Non English Robocalls: What are They Saying?

Sathvik Prasad, Bradley Reaves
North Carolina State University
{*snprasad,bgreaves*}*@ncsu.edu*

***Abstract*—Automated scam calls, also called robocalls, are one of the most widespread security problems affecting phone users in the United States. Some of the most egregious robocalls target vulnerable segments of our society. These segments often consist of non-English speaking phone users (e.g., international students, recent immigrants, tourists, etc.). Fraudulent robocall campaigns that target these populations often use languages like Spanish, Mandarin, Hindi, Arabic, etc., in the robocall audio. Regulatory authorities, enforcement agencies, and researchers investigating illegal robocalling campaigns lack automated tools to study such non-English campaigns, thereby struggling to extract meaningful insights from bulk robocall data. Relying entirely on manual analysis or expecting investigators to be fluent in numerous languages substantially limits action against illegal non-English robocalling campaigns. Furthermore, existing robocall audio analysis techniques focus solely on English robocalls. We propose developing a semi-automated robocall audio analysis pipeline to handle real-world non-English robocalls. We intend to build this pipeline using pre-trained multi-lingual speech transformer models.**

Introduction: Pre-recorded or machine-generated scam calls, also called *robocalls*, are often used as a medium to defraud phone users in the United States. They cause substantial harm to their targets by stealing personal information (Social Security Number, bank account details, credit card information, etc.) or defrauding them of large sums of money. Frustrated and victimized individuals continue to express their concerns to regulatory and enforcement agencies. Fraudulent robocalls are among the top consumer complaints to the FCC, the FTC, the FBI, and other state and federal enforcement agencies. Certain illegal robocallers [1], [2], [3] design their scheme so that they can specifically target US phone subscribers who are vulnerable to such scams. Such vulnerable segments include recent immigrants, international students, and tourists visiting the US. For example, malicious robocalling campaigns impersonate the Department of Homeland Security [8], the Customs and Border Protection agency [7], the Department of Homeland Security [8], and the US Citizenship and Immigration Services (USCIS) [6]. Many of these targets are non-native English speakers and are more fluent in other languages. Malicious robocallers take advantage of this fact and design their fraud schemes using languages like Mandarin, Spanish, Arabic, and Hindi. This enables the bad actors to establish a communication channel with the victims within the vulnerable population segment and eventually defraud them.

Problem: Existing techniques on analyzing bulk robocall audio data [5] focus primarily on English robocall audio recordings. Developing automated tools to ingest, process, and extract insights from non-English robocalls will empower stakeholders to take swift action against such illegal operations. Recent advances in multi-lingual speech processing techniques using transformer models [4] offer promising new directions toward developing robocall audio content analysis pipelines.

Proposed Work: We intend to leverage the dataset and open-source code released by Prasad et al. [5] and extend the SnorCall system to include multi-lingual capabilities. We hope‘ to explore recent breakthroughs in multi-lingual speech language models and study their effectiveness in developing multi-lingual robocall audio analysis pipelines.

1) **Dataset Creation:** Curate a dataset of real-world robocall audio recordings that spans multiple languages (Spanish, Mandarin, Arabic, English, etc.)
2) **Literature Survey:** Systematically identify open-source, multi-lingual speech transformer models available across various platforms (HuggingFace, GitHub, etc.)
3) **Evaluation:** Benchmark the capabilities of these open-source, multi-lingual models to compute audio embedding, speaker embedding, and other audio features from real-world robocall audio recordings.
4) **Pipeline Development:** Develop a robocall audio analysis pipeline using these models to extract named entities, callback numbers, and other operational attributes that can aid investigators in swiftly analyzing Non-English robocall data.

Preliminary Results: We have developed and deployed a telephony honeypot that automatically answers phone calls and collects call signaling information while recording the call audio. Through manual analysis, we have uncovered numerous non-English robocalling campaigns. However, a deeper investigation of these campaigns requires a robust robocall audio analysis pipeline.

Goal: Through this work, we explore the capabilities of speech transformer models to process audio data that is inherently noisy and lossy (due to the nature of the phone network). Since most speech-language models are trained on high-quality audio data, evaluating their performance on real-world robocall audio data enables researchers to adapt multi-lingual speech transformer models for further analysis.

Acknowledgements: The authors would like to thank the anonymous reviewers for their helpful comments. This material is based upon work supported by the National Science Foundation under grant number CNS-2142930. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation, other funding agencies or financial supporters.

References

[1] FCC. Chinese Americans Targeted in Consulate Phone Scam, 2022. https://www.fcc.gov/chinese-americans-targeted-consulate-phone-scam.

[2] FTC. Fake calls about your SSN, 2018. https://consumer.ftc.gov/consumer-alerts/2018/12/fake-calls-about-your-ssn?page=29.

[3] FTC. Scammers impersonate the Chinese Consulate, 2018. https://consumer.ftc.gov/consumer-alerts/2018/04/scammers-impersonate-chinese-consulate?page=3.

[4] HuggingFace. Pre-trained models for automatic speech recognition, 2024. https://huggingface.co/learn/audio-course/chapter5/asr_models.

[5] S. Prasad, T. Dunlap, A. Ross, and B. Reaves. Diving into robocall content with SnorCall. In *32nd USENIX Security Symposium (USENIX Security 23)*, pages 427–444, Anaheim, CA, Aug. 2023. USENIX Association.

[6] U.S. Citizenship and Immigration Services. Common Scams - USCIS, 2023. https://www.uscis.gov/scams-fraud-and-misconduct/avoid-scams/common-scams.

[7] U.S. Customs and Border Protection. CBP Phone Scam Continues to Target Citizens Callers seek information to bypass financial protocols , 2023. https://www.cbp.gov/newsroom/local-media-release/cbp-phone-scam-continues-target-citizens-callers-seek-information.

[8] U.S. Immigration and Customs Enforcement. Fraud alert: Scammers spoofing HSI telephone numbers to target victims , 2023. https://www.ice.gov/news/releases/fraud-alert-scammers-spoofing-hsi-telephone-numbers-target-victims.