## CORPUS INFORMATION FORM

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

1. **Name of the Corpus:**    <u>**CCC-VPR2C2005-1000X (**</u><u>CCC 2-Channel</u>

   <u>Corpus for Voiceprint Recognition 2005 – 11kHz</u>**)**

2. **IPR Holder:** _____

3. **Corpus Type:**

   - Speech Corpus       ( x )
   - Text Corpus       ( )

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

   - Purpose:
     - ASR       ( )
     - TTS       ( )
     - Other, please specify    <u>Channel Robust Voiceprint Recognition</u>
   - 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:    <u>11</u> k Hz
   - Sampling Precision:
     - PCM       ( x ), <u>16</u> bits per sample
     - A-law       ( )
     - Miu-law       ( )
     - Other, please specify    _____
   - Corpus size:    _____ hours <u>1000</u> speakers
   - SNR level:    <u>>30</u> 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:

This corpus contains speech data from 1,000 male speakers aged 18-23, each of whom uttered two segments of natural speech without disguising their voice.  For the first segment only the first 70

seconds of speech were recorded; and for the second segment only the first 30 seconds of speech were recorded. Each segment was uttered twice. The first time the utterances were recorded on a GSM mobile phone, and the second time the utterances were recorded on a landline telephone.