

# Automatic Transcription of Courtroom Recordings in the JUMAS project

Daniele Falavigna<sup>1</sup>, Diego Giuliani<sup>1</sup>, Roberto Gretter<sup>1</sup>, Jonas Lööf<sup>2</sup>, Christian Gollan<sup>2</sup>, Ralf Schlüter<sup>2</sup>, and Hermann Ney<sup>2</sup>

<sup>1</sup> FBK-Irst - Via Sommarive 18, 38050 Povo, Trento, Italy  
{falavi,giuliani,gretter}@fbk.eu

<sup>2</sup> Lehrstuhl für Informatik 6 - Computer Science Dept.  
RWTH Aachen University, Aachen, Germany  
{loof,gollan,schluter,ney}@cs.rwth-aachen.de

**Abstract.** In this paper we present ongoing work on speech recognition for the judicial domain, performed in the European project JUMAS (Judicial management for digital library semantics.) The specific challenges for courtroom speech recognition are discussed, and the development of speech recognition systems for Italian and Polish are described. The results achieved on the target domain are presented and discussed.

## 1 Introduction

This paper presents work performed in the context of the JUMAS project [1], a European Union project aimed at information extraction and indexing in the judicial domain, specifically for the processing of court recordings. As part of the project Polish, and Italian automatic speech recognition (ASR) system for the domain of court proceedings is being developed.

State of the art ASR systems allow to achieve good performance (approximately word error rates inferior to 10%) when the speech signal is acquired in controlled conditions. However, as demonstrated in recent DARPA evaluations [2] performance significantly decreases if transcription tasks include speech coming from non professional speakers, uttering their sentences in "free" conditions, where the recording environment is not optimal.

**Table 1.** *State of the art ASR word error rates for different application domains.*

	WER
connected digits	$\leq 0.5\%$
continuous dictation	$\leq 5\%$
studio broadcast news	$\leq 10\%$
telephone news reports	$\leq 20\%$
telephone conversations	$\leq 30\%$
meetings (head mounted microphone)	$\leq 30\%$
meetings (distant microphone)	$\leq 50\%$

Table 1 gives a general idea on the level of performance that can be presently reached with state of the art ASR systems. As can be seen from the Table, Word Error Rate (WER) increases as more degrees of freedom are allowed in the speech signal to recognize.













