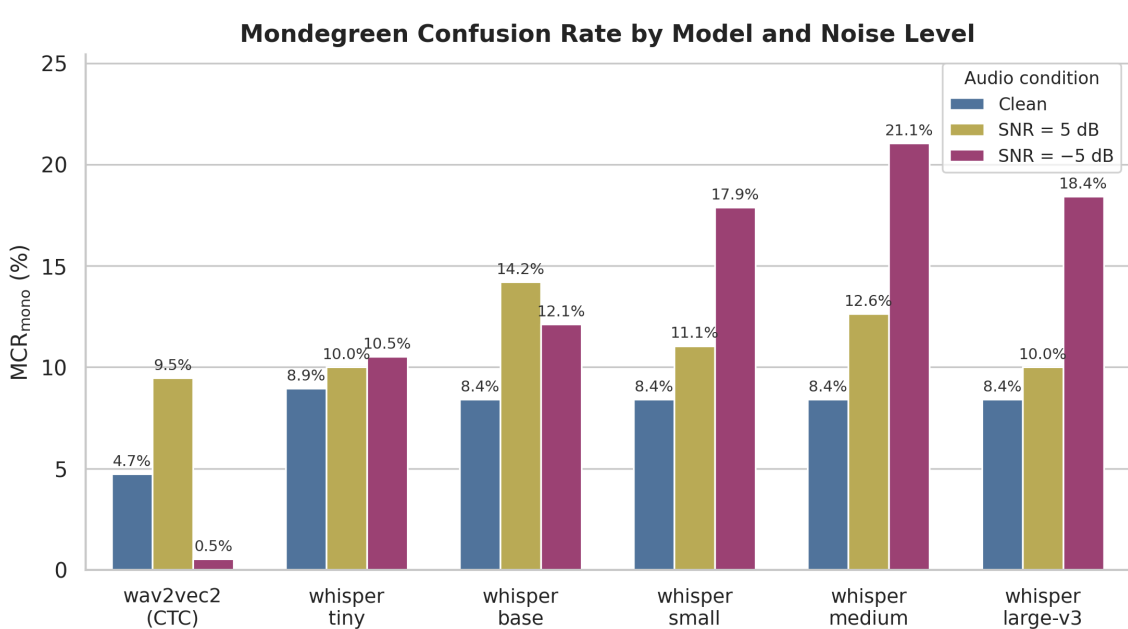
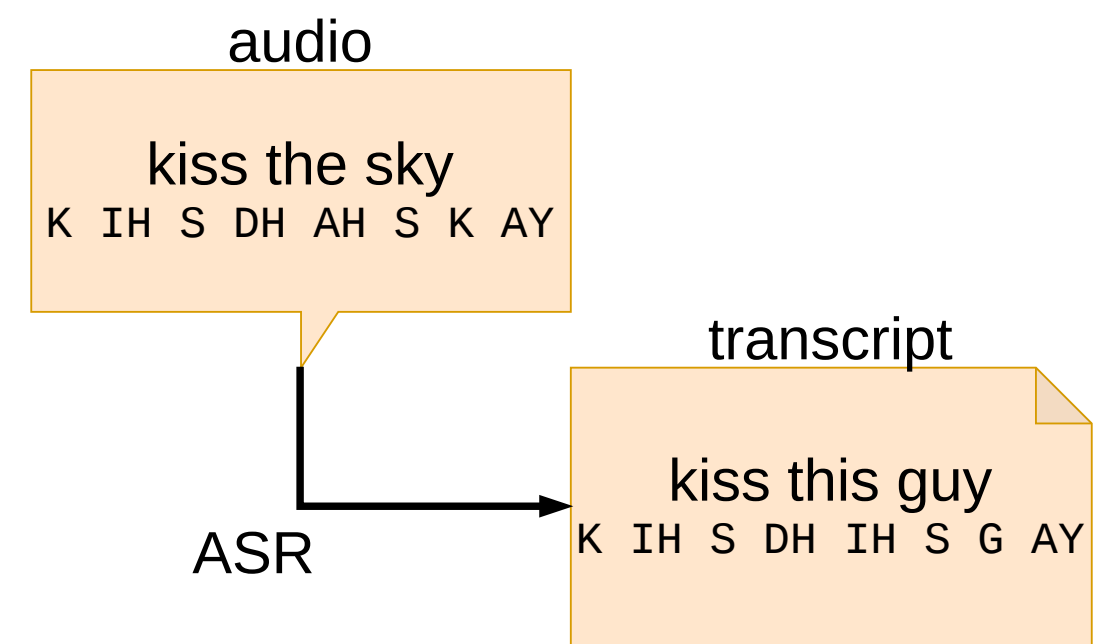


MondegreensEval: A Phonetic Benchmark for Measuring Language-model Bias in Automatic Speech Recognition

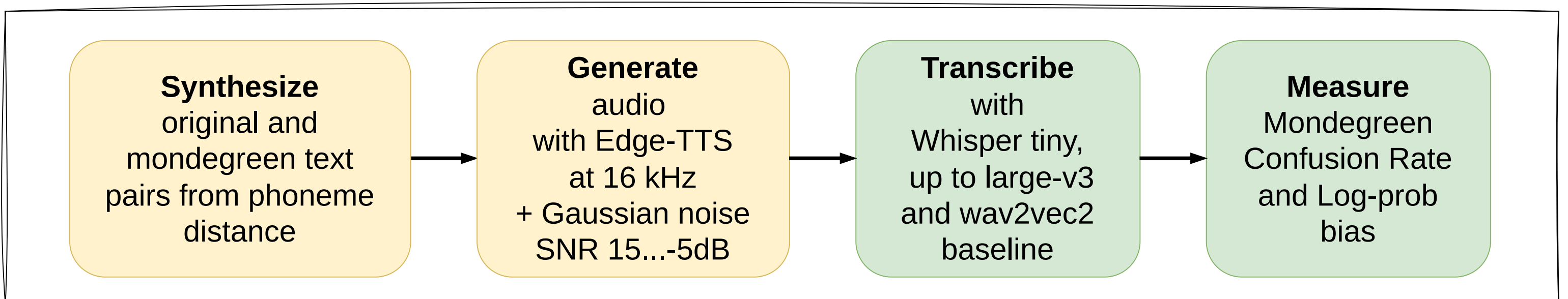
Wan Ju Kang / soarhigh0714@gmail.com

The 43rd International Conference on Machine Learning / July 6 - 11 / Seoul, Korea

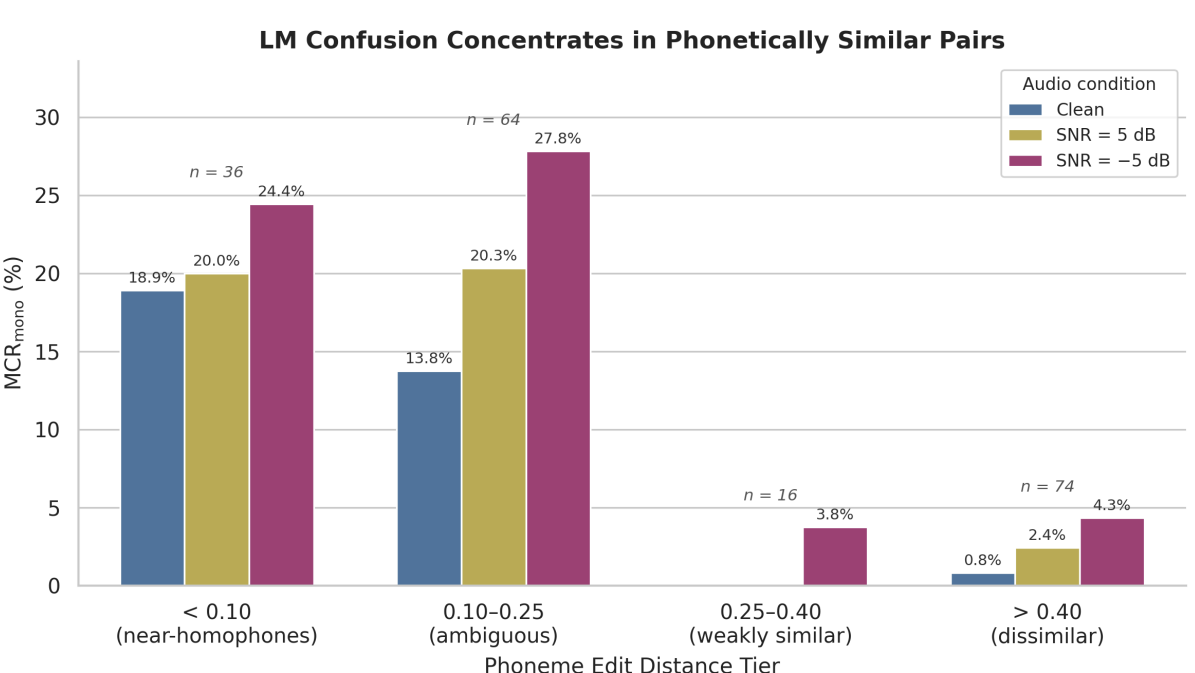
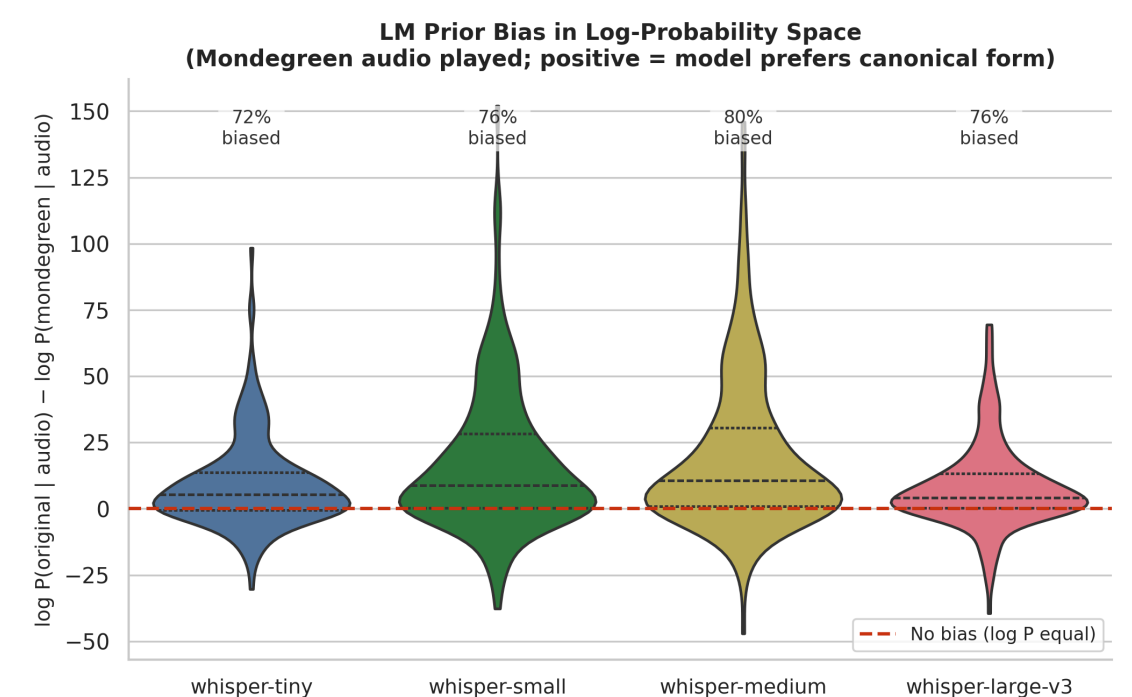
- A **mondegreen** is a phonetically near-identical (*i.e.*, not a homophone) phrase with a different meaning.
- Real acoustic cues survive, so a confusion reflects the prior, not genuine ambiguity.



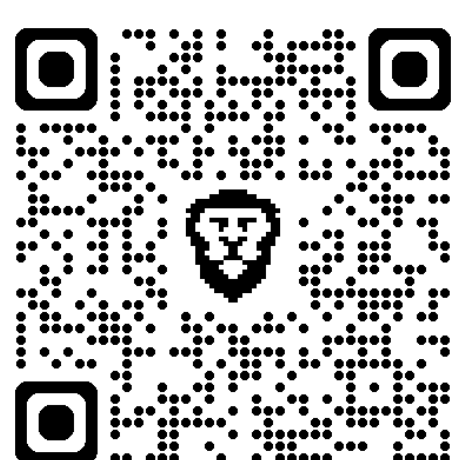
- ASR embeds a strong **language-model prior** that can override acoustic evidence.
- The result is a fluent and grammatically plausible phrase the waveform never supported - so **WER and CER stay blind**.



- 72-80% of (ground truth, mondegreen) pairs carry **positive log-prob bias** across every checkpoint.
- The logprob signal exposes the pull that the MCR metric misses: a **systematic architectural preference**, not an edge case.



- Confusion is tightly coupled to **phonetic proximity**.
- Near-homophone tier: MCR=19-21% clean, rising to 24-28% at -5dB degraded.
- Dissimilar tier stays below 4% across all conditions.



Generation and evaluation code



MondegreensEval dataset

