Challenges in Multi-Speaker Privacy

Anastasiia Korenevskaia

[1]Radboud University, Netherlands

anastasiia.korenevskaia@ru.nl

1. What is multi-speaker privacy?

This paper reflects on the meaning of the term "multi-speaker privacy" and discusses its different aspects. To better understand the scenario, we focus on the following questions:

- What information in a recording should be protected?
- How do people present on a recording relate to each other?
- Are people on the recording aware that they are being monitored, recorded, or analyzed?

By projecting existing works onto this three-dimensional perspective, we analyze potential research directions and highlight those that warrant further exploration.

1.1. Multi-speaker anonymization

Some works explicitly mention the term "multi-speaker privacy" in the sense of anonymizing every speaker present on the recording. For example, Miao et al. [1] introduce a benchmark for multi-speaker anonymization, designed to conceal speakers' identities while preserving content, naturalness, and speaker distinguishability. These objectives are closely related to those of the VoicePrivacy Challenge [1], although that initiative focuses on a single-speaker scenario and therefore does not include distinguishability as a utility goal. Target-speaker anonymization in multi-speaker recordings was also addressed in [2]. Ito et al. [3] present a dialogue-aware anonymization system that substitutes each speaker in the dialog with a pseudo-speaker. However, this area remains underexplored, as traditional speech privacy research mostly focuses on a single-speaker anonymization.

1.2. Privacy of a group

Anonymization is not the only possible privacy objective. A multi-speaker scenario may expose sensitive information about individual participants, relationships between them, or about the group of speakers in general. For individual information, Mairesse et al. [4] assess personal qualities (openness, agreeableness, etc.) based on audio and textual information. For interpersonal relationships, Matic et al. [5] identify forms of interactions between people at work based on their conversations and postures. For the group characteristic, Hung et al. [6] predict group cohesion based on spoken conversations. Inferring information about a group of people is often referred to as "Group privacy". Loi et al. [7] see negative consequences in revealing such group information as political affiliation, religious views, and controversial behaviors. To give one example of inferring group characteristics, Shen et al. [8] predict group gender (same or mixed) by investigating turn-taking patterns in the group. While inferential conclusions about groups are widely studied, they are often not examined from a privacy perspective.

1.3. Privacy of bystanders and stakeholders

Another important dimension of multi-speaker privacy lies in capturing background or incidental human-related information within a recording. A prominent example is the fine incurred by the LaLiga app for tracking users' locations and activating smartphone microphones to identify bars broadcasting soccer matches without a license [9]. This resulted in recordings not only without user consent but also capturing bystanders' voices and ambient sounds. Such scenarios, where background speakers are unaware of being recorded, pose significant privacy risks and lie at the core of multi-speaker privacy.

With the age of AI-powered tools, new complexities arise. Applications for recording and summarizing meetings proliferate widely. For instance, Google's NotebookLM [2], a tool powered by a large language model, can analyze and summarize uploaded audio, potentially including recordings of group discussions where not all participants have explicitly consented to this type of processing.

Yet another interpretation is drawn from [10]. Zhou et al. address the privacy of individuals who share smart devices with the primary owner but may have different expectations regarding the protection of their conversational data. They describe the scenarios where guests or visitors conversations are recorded by a smart speaker. Unlike passive bystanders, visitors are typically aware of the device, yet their privacy preferences can conflict with those of other stakeholders. The authors present a tool to facilitate negotiation of privacy settings in shared spaces.

2. Future work

Future research should focus on investigating privacy for multi-speaker settings in its wide, multi-faceted sense. To address this problem and also get familiar with the state-of-the-art anonymization tools, we will begin by examining the impact of multi-speaker anonymization on speech that contains backchannels. Specifically, we will investigate whether the presence of backchanneling in multi-speaker audio affects anonymization performance and determine if current speaker diarization and verification systems can reliably identify individuals based on their backchannel signals alone. Through this work, we aim to better understand the privacy threats in multi-speaker environments, particularly those related to the re-identification of both primary and incidental speakers.

[1]https://www.voiceprivacychallenge.org/

[2]https://notebooklm.google/

3. Acknowledgements

The work was supported by the Privacy for Smart Speech Technology joint doctoral programme (PSST!), funded by the European Union's Horizon Europe research and innovation programme under grant agreement No 101168193.

4. References

[1] X. Miao, R. Tao, C. Zeng, and X. Wang, "A benchmark for multi-speaker anonymization," *IEEE Transactions on Information Forensics and Security*, 2025.

[2] N. Tomashenko, J. Yamagishi, X. Wang, Y. Liu, and E. Vincent, "Target speaker anonymization in multi-speaker recordings," *arXiv preprint arXiv:2510.09307*, 2025.

[3] A. Ito and K. Itou, "Dialogue-pseudo: A speaker pseudonymization framework for privacy protection in dialogue speech data," in *2025 International Symposium on Multimedia (ISM)*. IEEE, 2025, pp. 156–163.

[4] F. Mairesse, M. A. Walker, M. R. Mehl, and R. K. Moore, "Using linguistic cues for the automatic recognition of personality in conversation and text," *Journal of artificial intelligence research*, vol. 30, pp. 457–500, 2007.

[5] A. Matic, V. Osmani, and O. Mayora-Ibarra, "Mobile monitoring of formal and informal social interactions at workplace," in *Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing: Adjunct Publication*, 2014, pp. 1035–1044.

[6] H. Hung and D. Gatica-Perez, "Estimating cohesion in small groups using audio-visual nonverbal behavior," *IEEE Transactions on Multimedia*, vol. 12, no. 6, pp. 563–575, 2010.

[7] M. Loi and M. Christen, "Two concepts of group privacy," *Philosophy & Technology*, vol. 33, no. 2, pp. 207–224, 2020.

[8] J. Shen, O. Lederman, J. Cao, F. Berg, S. Tang, and A. Pentland, "Gina: Group gender identification using privacy-sensitive audio data," in *2018 IEEE International Conference on Data Mining (ICDM)*. IEEE Computer Society, 2018, pp. 457–466.

[9] N. Lomas, "Laliga fined $280k for soccer app's privacy-violating spy mode," 2019. [Online]. Available: https://techcrunch.com/2019/06/12/laliga-fined-280k-for-soccer-apps-privacy-violating-spy-mode/

[10] H. Zhou, M. Goel, and Y. Agarwal, "Bring privacy to the table: Interactive negotiation for privacy settings of shared sensing devices," in *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems*, 2024, pp. 1–22.