**Experimental design**

1. **Recording scene:**

Choose a professional recording room or a quiet room to meet the background noise of no more than 28 dB. For ordinary indoor, a room with smaller space and low sound reflectivity should be chosen to avoid echo and noise; and avoid noisy places and hours such as traffic and human voices. Keep the indoor air fresh, and also pay attention to the environmental factors such as temperature and humidity of the collection scene to ensure the comfort of the speakers.

Recording locations:

√ Paris, France - LPP professional recording room

√ Leiden, Netherlands - student dormitory

√ Eindhoven, Netherlands - suburban villa bedroom

1. **Recording equipment:**

Choose a high-fidelity recording device with a condenser microphone + sound card.

When collecting speech data, keep the microphone radio device slightly away from the speaker's mouth and nose to prevent distortion of speech data due to airflow and excessive volume caused by mood swings. After wearing the recording device to the speaker and before starting, it should be debugged several times (but when the recording actually starts is hidden from the speaker) to obtain a clear and natural voice signal.The sampling rate used is 48 KHz, while quantization is set to 32 bits.

1. **Speaker:**

Recruit 4+16 participants, who are mainly Chinese students or employees from mainland China in Europe. Try to maintain a balanced gender ratio.

To maximize the relaxation of the speakers and obtain the most natural and spontaneous speech data, it is important to ensure that the participants are familiar with each other as much as possible. However, since device calibration is required for each recording and to prevent any interruptions during the data collection process, it is best for the data collector (myself) to be present and fully integrated into the conversation among the three parties to ensure the recording process runs smoothly and efficiently. Therefore, our final plan is to use the data collector (myself), M, as the social starting point and find at least 4 speakers A (A1, A2, A3, A4) who are close to me, and then use A as the social starting point to find 4 friends they are familiar with, B (A1-B1/B2/B3/B4...A4-B1/B2/B3/B4). Based on the specific participants, there are a total of 4x4=16 speech datasets.

The selection of speakers should also consider factors such as age, gender, region, and dialect. Participants should be able to produce speech naturally without significant pronunciation and language barriers. Detailed information about each speaker needs to be registered before the experiment. The specific information form of the speaker is at the end of the text.

In fact, considering the uncertainties of the speaker's performance, equipment failure, weather, strikes, etc., it is highly likely that some speech data may not be usable or meet the requirements. Therefore, we need to contact another A and a matching B group as a backup plan.

1. **Experimental plan:**

Before the end of the recording, we plan not to tell the speakers the true purpose of the experiment in order to avoid them focusing too much on the clarity of their speech output. Therefore, we have designed a "**cover story**". Of course, ethical requirements dictate that participants have the right to be informed, so we will inform the participants of the true purpose of the experiment after all recording tasks are completed. If the participants do not agree to the use of their speech data after knowing the truth, we will fully respect their wishes. Regardless, we will thoroughly anonymize the speaker's information when publishing the speech corpus. Specifically, when recruiting participants, we claim that we will be examining "**the relationship between oral grammar, rhetorical performance, and certain factors.**" Therefore, we need to record the speech and then transcribe it for subsequent discourse analysis, followed by grammar and rhetorical analysis.

\*Choosing this theme as a cover story was based on the following points. **Firstly,** in the complete system of language, speech is the material shell that wraps information, which is easily influenced by factors such as the speaker, topic, emotion, time, location, and environment, and can undergo significant changes (even if the same person speaks the same sound twice, it cannot guarantee complete consistency). The use of vocabulary is also subjectively controlled and filtered by the speaker under the influence of the aforementioned factors, while grammar is the deepest layer hidden in the language system, and even native speakers are hardly aware of the replacement of grammar in daily communication. Moreover, even if the subjects know that the grammar of their speech is under potential observation, it is difficult for them to adjust their expression in real-time. As a language expression mode that spans three levels of vocabulary, grammar, and pragmatics, rhetoric is also unstable and subject to active control by the speaker. However, inserting rhetoric here can further blur the real research object, disrupt the subjects' thinking, and make their spontaneous speech output less focused on a specific level.

**Secondly,** if we give up using linguistic elements such as grammar and rhetoric, which are inherent in speech, to mislead the subjects' attention and turn to other physiological and physical signals, it is difficult to truly obtain natural speech, and the experimental cost is also invisibly increased (it may be necessary to make some props to make the entire experimental process look more real). For example, if a camera is used to deceive the subject into focusing on video data, the subject needs a considerable amount of time to adapt to the experimental environment when communicating, otherwise, they will always be in a restrained state. Similarly, if a breathing belt or finger voltage is used to blur the subject's line of sight, it still does not help to improve the naturalness of speech and may even have a counterproductive effect. For the organizers of the experiment, carrying a large number of props is also inconvenient for cross-regional transfer of recording venues. In this way, it is a very reasonable and economical "cover story" to directly choose to explore the information of other levels of language system carried by the speech signal itself for investigation.

**Thirdly**, with this introduction, we can openly switch topics multiple times in the experimental setup without making the speaker feel strange. This is because we have informed them in advance that the experiment examines the factors that affect oral grammar and rhetoric, and topic is one of them. They can freely speculate and imagine the grammar and rhetoric changes that occur under various topics without deliberately changing their pronunciation habits and mannerisms due to topic switching.

**Moreover**, during my master's degree at the National University of Singapore, my advisor was indeed concerned with various grammar and pragmatic phenomena in spoken language, so I am very familiar with the steps of recording and analyzing discourse. These real research experiences can greatly reduce my psychological burden and prevent me from feeling too guilty about hiding or deceiving anyone. After all, in the experiment, I am also a participant in the conversation, so it is essential to ensure that I, as a fully informed person, can wholeheartedly engage in the dialogue.

**Last but not least**, the sociological background of this research objective is indeed relevant. Chinese people who have studied or worked in Europe for a long time often master a second language, such as English, French, Dutch, etc. They also need to use a foreign language for communication in various occasions, so it is highly likely that their Chinese grammar will become "impure," meaning that their Chinese expression is mixed with a lot of grammar from the Indo-European language. Studying this phenomenon is also valuable, so using this "cover story" can gain the understanding and trust of the participants to a considerable extent.

After laying the foundation in the first step, the subsequent experiment setup is expected to be relatively smooth. We do not need to consider too much about simulating equipment malfunctions midway through to trigger natural spoken communication from the speaker.

After the speaker is seated and all the equipment is calibrated, we can start recording from the self-introduction.

**Phase 1 (10 min):**

The three parties introduce their relationships, how they met, and their most profound impressions of each other. At the same time, I will explain the purpose of the recording and answer any doubts. The speech data in this part may not be natural, so it will not be included in the spontaneous speech corpus and is only used as an introduction.

**Phase Two (60-100 minutes):**

From a list of selected topics, I will choose 3-4 topics for free discussion with two speakers. However, it is important to note that I should speak as little as possible and instead focus on asking questions to facilitate a full discussion between the other two speakers. Silence should be controlled but not avoided. Also, the selected topics should be open-ended and allow for deep exploration and discussion. The list of potential topics for in-depth conversation can be found at the end of this document.

**Phase Three (15-30 minutes):**

The collection of spontaneous speech data has already been completed in Phase Two. Phase Three is mainly focused on collecting speech data from the same speakers, but this time reading the transcribed text in their own speaking style, for later comparison and analysis.

**Appendix 1** Speakers information

|  |  |
| --- | --- |
| **Personal Information** | **Description** |
| Name |  |
| Age |  |
| Gender | Male ( ) |
| Female ( ) |
| Native Language |  |
| Dialect/Accent |  |
| Education Level |  |
| Occupation |  |
| Health Condition | speech disorder: ( ) |
| Other: |
| Second Language Skill Level  (cover story) | English: |
| French: |
| Dutch: |
| Other: |
| Contact | Tel: |
| Wechat: |

**Appendix 2** Chat topic

(Here still need more adjustment and I will translate based on a better one to share with you later.)

↓

