

Sichuan dialect speech recognition with deep LSTM network

Wangyang YING, Lei ZHANG, Hongli DENG

Frontiers of Computer Science, DOI: [10.1007/s11704-018-8030-z](https://doi.org/10.1007/s11704-018-8030-z)

Problems & Ideas

- Problems of Sichuan dialect speech recognition
 - Sichuan dialect speech dataset is very scarce
 - There are many special words and oral language features in Sichuan dialect.
 - A corresponding speech recognition system need to be constructed for Sichuan dialect.
- Ideas
 - A Sichuan dialect speech dataset was created.
 - A Sichuan dialect lexicon was constructed to realize the mapping between the character and the phoneme sequence, and completes the recognition of the special words, polyphones and different pronunciations caused by regional differences in Sichuan dialect.
 - An HMM-LSTM model was used as acoustic model to obtain more context.

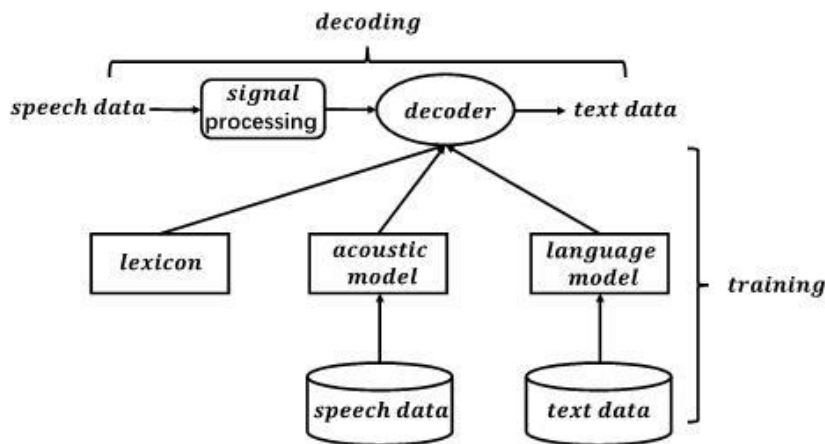


Fig. The architecture of speech recognition

Main Experiment Results

- **The CER of the triple**

model(triple)	parameters	CER(%)
HMM-GMM	-	14.91
HMM-DNN	20M	13.32
HMM-LSTM(no delay)	20M	12.52
HMM-LSTM(delay)	20M	11.34

- **The CER of the mono**

model(mono)	parameters	CER(%)
HMM-GMM	-	32.81
HMM-DNN	9M	27.38
HMM-LSTM(no delay)	9M	22.45
HMM-LSTM(delay)	9M	18.09

Conclusion: In the same condition, our model (HMM-LSTM) can get the best results in the experiments.