Contact details: Hareesh Mandalapu (hareesh.mandalapu@ntnu.no)

Thesis title: Cross-lingual Speaker Verification Using Deep Learning Approach

Background: Automatic Speaker Verification (ASV) systems highly depend on the language used in training and enrolling speakers. Language dependency makes the voice-based security systems less robust and generalizable to a wide range of applications. Speaker-specific patterns are more reliable than language-dependent features. In this work, novel deep learning methods are utilized to develop a unique language independent feature extractor for speakers. With the help of this, a speaker verification can be performed independent of the language used in enrolment. Experiments on multilanguage speaker datasets are performed to validate the proposed method.

Tasks: To develop skills in automatic speaker verification.
To implement machine learning and deep learning algorithms.
To design and develop a speaker specific feature extraction model.
To develop an automatic, language independent speaker verification system.
To perform experiments on SWAN dataset and obtain standardised results.

Reference:

1. Li, Lantian, Dong Wang, Askar Rozi, and Thomas Fang Zheng. "Cross-lingual speaker verification with deep feature learning." In *2017 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA ASC)*, pp. 1040-1044. IEEE, 2017.
1. Wang, Jianglin, and Michael T. Johnson. "Vocal source features for bilingual speaker identification." In *2013 IEEE China Summit and International Conference on Signal and Information Processing*, pp. 170-173. IEEE, 2013.
2. Snyder, David, Pegah Ghahremani, Daniel Povey, Daniel Garcia-Romero, Yishay Carmiel, and Sanjeev Khudanpur. "Deep neural network-based speaker embeddings for end-to-end speaker verification." In *2016 IEEE Spoken Language Technology Workshop (SLT)*, pp. 165-170. IEEE, 2016.
3. Misra, Abhinav, and John HL Hansen. "Spoken language mismatch in speaker verification: An investigation with nist-sre and crss bi-ling corpora." In *2014 IEEE Spoken Language Technology Workshop (SLT)*, pp. 372-377. IEEE, 2014.