

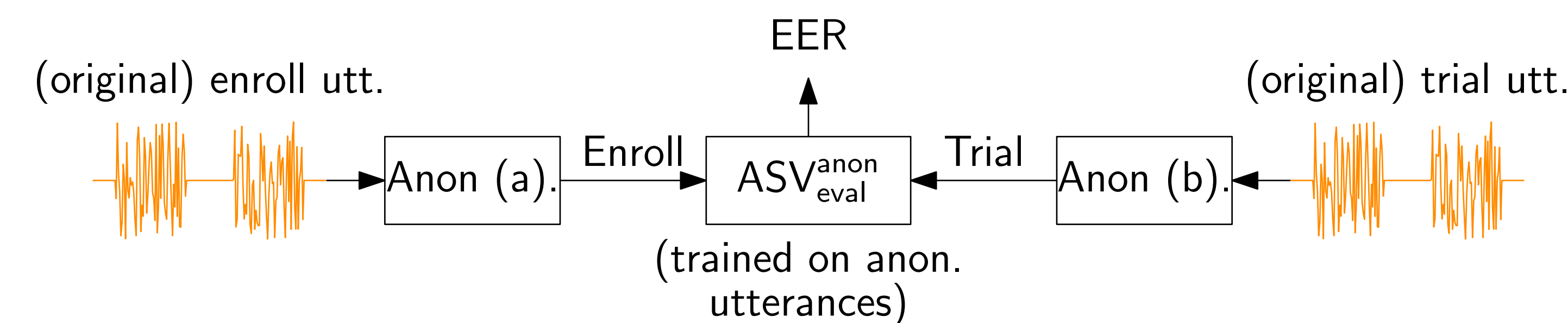
You Are What You Say: Exploiting Linguistic Content for VoicePrivacy Attacks

Ünal Ege Gaznepoglu, Anna Leschanowsky, Ahmad Aloradi, Prachi Singh
Daniel Tenbrinck, Emanuël A. P. Habets, Nils Peters



1. Introduction

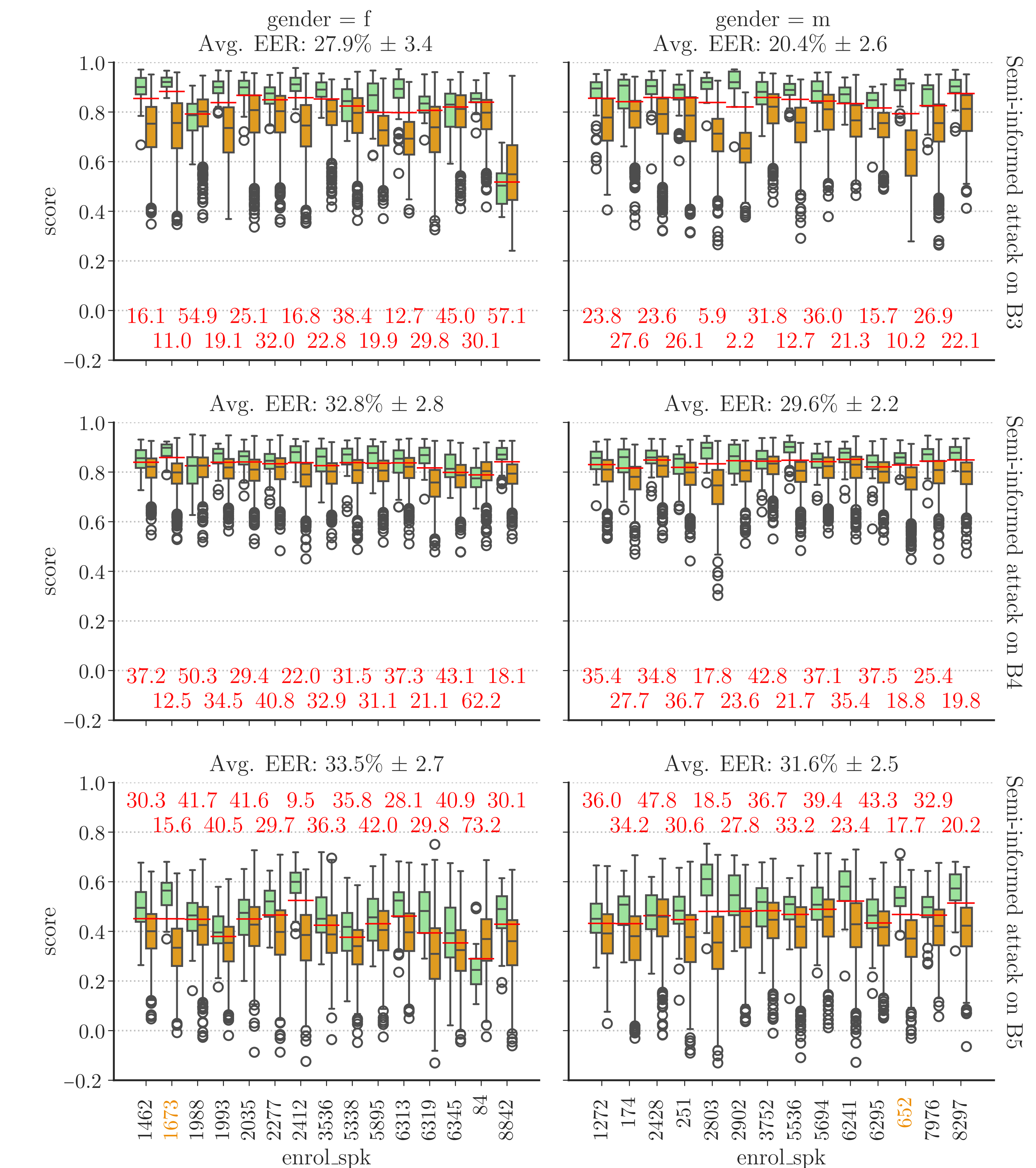
- Speaker anonymization systems process speech signals to hide the speaker characteristics, meanwhile retaining useful information for downstream tasks (e.g., ASR, SER, ...)
- Their privacy benefits are evaluated using ASV attacks (e.g., ECAPA-TDNN) per VoicePrivacy Challenge evaluation protocol [1]
 - Trained on LibriSpeech train-clean-360
 - Evaluated on LibriSpeech dev and test



2. Speaker-Level Analysis

We analyzed the Libri-dev semi-informed attacker scores

- Green bars: scores for positive pairs (matching speakers)
- Orange bars: scores for negative pairs (different speakers)
- Red lines: speaker-level equal error rate (EER) threshold
- Red numbers: corresponding EER [%]

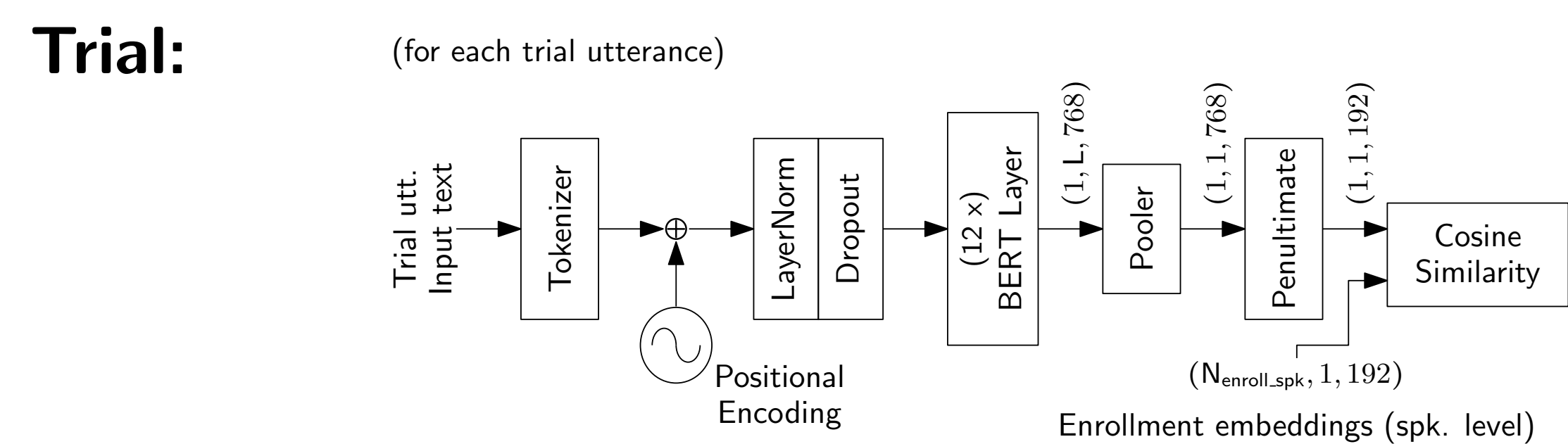
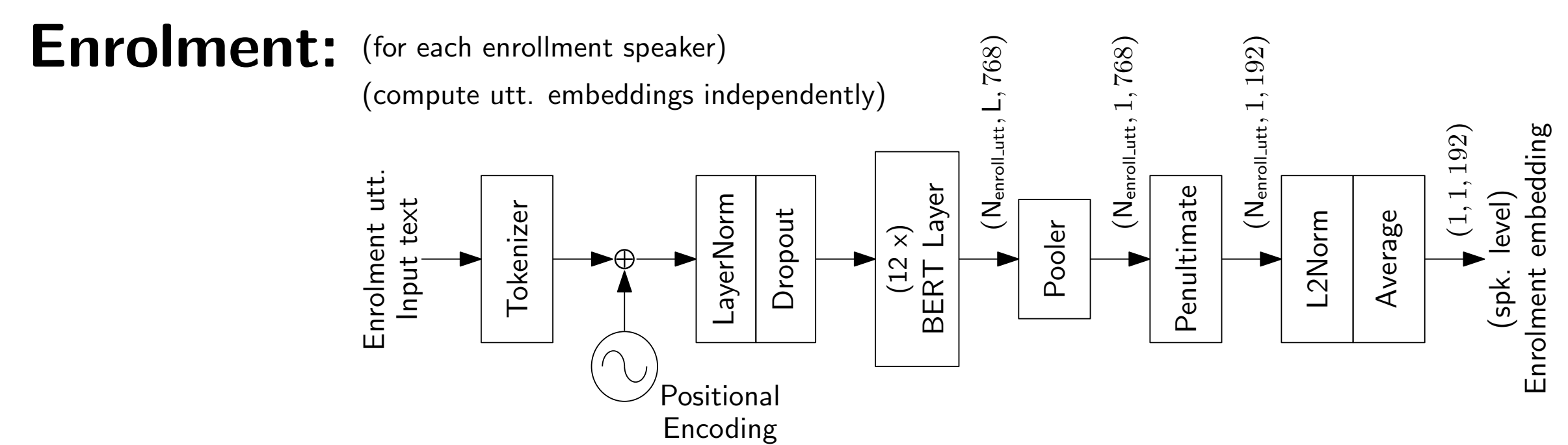
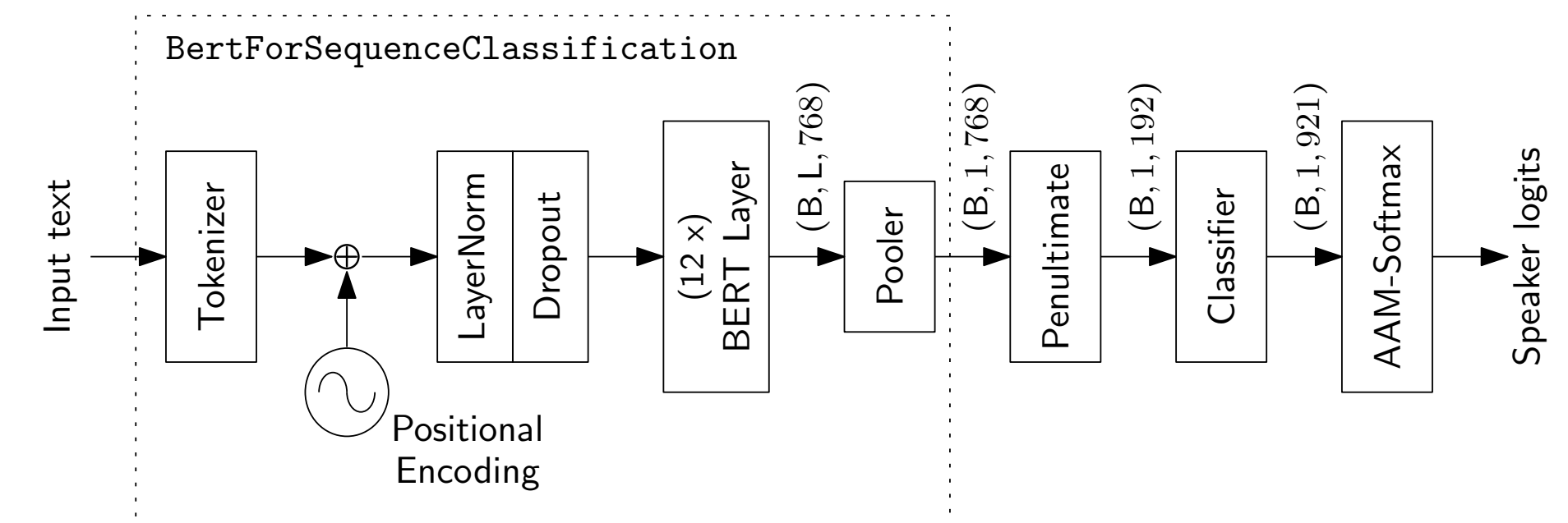


Unexpected findings:

- Performance across speakers varies greatly
- Global threshold (or average EER) overestimates the privacy benefits
- Certain speakers (e.g., 1673, 652) were not successfully anonymized by any of the evaluated anonymization systems (B3, B4, and B5) → WHY?

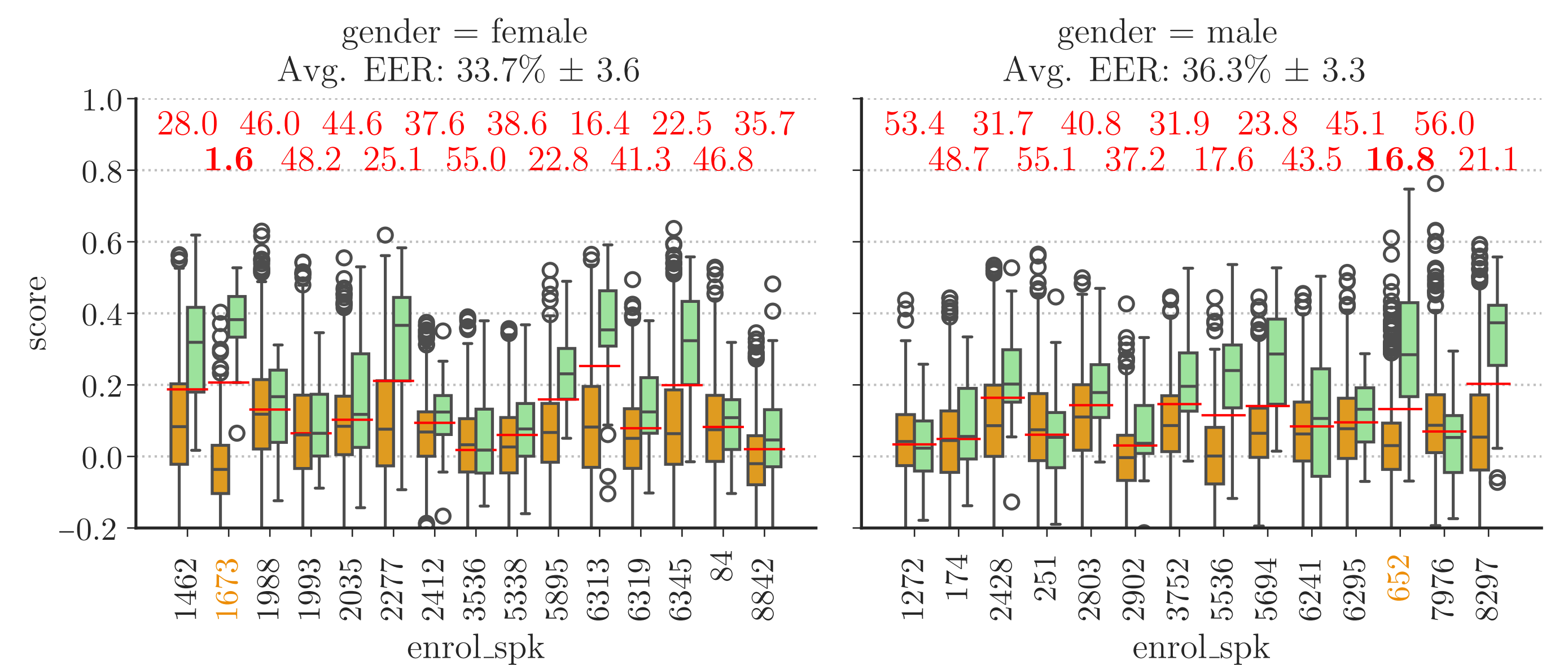
3. Text-Based Attack

Hypothesis: Distinctive linguistic features can reveal speaker identity
How to test: Via a text-based VoicePrivacy Attack based on BERT [2]
Training:



4. Results

VPC 2024 evaluation



Text content is sufficient to de-anonymize speakers 1673 and 652!

Explainability study (using Layer Integrated Gradients [3])

Speaker	True Label	Predicted Label	Attr. Label	Attr. Score	Word Importance
652	1	1 (0.64)	1	4.64	[CLS] then take it out of the roast ##ing pan and put it into a cass ##ero ##le after sp ##rin ##kling it with two ounce ##s of flour [SEP]
652	1	1 (0.55)	1	3.15	[CLS] we finally got him to select the one prized above all others and this is what chef sc ##hei ##ler gave us [SEP]
652	1	1 (0.63)	1	4.65	[CLS] ce ##ler ##y and nut salad cut enough ce ##ler ##y fine to measure two cups add one cup of finely shredded or shaved cabbage and one and one half cups of walnut meat ##s broken in small pieces but not chopped [SEP]
652	1	1 (0.59)	1	3.00	[CLS] cha ##bl ##is a white burgundy dry and of agree ##able aroma [SEP]
652	1	1 (0.59)	1	2.07	[CLS] when done take the meat out of the sauce [SEP]
7976	1	1 (0.50)	1	1.83	[CLS] gr ##eth ##el began to cry but it was all useless for the old witch made her do as she wanted [SEP]
7976	1	1 (0.63)	1	1.73	[CLS] go back to the regiment he said smiling all will be needed [SEP]
7976	1	0 (0.50)	0	1.75	[CLS] how can we feed our children when we have no more than we can eat ourselves [SEP]
7976	1	0 (0.49)	0	1.62	[CLS] long ago there lived a merchant who had three daughters [SEP]
7976	1	1 (0.66)	1	2.90	[CLS] fifteen officers of our little half regiment were dead or wounded [SEP]

Text-based attacker decisions can be traced to thematic keywords!

References

- [1] P. Champion et al., 3rd VoicePrivacy Challenge Evaluation Plan (v2.1), 2024. [Online]. Available: https://www.voiceprivacychallenge.org/vp2024/docs/VoicePrivacy_2024_Eval_Plan_v2.1.pdf.
- [2] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding," in Proc. Conf. North American Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL-HLT), 2019.
- [3] M. Sundararajan, A. Taly, and Q. Yan, "Axiomatic Attribution for Deep Networks," in Proc. Intl. Conf. on Machine Learning (ICML), 2017.