

Domain : Natural Language, Speech and Audio Processing

Domain extra : corpus linguistics, machine learning, psycholinguistics

Year : 2010

Starting : autumn 2010

Status : open

Subject : Automatic speech transcription error recovery using multiple knowledge sources

Thesis advisor : Martine Adda-Decker

Co-advisors : Ioana Vasilescu, LIMSI/CNRS

Collaborations : RWTH Aachen University Germany, KIT Karlsruhe Germany

Abstract :

Over the past decade, it has been firmly established that human listeners still significantly outperform machines on speech transcription tasks. Indeed, human native listeners generally do a very well job in handling many aspects of variation that are proper to speech, such as pronunciation variants, disfluencies, ungrammatical sentences, accents, noise and so forth. These observations are particularly true when large surrounding contexts (complete and long sentences) are available. However, ASR systems generally take their transcription decisions on relatively limited contexts (several words) and their handling of variation in speech still remains a big challenge for current automatic speech recognition (ASR) systems. ASR being an enabling technology for a large variety of advanced potential applications, such as multi-media information access or speech-to-speech translation, the impact of ASR errors on their performances will also be investigated.

Context (Current state of the art in the domain - max : 1000 chars) :

The handling of variation in speech still remains a big challenge for current automatic speech recognition (ASR) systems. In particular, the handling of casual interactive speech often results in high word error rates, which ask for specific error recovery strategies. The rich experimental environment of the Franco-German Quaero project (with annual ASR evaluations in multiple languages), provides a unique testbed for a systematic study of ASR errors. The proposed parallel between human and machine errors is then highly innovative and may push both our fundamental knowledge about human speech processing as well as basic techniques for automatic speech processing and error recovery.

Objectives (Scientific objectives - max : 1000 chars) :

The aim of the present proposal is to identify current obstacles that affect ASR performance, to propose a sound ASR error typology and to benchmark human vs ASR performances according to this typology, to design innovative mechanisms for error recovery, as well as to explore new solutions to improved spoken language modeling.

Work program (Main expected steps - max : 1000 chars) : ASR errors need to be investigated according to at least three axes : (i) perceptual experiments on selected materials to benchmark human performances.

(ii) proper names which produce errors which are further harmful to further processings, such as information access, translation or question answering (factors : frequency of occurrence, pronunciation variants, repetitions).

(iii) reduced pronunciations (modeling options : specific acoustic models, pronunciation dictionary).

Apply and evaluate the impact of different knowledge sources (named entities, POS, prosody, pronunciation variants).

Extra information (Extra information like a link to a full description of the subject - max : 500 chars) :

Prerequisite (Expected background) :

Expected funding : research contract

Status of funding : expected

Candidates (informations on potential candidates - max : 500 chars) :

user : martine.adda-decker@limsi.fr