Construction of Chinese Deceptive Speech Detection Corpus

Cheng Fan[1,a], Heming Zhao[1,b*], Xueqin Chen[1], Xiaohe Fan[1], and Shuxi Chen[1]

School of Electronics and Information Engineering, Soochow University, Suzhou, China

[a]fanchengrice@163.com, [b]hmzhao@suda.edu.cn

Keywords: Deception Detection, Speech Signal, Corpus Construction

Abstract. Nowadays, deceptive speech detection has raised more and more interests. The corpus is the foundation of researches in deceptive speech detection. While there are several corpora about English deception detection, few efforts have been put on Chinese which is quite different due to the culture divergence. In this paper, we construct a deceptive and non-deceptive Chinese speech corpus, the SUSP-DSD corpus. We first describe construction approach in detail, and then give a subjective and objective evaluation about the corpus.

Introduction

Deception is generally defined as "a deliberate attempt to mislead others" [3]. Deceptions cause many damaging influences in occasions such as law enforcement, government agencies, border crossings and so on. As a result, identification of deception has become more and more important. Existing researches have covered a lot technologies including heart rate, brain wave, facial expression and so on [4]. These methods usually need complicated equipment connecting to participants to collect signals. Compared to other techniques mentioned above, deception detection in speech only depends on voice signals and can be carried out using simple devices. Moreover, it can reduce influence caused by the complicated wires connection which can make the participants feel nervous. So, it is suitable in many circumstances.

The corpus is the foundation of researches in deceptive speech detection. While there have been several corpora on English deception detection, few efforts have been put on Mandarin which is quite different due to the pronunciation and culture divergence. In this paper, we design and construct a Chinese deception detection corpus which will be discussed in detail later.

Related Work

Recently, there have been a considerable amount of research works on deception detection in speech [6, 5, 7, 8]. Martin Graciarena et al. designed a collection paradigm to elicit within each participant deceptive and non-deceptive speech, from participants who had both financial incentive and motivation [6]. Those who successfully deceived the interviewer that they matched the profile would get $100. K. Gopalan and S. Wenndt obtained utterances of 'No' recorded of a male suspect under criminal investigation [5]. Kirchhübel. C did an experiment based on a mock-theft paradigm [7]. The participants received $5 for participanting with the chance of earning more money through the trial. Patton collected speech signals and videos of facial expression by a particular set of questions [8]. Roberto Cabrera Cosetl interviewed participants with questions from three dynamics, where each interview was recorded [2].

Chinese Deception Detection Corpus Construction

We design and construct a Chinese deception detection corpus, the Soochow University Speech Processing Researches-Deceptive Speech Detection (SUSP-DSD) Corpus. Our corpus contains three parts which will be listed below.

Collection Paradigm Design In part 1, participants only answer 'shi' or 'bushi'. By comparing the answer 'shi' from deceptive and non-deceptive speech of the same participant, we can eliminate the

Table 1: Naming rules of part 1

Part No.	Participant No.	Utterance No.
P1	F1-F20	T1-T8
	M1-M20	F1-F8

influences introduced by different utterances and different people. Thus, we can only concentrate on analysing the differences of deception and truth. In part 2, cards which are numbered from 1 to 10 are prepared, participants need to tell the number he or she has seen in Chinese. Number 1 to number 10 contains most of the vowels in Chinese phonetic, including 'a', 'e', 'i', and 'u'. In part 3, participants can say anything they want during the interview. So the recordings collect most of the common used pronunciations in Chinese.

Participants, Environment, Equipment and Parameters Forty native speakers of Chinese are recruited for the study. All the participants are students of Soochow University. The corpus collection is carried out in a recording room with almost little noise. All the recordings are collected using a Zoom H4n handy recorder. They are recorded to digital audio tape on two channels, sampled at 48kHz and bit depth of 16 bit, and then downsampled to 16kHz, 16bit coding quant and mono by cool edit.

Details of Recording Process Here, we give a detailed introduction about recording process of the three parts one by one.

Part 1: Participants need to answer 10 yes-or-no questions in this part. The interviewer read out one question at a time and the participant give his/her answer. The whole collecting process is carried out in two phases. First, the participants are paid 10 yuan to give the false answers. Second, the participants need to give the real answers this time. Both the two phases compose the deceptive and not-deceptive speeches which are necessary for deception detection. The questions prepared for participants should be brief, clear and understandable. Questions of ambiguity need to be excluded.

Part 2: We prepare cards numbered from 1 to 10, and each number appears four times. The cards are face-down and are shuffled at the beginning. The interviewer asks "which number is it?". Then he will turn over the card right after he finishes asking this question. By doing this, we can record the response time of the participants. The participants are paid 10 yuan to give 20 false answers and 20 true answers.

Part 3: Participants are asked to prepare a topic about their own experience which contains 20 sentences at least. The story is assigned true or false, and it can not be told to others. In the experiment, 34 participants are asked to prepare a fake story, while others need to tell a true story. On the right day of recording, before going into the recording room one by one, they are told that, two interviewers will ask related questions about their stories. The interviewers have no knowledge about whether their stories are true or not. And they can get 100 yuan as a reward if they can convince the two interviewers that they are telling the truth. If only one of the interviewers is persuaded, they can get half. Then, the recording begins. The participants firstly tell their stories. After that, interviewers ask related questions they are interested in and give a judgement whether the participant lies or not. Every two interviewers are responsible for 10 participants to ensure the accuracy of human judgement on site. Participants who tell a lie need to record a true story later in order to collect data for contrastive analysis.

After the interviews, 40 recordings of truth and 34 recordings of deception are obtained, each one lasts about 10 minutes. It is difficult to obtain recordings of lie, while our experiment is effective for this. Participants are motivated to deceive by financial reward in our experiment, it offers a relatively appropriate scenario.

Cut and label After recording, we now have lots of both deceptive and non-deceptive utterances of 'shi'. We cut the utterances and label them using naming rules shown in Table 1.

As we mentioned above, the interviewer asks "which number is it?" and turns over the card right after he finishes asking his question. We cut the utterances from the end of interviewers' voice to the end of the participants'. We label the utterances as Fig. 1 shows.

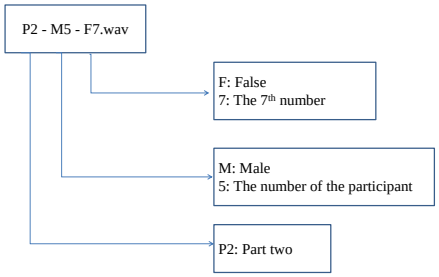

Fig. 1: An example of utterance's name in part 2

Table 2: Naming rules of part 3

Part No.	Participant No.	SU No.
P3	F1-F15	T1-Txx
	M1-M15	F1-Fxx

34 of the participants have interviews of both deception and truth. Excluding the exaggerating performance, we choose 30 of them, half of which are females. The speech is first segmented into sentence-like units (SUs), and labeled as Table 2 shows.

Subjective and Objective Evaluation

Subjective Evaluation To ensure the reliability and validity of our SUSP-DSD corpus, we hire 12 students to listen to the recordings. They are required to fill in a form about whether they think the recording is true or not. Many of them report that the first two parts are difficult to figured out whether they are deceptions or not. They give their judgement totally by wild-eyed guess. So, we only calculate the accuracy of part three. As we mentioned before, interviewers on site will also give their judgements, the accuracy of them is 56.25%. And the accuracy of the subjective listening experiment is 56.86%. A meta-analysis of individual differences in detecting deception [1] shows that people's ability to distinguish deceptive from non-deceptive speech is range from 40.42% (parole officers) to 65.40% (criminals). The accuracy of our corpus is close to the result claimed in [1] which is 54.20% after students' evaluation. That means our corpus is valid.

Objective Evaluation There has no literature suggests that any signal speech feature can distinguish deception from truth reliably and consistently so far. However, as we all know, duration, short-time energy, pitch and formant are important features for Emotion Recognition. And when someone is telling a lie, there is a complicated combination of several specific emotions. Here we list pitch related features and pitch track of P1-M1-F1, P1-M1-F2, P1-M1-T1 and P1-M1-T2 as shown in Table 3 and Fig. 2.

From (b) in Fig. 2 , we can see that the participant obviously hesitated when answering. And features of P1-M1-F2 are different from others. More researches need to be taken to find out which feature or combined feature is suitable to distinguish deception from truth.

Conclusion And Future Work

The construction of deceptive speech detection corpus is very complicated and tedious. In this paper, we design and record the SUSP-DSD corpus which solve the problem of lacking a cleanly recorded corpus of deception and non-deception in Chinese. Our corpus contains both single words and long sentences. The subjective and objective evaluation has proved the reliability and validity of our corpus.

Table 3: Pitch related features of 4 utterances in part 1

	P1-M1-F1	**P1-M1-F2**	**P1-M1-T1**	**P1-M1-T2**
P_{max}	120.3974	158.9337	120.5544	126.6514
P_{mean}	98.0392	119.7578	105.6343	100.6246
P_{rate}	5.2913	2.1910	5.6449	5.3809
P_{std}	12.9359	17.7769	10.4765	14.6127

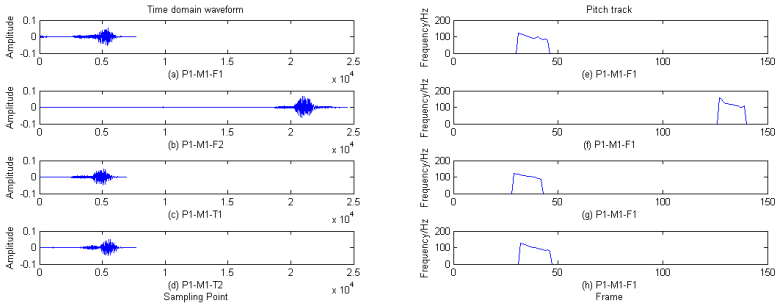

Fig. 2: Time domain waveform and pitch track of 4 utterances in part 1

As for future work, we will try to improve the size of our corpus and investigate the performance of some machine learning algorithms for deception detection when applying on our corpus.

Acknowledgment

This work is supported by the National Natural Science Foundation of China (Grant No. 61372146).

References

[1] M. G. Aamodt and H. Custer. Who can best catch a liar? a meta-analysis of individual differences in detecting deception. *Forensic Examiner*, 15(1):6--11, 2006.

[2] R. C. Cósetl and J. López. Voice stress detection: A method for stress analysis detecting fluctuations on lippold microtremor spectrum using fft. In *Electrical Communications and Computers*, pages 184--189. IEEE, 2011.

[3] B. M. DePaulo, J. J. Lindsay, B. E. Malone, L. Muhlenbruck, K. Charlton, and H. Cooper. Cues to deception. *Psychological bulletin*, 129(1):74, 2003.

[4] P. Ekman, M. O'Sullivan, W. V. Friesen, and K. R. Scherer. Invited article: Face, voice, and body in detecting deceit. *Journal of nonverbal behavior*, 15(2):125--135, 1991.

[5] K. Gopalan and S. Wenndt. Speech analysis using modulation-based features for detecting deception. In *Digital Signal Processing*, pages 619--622. IEEE, 2007.

[6] M. Graciarena, E. Shriberg, A. Stolcke, F. Enos, J. Hirschberg, and S. Kajarekar. Combining prosodic lexical and cepstral systems for deceptive speech detection. In *ICASSP*. IEEE, 2006.

[7] C. Kirchhübel and D. M. Howard. Detecting suspicious behaviour using speech: Acoustic correlates of deceptive speech--an exploratory investigation. *Applied ergonomics*, 44(5):694--702, 2013.

[8] M. Patton. Decision support for rapid assessment of truth and deception using automated assessment technologies and kiosk-based embodied conversational agents. 2009.