EDUCATION IN SPOKEN LANGUAGE ENGINEERING IN EUROPE

Phil Green[1], Carlos Espain[2] and the Spoken Language Engineering Working Group of the Socrates Thematic Network in Speech Communication Sciences

1: Speech and Hearing Research Group, Department of Computer Science, University of Sheffield, Regent Court, 211 Portobello St., Sheffield S14DP, UK.Tel. +44 114 222 1828, FAX: +44 114 278 0972, e-mail: p.green@dcs.shef.ac.uk 2: University of Porto.

ABSTRACT

We summarise the results of a survey of Spoken Language Engineering Education in Europe. We highlight the multidisciplinary nature of the field and the consequences this has for teaching. The survey indicates a wide variety in the breadth and depth of coverage of SLE.

1. INTRODUCTION

The Spoken Language Engineering Working Group is one of four groups active within the Socrates Thematic Network in Speech Communication Sciences. Its brief for the period January to July 1997 was to survey and report on education in SLE, primarily within the EU. The members of the group who contributed to this work are listed in table 2 at the end of this paper. The full report of the working group will be published in [1] and posted on **http://tn-speech.essex.ac.uk/tn-speech/** and should be available by the time this paper is published.

1.1. What is Spoken Language Engineering?

SLE is concerned with the design of devices and systems which process speech. The term 'Speech Technology' is synonymous. There are generally taken to be three main areas within SLE:

- **Speech Coding,** concerned with the efficient encoding, storage, transmission and playback of speech data.
- **Automatic Speech Recognition** (ASR), concerned with transcribing speech into its orthographic form, or with understanding the meaning of a spoken message. In addition to the question 'What is being said?', related work addresses questions such as 'Who is speaking?' (speaker recognition, speaker verification) and 'What language is being spoken?' (language identification).
- **Speech Synthesis**, concerned with generating speech. In Text-To-Speech (TTS), for instance, the goal is to produce a device capable of reading out loud.

Speech coding has been studied in connection with telephone systems for most of this century. Recognition and Synthesis have been researched for more than 40 years: almost as soon as computers existed there were attempts to give them ears and voices. The subject is both scientifically compelling and commercially attractive. The apparent ease with which people communicate by speech belies the intricacy and the subtlety of the processing which is taking place. Only in the last decade has sufficient progress been made for SLE to move into the market place, but it is now becoming part of everyday life, with a growing need for component technologies to be properly engineered into habitable systems.

1.2. Why is Education in SLE Important?

There is a need to focus on SLE education because:

- as shown in later sections, the multidisciplinary nature of SLE means that it is not normally offered as a subject in its own right in any one Department or School. Perhaps it should be.
- of the demand for employees in this area: a look at the number of employment opportunities advertised from both industry and academia on ELSNET shows the increasing demand for employees in SLE.
- of the demands from industry: many computing (e.g. Microsoft), electronics (e.g. Toshiba, NEC) and telecommunications (e.g. NTT, Nokia, Ericsson, Motorola) companies are now focussing on SLE. It is clear that with the convergence of computing SLE has a significant role. Human-computer interaction, mobile communications and computing will include many applications and products in SLE.
- of the European dimension: the numerous cultures and languages of Europe impact on both research and education which contrasts with the US/Japan. In Europe there is considerable research on machine (aided) translation (e.g. systran, Verbmobil), software localisation, and MultiLingual spoken dialogue systems. There is also a diversity of educational programmes in terms of structure and content. Any specialisation in SLE would help to bring together the variety of cultures/languages and educations in Europe.

1.3. The multidisciplinary nature of SLE

SLE is exceptional in the number of subjects a practitioner must, at the very least, know the basics of and, ideally, should be at ease with. They are

1.**Acoustics**: the physics of sound.

2.**Speech studies**: speech production, speech perception, hearing, psycho-acoustics, phonetics.

3.**Language studies**: grammar, semantics, pragmatics, computational linguistics, lexicography

4.**Mathematics,** especially integral and differential calculus, probability theory, Gaussian statistics.

5.**Statistical learning and decision-taking**, especially pattern recognition and artificial neural networks.

6.**Electrical Engineering,** especially signal processing and coding theory.

7.**Computer Science:** IT skills, programming ability, software engineering.

Students come into SLE courses from a variety of backgrounds: typically their degree is primarily in one of the above and they have already taken courses in a subset of the others. A computer scientist, for instance, has 7 above, and may have 3, 4 and 5. A phonetician has 2 and may have 1, 3 and 7. One of the problems facing teachers of SLE is the range of disciplines that come together in this subject. Furthermore, these disciplines differ markedly in the way they have traditionally been taught and researched, ranging from experimental phonetics through the mathematical treatment of signal processing to the specification and design paradigm of software engineering.

2. AN OVERVIEW OF SLE COURSES

SLE courses are found within a wide variety of curricula. Where a department has a group or an individual active in this research area it is likely that an SLE option will be offered within that department's degree schemes. A short overview of course types follows:

- To our knowledge, **no specialist SLE first degree exists** within Europe. The four-year MSc curriculum on Language, Speech and Informatics at Nijmegen (NL) perhaps comes closest.
- **1-year special course** following previous studies. Such courses can be identified more frequently. In Britiain we find postgraduate MPhil and MSc courses building on first degrees in other subjects (e.g. Cambridge, Edinburgh, UCL, Essex). It is worth noting that the problems caused by the interdisciplinary nature of SLE are at their most prominent for this kind of course, with students entering with a variety of first degrees from a variety of institutions.
- **one or more years SLE specialisation within a phonetics study.** This is found in the Netherlands and Germany and for instance the Essex MA in the UK. Sometimes SLE is offered as a separate strand, sometimes it is integrated into the phonetics curriculum. It is non-engineering because for instance speech coding is typically not included.
- **Option modules in Engineering, Computer Science or Phonetics studies.** This is the most common case and the courses can range from a few overview lectures to a fairly detailed series on communication, speech signal processing, synthesis and recognition of speech.

3. A SURVEY OF SLE COURSES

The TN-SLE working group is compiling a survey of SLE courses in Europe. At the time of writing, this survey is incomplete, but enough material has been collected to draw some conclusions.

Based on [2,3, 4], we are aware of some 64 European Institutions in which SLE is taught. These are offered by a mixture of Engineering, Phonetics/Linguistics and Computer Science departments. The broad geographical spread is as follows:

European Region	Phonetics	Engineering	Computer Science	Total
Central/ Northern	10	12	15	37
Southern	1	9	4	14
Eastern	2	7	4	13
Total	13	28	23	64

Table 1: Number of Institutions teaching SLE, by region and discipline.

We have collected profiles for SLE modules taught in 24 of these institutions. The questions asked, and a summary of responses, are given in the following sections:

3.1. To what level of students is the module offered?

SLE is predominantly taught either late in an undergraduate degree (at least year 3) or at Masters level. Where schemes permit, the same or nearly the same course is frequently offered to both cohorts. Only 2 of the institutions surveyed teach SLE at an earlier stage.

3.2. What is the proportion of a student's time occupied by the module (fraction of years)?

There is a wide range here, reflecting the varying depth to which SLE is taught. The specialist MSc courses include rather more than 0.5 years of SLE teaching, and also involve dissertations which may be on SLE topics. At the other end of the scale, about a third of the surveyed institutions use less than 0.1 student-years. The average is around 0.2 student-years.

3.3. What are the Aims and Objectives of the module?

Aims and objectives of the different modules in speech depend on the level they are targeted at, the proportion of the student time devoted to them, and the structure of the course they are included in. As a rule the module or modules are in general designed, at their own level, to cover all the main areas of speech technology, in what can be called a 'horizontal' topic axis, namely synthesis, recognition and, to a significantly less extent, coding. Exceptions are modules that are included in pattern classification areas, which are primarily targeted to recognition issues.Again as a rule, modules aim to cover, though at different levels, the entire field in a 'vertical' sense, from an introduction to perception and phonetics to natural language processing, reflecting that SLE is a multidisciplinary subject. It is also clear from the data that SLE is strongly applications oriented. Studying and building of practical systems are commonly presented as major aims of the modules: this is an engineering field. One word that appears frequently in the survey is 'skills', conveying the notion that SLE is an area where you should be prepared to be a practitioner, even when you go into research.

3.4. What are the module's Prerequisites?

The modules we have surveyed show a wide variety of prerequisites, reflecting the previous studies of the student cohorts they are designed for.

Several programmes, especially those which are language based rather than computer science or engineering based, are interested in a student's breadth of awareness of language in general and a basic ability to solve complex problems of any kind. The more technically oriented programmes generally require an undergraduate background involving computer science and mathematics (especially statistics) - these are, in any case, very specialist programmes focussed on areas such as signal processing. The difficulty here is that linguists and phoneticians get excluded from such courses - and of course vice versa.

Some institutions, especially in the UK, show a pattern of more broadly based courses (Sheffield, UCL, Cambridge, Edinburgh, Essex) - all teaching various areas which are required as prerequisites elsewhere. This enables students to plot their way among options which on the one hand build on an extend skills they may already possess and others which enable them to explore and develop new skills. This approach seems to fullfil the need for producing a new type of all-round graduate who is equally at home in all areas.

3.5. What is the SLE Syllabus?

The following sections summarise the material taught in SLE modules by topic:

3.5.1. Speech Signal Processing and Coding

Speech signal processing is a subject covered by the majority of the modules or series of modules. Most commonly referred sub-

jects are: speech production models, articulation mechanisms, speech perception models, feature extraction, formant and pitch analysis, endpoint detection and speech segmentation, space and time transforms, linear prediction, cepstral analysis, filter banks, mel-cepstral coefficients, vector quantization.

Coding as such is explicitly included in ten modules and is a major topic in three of them. Among the most frequently taught subjects are: vocoders such as channel and formant coders, waveform coders, including PCM, APCM, DPCM, ADPCM; sub-band coders and analysis-by-synthesis coders like CELP.

Electrical Engineering curricula normally include general purpose coding modules, which are not referred in the survey, but can easily be extended to include speech coding. The survey might therefore not properly reflect the situation in teaching coding techniques. The same specificity of Electrical Engineering curricula could explain the lack of references to topics such as channel equalisation, echo cancellation, or DSP algorithm implementation, which are commonly dealt with in more general, previous, modules, and are not taught within other curricula.

3.5.2. Speech Recognition

Recognition features in almost all the survey entries. Roughly half the sites provide a coverage which explains the techniques underlying the prevailing statistical approaches to ASR, for instance the Viterbi and Baum-Welch algorithms associated with Hidden Markov Models, corresponding neural-net techniques and the probabilistic formulation of acoustic and language modelling. At least three sites go beyond this, adding work on current research topics such as context-dependent phone units, HMM-ANN hybrids, out-of-vocabulary items and noise robustness. The remaining sites give a briefer overview of ASR techniques. Eight sites teach the related topic of speaker identification.

3.5.3. Speech Synthesis

These are the basic elements needed for a thorough treatment of speech synthesis:

- **Low level synthesis** (i.e. not text-to-speech): vocoders, formant synthesis (parametric synthesis) parallel vs. series arrangement of resonators; lpc; concatenated waveform synthesis (including PSOLA); articulatory synthesis; units - allophones, diphones, syllables, words, longer stretches (including the basic phonetic theory to classify and model these units). Basic acoustic theory which enables the above to work as practical ways of recreating the speech waveform.
- **High level synthesis** (i.e. text-to-speech, concept-to-speech) - need to include language model. Discussion of the philosophy: are we trying to recreate a perfect human-like speech waveform or are we trying to create the perceptual illusion of listening to human speech (the two are not necessarily the same). Orthography to phoneme conversion - rule based vs. neural network. Morpheme decomposition (for some systems). Dictionaries and what they might contain (syntactic information, phonological information) - not restricted just to exceptions but modern large-scale dictionaries (30000 words +) like DecTalk and SPRUCE. Elementary parsing (needed for accurate prosody assignment). Basic phonology (including prosodics). Inclusion of algorithms for improving naturalness based on modifications of neutral speech to introduce emotional and intentional content. Introduction of a module to process pragmatic and other semantic-related information to modify the phonological and prosody assignment algorithms (particular intonation and timing) to provide the proper acoustic correlates of intention and emotion - 'Pragmatic Phonetics'. Speech production theory (for things like coarticulation). Introduction of non-linear phonology and phonetics. Articulatory Phonology, Cognitive phonetics. Acoustic theory. Various complete systems: JSRU, INFOVOX, Laureate, DecTalk, Klatt, Festival, CNET, Elan, SPRUCE
- **Synthesis Applications** Prosthetic devices, General voice output systems, Interactive systems (dialogue systems), Resynthesis (copy synthesis).

Synthesis features in around two-thirds of the SLE courses surveyed. Two institutes cover the subject to approximately the detail above. Ten cover most but not all of these topics and five provide a briefer overview.

3.5.4. Spoken Language Processing

Natural Language Processing and/or Computational Linguistics are frequently covered in courses other than the SLE module itself. Such courses typically cover

- Introduction to Linguistics: linguistic theory, psycholinguistics, sociolinguistics.
- Syntactic and Semantic Analysis: grammatical formalisms, parsing, formal language theory, semantic theories, stochastic grammars.
- Pragmatics: discourse learning, dialogue modelling, intention and belief modelling.

3.5.5. SLE Applications

In a number of courses a short presentation in application systems and development tools is offered. The students are directed towards multilingual application environments, evaluation, and cross-comparison of spoken language technology products. Dictation Systems such as, DragonDictate, VoicePad, VoiceType are used to demonstrate the capabilities of probabilistic, neural and hybrid speech recognition methods. The CPK Generic Dialogue System, OSCON, Phillips Dialogue System, SCLU toolkit, SpeechMania, SUNDIAL, TRAINS, Vocalist provide a reference for the application of a comprehensive range of spoken dialogue systems. Large vocabulary ASR development tools such the HTK toolkit are sometimes used by the students in the laboratory, building their own ASR systems and solving development problems. Speech Synthesis systems are demonstrated using mostly JSRU and MITalk. Verbmobil is the only speech translation system that is mentioned in the survey.

3.6. What project work is given?

Most SLE courses have a significant project work component. For many institutions this involves using software such as Matlab, Waves, HTK, etc. in laboratories during course practical sessions. For others it means a more in depth study of a specific problem in SLE such as "dynamic time warping for isolated word recognition" or developing an "automated camera-man". In cases where a Master's degree is offered then there is obviously a much larger project which is written up as a dissertation or thesis.

Some projects are in collaboration with industry (e.g. the Cambridge, GB). There are cases where complete courses are actually projects (e.g. Erlangen, DE). Aalborg University (DK) has an education system where approx. 50% of the education and assessment is through group project work.

3.7. What resources are used?

3.7.1. Textbooks

The following texts are cited most frequently in the survey returns:

Speech

- Deller, J.R., Proakis, J.G. & Hanson, J.H.L. (1993) Discrete-Time Processing of Speech Signals. MacMillan. (4 mentions)
- Rabiner, L.R. & Huang, B.H. (1993) Fundamentals of Speech Recognition. Prentice Hall. (8 mentions)
- Holmes, J. (1993) Speech Synthesis and Recognition. Chapman and Hall (4 mentions)
- Owens, F.J. (1993) Signal Processing of Speech. MacMillan. (4 mentions)
- O'Shaughnessy, D. (1987) Speech Communication: Human and Machine. Addison Wesley (4 mentions)

Language

- Allen, J. (1995) Natural Language Understanding. Addison Wesley

General

- Cole, R.A. (ed.) (1996) Survey of the State of the Art of Human Language Technology. http://www.cse.ogi.edu/CSLU/HLTsurvey/

3.7.2. Software Packages

The following packages are mentioned in the survey returns: ESPS/Waves, MatLab and HTK.

4. Opinions of SLE Teaching

In addition to the factual information above, we also asked for opinions on what teaching SLE is like. Questions and response summaries are again given below: of course, these are the opinions of the teachers, not the students.

4.1. Are your SLE modules popular with students?

SLE is reasonably popular where students have to choose between modules. In this case, SLE is often in contrast with mainstream degree work and students 'like to do something different'.

4.2. What aspects of your modules appeal to the students?

The notion of having a speech interface. The perception that SLE is a new and exciting technology. The links between language, experimental work and engineering.

4.3. What aspects of your modules do students find difficult?

The mathematical aspects, particularly understanding signal processing and recognition algorithms. The large amount of material that is frequently covered in a short time. The diversity of this material.

4.4. What aspects of the modules are difficult to teach?

Generally those where the topic is outside the background of the students, or indeed of the lecturer: for instance teaching signal processing to phonetics students or speech perception experimentation to computer scientists.

4.5. How well do your students perform in assessment?

Most sites report adequate performance. Two sites report generally mediocre performance.

5. CONCLUSIONS

SLE is a difficult subject to teach, because of its multidisciplinary demands. It is hard to find a line between giving a brief overview which does not really explain anything and a full treatment whose understanding is beyond the capabilities of the audience. There is a need for greater support to be widely available, for instance in the form of internet tutorials as suggested in the companion paper [5], and for better teaching software.

6. ACKNOWLEDGEMENTS

This work is supported by the Socrates Thematic Network in Speech Communication Sciences. We thank everyone who provided information for our survey.

7. REFERENCES

[1] SOCRATES Thematic Network in Phonetics and Speech Communication, Year 1 Report, University of Utrecht, The Netherlands, to be published September1997.

[2] ERASMUS Database of courses in Phonetics and Speech Communication: http://www/kgw.tu-berlin.de/ERAS/eraschoice.html

[3] G. Bloothooft, V. Hazan, D. Huber, and J. Llisterri, J. (Eds.), "European Studies in Phonetics and Speech Communication", OTS publications, Utrecht, 1995.

[4] ELSNET European PhD Studies in Language and Speech, Utrecht, 1993.

[5] M. Huckvale et al, 'Opportunities for Computer-Aided Instruction in Phonetics and Speech Communication provided by the Internet', proc. Eurospeech 97.

Name	Institution	ISO Country Code
Phil Green (chair)	Uinversity of Sheffield	GB
Carlos Espain (deputy)	University of Porto	PT
Christian Benoit	Polytechnique de Grenoble	FR
Gerrit Bloothooft	University of Utrecht	NL
Gerard Chollet	ENST Paris	FR
Vagelis Dermatas	University of Patras	GR
Andrzej Drygajlo	Inst. of Technology, Lausanne	CH
Jurgen Haas	University of Erlangen	DE
Gernot Kubin	University of Technology, Vienna	AT
Jean-Pierre Martens	University of Gent	BE
Paul Mc Kevitt	Aalborg University	DK
Michael McTear	University of Ulster	GB
Georg Meyer	University of Keele	GB
Antonio Peinado	University of Granada	ES
Victoria Sanchez	University of Granada	ES
Mark Tatham	University of Essex	GB

Table 2: Contributing members of the SLE working group