



INESC-ID
Instituto de Engenharia
de Sistemas e Computadores
Investigação e Desenvolvimento
em Lisboa

Rua Alves Redol, 9
1000-029 Lisboa Portugal
Tel. +351.213100300
Fax: +351.213145843
Email: info@l2f.inesc-id.pt
www.l2f.inesc-id.pt/~imt/lectra/

Spoken language Systems Lab (L2F) Project LECTRA



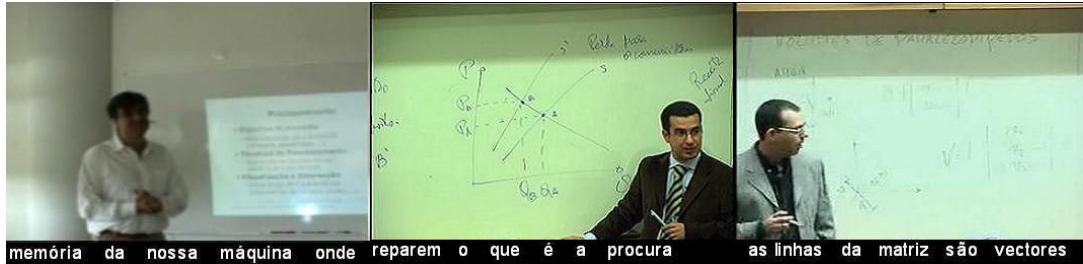
Producing automatic transcriptions of classroom lectures may be important for both e-learning and e-inclusion purposes.

The greatest research challenge is the recognition of spontaneous speech (error rate much higher than for read speech). Even human produced transcriptions would be very difficult to understand because of the absence of punctuation and the presence of disfluencies (filled pauses, repetitions, hesitations, false starts, etc.). Hence, one has to enrich the speech transcription by adding information about sentence boundaries and speech disfluencies.

The goal of the national project LECTRA - Rich Transcription of Lectures for E-Learning Applications - is the production of multimedia lecture contents for e-learning applications. We shall take as a pilot study courses for which the didactic material (e.g. text book, problems, viewgraphs) is already electronically available and in Portuguese. This is an increasingly more frequent situation, namely in technical courses. Our contribution to these contents will be to add, for each lecture in the course, the recorded video signal and the synchronized lecture transcription. We believe that this synchronized transcription may be especially important for hearing-impaired students.

The project will encompass 5 main tasks. In the first one we shall collect the training and test material (in terms of recorded audio-video signals and textual data) related to this course. In the second task we shall use this training data to adapt the acoustic, lexical and language models of our large vocabulary continuous speech recognizer to the course domain. The third task has as a goal to "enrich" this transcription with punctuation and structural metadata (e.g. marking sentence boundaries, disfluencies) that would render it more intelligible. The fourth task deals with integrating the recorded audio-video and corresponding transcription with the other multimedia contents and synchronize them according to topic, so that a student may browse through the contents, seeing a viewgraph, the corresponding part in the text book, and the audio-video with the corresponding lecture transcription as caption. The final task is user evaluation for which we intend to use a panel of both normal hearing and hearing impaired students.

Laboratório de sistemas de Língua Falada (L2F) Projecto LECTRA



A produção automática de transcrições de aulas pode ser importante tanto para aplicações em e-learning como e-inclusion (alunos com surdez).

O grande desafio do ponto de vista de investigação é o reconhecimento de fala espontânea (taxa de erro muito superior à obtida para fala lida). Por um lado, há que lidar com disfluências (pausas preenchidas, repetições, hesitações, etc.). Por outro lado, há que enriquecer a transcrição com pontuação (detectando automaticamente as fronteiras de frase), e capitalização.

Mais informações em:
www.l2f.inesc-id.pt

BOLSA DE INVESTIGAÇÃO

Encontra-se aberto concurso para atribuição de uma Bolsa de Investigação no âmbito do projecto POSC/PLP/58697/2004, designado por LECTRA - Transcrição de Aulas para Aplicações de E-Learning cofinanciado pela Fundação para a Ciência e Tecnologia e pelo FEDER através do Programa POS_C.

Duração: 5 meses, com início previsto para **1 de Agosto de 2007**, em regime de exclusividade.

Orientação Científica:

Prof. Isabel Trancoso, IST / INESC-ID

e-mail: Isabel.Trancoso@inesc-id.pt

Formação Académica:

Os interessados devem ser preferencialmente estudantes finalistas, Licenciados, Mestres ou Doutorados em Engenharia Electrotécnica e de Computadores, Engenharia Informática e de Computadores ou Linguística. Será dada preferência a candidatos que tenham conhecimentos e experiência na área do processamento da fala e/ou que queiram fazer doutoramento nesta área.

Remuneração: De acordo com a tabela de valores das Bolsas de Investigação no país atribuídas pela FCT.

As candidaturas devem incluir um curriculum vitae do candidato.