SINGAPORE HOKKIEN SPEECH RECOGNITION AND APPLICATIONS

Lee Estelle, Lim Zhi Yi Vanessa, Ang Hui Shan[1] and Lim Boon Pang[2]
[1] Raffles Girls' School (Secondary)
[2] Institute of Infocomm Research, 1 Fusionopolis Way, Singapore 138632

Institute for Infocomm Research
A*STAR

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

PROCEDURE

1. Generating Sentences for Applications

Set	# of Sent.	Example
Common Sentences	127	I have two umbrellas, I can lend you one.
Smart Homes	70	Switch on the lights.
Hospitals	50	Draw the divider halfway.

2. Eliciting Natural Hokkien Translations

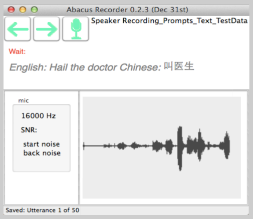
We asked language experts to translate the English sentences into Singapore Hokkien.

3. Collection of Audio Recordings

For test set: We collected 5 sentences each from 10 speakers, for a total of 50 sentences.

4. Lexicon Building

Our baseline lexicon was based on other researchers' work. We constructed 3 sets of data, shown in the table below, with sets A, B and C.



Abacus Recorder 0.2.3 (Dec 31st)

Speaker Recording_Prompts_Text_TestData

Wait:

English: Hail the doctor Chinese: 叫医生

mic

16000 Hz

SNR:

start noise
back noise

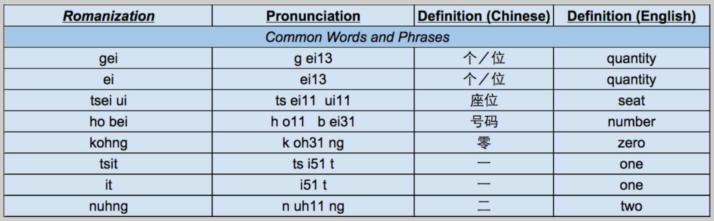
Saved: Utterance 1 of 50

Figure 2 - Recording Software Interface

Set	Corpus	Lexicon
A	N	Original
B	N, CS	Original
C	N, CS, CS2, HS, SmHm	New

Key:

N: Numbers	HS: Hospital Setting
CS: Common sentences	SmHm: Smart Homes
CS2: Common sentences 2	

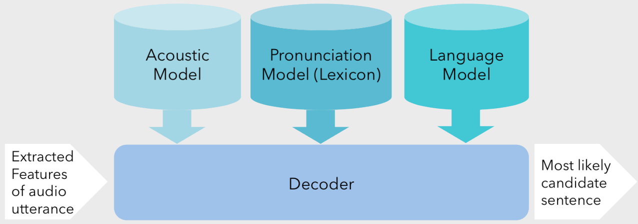


Romanization	Pronunciation	Definition (Chinese)	Definition (English)
Common Words and Phrases			
gei	g ei13	个／位	quantity
ei	ei13	个／位	quantity
tsei ui	ts ei11 ui11	座位	seat
ho bei	h o11 b ei31	号码	number
kohng	k oh31 ng	零	zero
tsit	ts i51 t	一	one
it	i51 t	一	one
nuhng	n uh11 ng	二	two

Figure 3 - Sample Section of Lexicon

HOW DOES AN ASR WORK?

- The ASR makes intelligent guesses using a probabilistic framework in to convert a given audio into text
- Main components: feature extractor and decoder



Acoustic Model
Pronunciation Model (Lexicon)
Language Model
Extracted Features of audio utterance
Decoder
Most likely candidate sentence

Figure 1 - ASR Decoder

The most likely sentence is found using the following equations:

$$S^{*}=argmax_{\forall S}P(S|X)$$

$$P(S|X)=P(W)P(W|\phi)P(\phi|X)$$

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

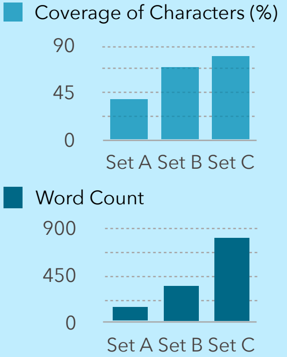
Experiment 3: Effectiveness of AM: Show improvement in performance of ASR with more data.

- Collected 50 utterances for testing
- Word Error Rate (WER)
  - Measure of inaccuracy of speech recogniser

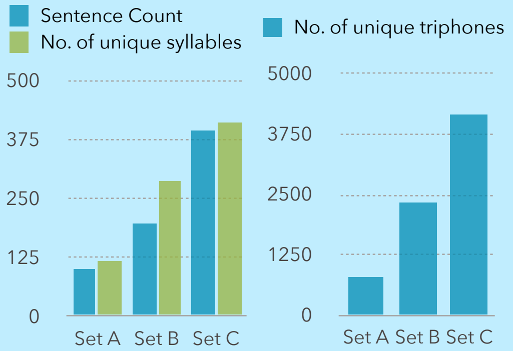
$$WER \triangleq \frac{I+D+S}{N}$$

- Sentence Error Rate (SER)
  - Proportion of sentences that do not match word for word
  - Suitable for assessing voice command applications

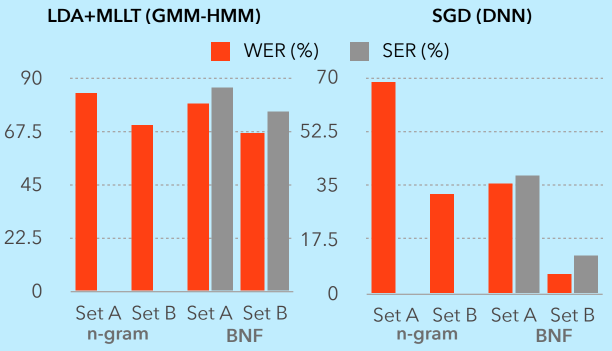
RESULTS



1. Lexicon Quality and Size



2. Corpus Quality and Size



3. Effectiveness of AM

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.