ABSTRACT
First viable Singapore Hokkien Automatic Speech Recognizer (ASR):
- Can benefit a large community in Singapore
- Builds on previous work
- Improved the language resources for training a Hokkien ASR, especially the lexicon and corpus
- Potential applications: Smart Homes and hospitals
- Experiments showed promising improvements in ASR performance

HOW DOES AN ASR WORK?
- The ASR makes intelligent guesses using a probabilistic framework to convert a given audio into text
- Main components: feature extractor and decoder

\[ S^* = \arg \max P(S | X) \]
\[ P(S | X) = P(W | \phi) P(\phi | X) \]

EXPERIMENTATION AND EVALUATION
Experiment 1: Lexicon Quality and Size: Show improvement over baseline using two metrics:
- Word count
- Percentage of Chinese characters covered by our lexicon

Experiment 2: Corpus Quality and Size: Show improvement over baseline corpus using three metrics:
- Sentence count
- Number of unique syllables
- Number of unique triphones

Experiment 3: Effectiveness of AM: Show improvement in performance of ASR with more data.
- Collected 50 utterances for testing
- Word Error Rate (WER)
  - Measure of inaccuracy of speech recogniser
  \[ \text{WER} = \frac{1 + D + S}{N} \]
- Sentence Error Rate (SER)
  - Proportion of sentences that do not match word for word
  - Suitable for assessing voice command applications

RESULTS
1. Lexicon Quality and Size
   - Coverage of Characters (%)
   - Word Count

2. Corpus Quality and Size
   - Sentence Count
   - No. of unique syllables
   - No. of unique triphones

3. Effectiveness of AM
   - LDA+MLLT (GMM-HMM)
   - SGD (DNN)

CONCLUSION
- We managed to expand our original lexicon and corpus extensively by almost doubling the word count, increasing the word diversity and providing more data in general that can be used for training.
- We showed that an ASR trained with a larger lexicon and corpus, showed much better performance,
- This suggests that an ASR system trained with Set C would probably perform even better.
- Our best trained system achieved a 6% WER and 12% SER, indicating that we have already have a highly usable voice command system for Hokkien.