Does Speaker Anonymization Really Work?

CNT 5410: Computer and Network Security

Nigama Annapurna Dendukuri

Kartheek Reddy Gade

Pavan Siva Sai Savaram

Sai Ram Varma Budharaju

Venkata Sai Karthik Metlapalli

Overview

- Introduction
- Background
- Approach
- Analysis of Papers
  - Anonymization models overview
  - Threat model
  - Comparision and selection
- Results
- Challenges
- Conclusion
- Future work

Introduction

- **Speaker anonymization** involves concealing the identity of the speaker, without affecting the intelligibility of the content.
  - Why: Ensures confidentiality by protecting privacy of individuals .
  - How: Can be done using physical or logical techniques.
- A good anonymization technique:
  - Has low similarity to original voice
  - Is intelligible
  - Sounds relatively natural
  - Is difficult to reverse engineer
- But the current voice anonymization methods lack robust security and requires balancing the need for efficient anonymization with processing speed

Background

- Voice Anonymization is the process of obscuring a speaker's identity while preserving speech content.
- Existing anonymization techniques use:
  - Physical techniques:
    - Noise addition
  - Logical techniques:
    - Voice Transformation -> Changing the pitch
    - Voice Conversion -> Converting from male voice to female.
    - Voice Synthesis -> Text to Speech
    - Voice Signal Processing -> Manipulating the acoustic signals

Understanding Related concepts

Key concepts to understand:

Automatic Speaker Verification (ASV): Verify the speaker's identity

Automatic Speech Recognition (ASR): Transcribing speech without identifying the speaker.

Metrics used:

EER (Equal Error Rate): Crucial for ASV, measuring the point where false acceptance equals false rejection.

WER (Word Error Rate):Important for ASR, assessing transcription accuracy.

Voice privacy challenge 2020:

Focused on advancing speaker anonymization systems and evaluating their efficacy.

Involved open-source models and comprehensive criteria, diverse datasets.

Approach

Thoroughly study speaker anonymization methods.

Analyze strengths and vulnerabilities.

Explore ethical implications.

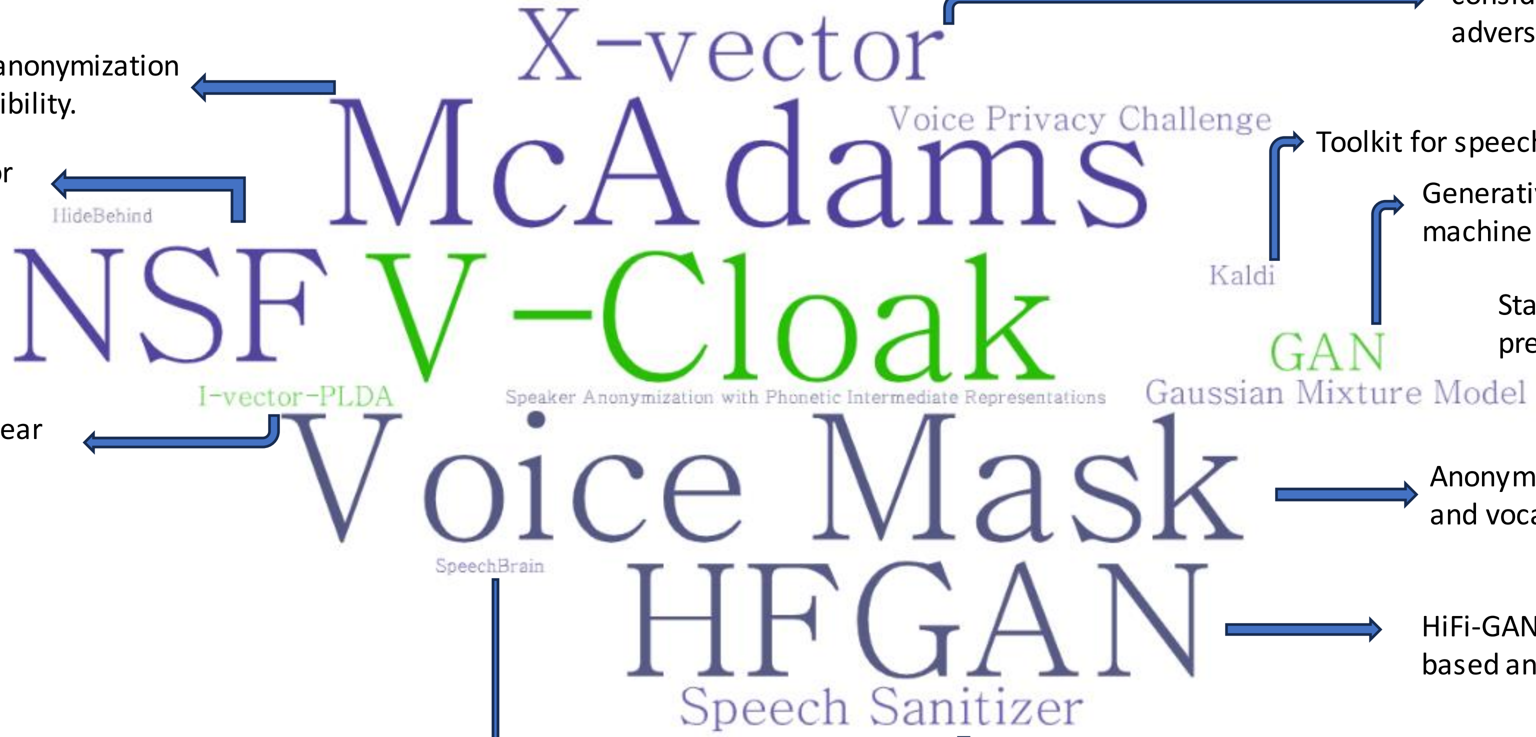
Engage in discussions about potential misuse and ethical responsibilities.

Related Work

Modifies spectral envelope for anonymization while preserving speech intelligibility.

Neural Source-Filter using x-vector based synthesis for speaker anonymization.

Uses i-vectors and Probabilistic Linear Discriminant Analysis for speaker recognition.



More recent speaker embeddings, considered robust but vulnerable to adversarial attacks

Toolkit for speech recognition.

Generative Adversarial Networks, a class of machine learning models.

Statistical model for representing the presence of a speaker.

Anonymization using frequency warping and vocal tract length normalization.

HiFi-GAN, a generative adversarial network-based anonymization technique.

State-of-the-art anonymization combining speech content desensitization and voice anonymization.

Toolkit for speech and audio processing.

Threat model

- The threat models defined to assess the effectiveness of anonymization techniques are:
  - Ignorant Adversary (A1)
  - Semi-Informed Adversary (A2)
  - Informed Adversary (A3)

- **NSF**: Effective against **A1** adversaries, but it is less effective against A2 and A3 adversaries.
- **HFGAN**: More effective against **A2** adversaries than NSF, but it is still less effective against A3 adversaries.
- **McAdams**: Effective against **A1 and A2** adversaries, but it is less effective against A3 adversaries.
- **VoiceMask**: More effective against **A3** adversaries than NSF, HFGAN, and McAdams.
- **V-Cloak**: Most effective technique against **all three** types of adversaries

Model Selection

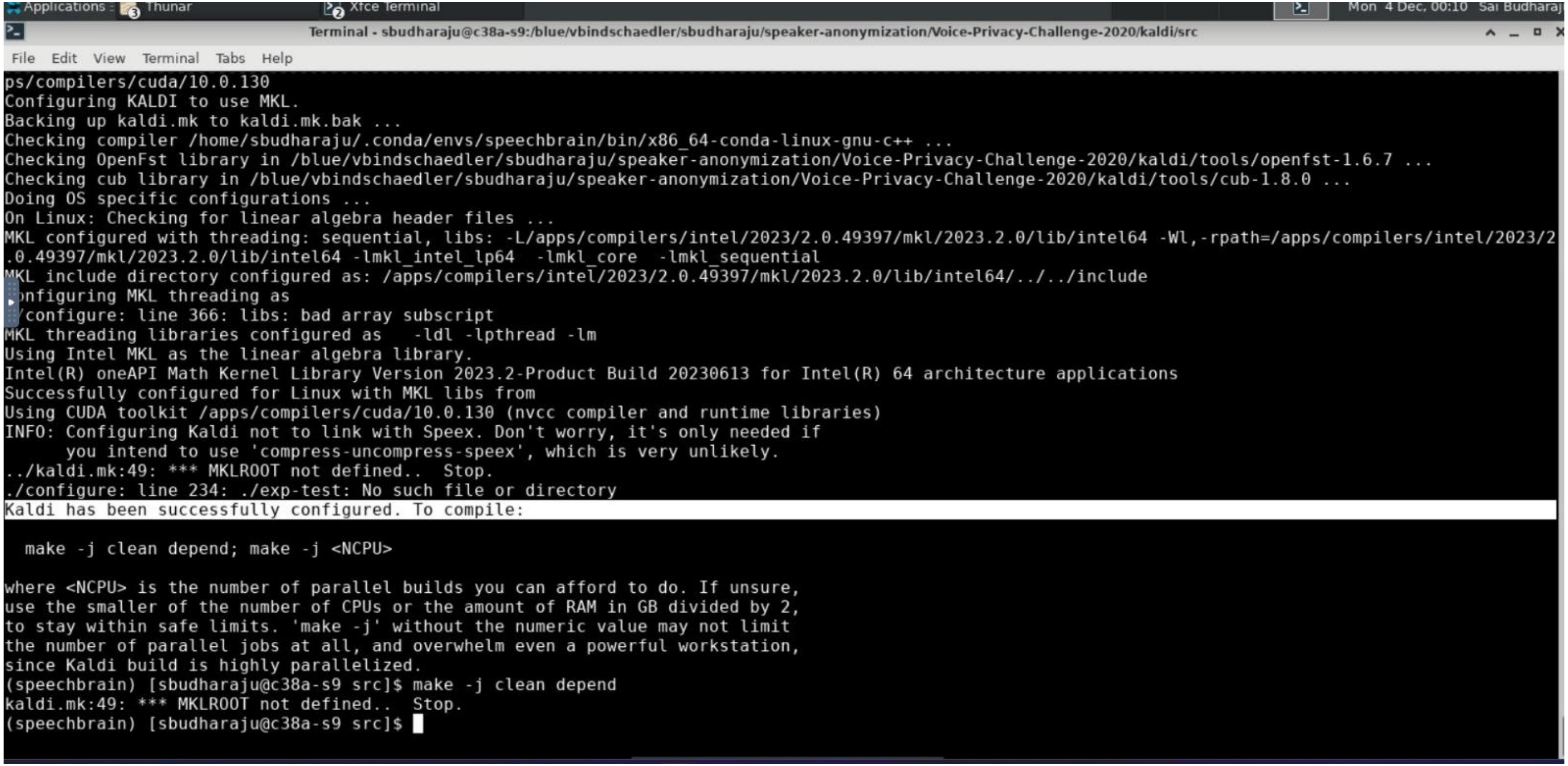
- Based on the performance of the models we have decided to perform in-depth analysis on V-cloak
- Real world effectiveness was assessed in ensuring high-level security
- V-cloak was evaluated to determine if it is the latest and most advanced approach.
- The challenges associated with V-cloak was also assessed

Model		B0 (%)	NSF (%)			HFGAN (%)			McAdams (%)			VoiceMask (%)			V-CLOAK (%)		
		EER	MMR	WMR	EER	MMR	WMR	EER	MMR	WMR	EER	MMR	WMR	EER	MMR	WMR	EER
ASV	EP	3.72	88.89	3.89	38.09	87.33	3.89	42.21	46.53	3.89	20.69	70.15	3.89	23.40	97.90	3.89	42.21
	XV	5.74	87.33	4.73	34.47	88.89	4.73	39.05	84.28	4.73	40.19	95.73	4.73	37.79	100.0	4.73	44.73
	DP	3.72	93.97	4.05	39.70	89.39	4.05	33.13	80.15	4.05	35.00	99.24	4.05	41.37	99.77	4.05	49.47
	AVG	4.39	90.06	4.22	37.42	88.54	4.22	38.13	70.32	4.22	31.96	88.37	4.22	34.19	99.22	4.22	45.47
	WCS	-	87.33	3.89	34.47	87.33	3.89	33.13	46.53	3.89	20.69	70.15	3.89	23.40	97.90	3.89	42.21

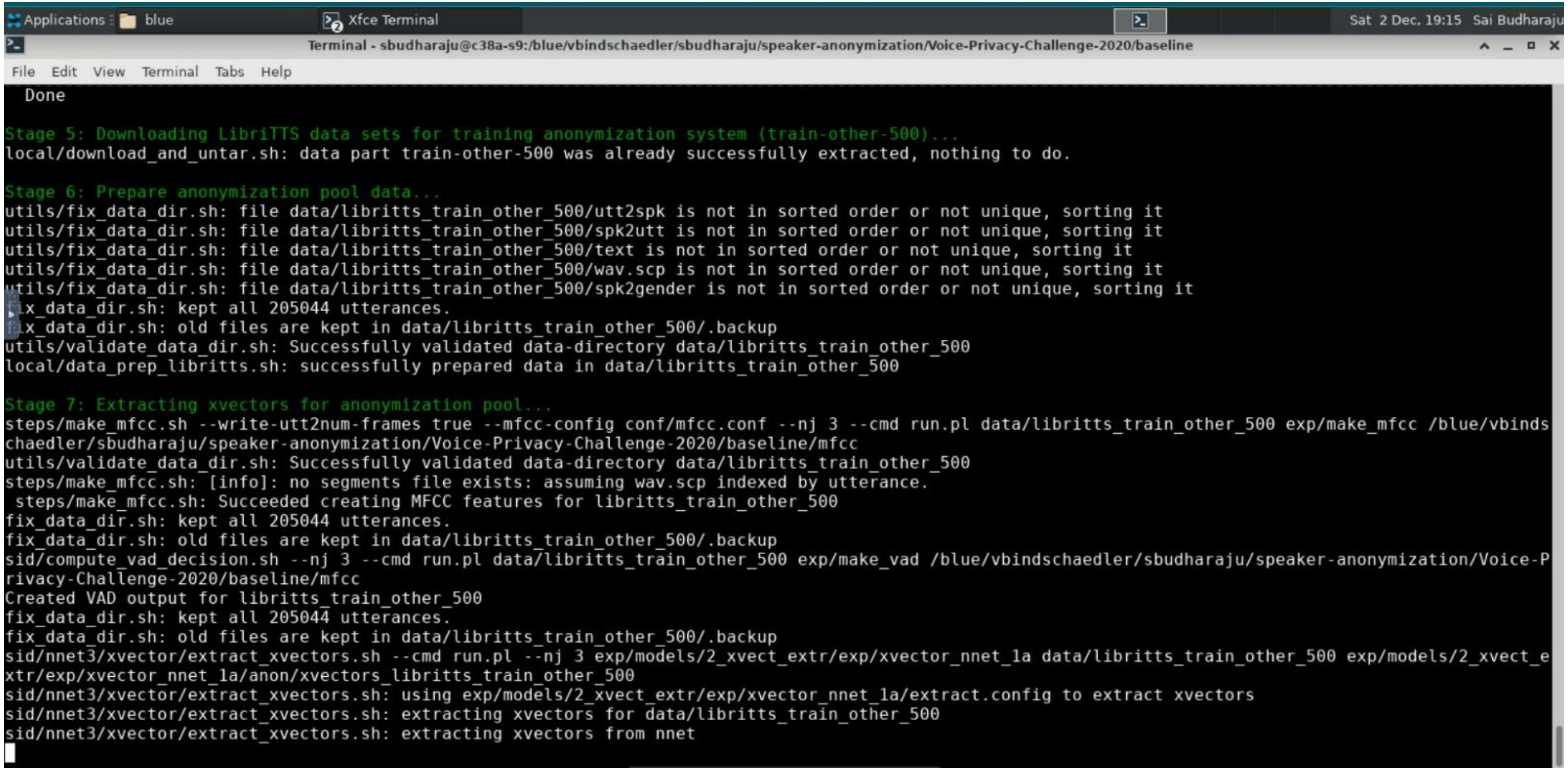
AVG: average, **WCS**: worst-case scenario. **EP**: ECAPA-TDNN, **XV**: X-vector, **DP**: DeepSpeaker.

Performance results for the models

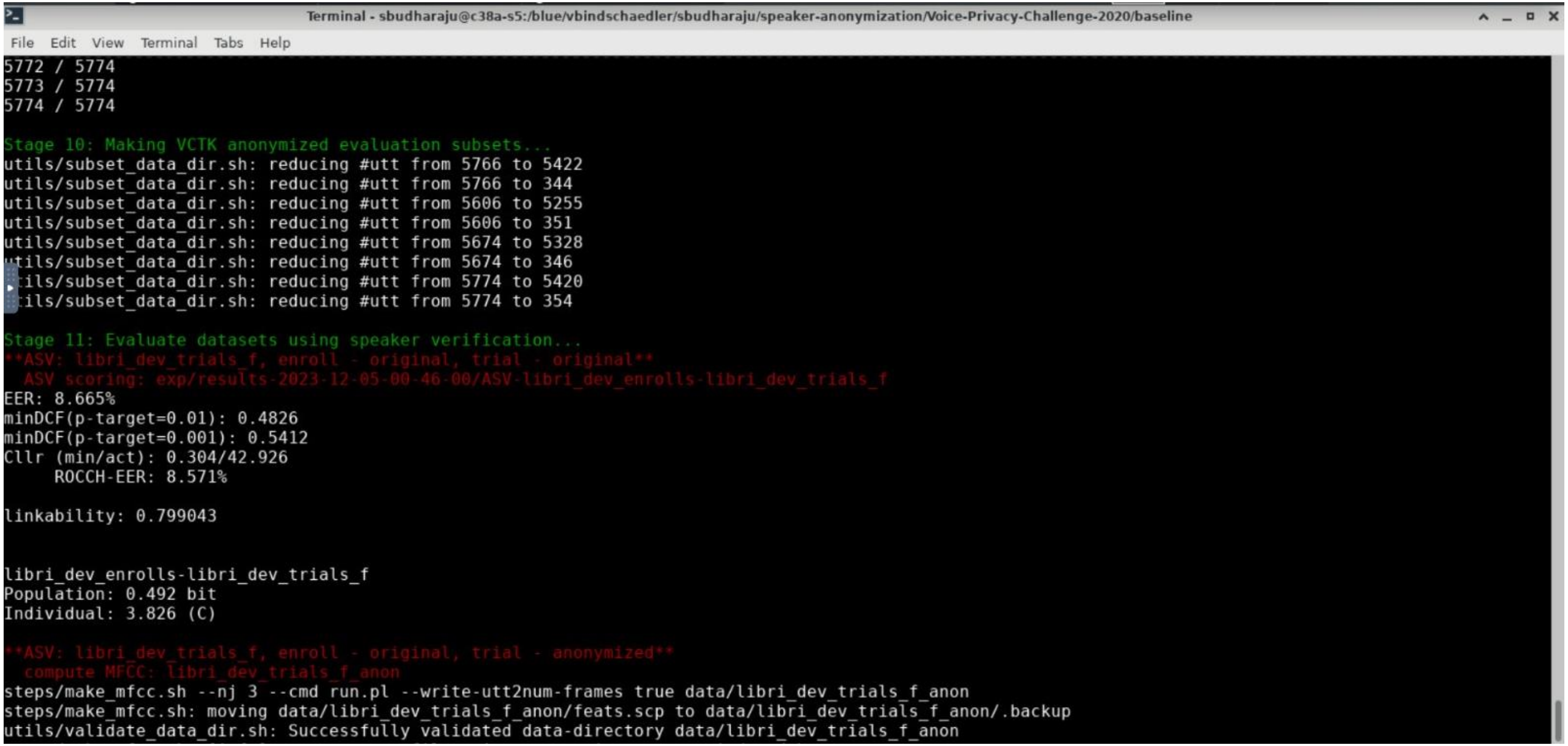
Kaldi configure setup success



X-vector extraction step in Voice Privacy Challenge



Output of Voice Privacy challenge ASR - McAdams



Results

- **Equal Error Rate (EER)**
  - 8.65%
- **Minimum Detection Cost Function (minDCF)**
  - 0.4826 for p(target=0.01)
  - 0.5412 for p(target=0.001)
- **Clustering Identification Rate (ClIR)**
  - 0.3044 (min action)
  - 0.4296 (overall)
- **Receiver Operating Characteristic Curve - Equal Error Rate (ROCCH-EER)**
  - 8.57%
- **Linkability**
  - 0.799043

Challenges

HARDWARE LIMITATIONS: REQUIRED GPU IN MANY MODELS. NEEDED ACCESS TO HIPERGATOR

TOOLKIT AND SOFTWARE DEPENDENCIES: KALDI AND SPEECHBRAIN WERE COMPLEX AND RESOURCE INTENSIVE

LINUX COMPATIBILITY: REQUIRED DUAL BOOTING THE SYSTEM.

Conclusion

- **Speaker Verification Effectiveness**
    - EER and ROCCH-EER rates indicate moderate system performance in verification tasks.
- **Accuracy and Cost Trade-off**
    - minDCF values point to a reasonable misclassification cost, highlighting potential for optimization.
- **Linkability vs Anonymization**
    - High linkability score suggests effective speaker linking, yet poses a challenge for stronger anonymization.
- **Key Takeaway**
    - The system shows promise but requires advancements in balancing verification accuracy with anonymization needs.

Future Work

Deep Learning Models

- Utilize advanced models trained on diverse datasets for improved accuracy.

- **Adversarial Training**
  - Train systems against adversarial attacks for enhanced security.
- **User-Centric Design**
  - Focus on user-friendly interfaces and feedback for practical use.
- **Ethical and Legal Considerations**
  - Address ethical implications and adhere to privacy regulations.
- **Collaboration with Research Community**
  - Engage with challenges and incorporate community-driven advancements.

Thank You!