rich sentences,
- phonetically-rich words,
- keywords (that is, those used in association with teleservices),
- numbers,
- dates,
- times,
- names,
- money amounts, and
- spelled words.

All calls had to have at least 95% of the defined items recorded correctly.

For each speaker, information such as age, sex, and accent region was provided and, for the SDB, additional details such as health, tiredness and stress levels were noted. The label files that accompany the speech files contain these data. Tables of call data are also provided to allow easy access to the type of call required – for example: female, aged around 37, from the London area.

To enable users to access the recorded information, each item was annotated to provide a written interpretation of what was actually said. At the same time, non-speech (for example, background noise) symbols were added. This information is also stored in the label file. A lexicon is produced based on the contents of this field. All uttered words are included in the lexicon and a phonemic transcription is given. Where there are different pronunciations for the same word – for example, read (r e d / r I: d) – the alternative versions are given. However, the variations caused by accent region have not been reproduced as phonemic alternatives. These phonemic entries can be used to identify test data for speech recognition testing of a particular phoneme, as required.

Other Related Projects

The Speechdat project work continues with Speechdat (CAR) which is collecting data in nine languages recorded from GSM◆ phones in motor vehicles. This project is also assessing voice control for in-car equipment (for example, radio). In addition, Speechdat (E) is collecting data from Eastern European nations, whilst SALA is a South American project collecting data to conform to Speechdat specifications.

Finally, another European collaborative project, COST 249, is reviewing continuous speech recognition over the telephone. This project is intended to obtain the best performance from speech recognizers, and the collaboration enables exchange of useful information as the testing is performed.

◆ Global Systems for Mobile Communications

Availability of the Databases

The Speechdat II databases are available through ELRA (European Language Resources Association), and sales of resources for the first quarter of 1998 were up six-fold against the equivalent period for 1997. All databases are distributed in identical format on CD. The British English FDB is contained on 20 CDs and the SDB on 8 CDs§.

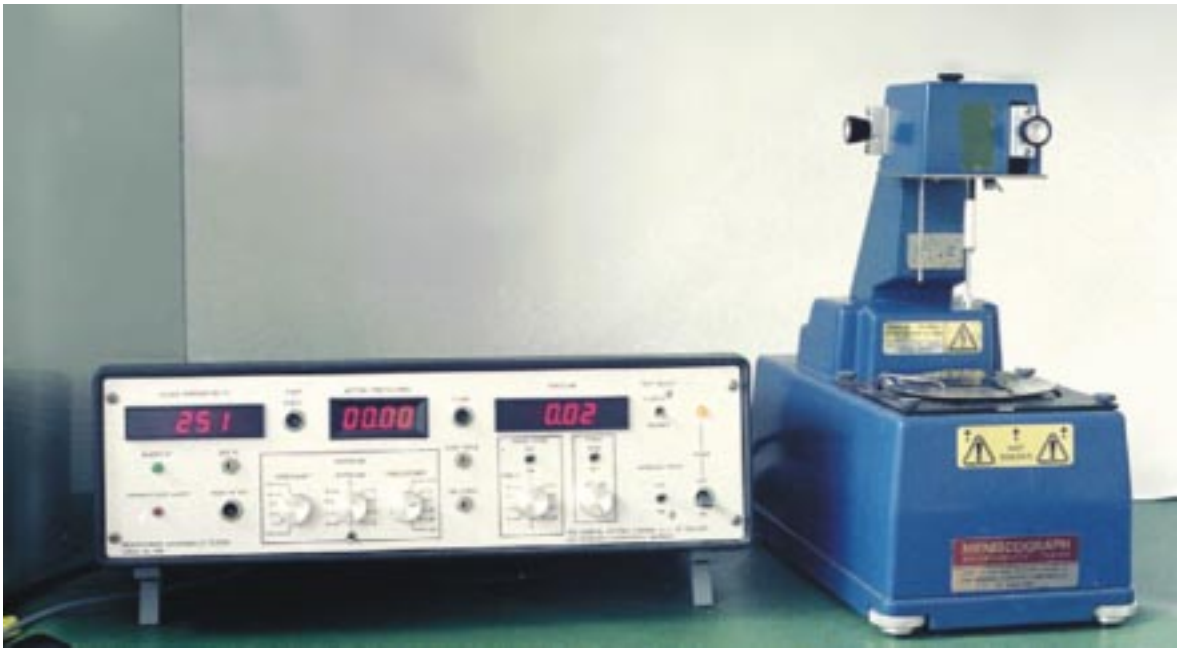
§ For further information on the database and speech recognition please contact Sharon Wheatley (sharon.wheatley@gecm.com) or Steve Ascham (steve.ascham@gecm.com), Marconi Research Centre, Marconi Technology Centres, Elstree Way, Borehamwood, Hertfordshire, WD6 1RX. For further information on the availability of the British English FDB and SDB databases, please contact Graham Fenner (Graham.Fenner@marconicomms.com), Marconi Communications Limited, Discovery Court, 551-553 Wallisdown Road, Poole, Dorset, BH12 5AG.

The GEC Meniscograph

The GEC Meniscograph is able to assess qualitatively and quantitatively the wetting of liquids on various substrates and provides a flexible development or quality process control tool, measuring the critical parameters of the wetting process. The equipment uses the wetting balance technique and records changes in the surface tension with time.

The GEC Meniscograph features are:

- A wide range of immersion speeds, depths, and times. Temperature control from ambient to 450°C.
- Sample contact detection by high-frequency electrical signal, or by sensing the upthrust caused by contact with the liquid.
- Windows software for data handling, and exporting to all popular spreadsheets. Facility to set the pass/fail criterion in the data handling software.
- Provides a complete picture of the way in which any liquid or substrate wets under selected conditions.



1 *The GEC Meniscograph*

- Suitable for use in product development, production, and goods inspection.
- The Meniscograph can be used to test the wettability of solders, adhesives (liquids, dispersions and hot melts), and surfactant systems.

For further information please contact:
Pete Hawkins, Chief Chemist,
Marconi Materials Technology Ltd.,
Marconi Technology Centres,
Caswell, Towcester, Northants, NN12 8EQ.
Telephone: +44 1327 350581
Fax: +44 1327 356775
Internet: www.gmmt.co.uk

VERA: an Experiment to Derive a Cost-effective Verification Process

The rising cost of testing is driving industry to revise its processes and techniques in order to meet the complexity of future systems in an economical way. Software testing is key to gaining confidence about the correctness of the code, but testing techniques tend to be labour-intensive and hence expensive. The aim of the VERA (Verification, Evaluation, Review and Analysis) experiment was to investigate the balance between the use of review and testing techniques to provide cost-effective error detection for real-time software systems.

VERA was a process improvement experiment (PIE) funded by ESSI (European Software System Initiative). The experiment started in 1997 and has lasted for eighteen months. It has been conducted by the Radar Systems Division (RSD) of what is now Alenia Marconi Systems (at Chelmsford) and the Marconi Research Centre, Great Baddow. Other verification and validation (V&V) practitioners within Marconi Electronic Systems have contributed to VERA via the three interactive workshops that have been held during the experiment. The final results from the VERA experiment were presented at EuroSTAR'98 in Munich, Germany, December 1998 and at the Marconi Software Engineering Technology Group (MSET) seminar, 'Improving the V&V Process', on the 8th of December 1998.

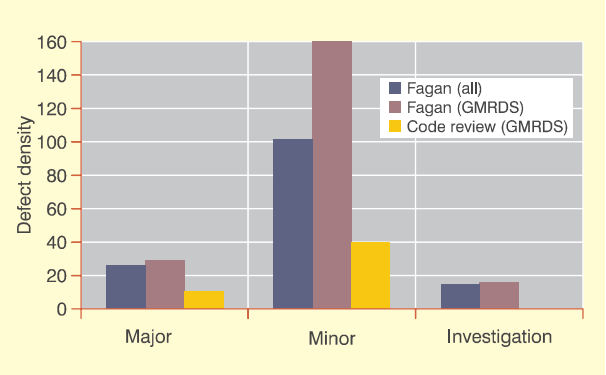
The project involved the analysis of existing metrics for code reviews and Fagan inspections, experiments with automated review techniques supported by Rational's† Ada Analyzer and experiments with automated testing techniques supported by Rational's TestMate™. The tool experiments were performed on a self-contained module of a real project within RSD and the results compared to the actual code review performed on that software. The techniques were assessed in terms of efficiency – to determine the effort involved in detecting defects, and effectiveness – to determine how many defects the technique detected.

† Rational Software Corporation, 18880 Homestead Road, Cupertino, CA 95014, USA. Phone: 00-1-408-863-9900. Fax: 00-1-408-863-4120. Email: info@rational.com. Web: www.rational.com/

Comparison between Code Review and Fagan Inspection

RSD took the decision in 1996 to adopt Fagan inspection as an alternative to the review process. The VERA experiment was able to analyse the metrics collected during Fagan inspections in order to evaluate the impact of inspections on the V&V process.

It was found that, on average, Fagan inspections find two-and-a-half times more major defects than code reviews (fig. 2). It is believed that this improvement can be attributed to the distinct roles for each inspector, the constraints to limit the rate of review, and the size of the document to be examined.



2 ***The average defect density per technique***

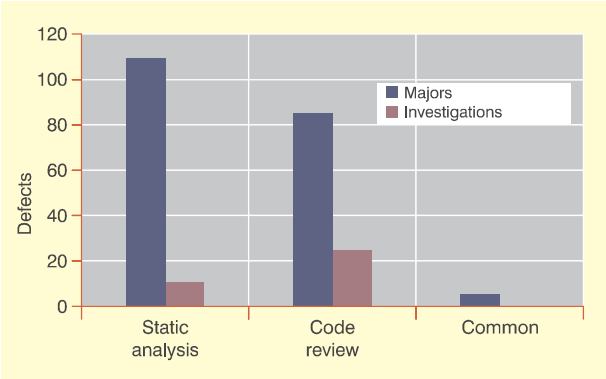
Finding defects early in the life cycle is paramount. Although the effort required to perform Fagan inspections is greater than the effort required to perform code reviews, the significant reduction in the predicted effort required for the integration test phase means that Fagan inspection is far more cost-effective overall.

The impact of individual engineers on any process is also extremely important. The VERA experiment results suggest that the more experienced engineers are in the language and application domain, the more defects are found during a Fagan inspection. This result contradicts the generally-accepted view that any competent engineer can review effectively.

Comparison between Static Analysis and Code Review

Static analysis is the activity of verifying software without executing it. Tool support for static analysis can (amongst other things) perform checks to detect coding violations and control-flow anomalies automatically (fig. 3).

With no prior knowledge in the application, the VERA experimenters completed the static analysis 34% faster than the reviewers reviewed the same code and detected 26% more major defects.



3 ***The total number of defects found per technique***

However, a significant number of defects were found, not by means of the tool, but by the experimenter whilst analysing the code. Code review and static analysis appear to be complementary, because only 5% of the defects found were common to both techniques. This is a major finding of the VERA experiment and is believed to be because the two techniques target different types of defects.

Automatic Test Case Generation

TestMate™ is still under development and various problems were encountered with the tool. In particular, the tool did not support certain data types (such as arrays and records of complex data types). Consequently it was only possible to use TestMate™ on approximately 7% of the units within the baseline project software. Although four major defects were detected, these required 120 hours of effort.

It is anticipated that many of the limitations encountered during the VERA experiment will be fixed in later releases of the tool.

Conclusions

The lack of a theoretical framework to determine objectively how much testing is required is problematic for the testing of complex software systems. The results of the VERA experiment provide guidance on which techniques are more cost-effective. From a technical point-of-view, the main lessons learned, are that:

- ***Fagan inspections are more effective than code reviews.*** Fagan inspections were shown to detect two-and-a-half times more major defects than code reviews.
- ***static analysis is effective at finding major defects and finds different types of defects to code reviews.*** Static analysis found 26% more major defects than code review of the same code. Only 5% of the defects were found by both processes.

- *automatic test case generation is currently difficult for the type of code used in the baseline project*. The versions of the TestMate™ tool used within the VERA experiment do not support certain Ada constructs. This meant that it was possible to test only about 7% of the units of the baseline project.

From a business point-of-view, the results of the VERA experiment show that:

- *the most cost-effective V&V process for software is static analysis coupled with Fagan inspections*. Performing static analysis and Fagan inspection on a typical piece of code is likely to result in the detection of three times as many defects as performing a code review. Although the effort required to perform static analysis and Fagan inspection is greater than the effort required to perform code reviews, the reduction in the predicted effort required for integration testing means that it is far more cost-effective overall.
- *the skills of the people performing a code review are vitally important to the effectiveness of the review*. The skills of the team members are less important during a Fagan inspection but do still have an impact.
- *static analysis, using 'Ada Analyzer', can be performed by competent engineers who have little or no domain knowledge*. This is an important finding because there is often a shortage of people with domain knowledge.
- *a formal process only produces consistent results if the rules of the process are followed*. Similarly, automating a process produces repeatable results only if use of the tool is controlled (for example by a code of practice).

Acknowledgments

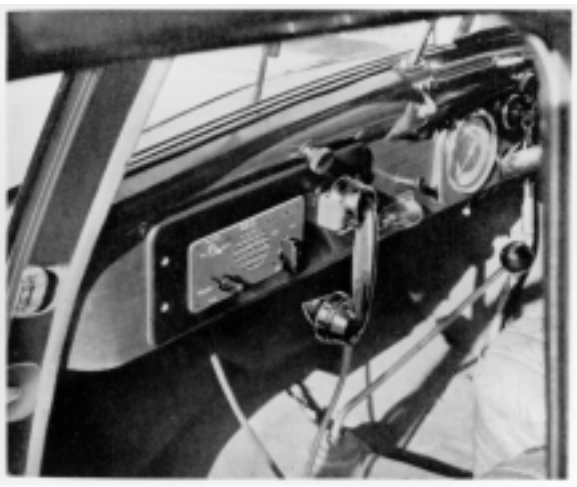
The VERA team would like to thank ESSI for the support to perform the project, Andy Warman (Alenia Marconi Systems, Farlington) who provided the Fagan Inspection Metrics Database and the participants of the MSET workshops for their inputs to the project.

For further information please contact:

Bruno Hémeury,
Marconi Research Centre,
Marconi Technology Centres,
Great Baddow, Chelmsford, Essex, CM2 8HN.
Telephone: +44 (0) 1245 242232,
email: bruno.hemeury@gecm.com

50 Years Ago

Typical mobile installation showing the transmitter/receiver in the luggage compartment of a car



Mobile controller installed on the dashboard of a car

(From 'Applications of V.H.F. radio' by E.W. Northrop, GEC Journal, XVI, 4, pp. 184-196, 1949)