CORPUS INFORMATION FORM

Check (x) or fill in the blank (_____) if appropriate.

1. Name of the Corpus: CADCC (Chinese Annotated Dialogue and Conversation Corpus)

2. IPR Holder: Phonetics lab., the Institute of Linguistics, **Chinese Academy of Social Sciences**

3. Corpus Type:

- ■ Speech Corpus (X)
- ■ Text Corpus ( )

4. If it is a speech corpus:

- Purpose:
  - ■ ASR (X)
  - ■ TTS ( )
  - ■ Other, please specify phonetic and prosodic analysis for spontaneous speech
- Language:
  - ■ Putonghua (SC) (X)
  - ■ Mandarin in Taiwan (TC) ( )
  - ■ Cantonese in HK (TC) ( )
  - ■ Other, please specify _____
- Style:
  - ■ Read speech ( )
  - ■ Spontaneous speech (X)
  - ■ Conversational speech (X)
  - ■ Other, please specify _____
- Channel:
  - ■ Close-talk Microphone (X)
  - ■ Telephone ( )
  - ■ Mobile phone ( )
  - ■ Other, please specify _____
- Sampling Rate: 16 k Hz
- Sampling Precision:
  - ■ PCM (X), 16 bits per sample
  - ■ A-law ( )
  - ■ Miu-law ( )
  - ■ Other, please specify _____

- Corpus size: 8 hours 22 speakers
- SNR level: _____ dB
- Transcriptions:
  - ■ Character tier (SC) (X)
  - ■ Character tier (TC) ( )
  - ■ Canonical Pinyin tier (X)
  - ■ Other canonical pronunciation tier, please specify _____
  - ■ Surface form IF tier (X)
  - ■ Surface form IPA tier ( )
  - ■ Surface form SAMPA-C tier ( )
  - ■ Other surface form tier, please specify prosodic annotation on break index tier and stress tier
  - ■ Other transcription, please specify _____
  - ■ Other transcription, please specify turn taking tier
  - ■ Other transcription, please specify Miscellaneous tier for no-linguistic and paralinguistic information

5. If it is a text corpus:

- Language:
  - ■ SC ( )
  - ■ TS ( )
  - ■ Other, please specify _____
- Domain:
  - ■ Culture ( )
  - ■ Economy ( )
  - ■ Military ( )
  - ■ News ( )
  - ■ Politics ( )
  - ■ Sciences ( )
  - ■ Sports ( )
  - ■ Other, please specify _____
- Other, please specify _____
  - ■ Other, please specify _____
  - ■ Other, please specify _____
  - ■ Other, please specify _____
  - ■ Other, please specify _____
- Corpus size: _____ Mega characters
- Tag Information:
  - ■ Word segmentation: ( )
  - ■ Part-of-Speech ( )
  - ■ Other, please specify _____
  - ■ Other, please specify _____
  - ■ Other, please specify _____
  - ■ Other, please specify _____

6. A brief Description of the Corpus:

- CADCC is a Chinese spontaneous dialogue and conversation corpus.
- Chinese Characters have been transcribed with spoken phenomena such as breathing, smacking and overlapping.
- Prosodic and segmental annotations with canonical as well as real pronunciation annotation.
- The phonetic annotation has not been finished yet. Only one hour of phonetic annotation is included in this release.

Specifications	Descriptions
Content	No limitation
Total length	8 hours
Speakers	11 male and 11 female speakers
Regional Accent	x
Chinese Character transcription	√
Linguistic annotation	Prosodic, segmental, syntactic
Sampling rate	16 KHz
Storage form	.wav files