CORPUS INFORMATION FORM

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

1. Name of the Corpus: ASCCD (Annotated Speech Corpus of Chinese Discourse)

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 ( )
  - ■ TTS (X)
  - ■ Other, please specify phonetic and prosodic analysis for discourse
- Language:
  - ■ Putonghua (SC) (X)
  - ■ Mandarin in Taiwan (TC) ( )
  - ■ Cantonese in HK (TC) ( )
  - ■ Other, please specify _____
- Style:
  - ■ Read speech (X)
  - ■ Spontaneous speech ( )
  - ■ Conversational speech ( )
  - ■ Other, please specify _____
- Channel:
  - ■ Close-talk Microphone (X)
  - ■ Telephone ( )
  - ■ Mobile phone ( )
  - ■ Other, please specify Laryngograph Channel
- Sampling Rate: 16 k Hz
- Sampling Precision:
  - ■ PCM (X), 16 bits per sample
  - ■ A-law ( )
  - ■ Miu-law ( )
  - ■ Other, please specify _____
- Corpus size: 9 hours 10 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 (to be finished)
  - ■ Other transcription, please specify _____
  - ■ Other transcription, please specify _____
  - ■ Other transcription, please specify _____

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:

- ■ Read discourses, 18 texts with 300-500 syllables each.
- ■ 5 female and 5 male speakers.

- The speech signal was recorded in two channels: speech waveform and the glottal impedance waveform through Laryngograph
- Segmental and prosodic annotations including canonical Pinyin and tone tier, initial / final tier of real pronunciation, sentence mode tier, and stress tier. These annotations are only available for 1 male and 1 female speaker in this release.