



International Conference on Computational Processing of Portuguese Language
Applications of Portuguese Speech and Language Technologies

Applications of Portuguese Speech and Language Technologies - Propor 2008 Special Session

Hosted by:



Universidade de Aveiro

Promoted by:



**Microsoft Language
Development Center**

Propor 2008 Special Session Committee

Special Session Chair

- **António Teixeira** - DETI/IEETA, Universidade de Aveiro, Portugal

Organising Committee

- **Daniela Braga**, Microsoft Language Development Center, Portugal
- **Miguel Sales Dias**, Microsoft Language Development Center, Portugal
- **António Teixeira** - DETI/IEETA, Universidade de Aveiro, Portugal

Programme Committee

- **António Teixeira** - DETI/IEETA, Universidade de Aveiro, Portugal
- **Daniela Braga**, Microsoft Language Development Center, Portugal
- **Vera Strube de Lima**, Pontifícia Universidade Católica do Rio Grande do Sul, Brasil
- **Luís Caldas de Oliveira**, INESC-ID/IST, Portugal

Editorial Board

- **Daniela Braga**, Microsoft Language Development Center, Portugal
- **Miguel Sales Dias**, Microsoft Language Development Center, Portugal
- **Luanda Braga Batista**, Microsoft Language Development Center, Portugal

BROADCAST NEWS SUBTITLING SYSTEM IN PORTUGUESE

João P. Neto

L2F - Spoken Language Systems Laboratory, INESC ID Lisboa
 Instituto Superior Técnico / Technical University of Lisbon
 VOICEINTERACTION
 R. Alves Redol, 9, 1000-029 Lisboa, Portugal
<http://www.l2f.inesc-id.pt/>
Joao.Netto@inesc-id.pt



The subtitling of broadcast news (BN) programs is starting to become a very interesting application due to the technological advances in Automatic Speech Recognition (ASR) and associated technologies as Audio Pre-Processing (APP). Also, there is a generic request from society and governments that are pushing the TV broadcasters to increase the amount and diversity of TV programs subtitled.

Over the last decade, the speech research community spent a large effort in the research and development of broadcast news (BN) systems, initially for English, and after for a variety of other languages. The development of a system for a new language is a challenging task due to the need of new acoustic training data, vocabulary definition, lexicon generation and language model estimation. Despite the good results very few systems were used for subtitling. The subtitling operation implies, besides real-time, an online operation.

In our lab, we have been working, for several years, on the development of a BN system for the European Portuguese language. This system was the result from a close cooperation with RTP (Rádio Televisão de Portugal), the Portuguese public broadcast company. As a result of this effort we create a product used for subtitling of recorded and live programs. The online system is used for live subtitling of the 1pm and 8pm o'clock news. This system was in text experimentally for more than 6 months and went public on last March 7th 2008. Since there we have received very good feedback from the hearing handicapped associations and from the public in general. The system works in less than Real Time but since it is a pipeline of several blocks have a total delay of 3.5 seconds. Since RTP is using the teletext to output the subtitles we have to compose full 2 lines subtitles which have a mean duration of 3 seconds. The total delay of the system is on mean 6.5 seconds, which it is an excellent performance.