UNDERGRADUATE EDUCATION FOR SPEECH TECHNOLOGY: AN INTRODUCTORY COURSE FOR ELECTRICAL ENGINEERS AND LINGUISTS

R.W. King and J.P. Vonwiller

Speech Technology Research Group
Department of Electrical Engineering
The University of Sydney

ABSTRACT - This paper describes a novel elective course in speech and language processing offered to undergraduate students from the Departments of Electrical Engineering and Linguistics. The course is presented by staff from the two Departments. The course content focusses on the components of speech and language technology systems and also on the importance of integrating them with high level language processing. The paper discusses the teaching methods employed for the course and provides brief comments on some of the educational issues which have arisen in presenting a course for students from the two, traditionally separate, cultures of Arts and Engineering.

INTRODUCTION

The multidisciplinary nature of speech technology requires the expertise of linguists, scientists and engineers who understand the nature of and the interfaces between the other contributing disciplines. Most current practitioners have gained their cross-disciplinary expertise in the course of carrying out speech technology research and development work - a rather informal process of 'on-the-job' education. One might suppose that more formal educational opportunities in speech and language processing and technology would lead to greater interest in the field, and thereby better sustain and increase research and application developments.

Of course, conference events with their tutorial days provide excellent opportunities for practitioners to formalise and broaden their knowledge and skills. Postgraduate courses, such that as offered in Computer Speech and Language Processing at Cambridge University (Univ. of Cambridge, 92) provide comprehensive coverage of the field for potential speech technology researchers. Such courses can be sustained only by major research groups and in very major centres. In the Australian setting it is more realistic to consider methods of introducing the principles of the discipline at the senior undergraduate level. We may note with some satisfaction that in engineering departments final year thesis projects in speech technology are very popular. Students taking these projects may be at a little disadvantage compared with their colleagues working in more mainstream areas, as they come to grips with the multidisciplinarity of speech and language specialisms. While research and project activities flourish, formal undergraduate courses in the area are rare. Were they to exist, they might well kindle more senior students' interests in the area, and this will enhance their capabilities in conducting project work, as well as introduce them to an intrinsically interesting area of science and technology.

In the past two years, the Department of Electrical Engineering and the Department of Linguistics at Sydney University have combined to present a course entitled 'Speech and Language Processing' at third year undergraduate level. The course is an optional one for students of both Departments. The multidisciplinary focus reflects a feature of our own research and development work, namely that major advances in the performance and scope of engineered applications will come about through greater understanding (and quantitative modelling) of all levels of speech and language. From a linguistics perspective, modern computational and signal processing techniques provide linguists with new analytical and synthetic tools on which to test hypotheses and models of language.

The paper describes the objectives, content and methodology of the course, which includes a project activity in which students work in inter-disciplinary pairs. The conclusion is a discussion of the educational issues which we have encountered in developing and running a course for students from the two traditionally separate cultural traditions of Arts and Engineering.

COURSE AIMS

Overall, the course aims to examine at an introductory level, those aspects of speech signal processing, computing and linguistic processing, which contribute to present and future speech technology systems. Figure 1, a schematic of the processes and resources used in an advanced speech response system, indicates (in an oversimplified way) how the processes are supported by a multi-level database of speech knowledge. This knowledge is realised from models of speech and language, derived from linguistic studies.

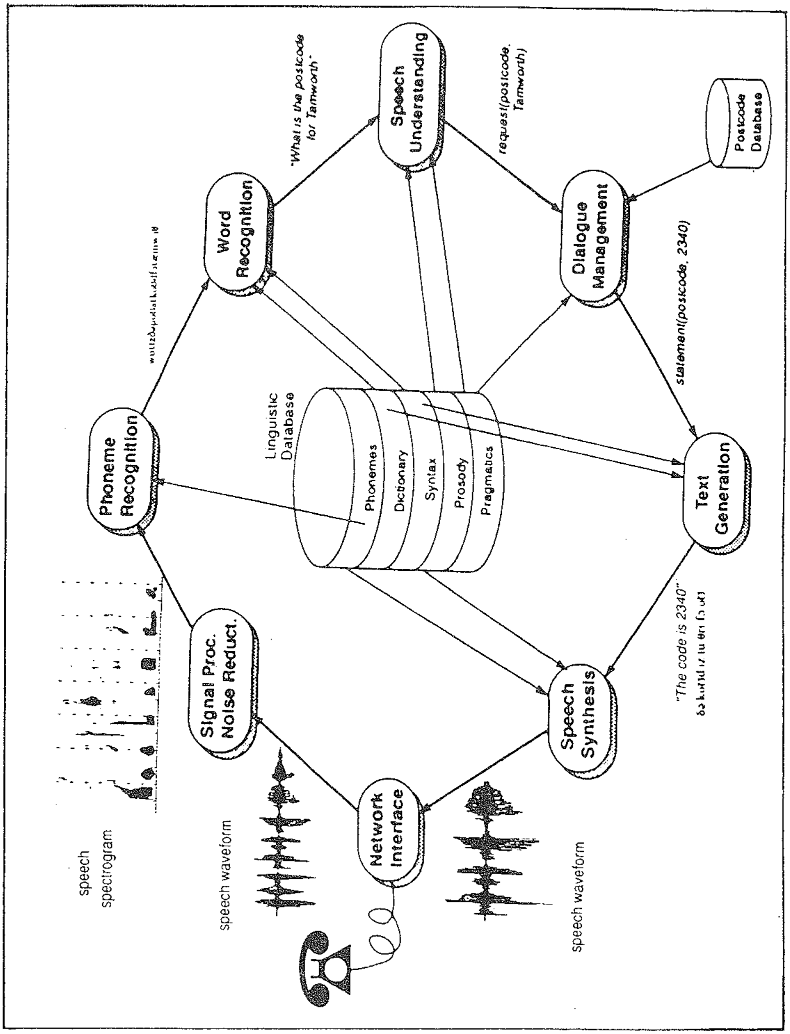

Figure 1 The processes and resources in an automated speech response system

More specifically, the course aims to provide students with:

- *knowledge* of the signal processing and computing technologies that can be applied in speech and language systems, and to enhance linguistic science;

- *knowledge* of how the phonetic, phonological, and lexico-grammatical descriptions of speech and language can be applied in technological systems, to enhance the performance of speech technology applications;

- *skills* in inter-disciplinary work, particularly in the conduct of a short project;

- *attitudes*, in which the values and methodologies of the two disciplines of Arts and Engineering are better understood and appreciated.

COURSE PROGRAMME

The course has two lectures and one tutorial per week for one 13-week semester. By this stage, the Electrical Engineers have a good grounding in signal theory and computing (and many of our double-degree students will have completed a B.Sc. in Computer Science). The Linguistics students have extensive understanding of lexico-grammar, semantics and some articulatory phonetics.

As outlined below, our course design aims to *compensate* for the two groups' basic deficiencies - of which the specialist terminologies and focus of the two disciplines are perhaps the most important - by teaching the two groups separately for the first three weeks. To take *advantage* of the different skills of the two groups, the students carry out practical/computing projects in small groups. Ideally we would have equal numbers of students from each discipline, achieved in the first year we ran the course with 26 in the class.

Within the very limited time constraints of the course, and the enormous scope of possibilities for theoretical and applied course content, we decided to focus on fundamental signal processing, linguistic concepts and computing methodologies for contemporary techniques in:

- automatic speech recognition
- speech synthesis (text-to-speech)
- speech (natural language) understanding
- text generation

These technological elements form key elements of the speech response system shown in Figure 1. In some ways the course could do no more than *introduce* and *demonstrate* principles and techniques. However, our access to up-to-date and good (and several interactive) demonstrations in the computing laboratory in Linguistics and the computing and speech signal processing facilities in the Speech Technology Research Group, allow students to gain insights relatively rapidly. The effectiveness of the more "hands-on" demonstrations, which took place during tutorials was due in considerable measure to the assistance of staff and research students from the two Departments. Brief descriptions of the lecture content, as it has evolved over the two years are given below.

Introductory material (Weeks 1 - 3)

The first time we ran the course we divided the two groups of students. Six lectures to the Linguists aimed to introduce the key engineering concepts of design as well as some basic engineering language. The only mathematical operations required are arithmetic, logarithmic and trigonometric. The concepts of speech signal representations and their practical realisation were presented though examples drawn from explanations of everyday experience of audio and telephone systems. Digital matters are introduced though coding and computers. The accompanying tutorial/demonstrations drew heavily on spectrum analysis of speech. The elements of computer programming with pseudocode, PASCAL and PROLOG are introduced through linguistically oriented examples.

The parallel introductory lectures for the engineering students introduced the ideas of phonetics, the set of sounds which make up a given language. The most important topic is to explain the articulatory phenomena which make up the sounds, and their realization in acoustic pressure, which in turn, can be analysed (signal processed) into spectral components. This introduction also included a discussion of the phonetic notation and phonological issues.

This divided approach left each group of students feeling they had missed out on something, so this year we reduced the introductory material somewhat, and provided more reading material. Another change was that we used the first tutorial to demonstrate six examples of speech technology systems, some from our research, others from commercial organizations. The students were required to summarize the main features and limitations of each demonstration, in order to give more focus to the material which followed.

To support the acoustic phonetics material, we required the students to record their vowels in the 'h - d' frame and plot them on an F1 - F2 vowel graph. The subsequent tutorial used these data to promote further analysis and discussion based on acoustic, phonetic and phonological knowledge, including, where possible, the influence of different accents.

Speech synthesis and recognition (Weeks 4 - 6)

The introductory text by Holmes (1988) is of a suitable level for much of this material. Emphasis is placed on the importance of simple modelling of the acoustic phenomena by separating the vocal source from the vocal tract model. The principles of speech synthesis are explained and demonstrated, using several electronic filter-based formant models of the vocal tract whose parameters can be adjusted sufficiently frequently to produce speech-like sounds. The material also included the principles of text-to-speech conversion and the computation of prosodic features for good quality synthesis.

The content on speech recognition includes signal processing techniques to extract acoustically and perceptually significant features from speech signals, together with the principles, examples and limitations of engineering implementations of automatic speech recognition. Template matching with dynamic time warping, hidden Markov modelling, and artificial neural network techniques are introduced with a minimum of mathematics. This introductory treatment compares the performance of phone and word based recognition systems, and reveals to the students the *linguistically* crude nature of speech recognition systems which lack semantic and grammatical processing.

Linguistics - lexicogrammar and parsing (Weeks 7 - 10)

This section commences with the concepts of concatenative and replacive morphology, using examples from English and other languages (including German, Aztec and Arabic). The syntax component deals primarily with how to construct sentences from noun phrases and verb phrases. Natural language parsing is dealt with using finite state machines, recursive transition networks and augmented transition networks. Unification models of grammar are also introduced, using examples including that of the unification of a sentence in an Aboriginal language.

Knowledge Representation and Text Generation (Week 11 -12)

The efficient computational representation of linguistic knowledge is a vital in complex systems dealing with natural speech and language. This section of the course introduces the formal methods of predicate calculus, and network based methods for knowledge representation. The topic of semantic networks is dealt with in some detail, and forms the basis of the accompanying tutorial exercises.

Text generation is discussed using the principles of systemic functional grammar. This description of language provides a set of hierarchically defined alternatives (choices) which is eminently well suited to computing. The lectures outline how a concept involving actors and an action can be expressed generically and realised a grammatically correct text in a large LISP-based software package which is under development.

Engineering speech systems - human factors (Week 13)

This final lecture reviews the problem of designing useful speech systems with current technology. Here the human factors are paramount, as even relatively good recognition performance can be confounded by poor systems design. The lecture draws on the recent literature on so-called 'Wizard of Oz' experiments (Fraser & Gilbert, 1991), and examples of good and bad systems.

PROJECTS

The project activity allows students to gain greater insight into one or more aspects of the area by working in interdisciplinary groups of two or three. The projects are expected to take about 4 hours per week for the latter half of the Semester. All of them contain some practical work, mostly of a computing nature, and the outcome submitted for assessment is a Report (one per group) and any relevant software.

Project proposals are made by the lecturing staff. Those from engineering tend to be specified in fairly general terms (like most engineering problems), so that the initial task of the group is to define a realistic set of goals. Project topics include:

- Phonetic parsing for recognition
- Perception of synthesized vowels
- Speech perception tests
- A Yes/No detector
- Database design for limited vocabulary speech recognition

The projects suggested by Linguistics staff tended to be formulated in more detail. The emphasis of these was on performing computational analysis and synthesis procedures, and have included:

- Syntax for agreement in French
- Russian case inflexions on nouns
- Syntax for information questions in English
- Modelling exchanges in dialogue

Most of the projects were very well done, with useful software being written in C and PASCAL. In one or two cases, the pseudocode was a satisfactory outcome. Some students were extremely resourceful. Where the partners found that they could build constructively on their background knowledge (as well as the course content), the project work was extremely good. One example of this was in the project on Phonetic Parsing for Recognition. The idea of this was to design a lexical matcher to find all the possible lexical translations of sequence of phonemes. These students investigated the effect of a number of phenomena such as phoneme deletion and substitution which exist in natural running speech. Thus they gained insights into articulatory and acoustic aspects of everyday speech, and their impact on automatic speech recognition, as well as designing and implementing quite a large software package to match hypothesized phoneme strings to the phonetic data in a Macquarie Dictionary.

COURSE ASSESSMENT

The project counted for 50% of the assessment. In only one case of a group of three was it necessary to assign marks other than equally to the group members. A final two-hour open-notes examination covered the lecture course content, other than the material of the introductory weeks. The paper was sectionalised into three 'engineering' and three 'linguistics' questions, and students were required to answer at least one from each section. The linguistics students were somewhat apprehensive of the examination, as examinations are not the normal form of assessment in their courses. In first run, all students passed the course.

EDUCATIONAL ISSUES

This course raised a number of interesting educational issues. At undergraduate level Arts and Engineering have distinctly different motivations and methodologies. The concept and realisation of 'creativity' are rather different between the two cultures. Engineers focus, ultimately on solving problems, and their core courses concentrate on the corresponding analytical tools. Accordingly, they are used to receiving knowledge presented as 'received wisdom', and the primary literature sources of are of little relevance. Arts courses are generally less objectively focussed, and examine primary and secondary literature in depth. Here re-analysis is common, and essays are a major creative activity. Linguistics combines science and the humanities in many respects. The study, description and modelling of speech and languages is highly scientific, and computational methods are becoming increasingly relevant. Nevertheless the didactic paradigm of core linguistics courses tends to follow that of the Arts tradition.

Both groups of students - engineers and linguists - found the approach of lecturers from the other discipline a little difficult. But both groups gained useful skills and attitudes from the experience. The engineers learned to work better from contemporary literature, and the projects introduced the linguists to a more practical style of work.

CONCLUSIONS

The student evaluation conducted at the end of the course was mostly positive about its concept and conduct. The second run of the course took into account the major criticism, and restricted the number of lecturers to four. As noted earlier, the introductory material was presented as in common in the second run.

Overall, we regard the course to be a success, and we intend to continue to run it. Possible future developments are short courses for practising professionals and an expanded course at the postgraduate level.

ACKNOWLEDGEMENTS

The concept, design and conduct of this course was the combined effort of many of the authors' colleagues. They particularly wish to acknowledge Prof. Bill Foley, Dr. Marilyn Cross, and Mr. Andrew Hunt for their contributions.

REFERENCES

Computer Speech and Language Processing, University of Cambridge Computer Lab & Engineering Department Handbook (1992).

Fraser, N.M. & Gilbert, G.N., (1991), *Simulating Speech Systems*, Computer Speech and Language, 5, 81 - 99.

Holmes, J.N. (1988), *Speech Synthesis and Recognition*, (Van Nostrand Reinhold: UK).