

CORPUS INFORMATION FORM

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

1. **Name of the Corpus:** SCSC (Syllable Corpus of Standard Chinese)

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 ()
 - Other, please specify coding and acoustic analysis for monosyllables
- 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 _____
- Sampling Rate: 16 k Hz
- Sampling Precision:
 - PCM (X), 16 bits per sample
 - A-law ()
 - Miu-law ()
 - Other, please specify _____
- Corpus size: _____ hours 15 speakers
- SNR level: _____ dB

- Transcriptions:
- Character tier (SC) (X)
 - Character tier (TC) ()
 - Canonical Pinyin tier ()
 - Other canonical pronunciation tier, please specify _____
 - Surface form IF tier ()
 - Surface form IPA tier ()
 - Surface form SAMPA-C tier ()
 - Other surface form tier, please specify _____
 - 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:

- 15 male speakers;
- 1,268 tonal syllables of Standard Chinese for each speaker;
- The corpus can be easily searched and edited and displayed with management

software.