

Modeling Surzhyk in Neural Automatic Speech Recognition

Automatic Speech Recognition is increasingly central to how people interact with technology, and getting it right matters, especially for communities that rely heavily on written communication tools. Yet linguistic resources are far from evenly distributed, and many languages, including Ukrainian, remain significantly underrepresented in ASR development. This gap has real consequences: speakers of underrepresented languages are effectively excluded from the full range of technologies that speakers of dominant languages take for granted.

Most existing models are trained on standardized language varieties and tend to struggle with non-standard or mixed speech. This study focuses on Surzhyk, a Ukrainian-Russian mixed variety with fluid lexical and phonological patterns and no standardized written form. Spoken predominantly in Ukraine, Surzhyk has gained additional sociolinguistic relevance in recent years as many speakers navigate a broader shift away from Russian. The core question is whether explicitly incorporating phonological information into ASR training improves transcription accuracy for Surzhyk, compared to baseline ASR models.

The methodology involves building a phonetically rich corpus of controlled sentences and a short narrative, designed to capture key Ukrainian phonemes alongside features typical of Surzhyk. Recordings are manually annotated, segmented with Montreal Forced Aligner, and used to fine-tune a pretrained neural ASR model under two conditions: a baseline trained on raw audio-text pairs, and a phonologically informed model using phoneme-level annotations. Performance is evaluated using Word Error Rate and Phoneme Error Rate, alongside qualitative analysis of phoneme-specific recognition patterns.

The hypothesis is that phonologically informed fine-tuning will improve both accuracy and robustness. Beyond the immediate results, the study aims to contribute to a growing body of research on integrating linguistic structure into speech technologies, and to make a case that low-resource and non-standard varieties deserve dedicated methodological attention, not merely adapted versions of tools built for majority languages.

Keywords: Automatic Speech Recognition(ASR), Code-switching, Ukrainian, Surzhyk, Neural ASR, Non-standard speech