Detection of Special Scams by Emotion Analysis

Shota Endo\textsuperscript{a}) and Koji Tanida
Nippon Telegraph and Telephone Corporation
Midori-cho, Musashino-shi, Tokyo, 180-8585, Japan
\textsuperscript{a}) syouta.endou.ns@hco.ntt.co.jp

Abstract: Telephone scams called "special scams" have become a major problem, with people being defrauded of nearly 30 billion yen every year in Japan. In this study, we used emotion analysis techniques based on speech analysis to detect whether victims were being deceived by analyzing multiple emotional parameters of victims in a special scam. In the verification, we first clarified the characteristics of the emotions generated in victims by two scam methods. Next, we improved the accuracy of detecting the risk of a victim being deceived by limiting the analysis section.

Keywords: special scam, emotion, speech analysis, emotion analysis

Classification: Network System

References

[1] Tu, Huahong, et al. "Sok: Everyone hates robocalls: A survey of techniques against telephone spam." 2016 IEEE Symposium on Security and Privacy (SP). IEEE, 2016.
[2] Maggi, Federico. "Are the con artists back? A preliminary analysis of modern phone frauds." 2010 10th IEEE International Conference on Computer and Information Technology. IEEE, 2010.
[3] Lee, Kong Aik, et al. "I4u submission to NIST SRE 2018: Leveraging from a decade of shared experiences." arXiv preprint arXiv: 1904.07386 (2019).
[4] Zhang, Lei, et al. "Voiceprint mimicry attack towards speaker verification system in smart home." IEEE INFOCOM 2020-IEEE Conference on Computer Communications. IEEE, 2020.
[5] Matsuo, Naoshi, Shoji Hayakawa, and Shouji Harada. "Technology to detect levels of stress based on voice information." Fujitsu Sci. Tech. J 51.4 (2015): 48-54.
[6] Stanley, Jennifer Tehan, and Fredda Blanchard-Fields. "Challenges older adults face in detecting deceit: The role of emotion recognition." Psychology and Aging 23.1 (2008): 24.
[7] Katharina Kircanski, Nanna Notthoff, Doug Shadel, et al. “Heightened Emotional States Increase Susceptibility to Fraud in Older Adults.” Stanford Center on Longevity. 9 Aug 2016.
[8] Shao, Jingjin, et al. "Why are older adults victims of fraud? Current knowledge and prospects regarding older adults’ vulnerability to fraud." Journal of elder abuse & neglect 31.3 (2019): 225-243.
[9] Ververidis, Dimitrios, and Constantine Kotropoulos. "Emotional speech recognition: Resources, features, and methods." Speech communication 48.9 (2006): 1162-1181.

[10] Chuang, Ze-Jing, and Chung-Hsien Wu. "Multi-modal emotion recognition from speech and text." International Journal of Computational Linguistics & Chinese Language Processing, Volume 9, Number 2, August 2004: Special Issue on New Trends of Speech and Language Processing. 2004.

[11] "Status of special scam awareness and arrests, “National Police Agency, https://www.npa.go.jp/bureau/criminal/souni/tokususagi/tokushusagi_toukei2020.pdf, 2021/6/6 (in Japanese)

[12] "Emotion AI in the age of the voice,” Empath, https://webempath.net/lp-jpn/, 2020/10/19 (in Japanese)

[13] Fletcher, G. C. L., et al. "The role of non-technical skills in anaesthesia: a review of current literature." British journal of anaesthesia 88.3 (2002): 418-429.

[14] Roseman, Ira J., Martin S. Spindel, and Paul E. Jose. "Appraisals of emotion-eliciting events: Testing a theory of discrete emotions." Journal of personality and social psychology 59.5 (1990): 899.

1 Introduction

Telephone scams called "special scams" have become a major problem, with people being defrauded of nearly 30 billion yen every year in Japan. A special scam is a general term for crimes in which criminals convince victims to trust them over the phone and transfer cash to the criminals’ accounts without meeting them in person. In many cases, the scammer will try to gain the victim's trust by impersonating a relative, acquaintance, bank, or public institution. Various research institutes and companies in Japan and abroad are researching countermeasures against special scams, but so far no effective countermeasures have been established and the number of victims has remained high. Therefore, countermeasure technologies against special scams urgently need to be established.

2 Purpose

There are two main ways to detect special scams: analysis of call control signals that control incoming and outgoing calls, and analysis of media signals. In the analysis of call control signals, a method has been proposed to analyze the time of day, frequency, and destination number of the scammer's calls [1]. However, this method is not practical in Japan, where prior consent of the user is required for data collection. Therefore, in this research, which aims to establish countermeasure technologies against special scams in Japan, we use media signal analysis to counteract special scams. Up to now, countermeasures against special scams have been proposed that use media signal analysis technology, such as context analysis of conversational content using voice calls, identification of individuals by voiceprint analysis, and detection of special scams using acoustic characteristics such as voice pitch and level. Although these previous studies have reported high detection accuracy, challenges remain. In context analysis [2], special scams are
detected by performing analysis on the basis of keywords learned in advance. However, this method may falsely detect normal conversations with similar content, so it needs to be combined with other analysis methods. In voiceprint analysis [3], a caller is identified by extracting the characteristics of the user's voice and comparing them with the characteristics registered in a database in advance. However, since this method analyzes the scammer's voice, the detection accuracy may decrease as the accuracy of voiceprint mimicry technology improves [4]. For detecting special scams on the basis of acoustic characteristics, a method has been proposed [5] to detect the victim's stress level by extracting acoustic characteristics such as voice height and level from media signals. However, special scam techniques have recently become more diverse, and scammers have found more ways to avoid stressing out victims. For example, in the refund-scam, the victims are directed to an ATM under the pretense that money will be returned to them and so are unlikely to be stressed. We consider that the detection of a single psychological state, as in previous studies, is insufficient against such diversified special scams. Therefore, in this research, we focused on emotion analysis as a technology that can be used to deal with various types of special scams. The reason for this is that elderly people, who are the most frequent victims of special scams, reportedly tend to have more emotional ups and downs [6-8]. Emotional analysis extracts acoustic characteristics from speech and outputs the emotional parameters contained in the conversation on the basis of these characteristics [9]. There are also reports on improving the accuracy of emotion analysis by combining it with context analysis [10]. The purpose of this study is to analyze multiple emotional parameters of victims in two special scam methods to determine whether the victims are being deceived or not.

3 Proposed Method

In this study, we determine the emotional parameters to be analyzed from the victim's voice in accordance with the special scam methods and determine whether the victim is being deceived or not. In this study, the special scams methods are based on the classification defined by the National Police Agency (NPA) of Japan. In this paper, the oreore-scam and refund-scam are selected for verification. We selected these two methods because the NPA's website has more audio data for them than other methods and they account for about 50% of all special scams [11]. Next, we hypothesized about the emotions of the victims who were deceived by each method. In the oreore-scam, a caller pretends to be a relative of the victim, confesses to the victim that he or she has caused a costly accident or mishap, and asks the victim for money to compensate a third-party. Therefore, we hypothesized that emotions such as sadness and anxiety would be observed. In the refund-scam, a scammer posing as a government or bank employee directs a victim to an ATM to transfer money under the pretense that a refund will be issued. Therefore, we hypothesized that emotions such as joy would be observed. We also hypothesized the victims would feel urgency because of the time pressure common to the two methods. In the next chapter, we verified the differences in the emotions of the deceived victims caused by the different methods.
4 Verification

4.1 Verification Procedure

In this verification, we used a commercially available emotion analysis engine (Empath API, Empath Inc.) that takes media signals as input and outputs five emotions (joy, sorrow, anger, calm, and energy) as independent scores from 0 to 50 [12]. We obtained 12 audio recordings of oreore-scams and 15 audio recordings of refund-scams from the websites of the NPA and prefectural police. Then, we verified the emotion analysis of the collected audio with the following procedure.

1) Collected speech is classified into Classes I and II on the basis of the following situation and keyword-based criteria.

*Oreore*-scam:
   - Class I
     - The scammer suggested the next action and the victim agreed.
     - The victim made a statement of concern about the incident or accident allegedly caused by the scammer.
     - The victim made a statement to the effect that he would lend money to the scammer.
   - Class II
     - The scammer or victim disconnected in the middle of a call.
     - The victim made a statement that he doubted the scammer's identity and the content of the conversation.

*Refund*-scam:
   - Class I
     - The scammer suggested the next action and the victim agreed.
     - The victim made a statement of gratitude to the scammer.
     - The victim asked a question about the refund procedure.
   - Class II
     - The scammer or victim disconnected in the middle of a call.
     - The victim made a statement that he doubted the scammer's identity and the content of the conversation.

2) Convert to a format that can be used as input for Empath API (11 kHz/16 bit/mono).

3) Extract the sections that contain only the victim's voice, and exclude the sections that contain the scammer's voice.

4) In accordance with the specifications of Empath API, if the audio chunk extracted in 3) exceeded 5.0 seconds, it was re-segmented at 5.0 seconds from the start point. However, chunks shorter than 0.33 second are excluded from the analysis.

5) Each chunk created in (4) was analyzed by Empath API to calculate the emotion score. Then, they were averaged to obtain a score per call sample.

4.2 Results and Discussion of the Verification

As a result of the validation in the *oreore*-scam, we focused on "sorrow," which showed the largest score difference between Classes I and II, and "calmness" as
the equivalent of the stress level in a previous study [5]. For "sorrow," the mean score in Class I was 209 % of that in Class II, so a clear tendency for higher sorrow was obtained. On the other hand, for "calmness," scores tended to be lower in Class I, but there was variation among the samples. From the obtained results, the ratio of "sorrow" and "calmness" scores in the oreore-scam is expressed as the risk level of being deceived by the oreore-scam (shown in (1) of Fig. 1). As a result, the average risk level in Class I was 200 % of that in Class II. Similarly, the ratio of "joy" and "calmness" scores in the refund-scam is expressed as the risk level of being deceived by the refund-scam (shown in (1) of Fig. 2). As a result, the average risk level in Class I was 235 % of that in Class II. We found that different emotions have to be detected in the oreore-scam and refund-scam. However, these results were not accurate enough to be used to perform two-class classification by machine learning. The reason for this is that redundant conversations were included in the analysis section. Next, we focused on the content of the conversation to improve detection accuracy. An example of Class I in the oreore-scam is shown in Fig. 3. At the two points where the "sorrow" score was highest, the "calmness" score was lowest. This change in emotion could be due to the influence of the scammer's speech that existed between each chunk. Scammers are reported to exert time pressure, which makes it difficult for the victim to make appropriate decisions [13]. In addition, situations involving punishment (e.g., compensating a third-party in the oreore-scam) cause distress, while situations involving rewards (e.g., a refund in the refund-scam) cause pleasure [14]. We selected the following scammer’s utterances and calculated the score using only the victim's chunks for that utterance (shown in (2) of Fig. 1).

- The scammer's statement that the deadline for the compensation payment is that day.
- The scammer's statement that he caused an accident/incident and needs to pay compensation.

Similarly, for the refund-scam, we selected the following scammer’s utterances and calculated the score using only the victim's chunks for that utterance (Shown in (2) of Fig. 2).

- The scammer's statement that the deadline for the procedural refund is that day.
- The scammer's statement that the victim needs to go through a procedure to receive a refund.

As a result, the average risk levels in Class I in the oreore-scam and refund-scam are 443 % and 1190 % of those in Class II, respectively. Therefore, by limiting the analysis to a specific section, the accuracy of calculating the risk level for each scam was improved. On the other hand, about 10 % of the samples need further improvement. As a solution to this, we consider that a more detailed analysis of a scam situation (the person the scammer is impersonating, relationship with the victim, number of calls) will improve accuracy.

5 Summary and Future Work

In this study, we proposed a method that combines context analysis and emotion
analysis to improve the accuracy of special scam countermeasures. As a result of the validation, we found that emotions differ depending on the scam method and that limiting the analysis section improves the detection accuracy. On the other hand, some samples need to be improved. In future work, we will try to improve the accuracy by further subdividing special scam situations.

**Fig. 1.** Risk level in oreore-scam (1) for all sections and (2) for only specific sections

**Fig. 2.** Risk level in refund-scam (1) for all sections and (2) for only specific sections

**Fig. 3.** Example of Class I in oreore-scam