

Research on methods for building a database for Vietnamese speech recognition

Pham Trung Kien

Abstract: This thesis investigates methods for constructing a speech database tailored for Vietnamese automatic speech recognition (ASR). Given the tonal nature and specific phonetic characteristics of Vietnamese, the development of a high-quality, representative corpus is critical for robust acoustic and language modeling. This research proposes a systematic framework for database design, encompassing criteria for phonetic balance, speaker diversity, and coverage of regional dialects. The methodology details the recording protocol, annotation schema, and quality assurance processes. A key contribution is the implementation of a semi-automated tool for phonetic segmentation and labeling, which significantly reduces manual effort while ensuring accuracy. Experimental evaluations using a standard ASR system demonstrate that the proposed database yields a substantial reduction in word error rate compared to existing, less-structured corpora. The findings underscore the importance of a principled approach to database construction, providing a scalable and replicable model for future Vietnamese speech resource development and contributing to the advancement of voice-enabled technologies for the language.

Keywords: Vietnamese speech recognition, database construction, Vietnamese language, speech database