

CORPUS INFORMATION FORM

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

1. **Name of the Corpus:** CCC-VPR27C2006-50 (CCC 27-Channel
Corpus for Voiceprint Recognition 2006 - 50 speakers)

2. **IPR Holder:** _____

3. **Corpus Type:**

- Speech Corpus (x)
- Text Corpus ()

4. **If it is a speech corpus:**

- Purpose:
 - ASR (x)
 - TTS ()
 - Other, please specify Channel Robust Voiceprint Recognition
- 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 ()
 - Telephone (x)
 - Mobile phone (x)
 - Other, please specify _____
- Sampling Rate: 8 k Hz
- Sampling Precision:
 - PCM (x), 16 bits per sample
 - A-law ()
 - Miu-law ()
 - Other, please specify _____
- Corpus size: _____ hours 50 speakers
- SNR level: >30 dB

- Transcriptions:
 - Character tier (SC) ()
 - 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:

This corpus contains speech data from 50 speakers, each of whom utters a total of 27 segments of speech, with each segment recorded on one of 27 handsets (11 handsets are GSM mobile phones, 10

handsets are CDMA mobile phones and 6 handsets are landline telephones). For each speaker, the transcripts of the 27 utterances are different, and each segment lasts at least 60 seconds. Each speaker uses each handset once.