





L²F — Spoken Language Systems Laboratory Brochure Contents

About us

Research Areas

- Semantic Processing of Multimedia contents
- Spoken / Multimodal Dialogue Systems
- Speech-to-Speech Machine Translation

Core Technologies

- AUDIMUS Portuguese Speech RecRecognizer
- DIXI Portuguese Text-to-Speech Synthesizer
- FACE Synthetic Talking Face
- Natural Language Processing Tools

On-Going Projects

- TECNOVOZ
- eCIRCUS
- VIDI-Video
- **LECTRA**

Researchers





L²F Spoken Language Systems Laboratory



INESC-ID, which stands for "Institute for Systems and Computer Engineering: Research and Development", is a non-profit institution dedicated to research in the field of information and communication technologies. Its mission is to develop tomorrow' technologies by excelling in research, today.

INESC-ID initiated activity in the year 2000 as a result of a reorganization of INESC. In 2004, INESC-ID was recognized by the Portuguese government, as an "Associated Laboratory", with intense activity in 5 thematic areas: Information and Decision Support Systems, Interactive Virtual Environments, Embedded Electronic Systems, Communications and Mobility Networks, and Spoken Language Processing. The latter area of activity is carried out at L2F. The L2F - Spoken Language Systems Laboratory is a lab created in 2001, bringing together several research groups of INESC and independent researchers that could potentially add relevant contributions to the area of computational processing of spoken language for European Portuguese.

The long term goal of L2F is to bridge the gap between natural spoken language and the underlying semantic information.

The main areas of activity of L2F are: semantic processing of multimedia contents, spoken/multimodal dialogue systems and speech-to-speech machine translation. The original focus on European Portuguese has now been broadened to encompass all varieties of Portuguese. Two application areas also deserve our special attention: technology enhanced learning and e-inclusion, namely in the development of alternative and augmentative communication tools for people with special needs.

The lab includes around 30 researchers, most of them either Professors (8) or graduate students at IST, and invited researchers from other universities in Portugal. Their background ranges from Electrical Engineering to Computer Science and Linguistics.

This strongly interdisciplinary group is actively involved in many core areas of spoken language research and development, including speech recognition, speech synthesis, speech coding, speech understanding, audio indexation, multimodal interfaces, language and dialect identification, among others. By bringing together researchers from the area of natural language processing, the lab acquired expertise in areas such as natural language database interfaces, natural language generation, alternative syntactic and semantic processing paradigms, etc.



L²F: Spoken Language Systems Laboratory

Team	PhD Diamantino A. Caseiro Luísa Coheur Nuno Mamede David Martins de Matos João Paulo Neto Luís C. Oliveira António J. Serralheiro Isabel M. Trancoso Christian Weiss Research Associates Renato Cassaca Hugo Monteiro Carlos Mendes Luís Neves Márcio Viveiros Luís Figueira Gustavo Coelho Helena Moniz	PhD Students Rui Amaral Gracinda Carvalho Fernando Batista Porfírio Filipe João Graça Ciro Martins Hugo Meinedo Cristina Mota Sérgio Paulo Joana Paulo Pardal Ricardo Ribeiro Paula Vaz Invited Researchers M. Céu Viana (CLUL) M. Isabel Mascarenhas (CLUL) A. Isabel Mata (CLUL) Jorge Baptista (UALG) Gaël Dias (UBI)
Ongoing Projects	 International ECIRCUS - Education playing Capabilities th VIDI-Video – Interacti machine learned audio COST 2102 - Cross-Ma (2006-2010) COST 2103 - Advanced ECESS - European Cernotational Tecnovoz (2006-2008) RiCoBA- Rich Content 	through Characters with Emotional Intelligence and Role- at Understand Social Interaction (2006-2008) ve semantic video search with a large thesaurus of -visual concepts (2007-2010) odal Analysis of Verbal and Non-verbal Communication d Voice Function Assessment (2006-2010) nter of Excellence on Speech Synthesis
	 LECTRA - Rich Transo 2007) NLE GRID - Natural L 2007) WFST - Weighted Fini Processing (2004-2006) DIGA - Dialog Interface Bilateral contracts LIFAPOR - Spoken Box 	cription of Lectures for E-Learning Applications (2005- anguage Engineering on a Computational GRID (2005- te State Transducers Applied to Spoken Language 5) ce for Global Access (2004-2006) poks in European and Brazilian Portuguese
Highlights	Some important landmarks in L speech synthesizer build from s first version of our large vocabi our language in 1997. Our first in 2005. This was also the year conference on spoken language award for the development of E with cerebral palsy.	² F's activity were the development of the first text-to- acratch for European Portuguese (DIXI) in 1991, and of the ulary continuous speech recognition system (AUDIMUS) for speech-to-speech machine translation system was ready in which we organized the most important international communication (INTERSPEECH'2005), and won a merit EUGÉNIO, a word predictor specially suited for children th the national industry: Vodafone. Portugal Telecom
More	Microsoft Portugal, Rádio-Tele	visão Portuguesa (RTP), Porto Editora, Texto Editora.
In Carry C.		—

or directly from the website http://www.l2f.inesc-id.pt/.



L²F – Spoken Language Systems Laboratory Semantic Processing of Multimedia contents



Nowadays there is a significant need to deal with large amounts of multimedia information. The use of advanced techniques for their segmentation, transcription and indexation makes it possible to access their contents.

In one of the services derived from this project, we give users the possibility of searching through the 8 o'clock broadcast news of RTP (*Telejornal*) for selected topics. Moreover, users may define which thematic areas they are interested, and receive at the end of the automatic processing of the whole broadcast news, an email alerting them to the news on their chosen topics. **Goal.** Use of audio, speech and language processing techniques for segmentation, transcription and indexation of multimedia data.

Summary. This activity represents a large framework for research and development in the area of semantic processing of multimedia information. It combines different audio, speech and language processing techniques in pipeline architecture in order to segment the multimedia data into homogeneous chunks, recognize them and classify them into topics. This process leads the way to a set of advanced applications such as selective dissemination of information, speech mining, and audio browsing.

Description. The complex pipeline process starts by an audio categorization and segmentation stage that divides the audio stream into coherent blocks, in terms of absence/presence of speech, acoustic background conditions and speaker characteristics. The blocks classified as containing speech are then fed through a large vocabulary continuous speech recognition system (*AUDIMUS.MEDIA*) that provides the corresponding transcription. This textual transcription and associated metadata derived from the segmentation stage is the input to an automatic topic indexation stage which classifies each block according to its topic and clusters contiguous blocks with the same topic. The overall process is a pipeline combination of different processing techniques that share a common XML structure description. At the end of the process, the multimedia document is available to be loaded into a database together with the associated XML information.

More information is available by email to <u>info@l2f.inesc-id.pt</u> or directly from the website <u>http://www.l2f.inesc-id.pt/</u>.



L²F: Semantic Processing of Multimedia contents

Tasks	The project is structured into the following tasks:			
	 Multimedia Database: identification, acquisition, definition and storage Acoustic data description and segmentation Automatic speech recognition Automatic update of vocabulary and language modeling Analysis of spontaneous speech Rich transcription Data block segmentation based on contents Data characterization based on transcription and indexation Data summarization Application interface and services 			
Team	João P. Neto, PhD Isabel Trancoso, PhD Hugo Meinedo, PhD Student Rui Amaral, PhD Student Ciro Martins, PhD Student Luís Neves, Eng.			
Applications	 This project is being used as a research and development platform for different applications: -Characterization of Broadcast News programs for selective dissemination of multimedia information -Automatic sub-titling of Broadcast News programs -Classroom lecture transcription -Meeting transcription -Court session transcription -News distribution service for mobile devices 			
Available Demo	In order to demonstrate the results and the potentialities of this project, a demo of a selective dissemination of information system associated to the 8 o'clock news program of RTP (<i>Telejornal</i>) was made available. Daily the news program is automatically collected directly from the cable network, loaded into a database, acoustically described and segmented, transcribed, segmented into news stories, and each story indexed in a set of topics. At the end of the process, the description data is also loaded into the database. This loading triggers a search process on the user profiles for the ones matching the same topics, and an alert message is sent to the selected users. A service interface makes it possible for any registered user to define which thematic areas they are interested in. After the automatic processing of the daily news broadcast, users with matched topics receive an email with the news summary and an indication of the website in which they can have access through streaming to the news story about the selected topics. The service is available from <u>http://ssnt.l2f.inesc-id.pt/</u> .			
More	By email to <u>info@12f.inesc-id.pt</u> or directly from the website http://www.12f.inesc-id.pt/.			



L²F – Spoken Language Systems Laboratory Spoken / Multimodal Dialogue Systems



Human-computer interface is an area where speech, integrated into a multimodal structure, creates a natural and easy way to establish communication. **Goal.** Building a platform for research and development of spoken dialogue systems integrated into a multimodal user interface.

Description. This is one of our mainstream lines of activity. We have been developing competences in several of the core technologies for Spoken Dialogue Systems, as Automatic Speech Recognition (AUDIMUS), Text-to-Speech (DIXI+), Synthetic Talking Face (FACE) and Dialog Management (DM). These technologies have been integrated into an Embodied Conversational Agent (ECA) with a unique and sophisticated user interface.

Applications. The spoken dialogue platform has been applied in different scenarios, illustrating the flexibility of our modular system:

- home environment: control of a wide range of home devices, such as, lights, air conditioning, hi-fi, and TV, based on X10 and IRDA protocols. We can extend the application to include any infra-red controllable device or any device whose control functions may be programmed by the X10 protocol;
- database access: allowing the user to inquire databases via voice. We have developed prototypes to request weather information, cinema schedules and bus trip information. This type of application can easily be extended to other domains;
- (iii) email access: through telephone queries based on a dynamic VoiceXML.

More information is available by email to <u>info@l2f.inesc-id.pt</u> or directly from the website <u>http://www.l2f.inesc-id.pt/</u>.

The main areas of activity of L²F are: semantic processing of multimedia contents, spoken/multimodal dialogue systems and speech-to-speech machine translation. Complementing these mainstream areas we are actively involved in many core areas of spoken language research and development.



L²F: Spoken / Multimodal Dialogue Systems



Nuno J. Mamede, PhD João P. Neto, PhD Luís Oliveira, PhD Isabel Trancoso, PhD Luísa Coheur, PhD David Matos, PhD Diamantino Caseiro, PhD

Joana Paulo, PhD Student Porfírio Filipe, PhD Student Márcio Mourão (got the M.Sc on 2005) Renato Cassaca, Eng. Márcio Viveiros, Eng. Filipe Martins, Student Pedro Arez, Student



The central component of the spoken dialogue system architecture is the *Dialogue Manager* which integrates several modules:

- *The Interpretation Manager* (IM) receives a set of speech acts from the Language Comprehension module and generates the correspondent interpretations and discourse obligations. Interpretations are frame instantiations that represent possible combinations of speech acts and the meaning associated to each object it contains. To select the most promising interpretation two scores are computed. The recognition score to evaluate the rule requirements already accomplished, and the answer score, a measure of the consistency of the data already provided by the user.
- *The Discourse Context* (DC) manages all knowledge about the discourse, including the discourse stack, turn-taking information, and discourse obligations.
- *The Behavioral Agent* (BA) enables the system to be mixed-initiative: regardless of what the user says, the BA has its own priorities and intentions. When a new speech act includes objects belonging to a domain that is not being considered, the BA assumes the user wants to introduce a new dialog topic: the old topic is put on hold, and priority is given to the new topic. Whenever the system recognizes that the user is changing domains, it first verifies if some previous conversation has already taken place.
- *The Generation Manager* (GM) receives discourse obligations from the BA, and transforms them into text, using template files. The GM uses another template file to produce questions that are not domain specific. For example, domain disambiguation questions, used to decide proceeding a dialogue between two or more distinct domains, or to clarify questions, are defined in this file.
- The *Service Manager* is the interface between the spoken dialogue platform and a set of heterogeneous domain devices. Each device description is composed by slots and rules. Slots define domain data relationships, and rules define the system behavior. A rule or service represents an user possible action.

More Information

By email to <u>info@l2f.inesc-id.pt</u> or directly from the website <u>http://www.l2f.inesc-id.pt/</u>.



L²F – Spoken Language Systems Laboratory Speech-to-Speech Machine Translation



Spoken language translation provides a means for communication using the speaker native language

您们 是受欢迎 的对 葡萄牙

Welcome to Portugal

Benvindos a Portugal

The main areas of activity of L²F are: semantic processing of multimedia contents, spoken/multimodal dialogue systems and speech-to-speech machine translation. Complementing these mainstream areas we are actively involved in many core areas of spoken language research and development. **Goal.** This is one of our emerging areas where we are building a platform for research and development of spoken language translation.

Description. Speech-to-Speech translation attempts to cross the language barriers between people speaking different languages. It provides a means to communicate using their native language, ideally using their own voice. Spoken language translation has to deal with hard problems on automatic speech recognition, machine translation and text-to-speech synthesis.

Our research on spoken language translation has focused on:

- Tight integration of speech recognition with machine translation to provide robustness against speech recognition errors;
- Cross domain adaptation, so that a system trained on parliamentary debates can be adapted to translate broadcast news;
- Using new knowledge sources, such as syntactic and semantic information, to improve machine translation quality.

Weighted Finite-State Transducers provide a unifying framework for integrating spoken language technologies and new knowledge sources. Our work in this area was initiated in the scope of national project WFST – *Transdutores de Estado Finito Ponderados Aplicados ao Processamento da Língua Falada*.

More information is available by email to <u>info@l2f.inesc-id.pt</u> or directly from the website http://l2f.inesc-id.pt/.

Try our demonstrations on the website.





L²F – Spoken Language Systems Laboratory Digital Talking Books



Digital Talking Books is a spreading commodity amongst users. They provide fast access to the audio or written texts of multimedia documents.

As an output of this activity, L2F has a platform to time align text and speech files in one passage, faster than real time. Audio files can be reasonably long (over 2 hours), avoiding partitioning the narration into smaller files. **Goal.** To allow Digital Talking Books (DTBs) to be used by a wider audience of users, by providing a browsing interface for navigation in multimedia documents. DTBs are thus an important tool for e-learning and e-inclusion.

Summary. This activity is the result of two nationally funded research projects, IPSOM (form Nov 2000 to Nov 2004) and RICOBA (still undergoing) that started with a consortium with L2F (coordinator), LASIGE of the Faculty of Sciences of the University of Lisbon and the National Library. The undergoing idea is to have the audio and text files synchronized at the word/topic/sentence level, thus allowing the two instances of a book (spoken and written) to be accessed simultaneously. This concept was further extended to encompass different varieties of Portuguese (project LIFAPOR with UFRGS, Brazil) and also parallel texts in other languages.

Description. Audio Books (AB) started early as vinyl or cassette recordings. However, looking for any information in these recordings was, at least, a trial-and-error experience. Computers, with their powerful storage, indexing and retrieval capabilities, were the obvious choice for AB, thus transforming them into Digital Talking Books, by the inclusion of metadata. Forced-alignment speech recognition technologies and Weighted Finite State Transducers, allowed us to synchronize (time align) simultaneously the audio and the text versions of the DTBs. The audio version of a DTB, previously recorded and manually edited to remove reading errors, is time aligned against its written version in less than real-time. A multimodal navigation and information retrieval interface, plus the audio and text files comprise the DTBs that can be further complemented with short videos or other multimedia data.

More information is available by email to <u>info@l2f.inesc-id.pt</u> or directly from the website <u>http://www.l2f.inesc-id.pt/</u>.





L²F – Spoken Language Systems Laboratory AUDIMUS – Portuguese Speech Recognizer



Automatic Speech Recognition (ASR) systems transform a speech message into text. It is a significant and powerful tool. Increases productivity in a large number of situations and gives the possibility of new applications.

The domains where we are applying our system require different capabilities from the ASR. There are simple actions of command and control, as turning on/off the lights or dialing automatically a telephone number, requiring small vocabularies of less than 1.000 words; medium applications of dictation, as writing reports, letters or emails, requiring vocabularies of 20.000 words; and difficult tasks as transcription of broadcast news, as the 8'o clock news, requiring vocabularies of 100.000 words. For all these tasks AUDIMUS is the appropriate solution.

Goal. Build a platform for research of new techniques and development of new applications for Automatic Speech Recognition in the Portuguese language.

Summary. AUDIMUS is the name of a generic platform for an Automatic Speech Recognition System specifically tailored to the European Portuguese language. This platform is used as a research base for new techniques of the different components of a speech recognition system. The improvements that result from the research work are fully integrated in AUDIMUS in order to give rise to new and better applications.

Description. *AUDIMUS* is a hybrid speech recognizer that combines the temporal modeling capabilities of Hidden Markov Models (HMMs) with the pattern discriminative classification capabilities of multilayer perceptrons (MLPs). This same recognizer is being used for different complexity tasks based on a common structure but with different components.

MLPs are used to estimate the context-independent posterior phone probabilities given the acoustic data at each input frame. The phone probabilities generated at the output of the MLP classifiers are combined using an appropriate algorithm. The acoustic models are dependent on the input facilities since we are using separated models for telephone speech, for microphone speech or broadcast news. The same occurs with the lexical and the language models which are dependent on the specific application domain.

The AUDIMUS decoder is based on a weighted finite-statetransducer (WFST) approach to large vocabulary speech recognition.

More information is available by email to <u>info@l2f.inesc-id.pt</u> or directly from the website <u>http://www.l2f.inesc-id.pt/</u>.



L²F: AUDIMUS – Portuguese Speech Recognizer

Team	João P. Neto, PhD Diamantino Caseiro, PhD Hugo Meinedo, PhD Student Ciro Martins, PhD Student Renato Cassaca, Eng. Márcio Viveiros, Eng.
Availability	 Generically AUDIMUS is available under three concepts: AUDIMUS.API – An Application Program Interface (API) available in JAVA and C++ under a DLL and LIB formats to be integrated in different applications. AUDIMUS.SERVER – A standalone server package accessible through RMI with interfaces in MRCP and JAVA. AUDIMUS.TRAIN – A package with several tools for training acoustic and language models, vocabulary selection and lexicon building. AUDIMUS is configurable through XML files for the architecture description and parameters initialization. Previous concepts are available for MSWindows and Linux OS.
Applications	 Through our work we have been developing different acoustic models, different vocabularies, different language models and applying the system to different tasks. Based on this work we created four main applications: <i>AUDIMUS.DICTATE</i> – A dictation system to be used in MSWindows and connected to MSWord. This system makes use of <i>AUDIMUS.AP1</i> and <i>AUDIMUS.TRAIN</i> tools to make adaptation of the acoustic models to the user and to use text files to adapt the system to specific tasks. Vocabularies are normally from medium size 10.000-30.000 words. <i>AUDIMUS.MEDIA</i> – A special architecture of AUDIMUS to automatically transcribe multimedia data with focus in broadcast news programs. This architecture is based on a parallel combination of different acoustic models trained on a different set of acoustic features and combined using an appropriate algorithm. Large vocabularies ranging from 50.000-100.000 words. <i>AUDIMUS.TELEPHONE</i> – A system trained for telephone speech and for small vocabularies tasks of 1.000 words. The system works on a small fraction of real-time and is easily scaled to a large number of instantiations. <i>AUDIMUS.MOBILE</i> – Specifically tailored for small vocabularies and for mobile devices as PDAs using a remote <i>AUDIMUS.SERVER</i> and a very light interface in the mobile device only for internet access to the remote server. These are just a few of possible applications rising from AUDIMUS. We are available for other ideas, just contact us.
Available Demo	In order to show the potentialities of <i>AUDIMUS</i> a demo using the <i>AUDIMUS.SERVER</i> and a web page application is available. It is possible to choose the task and using a local microphone the user could speak and see immediately the result on a text editor. For other applications it is possible to supply a way file to be recognized off-line.
More	By email to <u>info@l2f.inesc-id.pt</u> or directly from the website http://www.l2f.inesc-id.pt/.



L²F – Spoken Language Systems Laboratory DIXI - Portuguese Text-to-Speech Synthesizer



The synthesizers developed at L2F have been integrated in a wide range of applications: a dialogue system for home automation, a word predictor for children with special needs, a search engine for Portuguese, new voices for synthetic characters in children's games, etc.

The DIXI synthesizer resulted from a joint effort with the Center of Linguistics of the University of Lisbon (CLUL). This interdisciplinary team built the first synthesizer from scratch for European Portuguese in 1991. DIXI was a rule-based formant synthesizer, unlike its successor,.DIXI+, which is a concatenative-based synthesizer. **Goal.** Text-to-speech conversion for all applications requiring synthetic voices.

Description. DIXI+ is a concatenative-based text-to-speech synthesizer that may cope with either fixed-length (diphone) or variable-length segments of pre-recorded voices. It is based on the Festival Speech Synthesis System, a free software toolkit.

L²F's research in text-to-speech synthesis involves several trends:

- Development of tools to automate the process of creating new voices for both limited and unlimited domain synthesizers.
- Expressive speech synthesis Adding the ability to synthesize emotions, and mimic different speech styles becomes important for an increasing number of applications.
- Voice transformation Voice customization (with a minimum amount of recordings) is important not only for speech-to-speech translation systems but also to provide specialized voices for synthetic characters.
- Audio-visual synthesis Our research in this area has been aimed at developing talking faces, and also at developing time-scale transformations for lip synchronization in film dubbing applications.
- Flite (Festival light) Our group has financially supported the development of this small footprint synthesis engine.

More information is available by email to <u>info@l2f.inesc-id.pt</u> or directly from the website <u>http://www.l2f.inesc-id.pt/</u>.



L²F: DIXI – Portuguese Text-to-Speech Synthesizer

Team	Luís C. Oliveira, PhD Christian Weiss, PhD M. Céu Viana (CLUL), PhD Isabel Trancoso, PhD Sérgio Paulo Carlos Mendes Luís Figueira
Applications	 EUGÉNIO – One of the first applications to integrate DIXI+ was a word predictor, specially developed for children with cerebral palsy. The system resulted from a joint effort with ESTIG and CPCB, and worn a merit award in 2005. This public domain software has been downloaded by many different users not only in Portugal but also across the Atlantic, and is also very often used by children with dyslexia. SVIT (Vocal Service for Information Systems) – The very first contract of the speech processing group involved building a system for TLP (Telefones de Lisboa e Porto) that synthesized telephone numbers with perfectly natural quality, thus being our first demo of limited domain synthesis. TUMBA (Search engine for the Portuguese Web) – Following a request from the developers of TUMBA, we added a voice to this search engine, allowing the user to click on each word of the search terms and listen to how they sound. AMBRÓSIO – The virtual butler of the Home of the Future exhibit at Museu das Comunicações was one of the first synthetic characters to speak Portuguese.
International Projects	 ECIRCUS – This European project investigates educational role-play using autonomous synthetic characters and involving the child through affective engagement, including the use of standard and highly innovative interaction mechanisms. Synthetic characters express emotions in two different game scenarios: one on anti-bullying education, and another one on intercultural empathy. ECESS (European Center of Excellence on Speech Synthesis) - As partner in this network where major industry partners such as IBM, Siemens and Nokia as well as European key research institutes are connected, L²F plays an active role in the development and research activities. We are responsible for the intelligibility and naturalness of the synthesized utterances while developing the acoustic module of the speech synthesis software. ECESS activities are evaluated by the global research community in regular evaluation campaigns. COST Action 2103 (Advanced Voice Function Assessment) - The main objective of the Action is to combine previously unexploited techniques with new theoretical developments to improve the assessment of voice for as many European languages as possible, while acquiring in parallel data with a view to elaborating better voice production models. Progress in the clinical assessment and enhancement of voice quality requires the cooperation of speech processing engineers and laryngologists as well as phoniatricians. Specifically, this Action is a joint initiative of speech processing teams and the European Laryngological Research Group (ELRG).
More Information	By email to <u>info@l2f.inesc-id.pt</u> or directly from the website <u>http://www.l2f.inesc-id.pt/</u> .



L²F – Spoken Language Systems Laboratory FACE - Synthetic Talking Face



Human Computer Interface is an area where audio, text, graphics, and video are integrated to convey various types of information. The goal is to provide a more natural interaction between the user and the computer. One approach to achieve this goal is to display an animated character on the computer screen, with the ability to make head movements, facial expressions and emotions.

> We developed a Synthetic Talking Face system that is closed integrated with base technologies as Automatic Speech Recognition, Text-to-Speech and Natural Language Processing on generic Spoken Dialogue Systems platform.

Goal. Developments of a synthetic talking face to be integrated in an Embodied Conversational Agent platform.

Summary. Human-computer conversation is a broad research goal that is starting to be implemented through a new genre of Embodied Conversational Agents (ECA). We have been working on base technologies, as Automatic Speech Recognition (ASR), Text-to-Speech (TTS) and Natural Language Processing (NLP). In order to create an ECA platform we developed a Synthetic Talking Face that it is closed integrated with these base technologies. When associated with speech, the overall facial expressions constitute one of the most important communication channels used in human interactions. This ECA platform is serving as the interface to generic Spoken Dialogue Systems.

Description. Facial expressions allow the exposure of these emotions that play an important role in the context of human communication. The ability of expressing feelings like sadness, happiness or hanger allows the machine to emulate human emotions, acquiring capabilities only seen in humans, and bringing more reality to human-computer interactions. Also the synchronized movement of lips increases considerably the perception of the synthetic speech. This synchronization is based on a set of time stamps resulting from the TTS. The system in order to be able react to the environment receives from the ASR a set of acoustic features that represents the awareness of that environment, as silence, speech, music, noise, etc.

More information is available by email to <u>info@l2f.inesc-id.pt</u> or directly from the website <u>http://www.l2f.inesc-id.pt/</u>.



L²F: FACE - Synthetic Talking Face





L²F – Spoken Language Systems Laboratory Natural Language Processing Tools



Language processing can take advantage of several tools, such as syntactic and semantic analyzers. In order to perform their tasks, some of these tools use linguistic information (for instance, dictionaries and grammars), making natural language processing by computers closer to the human process.

We are using natural language processing tools in many of our applications, namely in dialog management, automatic summarization, information retrieval, question answering, discourse analysis, term and emotion extraction. Besides applying these tools to text we are applying them to automatic transcriptions of spoken documents, leading to new challenges. **Goal.** Build a library of state of the art natural language processing tools, that can be used in many different applications.

Description. Natural Language processing tools can be grouped as:

- (a) Morphological Tools are responsible for the first steps in the processing chain, such as splitting sentences, detecting composed terms (e.g. "guarda-chuva", "Presidente da República") and classify or disambiguate words (e.g. "canto": a noun or a verb? "a": a preposition or an article?). Many of these tools take advantage of dictionaries and hand-crafted or data-driven rules;
- (b) Syntactic Tools parse texts and return sentences organized in phrases (e.g. nominal or verb phrases) or dependency structures relating words. Typically they use grammars that can be inferred or built by linguistic experts;
- (c) Semantic Tools perform the last step that enable a sentence to be understood by a computer. This step uses the syntactic information previously obtained and the semantic information associated with words and/or groups of words. These tools generate a formal representation (such as a frame or a formula in some logic) for each sentence.

These basic tools can be integrated in complex architectures to achieve more sophisticated tools, such as, for instance, identifying who is talking in children stories.

More information is available by email to <u>info@l2f.inesc-id.pt</u> or directly from the website <u>http://www.l2f.inesc-id.pt/</u>.



L²F: Natural Language Processing Tools

Team

David Matos, PhD Nuno J. Mamede, PhD Luísa Coheur, PhD Jorge Baptista, PhD Ricardo Ribeiro, PhD Student Joana Paulo, PhD Student Fernando Batista, PhD Student João Graça, PhD Student

Tools N

(Grouping is more or less ad-hoc, but intends to reflect the level of processing for which the tool is useful) Morphological Tools

MARv - a morphossyntactic disambiguation tool;

monge - a word form generator;

PAsMo - a rule-based morphology processor, tag converter, and sentence splitter;

RuDriCo - a rule-based morphology processor;

SMorph - a morphological analyser;

XA - a morphological analyser similar to ispell and jspell;

YAH - yet another hyphenator.

Syntactical Tools

ParVO - a C++ implementation of Earley's algorithm with attribute unification;

SuSAna - a chunk analyzer;

TiraTeimas - verifies if a set of chunks satisfies a set of constraints;

Algas - establishes dependency relations between chunks and words.

Syntax/Semantics Interface

Ogre - transforms a structure where both chunks and words are connected into a dependency structure;

AsDeCopas - applies contextual rules (possibly hierarchically organized) to a graph.

Discourse Analysis

DID – identifies indirect and direct speech in children stories. It also attributes a character to each direct speech utterance.

Multi-purpose

Galinha - a portal for building and running applications;

LRDB - a language resources database and access framework;

FSTk - a finite-state transducer library;

ShReP - a framework for simplifying the process of constructing NLP systems.

More Information

By email to <u>info@l2f.inesc-id.pt</u> or directly from the website <u>http://www.l2f.inesc-id.pt/</u>.



L²F – Spoken Language Systems Laboratory TECNOVOZ



TECNOVOZ Speech Recognition and Synthesis Technologies

Funded by PRIME (3.1a) Project Number 03/165





The Tecnovoz consortium was built with the goal of creating a national technological centre capable of industrialising innovative systems based on speech technologies, at the same rhythm, and at equal technological level, that occurs in others countries. Simultaneously, it is intended that Portugal affirms itself as a real actor in the development of voice and speech technologies and a defender of its application to the Portuguese Language.

This consortium brings together 4 research institutes / universities, where INESC ID is included, and 9 companies from different market segments.

The voice and speech technologies are still in an ascending curve of its development and there are few applications for European Portuguese. However the market is attractive and the evolution process of these technologies is promising. In the scope of this project a set of technology modules (9) are created to be used in diverse demonstrators (13) from the companies in the project.

The total budget of Tecnovoz is 6.36 M \in with a partition of 47% for the research institutes and 53% to the companies.

Tecnovoz will create and develop new competences on voice and speech Technologies powering new market opportunities, through the extension of the lifecycle of the existing products and the emergence of new products and systems. These products and systems will supply efficiency and productivity gains on the Organizations, increasing the competitiveness on the short and medium period. To build these new products and systems it is important the cooperation between the research institutes and companies in order to have a technology transfer.

Duration: July 2006 - June 2008



L²F: TECNOVOZ project

Partners	×	ECCO Healthcare Solutions					NDI		-			
Global view of the project	There are 13 de Each demonstra several demonst	emonstrators and 9 technolog ator is based on several tec trators, as illustrated in the fo	y m chno llow	odul logy ving	es (4 mo matr	APIs dule ix:	s) res	sultir d ea	ng fro nch 2	om t API	he P is u	roject. sed in
				U								
		Áreas Tecnológicas	1	2		3		4	4		5	
		APIs	1.1	2.1	3.1	3.2	3.3	4.1	4.2	5.1	5.2	
		ESCTN	⊥	UMinho	INESC-ID	INESC-ID	INESC-ID	INESC-ID	INESC-ID	INESC-ID	INESC-ID	
	Empresas	Produto / Serviço (PPS)	01	02	04	05	06	08	09	11	12	
	Edisoft	14 Sistema de Escutas Telefónicas		S	S							
	CDCUS	15 Arquivo de Voz e Relatórios		S		S		S	S			
	Сгопо	17 Desktop Médico	S	S		S		S	S			
		18 Soluções de Corretagen		S	S			S	S	S		
	Promosoft	19 Soluções Bancárias		S	S			S	S	S		
		21 Soluções de Seguros			S			S	S	S		
	Datelka	22 Sistema Controlador	S	S								
	INOV	24 LISA - Secretária Digital			S			S	S	S		
	Priberam	27 Relatórios de Bloco Operatório				S						
		28 Legendagem Automática					S					
	RDP	29 Sistema de Transcrição Automática					s				S	
	Tecmic	31 Sistema Embarcado	S					S				
	Anditec	33 Sist. Comunicação Aumentativa	S			S		S				
INESC ID participation	INESC ID parti Speech to Text API 3.1 – Conti API 3.2 – Conti API 3.3 – Conti Text to Speech API 4.1 – Speec API 4.2 – Speec Support technol API 5.1 – Dialo API 5.2 – Segm	cipation is on the following ta conversion nuous Speech Recognition for nuous Speech Recognition for nuous Speech Recognition for conversion ch Synthesis for limited vocal ch Synthesis for limited vocal ch Synthesis from text applica ogical modules gue Management ientation, Indexation, Summa	asks or sn or m or la bula atior	nall v ediur rge v ry ar n ind tion	vocal m vo vocat nd ap epen and 1	bula ocabi oular oplic: dent	ries ulari ries a atior t	and t es ar and b n dep of M	elep nd clo oroad eendd	hone ean s lcast ent med	e spec speec new	ech h 's ta
More Information	By email to <u>inf</u> directly from th	<u>ow121.inesc-id.pt</u> or he website <u>http://www.12f.ines</u>	<u>sc-ic</u>	<u>1.pt/</u> .								



L²F – Spoken Language Systems Laboratory eCIRCUS – Voices for Synthetic Characters



eCIRCUS aims to develop novel conceptual models and innovative technology to support social and emotional learning through role-play and affective engagement for Personal and Social Education involving complex social situations. www.e-circus.org

eCIRCUS (Education through characters with emotional intelligence and role playing capabilities that understand social interaction) intends to further develop the FearNot! technology and carry out on a number of large-scale longitudinal psychological evaluations in schools using both the FearNot! software and the purpose built ORIENT software to be developed in the course of the project. The L2F - Spoken Language Systems Laboratory is a participant in the European eCIRCUS project. Its main task is to provide synthesized voices for the synthetic characters of the FearNot! demonstrator. FearNot! provides children with various scenarios about bullying behaviour, that promote engagement and believability with synthetic characters in a social interaction

The child user interacts with one physical bullying scenario and one relational scenario. After the introduction of the characters, school and situation, users view the first bullying episode, followed by the victimised character seeking rescue in the school library, where it starts to communicate with the child user. Within the initiated dialogue the user selects an advice from a list of coping strategies (shown as a drop down menu). The user also explains his/her selection and what he/she thinks will happen after having implemented the selected strategy, by typing it in.

The next episode then starts. The content of the final episode depends on the choices made by the user concerning the coping strategies: Paul, the bystander in the physical bullying scenario, might act as a defender for John (the victim), in case the user has selected a successful strategy, i.e. "telling someone"; or Martina (the bystander) might offer Frances (the victim) help. However, if the user has selected a unsuccessful strategy, i.e. "run away", the victim rejects the help in the final episode. At the end of the scenario, a universal educational message is displayed pointing out that "telling someone you trust" is usually a good choice. This universal message had to be incorporated as all teachers had strong preferences for children to finish the interaction with a positive feedback message.

More information is available by email to <u>info@l2f.inesc-id.pt</u> or directly from the website http://www.l2f.inesc-id.pt/.



L²F: eCIRCUS – Voices for Synthetic Characters

Main researchers (L2F)	Luís Caldas de Oliveira Christian Weiss Carlos Mendes Sérgio Paulo Luís Figueira
Terminology	Synthetic characters : A synthetic character is an autonomous character driven by an intelligent architecture whose interactions are not pre-scripted.
	Emergent narrative : Emergent narrative builds upon the model of improvisational drama rather than authored stories: an initial situation and characters with well defined personalities and roles produces an unscripted interaction driven by real-time choices.
	VICTEC (Virtual ICT with Empathic Characters) a European framework V project was carried out between 2002-2005. The project considered the application of 3D animated synthetic characters and emergent narrative to create improvised dramas to address bullying problems for children aged 8-12 in the UK, Germany and Portugal. Like VICTEC, eCIRCUS is to support social and emotional learning through role-play and affective engagement for Personal and Social Health Education (PSHE) involving complex social situations. The VICTEC characters had no voice: they communicated by textual messages. In eCIRCUS the virtual characters are able to speak to each others.
Bullying	Bullying behaviour has generated research interest among psychologists and educationalists over the past 10-15 years because of the amount reported to take place and the negative consequences for its victims.
	Bullying involves a wide range of behaviours which have been divided into a number of categories by researchers. Direct physical bullying includes actions such as being hit, kicked or punched, and taking belongings. Verbal bullying involves name calling, cruel teasing, taunting or being threatened. Finally, relational or 'indirect' bullying refers to behaviours such as social exclusion, malicious rumour spreading, and the withdrawal of friendships.
	Research has also identified a number of different roles in bullying, including the victim, bully, reinforcer of the bully, assistant of the bully, defender of the victim and outsider.
	Many different intervention initiatives have been tried in attempts to counteract and reduce bullying problems in schools. Examples include the whole-school approach to bullying, the no-blame approach for the bully and class activities such as 'circle time' and peer mediation techniques. However, all of these strategies have reported limited long-term success rates. VICTEC aimed to provide a novel, innovative approach to help deal with bullying problems in a fun and exciting environment.
More	About L2F: by email to info@l2f.inesc-id.pt or website http://www.l2f.inesc-id.pt/.



L²F – Spoken Language Systems Laboratory The VIDI-Video European Project



Video is vital to society and economy. It plays a key role in the news, cultural heritage documentaries and surveillance, ant it will soon be the natural form of communication for the Internet and mobile phones. Hence the need for video search engines that can cope with the peta bytes of future video archives.

Search concepts are categorized as types of scenes, types of objects, people, and events. Each concept can be specific: *Cavaco Silva*, or generic: *a happy person*. Some concepts will imply an audio analysis (e.g. *a cheering audience*), visual analysis (e.g. *an office scene*), style analysis (e.g. *a monologue*), or multimodal analysis (e.g. *a goal in soccer*).

VIDI-Video

Interactive semantic video search with a large thesaurus of machine-learned audio-visual concepts

Goal. Boosting the performance of video search engines by forming a 1000 element *thesaurus* detecting instances of audio, video or mixed-media content.

Description. The project will apply machine learning techniques to learn many different detectors from examples, using active one-class classifiers to minimize the need for annotated examples. The project approach is to let the system learn many, possibly weaker detectors describing different aspects of the video content instead of modeling a few of them carefully. The combination of many detectors will render a much richer basis for the semantics. The integration of audio and video analysis is essential for many types of search concepts.

Application scenarios:

- broadcast news
- cultural heritage
- surveillance

Duration: February 2007 - January 2010



researchers

L²F: The European Project VIDI-Video

UvA - Universiteit van Amsterdam, the Netherlands (coordinator) **Partners** UNIS - University of Surrey, UK UNIFI – Universita degli Studi di Firenze, Italy INESC-ID – Instituto de Engenharia de Sistemas e Computadores . Investigação e Desenvolvimento em Lisboa, Portugal CERTH – Centre for Research and Technology Hellas, Greece CVC - Centroi de Vision por Computador, Spain B&G – Stichting Netherlands Instituut voor Beeld & Geluid, the Netherlands FRD - Fondazione Rinascimento Digitale, Italy Subcontracting UoM - University of Modena e Reggio Emília, Italy IIT - Indian Institute of Technology, India **WorkPackages** Learning system WP4 Visual Analysis WP2 Video Processing _ . . (2b) **2**c WP3 Audio Analysis 2d WP8 Data and queries WP9 Dissemination WP5 Learning integrated feature detectors (1) \bigcirc 4 Run time system 6 WP6 Software WP7 Demonstrators development and applications 5 WP1 management 13 **INESC-ID** Task 2.3 Audio Segmentation Segmentation of the audio stream into homogeneous regions according to **Participation** background conditions and speakers. Task 3.1 Detection of Audio Events Perform audio event detection by machine learning techniques(goals in sport matches, shooting, explosions, car or helicopter noises, cries, screams, laughter, cheering, ... Task 3.2 Speech Recognition Build a large vocabulary continuous speech recognition system tailored to broadcast news **INESC-ID** Isabel Trancoso Main João Paulo Neto



L²F – Spoken Language Systems Laboratory Project LECTRA



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





L²F – Spoken Language Systems Laboratory **Researchers**



L2F was created in 2001, bringing together several research groups of INESC and independent researchers that could potentially add relevant contributions to the area of computational processing of spoken language for European Portuguese. The lab includes around 30 researchers, most of them either Professors (8) or graduate students at IST, including 2 post-doc researchers, and 4 invited researchers from other universities in Portugal. Their background ranges from Electrical Engineering to Computer Science and Linguistics. This strongly interdisciplinary group is actively involved in many areas of spoken language research and development, including speech recognition, speech synthesis, speech coding, speech understanding, audio indexation, multimodal dialogue systems, language and dialect identification, and speech-to-speech machine translation, among others.

In the early nineties, a formal protocol agreement was established between the former Speech Processing Group and the Center of Linguistics of the University of Lisbon (CLUL). This agreement, headed by Dr. Céu Viana on CLUL's behalf, broadened the expertise of the group to include different areas of linguistics, More recently, the L2F team was extended to include invited researchers from the Univ. of Algarve (Prof. Jorge Baptista) and Univ. of Beira Interior (Prof. Gaël Dias).



L²F: Researchers

Diamantino Caseiro



Licenciatura – Informatics and Computer Engineering, IST, 1994 MS - automatic language identification, IST, 1998 PhD – Computer Science (finite-state methods in automatic speech recognition), IST, 2003

www.l2f.inesc-id.pt/~dcaseiro

Diamantino Caseiro has been a lecturer at Instituto Superior Técnico, Lisbon (IST), since 2000, teaching human computer interfaces, advanced algorithms and compilers. He has been a researcher at the Speech Processing Group of INESC since 1996. He is currently a senior researcher at the Spoken Language Systems Lab (L2F) of INESC-ID that resulted from the reorganization of the previous group.

His current research interests are: automatic speech recognition, statistical machine translation (in particular speech-to-speech translation), and in general integrating knowledge-based information with data-driven methods in different fields of natural language processing. In the past, he has participated in several European and National projects, and is currently the principal investigator of a national project on applying weighted finite-state transducers to spoken language processing. He is a member of ISCA (International Speech Communication Association), ACM and the IEEE Computer Society.

Auxiliar Professor

Licenciatura - Applied Mathematics and Computation, IST, 1994 MS - natural language database interface, IST, 1997, PhD – syntactic/semantic interface based on hierarchically organized rules, IST and Université Blaise-Pascal, France. www.l2f.inesc-id.pt/~lcoheur

Luísa Coheur She has been a lecturer at Universidade Autónoma (from October 2004 to September 2005), Universidade Lusófona de Humanidades e Tecnologias (from October 2005 to February 2006), and at IST since March 2006, where she has been teaching Artificial Intelligence and Algorithms and Data Structures. She has started her research at INESC in 1995, in the Telematic Services and Systems group. In 1998 she became a member of the Telematic and Computational Systems Center from IST. In 2001 she became a member of the newly created Spoken Language Systems Lab (L2F). Her current research interests are: question answering, dialogue systems and machine translation.

Auxiliar Professor

Licenciatura - Electrotechnical and Computer Engineering, IST, 1990 MS - ECE (object-oriented programming in distributed systems), IST, 1995 PhD – (natural language generation), IST, 2005 www.l2f.inesc-id.pt/~david

David Matos has been teaching at IST since 1993 (on logic and functional programming, object-oriented programming, algorithms and data structures, compiler construction, computer architecture, distributed systems, computer graphics, and natural language processing). He has been a researcher at INESC since 1988 in the Distributed Systems, Telematic Services and Systems, and Software Engineering Groups. In 1998 he became a member of the Telematic and Computational Systems Center (IST), until 2001, when he became a member of the newly created Spoken Language Systems Lab (L2F). In the past, he has participated in several European and National projects, as well as projects in the private sector (banking and telecommunications industries) both national and international. He is a member of the ACM, the API, the Order of Portuguese Engineers and the IEEE Computer Society.







David Martins de Matos



Licenciatura - ECE, IST, 1981



L²F: Researchers





Nuno Mamede has been a lecturer at Instituto Superior Téncico since 1982 since

Auxiliar Professor

MS – ECE, IST, 1991 PhD – ECE, IST, 1998 www.l2f.inesc-id.pt/~jpn

Licenciatura - ECE, IST, 1987

Associate Professor

MS - ECE, IST, 1985 PhD - ECE, IST, 1992 www.l2f.inesc-id.pt/~njm

his graduation and he has taught courses on Digital Circuits, Microprocessors, Logic Programming, Artificial Intelligence, Object Programming and Natural Language Processing. He is a researcher at INESC, in Lisbon, since its creation in 1980 and participated in the foundation of L2F where hold a position in the Executive Board. His activities have been in the areas of Natural Language Processing, namely. on Speech Dialog Systems and text processing. He worked on several european and national research projects. He is the responsible researcher of the DIGA project for the development of a Spoken Dialog System. He has authored a significant number of scientific papers. He is a member of AAAI, ACM and ACL.

João Paulo Neto's PhD topic was speaker-adaptation in the context of hybrid Artificial Neural Networks and Hidden Markov Models continuous speech recognition systems. In 1991 he started as lecturer and since 1998 he holds a position of Assistant Professor in Instituto Superior Técnico, where he has taught

have been in the areas of neural networks (training algorithms and applications) and speech recognition (HMM/neural network based). He has authored a

significant number of scientific papers. He is a member of IEEE and ISCA.

João Paulo Neto



signal theory, discrete signal processing, control systems and neural networks. He joined the Neural Networks and Signal Processing group of INESC in 1987, having participated in the European projects PYGMALION, WERNICKE, SPRACH and ALERT and in several national projects. In 2001 he participated in the foundation of L2F where hold a position in the Executive Board. His activities

Luís Caldas de Oliveira



Auxiliar Professor Licenciatura - EE, IST, 1985 MS - ECE, IST, 1989 PhD - ECE, IST, 1997 www.l2f.inesc-id.pt/~lco

Luís Caldas de Oliveira has been a lecturer at Instituto Superior Técnico since his graduation and he has taught courses on circuit theory, signal theory, signal processing and speech processing. He is a researcher of INESC, in Lisbon, since 1985, and he har been working in the area of speech processing, and particularly, in speech synthesis. From 1991 to 1993 he worked at AT&T Bell Laboratories, on modelling and analysis of the glottal flow and its applications to text-to-speech synthesis. He worked on several european national research projects. He is the responsible researcher of the DIXI+ project for the development of a text-to-speech system for European Portuguese. He is a member of IEEE, ISCA and Ordem dos Engenheiros.



L²F: Researchers



Isabel Trancoso



Associate Professor Licenciatura – EE, IST, 1978 MS – ECE, IST, 1984 PhD – ECE, IST, 1990 www.l2f.inesc-id.pt/~ajs

António Serralheiro was a lecturer at Instituto Superior Técnico from 1977 to 2002. He is, since 2002, Associted Professor at the Military Academy, teaching Signal and Systems, Fundamentals of Telecommunications and Introduction to Power Systems courses. He is also a senior researcher at INESC ID Lisbon, since 2000. His first research topic was speech recognition (isolated words) using stochastic models. From October 1986 through September 1987, he worked on these topics at AT&T Bell Laboratories, Murray Hill, New Jersey. His current interests spans from speech recognition to fractional system modelling. He belonged to the Organizing Committee of the INTERSPEECH'2005 Conference that took place in Septembre, in Lisbon. He is a member of IEEE and Ordem dos Engenheiros.

Full Professor Licenciatura – EE, IST, 1978 MS – ECE, IST, 1984 PhD – ECE, IST, 1987 Agregação – ECE, IST, 2002 www.l2f.inesc-id.pt/~imt

Isabel Trancoso has been a lecturer at Instituto Superior Técnico, since 1979, having coordinated the ECE course for 6 years. She is currently a Full Professor, teaching speech processing courses. She is also a senior researcher at INESC ID Lisbon, having launched the speech processing group, now restructured as L2F, in 1990. Her first research topic was medium-to-low bit rate speech coding. From October 1984 through June 1985, she worked on this topic at AT&T Bell Laboratories, Murray Hill, New Jersey. Her current scope is much broader, encompassing many areas in speech recognition and synthesis, with a special emphasis on tools and resources for the Portuguese language. She was a member of the ISCA (International Speech Communication Association) Board (1993-1998), the IEEE Speech Technical Committee (since 1999) and the Permanent Council for the Organization of the International Conferences on Spoken Language Processing (since 1998). She was elected Editor in Chief of the IEEE Transactions on Speech and Audio Processing (2003-2005), Memberat-Large of the IEEE Signal Processing Society Board of Governors (2006-2008), and Vice-President of ISCA (2005-2009). She chaired the Organizing Committee of the INTERSPEECH'2005 Conference that took place in September 2005, in Lisbon.

Christian Weiss



Post-Doc

M.A. – Computational Linguistics, Computer Science and Sociology, Univ. Heidelberg, Germany, 2002 PhD – IKP, Univ. Bonn, Germany, 2006

https://www.l2f.inesc-id.pt/wiki/index.php/Christian_Weiss

Christian Weiss's PhD thesis was on "Adaptive Audio-Visual Synthesis" covering statistical learning for video-realistic audio visual synthesis. He was working in a German Research Foundation <u>DFG</u> granted project on automatic training strategies for unit-selection based speech synthesis. Previous he worked for IBM European Speech Research and was a DAAD/JSPS Fellow at the Nagoya Institute of Technology, Japan <u>Nitech</u>. His research interests are in speech synthesis, audio-visual synthesis and statistical learning. He is a member of <u>GI</u> and <u>ISCA</u>.