

CHAPTER 3

SPOKEN LANGUAGE ENGINEERING

Carlos Espain

University of Porto, Portugal

Gerrit Bloothoof

Utrecht University, The Netherlands

Antonio Bonafonte

Polytechnic University of Catalunya, Barcelona, Spain

Klaus Fellbaum

Brandenburg Technische Universität, Cottbus, Germany

Jürgen Haas

Friedrich-Alexander-Universität, Erlangen, Germany

Ruediger Hoffman

Technische Universität, Dresden, Germany

Georgios Kouroupetroglou

National and Capodistrian University of Athens, Greece

Paul Mc Kevitt

Aalborg University, Danmark

Michael Mc Tear

University of Ulster, Northern Ireland

Victoria Sanchez

University of Granada, Spain

Kyriakos Sgarbas

University of Patras, Greece

1. Introduction

In the first book of the Thematic Network an analysis of the current situation in Spoken Language Engineering (SLE) teaching across Europe [Bloothoof, 1997] has been presented. Based on that analysis we now present a proposal for the contents of studies in SLE in section 3.

In addition, we aimed to get a better understanding of the skills and knowledge as needed by SLE professionals working in the field. To this end a questionnaire has been sent to the SLE community. The results are presented in the next section.

2. The Questionnaire

To investigate the relation between the education received and the job requirements for professionals in spoken language engineering, we have distributed a questionnaire among the SLT community. The questionnaire addressed the academic and professional qualifications, main area of work, expertise, skills and job requirements, professional functions, type of contract and asked for personal views on SLE education and market trends.

Forty three answers were received from everywhere in Europe, which is not enough to claim a representative sample, as the total number of employees in SLE must be estimated far over thousand. However, some trends can be obtained from an informal analysis of the data.

2.1 Characterisation of respondents

>From the 43 questionnaires returned, 26 come from respondents working in speech recognition, 10 in synthesis, 5 in dialogue modelling, 4 in speech analysis, 4 in language engineering, 3 in human-computer interfacing, 3 in speaker verification, 1 in coding, 1 in phonetics, 1 in assessment and 1 in translation. A few respondents mentioned more than one area of work. Only 14 respondents have short-term contracts, which, within the sample, shows a steady job market.

Concerning the background education, 17 respondents pointed out electrical engineering, 7 physics, 6 computer science, 5 telecommunications engineering, 2 linguistics, 1 artificial intelligence, 1 informatics, 1 electronics engineering, 1 humanities. The information was not given by all. The group comprises 13 respondents with a Ph.D. degree while 5 mention a Masters.

Each respondent was asked to scale the own professional activities in terms

of the following categories: consultant, technician, researcher, applications developer, system integrator, research director or manager. In addition there was a category where one could specify other activities and 11 respondents mentioned either 'teacher' or 'professor' there.

Table 3.1 shows the percentage devoted to each activity, averaged across all respondents. It shows that on average all activities were about equally represented.

Table 3.1.: Average percentage of time devoted to activities across all respondents

	%
consultant	13
technician	13
researcher	19
application development	16
system integration	13
research director	14
manager	13

2.2 SLE education and job requirements

To get an overview of the knowledge and skills necessary in SLE, 16 topics were considered: signal processing, statistics, neural nets, DSP programming, phonetics, speech production, speech perception, acoustics, channel equalisation and echo cancellation, artificial intelligence, linguistics, speech analysis, language modelling, human-computer interfacing and dialogue modelling. These subjects range from basic tools to advanced matters. For each subject the following questions were posed:

- *How would you judge your expertise in the subject?*
- *Where did you get your education on the subject?*
 - At the University?*
 - On the job?*
 - By personal effort?*
- *How relevant has been this expertise for your professional activities?*

To each question the respondent could answer with a number from one to three with three reflecting the greatest importance¹. Table 3.2 shows for each topic in the first column the average score for the relevance of the topic for the professional activities and in the second column the average score for the importance the topic had in the back ground education, here summarised by 'University education'. Since these averages encompass all

¹In a first version of the questionnaire the reverse interpretation has been used.

possible professional profiles the two columns only give a general idea of the relevance of the topic.

Table 3.2.: Average score [max. = 3] on university education and professional relevance for a series of topics. The correlation between the two items is given in the third column

	University education	Professional Relevance	Correlation coefficient
signal processing	2.3	2.3	0.4
statistics	2.3	1.7	0.0
neural nets	1.4	1.5	0.0
DSP programming	1.6	1.5	0.3
computing skills	2.3	2.6	0.1
phonetics	1.7	2.1	0.1
speech production	1.7	1.8	0.5
speech perception	1.5	1.9	0.3
acoustics	1.9	1.6	0.2
channel eq. and echo cancellation	1.3	1.4	0.2
AI	1.7	1.7	0.2
linguistics	1.6	1.9	0.2
speech analysis	1.8	2.3	0.3
language models	1.3	2.0	0.0
Human-Computer interfacing	1.3	2.0	0.1
dialogue modelling	1.3	1.9	-0.2

In general, most important in the University education were signal processing, statistics and computing skills. Least attention has been given to neural nets, channel equalising and echo cancellation, language models, human-computer interfaces and dialogue modelling. For the current job, the major requirements were with computing skills, signal processing and speech analysis. Least important are channel equalising and echo cancellation, neural nets and DSP programming. Still, it should be remembered that these averages are biased because more than half of the respondents worked on speech recognition issues and for instance only one on speech coding. If we compare the two columns per topic, major mismatches are found with statistics which is less relevant for the job than it is in education. Statistics and acoustics are the only topics where education scored higher than job relevance. The reverse was the case with spoken dialogue systems oriented topics such as language models, human-computer interfacing and dialogue modelling, and slightly less with speech perception, phonetics and linguistics. Especially the topics related to spoken dialogue systems show that university education lags application developments and needs. Also the need for more attention for both language and speech processing in university studies is reflected in the table.

An additional way of analysing the data is to compute per topic the correlations between the scores for relevance during university education and the scores reflecting the professional relevance [third column in Table 3.2]. In general the correlation coefficients are low, which means that professionals often work in areas in which they have not received specific education and training at the University². This may be due to the fact that there is no specific curriculum in SLE. But it also may be related to the fact that SLE employees often have very different backgrounds and that in smaller teams a professional with, for instance, a background in electrical engineering can be asked to do work in phonetics. It may also be, however, that this lack of mapping is a very general property for academic graduates in many scientific fields. Broad academic values may in general prevail over job dependent skills and knowledge.

From the questionnaire it becomes thus apparent that the market currently demands a professional who is called to work in other areas than for which he or she has received academic training. If a curriculum in SLE is to be established, it must have the flexibility to allow for entrances from people with different backgrounds. This seems to point to the implementation of a curriculum at Masters level on top of other curricula and where credits can be inherited from these previous background curricula [see Chapter 4 for a proposal for a European Masters in Language and Speech on top or within current education schemes].

2.3 The background of the professional profile

The respondents have been asked about their personal professional profile. For the various topics they were asked to score their expertise. They were also asked to score to what extent they had obtained the expertise through (1) their back ground studies, (2) on the spot, (3) by personal effort. Finally the relevance of this expertise for their current job had to be indicated.

The correlations between the scores for the various items have been computed across all respondents and all topics. Table 3.3 shows these coefficients and their 95% confidence interval.

Table 3.3.: Correlation coefficients between own expertise, the relevance of the expertise for the current job, and the importance of the way the expertise has been acquired (1) through university education, (2) during the job, (3) by personal effort.

	correlation	95%
--	-------------	-----

²A low correlation coefficient may also arise when there is little spread in the scores for both items. This was not the case in our sample.

	coefficient	confidence interval
university education * relevance for job	0.28	0.08
during the job * relevance for job	0.49	0.10
personal effort * relevance for job	0.43	0.11
university education * own expertise	0.50	0.11
during the job * own expertise	0.41	0.08
personal effort * own expertise	0.43	0.11
own expertise * relevance for job	0.70	0.05

Two conclusions can be drawn from Table 3.3. First, there is a relatively high and significant correlation between the self-judged expertise and its relevance for the current job, which means that the community shows a strong, healthy, self-confidence! Second, the correlation between relevance and university education is relatively low, which agrees with the observations from Table 3.2. As said before, this is of course due to the fact that most of the current studies prepare students for a wider range of job profiles than SLE alone. Whether it is time or not to have a specific curriculum on SLE depends on the size of the SLE job market as compared to that range. Nevertheless, universities clearly could do better on preparing SLE professionals.

2.4 Views from respondents

27 suggestions for SLE teaching and practice were received. The majority of the answers stressed the multidisciplinary nature of SLE. 5 answers spontaneously referred the need of a specific curriculum in SLE, of which 2 suggested it should be implemented at undergraduate level.

Some pointed to the need of better education in specific subjects. Somewhat surprisingly if one reminds the job requirements mentioned in Table 3.2, the subjects were phonetics (5 references), statistics (4), acoustics (3), mathematics (3), linguistics (3), speech processing (3), artificial intelligence (2), programming methods (1), communication theory (1), cognition sciences, (1) and machine learning (1). Three answers mentioned the importance of laboratory work, and two, the need for a broader vision of SLE applications. A respondent complained about the lack of education on how to give an oral presentation and written reports. From a broader point of view, another respondent asked for the increase of co-operation between different laboratories and a better synchronisation of European programmes.

A few strategic lines were suggested. Some, giving guidelines:

Education in such a multidisciplinary area should be based in one of the disciplines, but give an elementary exposure to the relevant other areas. For example, signal processing students should be trained to be signal processing experts, but need elementary courses on topics such phonetics, acoustics, etc..

Others, very direct and clear:

Have a broader education! Do look at signals! Do hand labelling! Understand all levels of language and speech! Communicate with experts of other disciplines! Do more error analysis! ('Why' rather than 'how much'). Do more crazy experiments!

Roughly one third of these 27 people gave hints on future market trends. All the respondents believe the market will grow substantially in the near future. Two of them underline a certain delay between EU and USA markets, but both agree this delay will be overcome soon. One respondent states that the next improvements will come from the areas of linguistics and dialogue systems as for him speech signal processing has reached its limits. Another believes that the future area of research will be dialogue systems with spontaneous speech. However no less than three people think these developments will depend on the growing of artificial intelligence as a tool. Four respondents defend that multidisciplinary human-machine interfacing, including speech, is the horizon. Voice control, rehabilitation engineering, telecommunication applications, dictation systems, translation systems, telematics and mobile communications are also mentioned as promising areas. One response stresses the need for more flexible systems particularly with multilingual processing. Two people refer to the difficulties for market expansion that rise from the fact that general public is sometimes reluctant to use voice interfacing with machines. Three responses are however slightly pessimistic. Though not denying the growth of the market, they point out that the breakthrough in speech technology that is widely expected is still not at reaching distance.

3. The SLE Studies

3.1 Guidelines and Structure

The contents of studies in Spoken Language Engineering must be grounded in a consensus of what a SLE professional is. It is somewhat restricting to define such a profile in an area which is changing very fast. A SLE professional is no doubt seen as an engineer, someone who masters the theoretical issues in speech technology and is able to produce end products. His³ work concerns technological areas such as synthesis, recognition, coding, but his education includes, besides engineering tools such as signal processing and programming skills, disciplines which a normal engineer seldom deals with, like phonetics and linguistics. Therefore, he must have background education in the latter areas also, not only as a fundamental basis of the own work but also to be able to cooperate with phoneticians and linguists. He is the one who is in charge of processing whatever information he receives from experts in many areas, and to incorporate it in systems that work. He is a researcher, an application developer and a system integrator.

³Reads her or she equally

The proposal for the Contents of SLE studies that follows has various goals and many restrictions as well. The proposal should be read mainly as a repertoire of the knowledge and skills a SLE professional should acquire during academic education and training. This means that we believe the contents has an academic level and should be taught at universities. We do not believe that in-house training in companies would be the right and cost effective solution to educate and train employees, with the exception perhaps of very specialised topics or skills.

At this point no precise indication is given as to whether this Contents of SLE studies should be offered as a full degree at any university, although it can be the basis for such an implementation. The Contents can be understood as a reference on the basis of which departments can decide whether and what kind of specialisation in SLE they want to offer to their students. Along the same line, the Contents can also be thought as a guide for the implementation of a Masters Degree that can be implemented on top of other degrees. In many respects the following contents will be part of the contents of the European Masters in Language and Speech which will be described in Chapter 4. Finally, for the professional, for the lecturer, for the student and for the team research manager the Contents can serve as a reference guideline covering all aspects of current SLE consolidated theory and practice.

For the sake of clarity and to facilitate the definition of access requirements for students from different backgrounds, the curriculum will be stratified into 7 levels, labelled from A to G, according to the nature of the subjects and their interdependence, their character as general knowledge as opposed to SLE specific knowledge, the level of practical and theoretical work and the role of a project and research attitude. It is not recommended that a student enters into any level before he/she has passed all the subjects from the previous level.

Levels A, B and C comprise those subjects which should be taught before entering speech technology itself. Level A refers to a basic background, level B to some general tools in informatics, mathematics and engineering and level C to specific tools and knowledge that are particularly useful for dealing with speech technology issues. These levels will be presented in a very loose manner since they are common to other areas of technology and science.

Level D handles those subjects of SLE that are common to all main areas of work, while level E goes into those areas. Subjects in both levels will be presented in a detailed form as fully specified modules. Level F addresses SLE applications and level G current research areas. Both will also be

presented as single modules.

Although the Contents should be integrated in the European credits system, no indication of credits or even student hours of work is presented aiming to give full flexibility for the different European education systems.

3.2 Basic subjects

Level A

Level A encompasses subjects that give a solid formation in Mathematics and Physics at the level of an engineering degree.

Level B

Informatics

- Operating Systems
- Programming [C,...]

Signal Processing

- Sampling and digitisation of signals
- Digital filters
- Fourier series and transforms
- Cepstrum
- Linear prediction

Statistics

- Probability Theory
- Random Processes
- Vector-valued random processes

Telecommunications

- Information Theory of Sources and Channels
- Source Coding
- Channel Coding

Electrical Engineering and Electronics

- Recording instrumentation

Level C

Advanced informatics and computer science

- System's programming
- Algorithms and complexity
- Introduction to artificial intelligence
- Logic programming

- PROLOG programming
- DSP programming
- Real-time systems

Acoustics

- The physics of sound
- Subjective and objective properties of sound
- Complex sound waves
- Sound quality
- Principles of microphones

Basic Phonetics

- sounds, patterns, features and rule systems
- symbolic phonetics
- phonetic transcription
- phonology

Linguistics

- morphology and lexicography
- syntax
- semantics
- pragmatics
- formal language theory
- parsing

Pattern recognition

- distance measures
- maximum likelihood classification
- feature selection and probabilistic separability measures
- clustering algorithms
- entropy concepts in pattern recognition

Neural nets

- the artificial neuron
- network principles and paradigms
- layered networks
- the multilayer perceptron
- learning vector quantizer

3.3 Core subjects

Core subjects are presented as example modules, with the headers: (1) Title, (2) Objectives, (3) Syllabus, (4) Pre-requisites, and (5) Best practice, reading material and Internet support.

LEVEL D

Title: **Introduction to Speech Communication and Speech Technology**

Objectives: To formulate a general but integrated overview of speech communication principles, parts, methods, technologies and applications.

Syllabus:

- The importance of: a) human speech communication, b) human-machine interaction by speech and c) computer mediated interpersonal communication by speech.
- **Speech communication by humans:** basic modelling, acoustic and articulatory phonetics.
- **Acoustic properties of speech sound**
- **Speech production** (from anatomy to modelling).
- **Speech hearing and perception.**
- **Speech signal analysis:** aim, classification of non-parametric and parametric methods, auditory based analysis, speech time-frequency representations, comparison of techniques and methods.
- **Speech coding:** scope, application areas
- **Speech synthesis:** classification of methods, comparison of methods, an advanced speech synthesiser (in block diagram), application areas.
- **Speech recognition:** dimensions of difficulty, classification of methods, comparison of methods, speech databases in recognition, application types.
- **Human-machine interaction by speech:** dialogue modelling, the speaking environment, noise problems, speech quality assessment, speaker and language recognition, multimodal interaction, human factors.
- **Speech technology applications:** state of the art in speech coding, synthesis and recognition products, applications in telecommunications, speech technology in office automation, speech technology for command/control and in manufacturing, applications for disabled and elderly people and voice therapy, multimedia applications, speech technology in education and training, other application areas.

Pre-requisites:

As this module is general and introductory the pre-requisites can be covered (not in detail) inside the module.

Acoustics: basics of sound waves - the physics of sound, subjective and objective properties of sound, sound intensity, sound pressure level, dBel, neper, loudness, phone, sone, frequency and pitch, complex sound waves, sound quality, noise, acoustic resonance, principles of microphones.

Phonetics: relation between written and verbal language, phonemes and

allophones, IPA, coarticulation, prosody and intonation, accent and dialect.

Signal processing: Sampling and digitisation of signals, Filters, Fourier Transforms, FFT,

Best practice, reading material and Internet support:

Parts/chapters from the following books

G.Borden, K. Harris and L. Raphael (1994). *Speech Science Primer - Physiology, Acoustics and perception of Speech*. Williams & Wilkins, ISBN 0-683-00944-3

P.Denes and E. Pinson (1993). *The Speech Chain*. W.H. Freeman and Company, ISBN 0-7167-2344-1

J. Holmes (1988). *Speech Synthesis and Recognition*. Von Nostrand Reinhold, ISBN 0-278-00013-4

Owens (1993). *Speech Processing of Speech*. MacMillan, ISBN 0-333-51922-1

Furui, S. (1989). *Digital Speech Processing, Synthesis and Recognition*. Marcel Dekker, Inc, ISBN 0-8247-7965-7

Sydral, R. Bennett and S. Greenspan, Eds. (1995). *Applied Speech technology*. CRC Press, 1995, ISBN 0-8493-9456-2

Roe and J. Wilpon, Eds. (1994). *Voice communication between humans and machines*. National Academy Press, Washington DC, ISBN 0-309-04988-1

Title: **Speech analysis**

Objectives: The module deals with fundamental knowledge in signal analysis with strong emphasis to speech signals.

Syllabus:

- **Speech signal analysis in the time domain:** Amplitude, energy, zero crossings, statistic parameters, autocorrelation function
- **Speech signal analysis in the frequency domain:** Short time Fourier transform and windowing, spectrogram representations (wide and narrow band), general time-frequency descriptions (Wavelets, etc.), filter banks for speech analysis
- **Homomorphic speech processing:** Cepstrum and its applications
- **Linear prediction of speech:** Applied general prediction theory to speech, calculation methods (autocorrelation and covariance methods), application in speech signal modelling
- **Determination of voice source parameters:** Mode of excitation, pitch and fundamental frequency determination
- **Determination of vocal tract parameters:** Formant and anti-formant analysis, estimation of vocal tract geometry

Pre-Requisites :

Signal processing: Basic concepts, Fourier series and transforms

Digital filter theory: Fundamentals of statistics, Advanced signal processing, Cepstrum, Linear prediction

Advanced statistics: Stochastic processes

Speech perception and production: Psychoacoustics (critical bands)

Speech production models

Best Practice, Reading Material And Internet Support:

The course should, among others, offer possibilities for the student to do computer experiments with speech signals in the time and frequency domain.

Rabiner, L. R.; Schafer, R. W. (1978). *Digital Processing of Speech Signals*. Englewood Cliffs: Prentice Hall, ISBN 0-13-213603-1

Deller, J. R.; Proakis, J. G.; Hansen, J. H. L. (1993). *Discrete-Time Processing of Speech Signals*. New York: Macmillan, ISBN 0-02-328301-7

Hess, W. (1983). *Pitch Determination of Speech Signals*. Berlin: Springer.

Title: **Natural Language Processing**

Objectives: The student should be able to acquire a general overview of the function of modern NLP systems. He will understand the strong and weak points of several NLP methods and be informed on the new developments currently being pursued. At the end of the course he will be able to implement simple NLP systems.

Syllabus:

- Aim of NLP [1.Ch1], [2.Ch1]
- Applications of NLP
- Machine Translation
- Grammar Checkers
- Dictation
- Automatic Document Generation
- NL Interfaces
- Basic Tools
- Finite State Automata [2.Ch2]
- Recursive Transition Networks [2.Ch3]
- Augmented Transition Networks [2.Ch3]
- Lexical Analysis
- Features [2.Ch7]
- Directed Acyclic Word Graphs [2.Ch7]
- Morphology [1.Ch3.7], [3]
- Syntactic Analysis

- Grammars [1.Ch3,4,5], [2.Ch4]
- Formal Grammars
- Type-0 Grammars
- Context-Sensitive Grammars
- Context-Free Grammars
- Regular Grammars
- Definite Clause Grammars [1.Ch4.7], [2.Ch4.5]
- Transformational Grammars [7]
- Systemic Grammars [12]
- Case Grammars [6]
- Unification Grammars [21.Ch15.2.3]
- Functional Unification Grammars [15]
- Generalised Phrase Structure Grammars [9]
- Head-Driven Phrase-Structure Grammars [19]
- Lexical Functional Grammars [5]
- Categorical Unification Grammars [11]
- Tree Adjunction Grammars [13]
- Parsing [1.Ch3,6], [2.Ch5]
- Top-Down
- Bottom-Up
- Chart Parsers [1.Ch3.4,6], [23]
- Earley Algorithm
- CYK Algorithm
- Partial Parsing [1.Ch6.5]
- Semantic Analysis [1.Ch8], [2.Ch8]
- Logical Forms [1.Ch8.1]
- Ambiguity Resolution [1.Ch8,Ch10]
- Semantic Networks [1.Ch10.3]
- Semantic Parsers [1.Ch11.4]
- Procedural Semantics [1.Ch13.7]
- Montague Semantics [21.Ch15.3.4]
- Pragmatic Analysis [2.Ch10]
- Knowledge Representation [1.Ch13]
- Reasoning [1.Ch13]
- Plan/goal recognition [21.Ch15.4.3]
- Speech Acts [22], [21.Ch15.4.4]
- Belief Models [21.Ch15.4.2]
- Discourse [1.Ch14,16], [14]
- Reference [1.Ch16.3]
- Natural Language Generation [2.Ch10.8], [8], [17]
- Content Determination
- Sentence Planning
- Surface Realisation
- Other Approaches
- Statistical NLP [1.Ch7], [24]

- Connectionist NLP [20], [16]

Pre-requisites: The student should have a previous experience in Artificial Intelligence and Prolog Programming.

Best Practice: The student should implement a small project during the course.

Reading Material: Two types of reading material are shown: the Main reading material is adequate enough to generally cover the course. The Additional reading material provides detailed insight to special course subjects.

Main:

- [1] Allen J. (1995). *Natural Language Understanding*. 2nd ed., The Benjamin/ Cummings Publishing Company, Inc.
- [2] Gazdar G. and Mellish, C. (1989). *Natural Language Processing in Prolog*. Addison-Wesley Publishing Co.

Additional:

- [3] Antworth, E. (1990). PC-KIMMO: A Two-Level Processor for Morphological Analysis. Occasional Publications in *Academic Computing*, **16**, Summer Institute of Linguistics, Inc., Dallas, Texas.
- [4] *Artificial Intelligence Journal*, (1993) vol.63, Special Issue on Natural Language Processing.
- [5] Bresnan J. (ed.) (1982). *The Mental Representation of Grammatical Relations*. MIT Press, Cambridge.
- [6] Bruce B. (1975). Case Systems for Natural Language. *Artificial Intelligence*, 6(4):327-360.
- [7] Chomsky N. (1965). *Aspects of the Theory of Syntax*. MIT Press, Cambridge.
- [8] Dale R., Hovy E.H., Rosner D., and Stock O. (eds) (1992). *Aspects of Automated Natural Language Generation*. Lecture Notes in AI #587, Springer-Verlag, Heidelberg.
- [9] Gazdar G., Klein E., Pullum G.K., and Sag I. (1985). *Generalized Phrase Structure Grammar*. Harvard University Press, Cambridge.
- [10] Grosz B., Spark Jones K., and Lynn Webber B. (eds) (1986). *Readings in Natural Language Processing*. Morgan Kaufmann Publishers, Los Altos, CA.
- [11] Haddock J.N., Klein E., and Morrill G. (1987). *Unification Categorical Grammar, Unification Grammar and Parsing*. University of Edinburgh.
- [12] Halliday M.A.K. (1985). *A Short Introduction to Functional Grammar*. London: Arnold.

- [13] Joshi A.K., and Schabes Y. (1992). Tree-Adjoining Grammars and lexicalized Grammars. in *Tree Automata and LGS*, Elsevier Science, Amsterdam.
- [14] Kamp H. and Reyle U. (1993). *From Discourse to Logic*. Kluwer, Dordrecht.
- [15] Kay M. (1984). Functional Unification Grammar: A Formalism for Machine Translation. in *Proc. 10th Intern. Conf. On Computational Linguistics*, Stanford University, California, ACL.
- [16] Mikkulainen R. (1993). *Subsymbolic Natural Language Processing: An Integrated Model of Scripts, Lexicon and Memory*. MIT Publisher.
- [17] Paris C.L., Swartout W.R., and Mann W.C. (eds) (1991). *Natural Language Generation in Artificial Intelligence and Computational Linguistics*. Kluwer Academic, 1991.
- [18] Pereira F. and Shieber S. (1987). *Prolog and Natural Language Analysis, Center for the Study of Language and Information*. Stanford, CA.
- [19] Pollard C., and Sag I. (1994). *Head-Driven Phrase Structure Grammar*. CSLI, Stanford Univ. Press and Chicago Univ. Press.
- [20] Reilly G., and Sharkey N.E. (eds) (1992). *Connectionist Approaches to Natural Language Processing*. Earlsdale.
- [21] Rich E., Knight K. (1991). *Artificial Intelligence*. 2nd ed., McGraw-Hill.
- [22] Searl J.R. (1969). *Speech Acts*. Cambridge University Press.
- [23] Winograd T. (1983). *Language as a Cognitive Process: Syntax*. Addison-Wesley, Reading.
- [24] Young S., and Bloothoof G. (Eds.) (1997). *Corpus-based Methods in Language and Speech Processing*. Kluwer Academic Publishers, The Netherlands.

Internet Support:

- {1} The ACL NLP/CL Universe:
<http://www.cs.columbia.edu/~radev/u/db/acl/>
- {2} Survey of the State of the Art in Human Language Technology:
<http://www.cse.ogi.edu/CSLU/HLTsurvey/HLTsurvey.html>
- {3} Natural Language Processing on the WEB:
<http://www.cs.cmu.edu/afs/cs/user/ralf/pub/WWW/nlp.html>
- {4} Software Tools for NLP: http://n106.is.tokushima-u.ac.jp/member/kita/NLP/nlp_tools.html

Title: **Speech production and perception**

Syllabus:

Speech production

- The sound source [C&Y 6.4 - 6.6, 7.11 - 7.12]

Respiration during speech

Phonation: structure and functioning of the vocal folds

Periodic, aperiodic and non-periodic voice

- Articulation [C&Y 6.7 - 6.11]

Anatomical structures of the vocal tract

Vowels (height, place, lip activity)

Frication and affricates

Stop production (occlusion, explosion, aspiration)

Other consonants (place, manner, voicing)

Acoustics phonetics

- Acoustic theory of speech production: Fant's source-filter model [C&Y 7.10 & 7.18]

Glottal wave shape models, Source spectrum, Area functions

Poles and zeros, Formants (frequencies, amplitudes and bandwidths)

- Experimental methods and tools [C&Y 7.1 - 7.8]

Waveforms

Spectrum and spectrogram (wide-band and narrow-band)

C&Y 7.14)

Acoustic properties of speech sounds

- Vowels [C&Y 7.15]

Formant structure

Diphthongs, triphthongs, glides

- Consonants [C&Y 7.16- 7.17]

Stops

Nasals

Fricatives

Affricates

Laterals

Approximants

Coarticulation and transitions

- Prosody [C&Y 7.19]

Stress

Rhythm

Intonation

Timing (Tempo, Final lengthening, Accentual lengthening,

Pausing)

Speech perception

- The auditory system [C&Y 8.2]

Anatomy, physiology and acoustic processing of the ear

Neurophysiology of hearing

- Psycho-acoustics [C&Y 8.3]

- Perception of speech units [C&Y 8.5 - 8.11]

Categorical perception

Acoustic cues

Textbooks: See Chapter 2 on Phonetics with series of books on the various topics. As a reference for the key words above we have used:

Clark, John & Yallop, Colin (1995). *An Introduction to Phonetics and Phonology, Second edition*. Oxford & Cambridge, MA: Blackwell.

LEVEL E

Title: **Speech Coding**

Objectives: The student should be able to design speech coders for different bit rates and speech quality requirements and become familiar with the current speech coding standards.

Syllabus:

Introduction

Quantization

- Scalar Quantization
- Vector Quantization

Waveform Coding

- Time Domain Coding
 - PCM
 - DPCM
 - ADPCM
- Frequency Domain Coding
 - Sub-band Coding
 - Transform Coding
- Standards

Parametric Coding

- LPC vocoders
- Sinusoidal coders
 - Sinusoidal Transform Coding
 - Multiband Excitation Coding
- Waveform-interpolation coders
- Standards.

Hybrid Coding

- Multipulse Coding
- Code-excited Linear Prediction Coding
- Standards

Applications

Pre-Requisites :

Informatics:

- Operating Systems
- C programming

- DSP programming

Signal Processing:

- Digital Filters
- Transforms

Statistics:

- Probability Theory
- Random Processes

Telecommunications:

- Information Theory of Sources and Channels
- Source Coding
- Channel Coding

Best Practice: The module should dedicate half of its credits to practical work and these should be devoted to the design, implementation and testing of speech coders and speech coding standards.

Reading material and internet support:

Deller, J; Proakis, J.G.; Hansen, J.H.L. (1993). *Discrete-Time Processing of Speech Signals*. Macmillan Publishing Company.

Furui, S. and Sondhi, M.M. (1992). *Advances in Speech Signal Processing*. Marcel Dekker.

Gersho, A. and Gray, R.M. (1992). *Vector Quantization and Signal Compression*. Kluwer Academic Publishers.

Jayant, N.S. and Noll. P. (1984). *Digital Coding of Waveforms*. Prentice Hall.

Kleijn and Paliwal (1995). *Speech Coding and Synthesis*. Elsevier.

O'Shaughnessy, D. (1987). *Speech Communication; Human and Machine*. Addison Wesley.

<http://wwwdsp.ucd.ie/speech.htm>

Title: **Speech Synthesis**

Objectives: Thorough knowledge in the principles, problems and applications of speech synthesis

Syllabus:

Introduction

- Definition of speech synthesis, text-to-speech and concept-to-speech, comparison to voice store and forward, general components of a speech synthesis system (block diagram), application areas.
- Basic Principles of Speech Synthesis

Speech synthesis

- Human speech production (excitation, articulation, radiation), summarised repetition of the module "Speech Production"

- Waveform synthesis
 - Speech units (phonemes, diphones, demisyllables)
 - Parameterisation and storage
 - Concatenation
 - Linear prediction
- Spectrum synthesis
 - Spectral representation (FFT spectrum, spectrograms, visible speech diagrams)
 - Filterbank synthesis
 - Formant synthesis
- Articulation synthesis
 - Webster's equation, acoustic radiation
 - Excitation sources
 - Vocal tract models
- Hybrid versions of synthesis
 - PSOLA principle (FD and TD PSOLA)
 - Other hybrid versions

Text-to-Speech Synthesis

- Input processing
 - Pre-processing of text and graphic pages, layout analysis
 - Grapheme to phoneme conversion
 - Linguistic rules on word level
 - Coarticulation aspects
 - Word-level prosody
- Dictionaries for speech synthesis
 - Lexical databases
 - Lexical entries vs. phonological rules
 - Names and abbreviations
 - Transcription of number expressions
- Linguistic processing
 - Syntactic parsing
 - Semantic interpretation
 - Pragmatic aspects
- Prosodic processing
 - Prosodic components (focus, stress, duration etc.)
 - Prosodic phrasing
 - Intonation, the role of the fundamental frequency
- Multilingual text-to-speech synthesis

Applications for Speech Synthesis

- Speech-based communications
 - Voice mail
 - Database-query
 - Voice over Internet
 - Spoken email
 - Spoken traffic information

- Aids for the Handicapped
 - Reading machines and text processing systems for blind persons
 - Speaking aids for speech-impaired persons
- Education
 - Demonstration of artificial speech production and manipulation of relevant parameters like pitch, formant frequencies and prosodic parameters

Pre-Requisites :

Speech analysis
Speech perception and production
Linguistics

Best Practice: The student should have experimental practice in speech segmentation and speech segment concatenation with auditory control. The influence of parameter variations (pitch, amplitude, duration, concatenation manipulations etc.) on the speech quality should be carefully investigated.

Reading Material:

- Keller, E. (Ed.) (1994). *Fundamentals of Speech Synthesis and Speech Recognition*. Chichester: Wiley.
- Santen, J. P. H. van, et al. (Eds.) (1995). *Progress in Speech Synthesis*. Berlin etc.: Springer.
- Dutoit, T. (1997). *An Introduction to Text-to-Speech Synthesis*. Dordrecht: Kluwer, ISBN 0-7923-4498-7
- Sproat, R. (Ed.) (1998). *Multilingual Text-to-Speech Synthesis*. Dordrecht: Kluwer.
- Hess, W. (1983). *Pitch Determination of Speech Signals*. Berlin etc.: Springer.

Title: **Speech Recognition I**

Objectives: After the end of the module the student should be able to define the best technical solution for a simple isolated word recognition problem, to mount the basic components of the solution, either by writing his/her own algorithms or by using available tools and components.

Syllabus:

Introduction (scope and state of the problem, fundamentals, dimensions of difficulty, data-based and knowledge-based approaches)
Pattern comparison techniques and Dynamic Time Warping (DTW)
Hidden Markov Models (HMM)
Isolated Word Speech Recognition -IWSR (DTW applied to IWR, HMM)

based IWR)

Software Tools and databases for IWSR
Evaluation of IWSR speech recognition systems
Application types of IWSR

Pre-requisites :

Introduction to Speech Communication and Speech Technology
(module of level D)

Speech perception and production: (module of level D)

Speech analysis: (module of level D)

Stochastic processes: Probability spaces, random variables, random processes, vector-valued random processes.

Pattern recognition: distance measures, maximum likelihood classification, feature selection and probabilistic separability measures, clustering algorithms.

Information Theory: Definitions, random sources, entropy concepts in pattern recognition.

Reading material:

Basic

L.Rabiner, L. and Jung, B.-H. (1993). *Fundamentals of Speech Recognition*.
Prentice Hall.

Deller, J. , Proakis, J. and Hansen, J. (1993). *Discrete Time Processing of
Speech Signals*. MacMillan Pub.Co..

Markowitz, J. (1996). *Using Speech Recognition*. Prentice Hall.

Advanced

Keller, E. (1994). *Fundamentals of Speech Synthesis and Speech recognition*.
J. Wiley.

Title: **Speech Recognition II**

Objectives: After the end of the module the student should be able to define the best technical solution for a simple continuous speech recognition problem, to mount the basic components of the solution, either by writing his/her own algorithms or by using available tools and components.

Syllabus:

Introduction (scope and state of the problem of Continuous Speech Recognition (CSR))

Continuous Speech Recognition -CSR (DTW applied to CSR, HMM based CSR)

Language modelling in speech recognition

The application of Artificial Neuron Networks in speech recognition
Speaker adaptation and modelling
The speaking environment
Software Tools and databases for CSR
Evaluation of speech recognition and performance analysis
Application types of continuous Speech recognition systems.

Pre-requisites :

Speech recognition I (module of level E)

Neural nets: the artificial neuron, network principles and paradigms, layered networks, the multilayer perceptron, learning vector quantizer.

Advanced informatics and computer science: system's programming, algorithms and complexity, introduction to artificial intelligence, logic programming, real-time systems.

Advanced related modules: speech dialog modelling and human-machine interaction, dictation systems, telephone applications

Reading material:

Basic

L.Rabiner, L. and Jung, B.-H. (1993). *Fundamentals of Speech Recognition*. Prentice Hall.

Deller, J. , Proakis, J. and Hansen, J. (1993). *Discrete Time Processing of Speech Signals*. MacMillan Pub.Co..

Markowitz, J. (1996). *Using Speech Recognition*. Prentice Hall.

Janqua, J.-C. and Haton, J.-P. (1996). *Robustness in Automatic Speech Recognition*. Kluwer Academic Pub..

Advanced

Keller, E. (1994). *Fundamentals of Speech Synthesis and Speech recognition*. J. Wiley.

Title: **Spoken Dialogue Modelling**

Objectives: On completion of this module the student should

- appreciate the communication issues that arise in spoken dialogue systems
- be aware of the approaches taken in some currently available systems to the design of advanced dialogue interfaces
- be familiar with a number of working systems
- be able to develop a functional specification for a spoken dialogue system, based on an understanding of relevant theories of dialogue and communication

Syllabus:

Introduction to spoken dialogue systems

- Basic architectures and components
- Comparisons with natural language interfaces to databases, interactive voice response systems, conversational systems.
- Functions of spoken dialogue systems: information retrieval (databases, Web pages), service provision (email, appointment scheduling), problem solving (route planning, equipment repair).

Theories of dialogue modelling

- Speech acts
- Gricean maxims and theory of co-operative human-machine dialogue
- Conversational Analysis
- Discourse structure theory
- Plan recognition
- Belief models

Case studies: overview and comparison of some representative systems

- Research systems: e.g. SUNDIAL, SpeechActs systems, Circuit Fix-it Shop, TRAINS.
- Commercial systems: e.g. Philips Train Timetable System, Vocalis systems, AT&T's Voice Recognition Call Processing (VCRP) system

Methodologies for development of spoken dialogue systems

- Dialogue development in terms of a software development lifecycle
- Analysis methods: Wizard of Oz studies; scenarios; use of corpora
- Design guidelines, design tools
- Evaluation methods: core metrics for the evaluation of spoken dialogue systems, the PARADISE evaluation method

Dialogue control and dialogue modelling

- State transition networks and dialogue grammars
- Self-organising control
 - Frame-based systems
 - Task and plan-based approaches
 - Conversational agency
 - Event-driven dialogue control
- Confirmation and repair strategies

Toolkits and development environments

- The CSLU toolkit
- The Danish Dialogue Project: Dialogue Description Language
- SpeechMania (Philips)

Future developments and applications

- Dialogue modelling for multi-modal systems
- Spoken dialogue interaction on the Internet
- Multi-lingual systems: spoken language translation
- Dialogue systems for disabled users

Pre-requisites: The student should have a previous experience in Linguistics,

particularly in Pragmatics.

Best Practice: The student should implement a small project during the course, using a toolkit such as CSLU's RAD.

Reading material and Internet support: There are currently no textbooks on spoken dialogue modelling. The following list includes some relevant books, conference proceedings, journal publications, and Internet sites.

Books

Luperfoy, S. (Ed.). *Automated Spoken Dialog Systems*. Cambridge, Mass.: MIT Press.

Roe, D.B. & Wilpon, J.G. (Eds.). *Voice communication between humans and machines*. Washington, D.C.: National Academy Press

Smith, R.W. & Hipp, D.R. (1994). *Spoken Natural Language Dialog Systems: A Practical Approach*, New York: Oxford University Press.

Proceedings

Dalsgaard, P., Larsen, B., Boves, L. & Thomsen, I. (Eds.), *Proceedings of the ESCA Workshop on Spoken Dialogue Systems*, Vigso, Denmark,

LuperFoy, S., Nijholt, A. & Verdhuijzen van Zanten, G. (Eds.), *Proceedings of the eleventh Twente Workshop on Language Technology (TWLT 11): Dialogue Management in Natural Language Systems*, Universiteit Twente,

Maier, E, Mast, M. & Luperfoy, S. (Eds) (1997) *Dialogue Processing in Spoken Language Systems*. Springer Verlag.

Journals publications and book chapters

Allen, J.F., Schubert, L.K., Ferguson, G., Heeman, P., Hwang, C.H., Kato, T., Light, M., Martin, N., Miller, B.W., Poesio, M. & Traum, D.R. (1995). The TRAINS project: a case study in building a conversational planning agent. *Journal of Experimental and Theoretical Artificial Intelligence*, 7, 7-48.

Aust, H., Oerder, M., Seide, F. & Steinbiss, V. (1995). The Philips automatic train timetable information system. *Speech Communication* 17, 249-262.

Bernsen, N.O., Dybkjaer, H. & Dybkjaer, L (1998). *Designing Interactive Speech Systems: From First Ideas to User Testing*. New York: Springer Verlag.

Fraser, N. (1997). Assessment of interactive systems. In D. Gibbon, R. Moore & R. Winski (eds.) *Handbook of Standards and Resources for Spoken Language Systems*. New York: Mouton de Gruyter, 564-614.

Giachin, E. & McGlashan, S. (1997). Spoken Language Dialogue Systems. In S. Young. & G. Bloothoof (eds.) *Corpus-based methods in language and speech processing*. Dordrecht: Kluwer Academic Publishers, 69-117.

Grosz, B.J. & Sidner, C. (1986). Attention, intention, and the structure of discourse. *Computational Linguistics*, 12,3, 175-204.

Shirai, K. & Furui, S. (eds.) (1994). *Speech Communication*. Special Issue on Spoken Dialogue (eds.) Volume 15, No. 3-4, December 1994

Walker, M.A., Litman, D., Kamm, C. & Abella, A. (1997). PARADISE: a general framework for evaluating spoken dialogue agents. *Proceedings of the 35th Annual General Meeting of the Association for Computational Linguistics, ACL/EACL*, Madrid, 271-280.

Internet sites

- <http://sls-www.lcs.mit.edu/SLShome.html> - Spoken Language Systems Group (MIT)
- <http://www.cs.duke.edu/~msf/voicelab/circuit/> - Circuit Fix-it Shop, Duke University
- <http://www.cs.rochester.edu/research/trains/> - TRAINS Project Home Page (University of Rochester)
- <http://www.cse.ogi.edu/CSLU/cslu.html> - CSLU Home Page (Center for Spoken Language Understanding, Oregon)
- <http://www.dfki.uni-sb.de/verbmobil/overview-us.html> - Verbmobil (Large project based in Germany on spoken language and dialogue)
- <http://www.elsnet.org/disc/tools/> - DISC - Spoken Language Dialogue Systems and Components Best practice in development and evaluation
- <http://www.kom.auc.dk/CPK/Speech/> - Center for PersonKommunikation (CPK), Aalborg, Denmark - member of the Danish Dialogue Project
- <http://www.limsi.fr/Recherche/TLP/projects.html> - LIMSI: Projects on spoken language (France)
- <http://www.mip.ou.dk/~laila/index.html/index.html> - Danish Dialogue Project - Guidelines for design of dialogue systems
- <http://www.speech.kth.se/waxholm/waxholm2.html> - Waxholm dialog project (Sweden)
- <http://www.sunlabs.com/research/speech/> - Speech Applications Project (Sun Microsystems)

Title: Language resources

Objectives: Thorough knowledge in the needs and applications of speech and language corpora, design, collection and labelling strategies.

Syllabus:

Introduction

- Application of spoken language resources
- Sources of information
- Speech standards

Specification of the linguistic content

- Types of speech data
- Speakers specifications

Spoken language corpus collection

- Data collection dimensions

- Recording environment
- Interaction and control
- Monitoring and validation
- Multi-channel recording (speech and non speech signals)
- Equipment
- Wizard of Oz

Spoken language representation

- Symbolic representation of speech
- Recording script
- Orthographic transcription
- Morphological, syntactic, semantic and pragmatic representations
- Phonemic representation
- Broad and narrow phonetic transcription
- Acoustic-phonetic transcription
- Prosodic transcription
- Non-linguistic phenomena (filled pauses, speaker and non speaker noises, truncation, etc.)
- Transcription of spontaneous speech
- Transcription of dialogue

Lexical and textual corpus for spoken engineering

- Introduction
- Lexical information (Surface, morphological, grammatical, lexical content)
- Textual corpus for language modelling

Tools

- Grammars
- Software

Available linguistic resources and providers

Pre-Requisites :

Basic statistics

Spoken language technology: continuous speech recognition, speechsynthesis and dialogue systems.

Basis of phonetic and prosody.

Basis on electronic and recording instrumentation

Programming abilities

Reading Material And Internet Support:

Internet

<http://www.icp.grenet.fr/ELRA/home.html>

<http://www ldc.upenn.edu/>

Bibliography

Gibson, D., Moore, R. and Winski, R. (Eds.) (1997). *Handbook of Standards and Resources for Spoken Language Systems*. Mouton de Gruyter, 1997.

ISBN: 3-11-0153661, <http://www.degruyter.de/EAGLES/degruyt/eagbook.html>

3.4 Applications

LEVEL F

Level F is primarily thought as a collection of examples for a useful application of speech technology. Starting with a short summary of speech processing forms (coding, recognition, synthesis) and today's state of the art, it mainly aims at aspects of human factors and ergonomics. It is well known that the number of real (commercially used) applications of voice technology is far behind the expectations of the market, although voice systems now have a sufficient state of maturity. It is, therefore, very important to discuss reasons for hindrances or even refusals and, on the other hand, to collect convincing applications showing that voice technology is very advantageous in comparison to 'conventional' techniques.

Finally, the speech (or more general the acoustic) component in connection with visual information (graphics, still and moving pictures), i.e. the multimedial aspect is discussed. It will be obvious that speech information can support visual presentations and vice versa and that there is a high synergy between both communication forms.

Syllabus:

Introduction

Characteristics of voice communication, some aspects of man-man- and man-machine communication, state of the art in speech recognition, synthesis and coding (summary only).

Application scenarios

- Office (dictation systems, acoustic dialling,...)
- Industrial environments (quality control, parcel sorting, robot control ...)
- Public transportation (luggage sorting, information and emergency announcements on railway stations and airports)
- Bank transactions based on speech and speaker recognition
- Medical areas (speech enrolment, voice-controlled systems like operation microscopes)
- Aids for the Handicapped, rehabilitation engineering
- Voice therapy
- Smart home applications

Multimedia aspects

- Characteristics and comparison of communication forms (speech, graphics, still and moving pictures)
- Interworking and synergy effects between different communication forms

Multimedia Applications Focused on Acoustic Components

- Design of a multimedial tutorial in speech processing. This could be an examination or seminar work.
- Multimedia database access
- Virtual shopping
- Multimedial Internet applications

Pre-Requisites: The module requires deep knowledge in most of the areas of speech processing. It should be one of the last courses before the Master's degree.

Best Practice, Reading Material And Internet Support: Information sources are proceedings of scientific conferences (ICASSP, EUROSPEECH) or commercial workshops (e.g. SPEECH TECHNOLOGY). In general, detailed descriptions of speech-based applications are difficult to find (this has mainly commercial reasons) and very often they have an advertising touch. It is recommended to browse through technical magazines and the Internet. Finally, there are scientific and technical reports on speech projects and some of them are freely available (not confidential).

3.5 Introduction to current research in SLE

LEVEL G

Level G is proposed as advanced-level education, e.g. master's work, starting point for a PhD thesis or other research work. Lectures or other teaching activities (from the teaching staff) are widely replaced by a supervised independent research-oriented work combined with seminaristic presentations.

The student (or doctorand) has to prove that he or she is able to cope with a new and scientifically demanding task in the spoken language area by analysing scientific literature (books, proceedings, reports) and developing new ideas. The student should present papers on national and international conferences, should visit other (preferably European) research institutions and apply for scientific collaboration.

Syllabus:

Areas of current research (selection)

Hybrid models (HMM plus Neural Nets)

New structures of Neural Nets

Augmental models

Semantic analysis for speech recognition

Semantic text analysis for speech synthesis

Robust speech recognition, noise reduction strategies
Advanced studies in man-man- and man-machine dialogs
Recognition of spontaneous speech
Automatic speech translation
Multimodal dialogs
High-quality speech and audio coding using psychoacoustic phenomena
Speech recognition based on voice and visual information (gestures, mimic, lip reading etc.)
Aids for the Handicapped

- Blind persons
- Reading machine with advanced text and graphic pre-processing of the input pages
- Text processing systems using speech synthesis, natural speech and acoustic sounds for error correction; design of acoustic cursors
- Deaf and hard of hearing persons
- Adequate visual representation of spoken utterances
- Speech recognition of disordered speech combined with speech synthesis
- Automatic sign language recognition and speech synthesis output
- Speech and language therapy applications

Pre-Requisites: The module requires an extended knowledge of all spoken language subjects, mainly in spoken language engineering. It should be one of the last courses or a post-master course.

Suggested Reading: Depends on the special subject. As a general suggestion, a careful reading of the proceedings of recent conferences (above all, ICASSP and EUROSPEECH) is recommended. Additionally, an Internet search is useful in order to find research institutions, to make contacts with experts and to get relevant text material.

References

Bloothoof, G., van Dommelen, W., Espain, C., Green, P., Hazan, V., Huckvale, M. and Wigforss, E. (Eds.) (1997). *The Landscape of Future Education in Speech Communication Sciences. 1. Analysis*. Utrecht: OTS publications.