Spoken Language Identification

Overview

Automatic spoken language identification (LID) copes with the problem of identifying the language spoken by an unknown speaker. During the project LID experiments were conducted on television news and talk-show videos broadcast in four languages, namely: Dutch, English, German and Italian. Experiments showed a LID error rate of 3.5% and 2.5% in “open-set” and “closed-set” conditions, respectively.

In depth description

Two different systems were developed compared on project data. The first one makes use of a separate Gaussian Mixture model for modelling acoustic features for each language. The second one is a multilingual speech recognizer that exploits a multilingual set of triphone hidden Markov models and a multi-language N-gram LM to decode speech in an unknown language. The second system is more accurate while the first one is easier to configure and train in order to include new languages. The developed systems process an audio recording by performing audio partitioning and clustering of speech segments and return for each identified speech segment in the audio recording a language label.

Potential fields of application

The technology for spoken language identification developed in TOSCA-MP has a great potential for application in many application areas such, for example, automatic metadata generation for video material and automatic transcription (for properly routing speech utterances in a given language towards a proper speech recognizer).

Possibilities for exploitation

Exploitation of the result is on a bilateral licensing agreement basis.

Further Information

Further technical information is available in TOSCA-MP Deliverables D2.1, D2.2 and D2.3 “Automatic Metadata Extraction and Enrichment”.

Contact Person

Diego Giuliani
Fondazione Bruno Kessler
Human Language Technology research unit
Via Sommarive 18, I-38123
Trento – ITALY
giuliani@fbk.eu