

>_ Evaluating the performance of TDNN-BLSTM on Mandarin Speech

1. Background

- The limitation of Automatic Speech Recognition (ASR) systems is that they cannot recognize words not present in the training lexicon.
- Phonemes can be used to distinguish between words as they are the basic unit of sound.
- Automatic Phoneme Recognition (APR) systems are a variation of ASR systems that use phonemes instead of words.

4. Results

- **35.38%** PER on spontaneous speech.
- **45.31%** PER on clear speech.
- Architectural differences of TDNN-BLSTM do not impact the error rates of individual phonemes.
- **~30%** of the errors are caused by mismatched tones.
- Tonal phonemes are more error-prone in spontaneous speech.
- Non-tonal phonemes are more error-prone in clear speech.

2. Motivation

- Recent research on Dutch PR indicates that TDNN-BLSTM is currently one of the best performing APR architectures [1].
- How does phoneme recognition using TDNN-BLSTM perform on a tonal language such as Mandarin.

3. Objectives

- Analyze the performance of phoneme recognition using TDNN-BLSTM on Mandarin speech.
- What is the difference in the Phoneme Error Rate (PER) between clear and spontaneous speech ?
- Determine the impact of tones for phoneme recognition using TDNN-BLSTM by comparing the results with previous research on Dutch PR.

5. Future Work

- Consider using feature vectors capable of capturing tonal information when present.
- How does the distribution of (non-)tonal phonemes between different speech styles compare to other tonal languages.