

EXHIBIT A

EXHIBIT A: DESCRIPTION OF THE DATASET

OC16-CE80 Chinese-English mixlingual Database			Development Set			Training Set			Testing Set		
Language	Channel	Utterance/Speaker	No. of Speaker	Total Utterances	Recording Hours	No. of Speaker	Total Utterances	Recording Hours	Nos. of Speaker	Total Utterances	Recording Hours
Chinese-English Mixed ASR Database	Mobile	50	140	6974	7.76	1163	58132	63.8	142	7099	7.93

Note:

- * Male and Female are mostly balanced;
- * Original scripts are transcription available;
- * Lexicon is available;
- * There might be a minor discrepancy with the numbers of total utterances;