



Radboud–Aalto Submission to the Voice Privacy Challenge 2026

Jiusi Zheng^{1, 2}, Martha Larson¹, Tom Bäckström²

¹ Radboud University, the Netherlands

² Aalto University, Finland

Presenter: Jiusi Zheng

Date: September 26, 2026

Motivation

VPC 2026 evaluation protocol:

- Utility Evaluation:
 - LibriSpeech (Intelligibility WER)
 - IEMOCAP (Emotion UAR)
- Privacy Evaluation:
 - LibriSpeech (Speaker Verification EER)
- Attacker:
 - Fine tuned on LibriSpeech

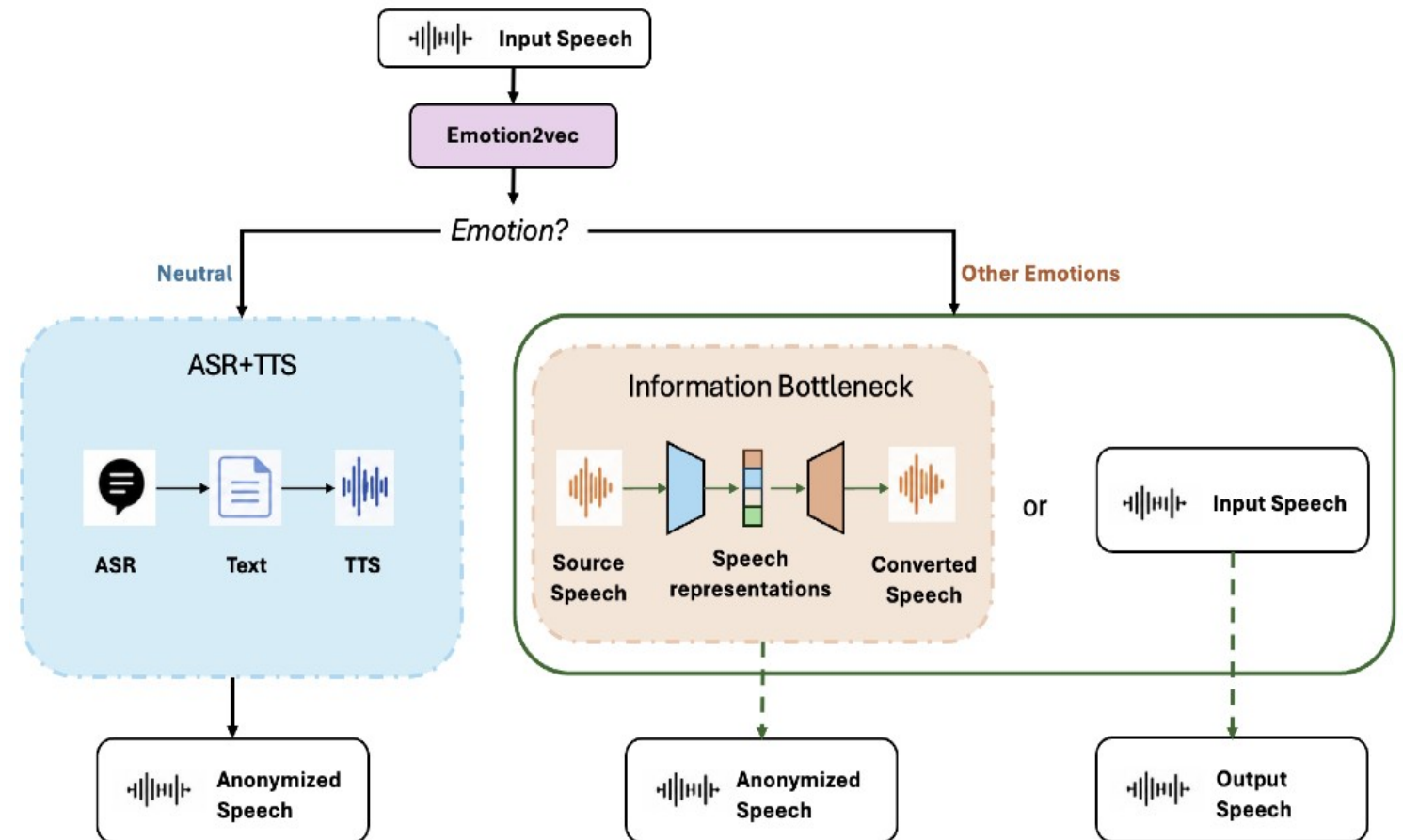
Observation:

- The emotion distribution is imbalanced across different test sets: The percentage of LibriSpeech-dev utterances classified as emotional (non-neutral) by emotion2vec+base model is 9.3%, and for IEMOCAP-test, the ratio is 64.1%.

Proposed System

Main Idea: we dynamically selects the voice anonymization strategy based on the **predicted emotional state** of the input speech.

Module	Open-source link
Whisper [9]	https://github.com/openai/whisper
VITS [10]	https://rhasspy.github.io/piper-samples/#en_US-libritts_r-medium
UUVIC [11]	https://github.com/b04901014/UUVIC
Emotion2vec+base [8]	https://huggingface.co/emotion2vec/emotion2vec_plus_base



Key results:

Our system achieves a UAR of 64.69% on the IEMOCAP-test and 63.08% on the IEMOCAP-dev. It also achieves an EER of 33.61% and 36.41% on the LibriSpeech test and development sets, outperforming all official baselines.

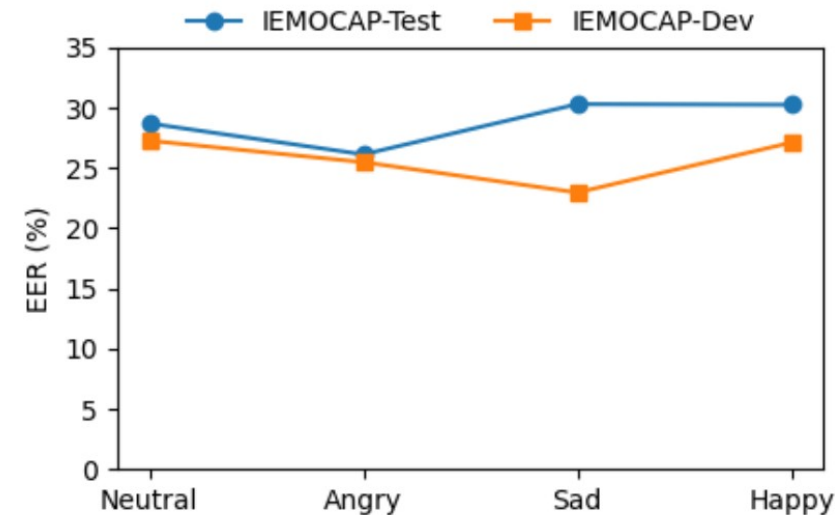
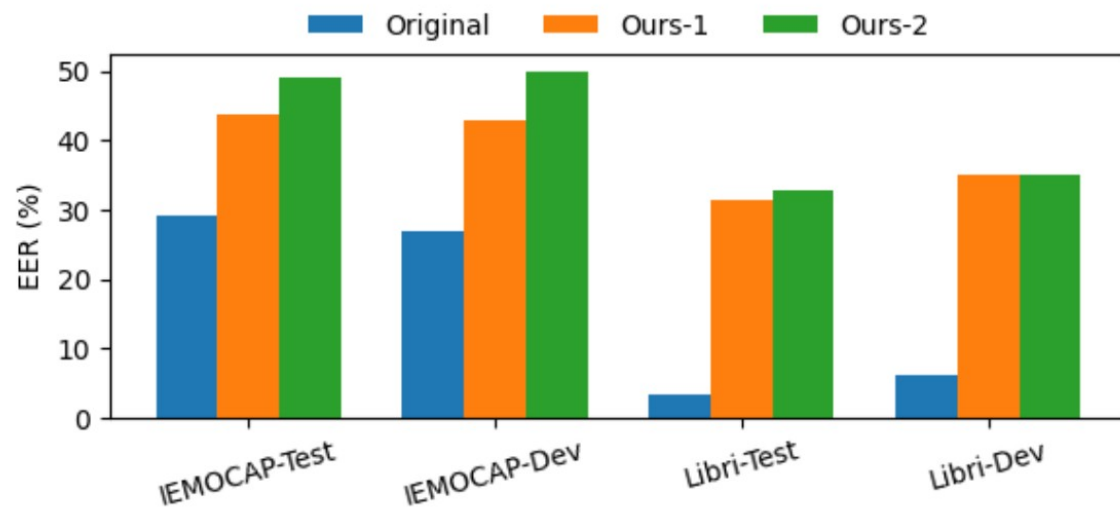
	Libri-dev		Libri-test		IEMOCAP-dev					IEMOCAP-test				
	WER	EER	WER	EER	sad_acc	ang_acc	hap_acc	neu_acc	UAR	sad_acc	ang_acc	hap_acc	neu_acc	UAR
Orig.	1.80	6.08	1.84	3.41	63.63	79.78	66.95	65.97	69.08	72.57	72.82	67.19	71.66	71.06
B2	10.44	4.23	9.96	3.72	31.14	64.44	67.09	57.97	55.66	32.78	56.97	57.98	66.23	53.49
B3	4.31	16.32	4.31	14.24	0.42	70.45	47.45	34.00	38.08	0.30	61.07	39.78	39.79	35.24
B4	6.16	15.28	5.90	11.13	-	-	-	-	41.97	-	-	-	-	42.78
B5	4.91	20.05	4.44	17.05	8.26	64.51	37.27	50.29	40.08	4.57	58.05	33.81	56.58	38.25
Ours-1	3.87	34.94	3.79	31.49	66.17	66.07	45.43	51.46	57.28	68.72	64.06	42.15	49.81	56.18
Ours-2	3.89	35.16	3.72	32.82	59.35	81.04	61.68	50.23	63.08	68.76	75.54	62.89	51.57	64.69

Limitation and Future Work:

Limitations

- Privacy and emotion preservation are evaluated on different datasets, with substantially different emotion distributions.
- Non-neutral speech receives weak or no anonymization, which may expose speaker identity.
- Domain mismatch may weaken the attacker and lead to overestimated privacy protection.
- Emotion alone does not explain the attack-performance gap across datasets.

Future work: Develop unified evaluation protocols and stronger cross-domain attackers for diverse and emotionally expressive speech.



*Thanks for your
attention!*

I would be happy to answer your questions and discuss the paper further. Please feel free to contact me at **jiusi.zheng@ru.nl**