

4th INTERNATIONAL CONFERENCE ON
AI ML, DATA SCIENCE AND
ROBOTICS

NOVEMBER 28-29 | PORTO, PORTUGAL



Razieh Khamsehashari*, Sebastian Möller, Amin Jaber

Quality and Usability Lab, Technische Universität Berlin

Enhancement of a Multilingual Model for Automatic Speech Recognition Analysis of Time and Frequency Domain Data Augmentation

Abstract: In recent years, significant progress has been made in automatic speech recognition (ASR) systems, especially for languages with abundant transcribed speech data. However, freely accessible models for low-resource languages, particularly those other than English, are still scarce. This study aims to enhance the performance of the prominent multilingual ASR model, Whisper, through fine-tuning and data augmentation techniques. Comprehensive and systematic experiments are conducted on time, frequency, and time-frequency domain augmentation strategies across multiple low-resource languages (German, Farsi, Arabic, Russian) to investigate whether these techniques could improve ASR performance in low-resource settings. The study sheds light on the feasibility and limitations of various data augmentations, with a specific emphasis on classic augmentation approaches including audio and spectrogram-based techniques. The findings indicate that audio-based augmentation generally outperforms other evaluated methods. Furthermore, the study explores understanding the effects of classic data augmentation on model regularization and learning behavior, an aspect that, to our knowledge, has not been extensively explored in ASR. Our findings reveal different behaviors of these techniques not only in the learning process but also in model regularization and generalization. This insight is crucial for integrating these methods into more sophisticated deep learning models, particularly those employing aligned data augmentation in conjunction with classic approaches.

Keywords: automatic speech recognition, data augmentation, low-resource languages.