This thesis describes an online speaker recognition as part of an immersive system for teleconferencing, developed at the Institute for Data Processing. With this conference system, it is possible to create an individual audio stream for every speaker in a conference room by using only a device on the conference table. The current active speaker in a teleconference is identified online. Therefore, a model of each conference participant is constructed by using short time spectral features. In the system approach, Gaussian Mixture Models (GMMs) are applied for modeling individual speakers. Features are extracted from the incoming audio stream, constituting a likelihood score for every model and thus identifying the present speaker. Mel-frequency cepstral coefficients (MFCCs) are chosen as features to represent spectral attributes of different sounding voices. Due to the required low computational complexity of an online speaker recognition task, a Universal Background Model (UBM) in combination with maximum a posteriori (MAP) adaptation is used to provide a very fast creation of speaker-dependent GMMs, with only few training data needed. Experiments are...
carried out on the AMI meeting corpus the active speaker is correctly determined with a recognition accuracy of 78%.

Dewey Dezimalklassifikation:
620

Beauftragende Einrichtung:
Technische Universität München, Lehrstuhl für Datenverarbeitung

Jahr:
2014

Jahr / Monat:
2014-04

Monat:
Apr

Sprache:
en

TUM Einrichtung:
Lehrstuhl für Datenverarbeitung

Format:
Text

CC-Lizenz:
by, http://creativecommons.org/licenses/by/3.0/de

Occurences:
· Einrichtungen > Fakultäten > Fakultät für Elektrotechnik und Informationstechnik > Lehrstühle und Professuren > Datenverarbeitung (Prof. Diepold) > 2014

entries: