

# Anonymization of Stuttered Speech - Removing Speaker Information while Preserving the Utterance

Jan Hintz<sup>1\*</sup>, Sebastian Bayerl<sup>2\*</sup>, Y. Sinha<sup>1</sup>, S. Ghosh<sup>3</sup>, M. Schubert<sup>1</sup>, S. Stober<sup>3</sup>, Korbinian Riedhammer<sup>2</sup>, Ingo Siegert<sup>1</sup>

<sup>1</sup>Mobile Dialog Systems, Otto-von-Guericke-University Magdeburg, <sup>2</sup>Technische Hochschule Nuremberg, <sup>3</sup>Artificial Intelligence Lab AI-LAB, Otto-von-Guericke-University Magdeburg

## Motivation

### Anonymization:

- Identity concealment in different contexts.
- Data collection
- Re-identification through speech pathologies

### E-Health:

- Remote treatment scenario for speech pathologies
- Accessibility

Need for evaluation and more inclusive models!

## Evaluation

### Speaker Recognition:

- Pretrained ECAPA-TDNN from Nemo Toolkit
- Feature vectors are compared using euclidean distance
- Evaluate level of anonymity by Equal Error Rate (EER)

### Stuttering Classification:

- KSF-C Challenge Task (8 class classification task)
- Fine-tuned wave2vec 2.0-based classification system
- Finetuning and evaluating on original and anonymized data

Details see:



## Dataset

- German Kassel State of Fluency (KsoF) dataset with 37 speakers
- KSF-C classification task (ComParE challenge 2022)
- 5597 three-second .wav clips
- Verbatim transcripts used for Re-synthesis

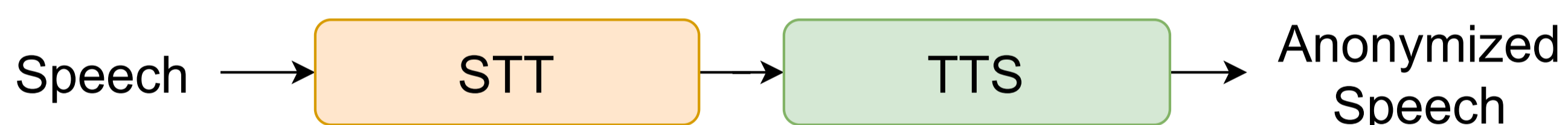
### Dataset



## Anonymization

### Re-synthesis:

- Speech Recognition with Whisper medium model
- Synthesis with FastPitch and HiFi-GAN from Nemo Toolkit
- Fine-tuning on HUI-Database (model available)



### Voice Conversion:

- StarGAN v2 as baseline
- Additional losses to capture specifics of stuttering
- Changes in pitch and pitch modulation
- Changes in intensity

## Results

Equal Error Rate (EER) and Unweighted Average Recall (UAR)

	EER (%)	UAR (%)	UAR* (%)
KsoF original	8.24	62.78	-
StarGAN [6] baseline	27.25	35.50	52.66
StarGAN [6] Delta	32.63	35.39	46.38
Re-Synthesis	48.27	14.15	22.25

\*Fine-tuned on anonymized data

Dysfluency prediction on KSF-C

	Block	Fillers	Garbage	Mod.	Prolong.	Sound Rep.	Word Rep.	No Disfl.
Block	0.48	0.00	0.05	0.03	0.05	0.28	0.00	0.11
Fillers	0.01	0.87	0.02	0.02	0.01	0.02	0.01	0.02
Garbage	0.00	0.00	0.31	0.00	0.06	0.12	0.00	0.50
Mod.	0.00	0.00	0.00	0.91	0.02	0.01	0.00	0.05
Prolong.	0.12	0.01	0.00	0.04	0.65	0.15	0.00	0.03
Sound Rep.	0.04	0.01	0.01	0.01	0.07	0.83	0.01	0.03
Word Rep.	0.17	0.06	0.00	0.17	0.00	0.11	0.22	0.28
No Disfl.	0.08	0.00	0.02	0.10	0.03	0.03	0.00	0.74

Dysfluency prediction on Data anonymized with StarGAN F0 Delta

	Block	Fillers	Garbage	Mod.	Prolong.	Sound Rep.	Word Rep.	No Disfl.
Block	0.16	0.00	0.67	0.01	0.01	0.12	0.00	0.03
Fillers	0.10	0.17	0.54	0.04	0.02	0.06	0.00	0.07
Garbage	0.00	0.00	0.81	0.00	0.00	0.06	0.00	0.12
Mod.	0.01	0.00	0.40	0.45	0.00	0.04	0.00	0.10
Prolong.	0.06	0.01	0.49	0.02	0.11	0.29	0.00	0.02
Sound Rep.	0.02	0.01	0.28	0.01	0.01	0.63	0.00	0.05
Word Rep.	0.00	0.00	0.39	0.06	0.00	0.17	0.11	0.28
No Disfl.	0.05	0.01	0.44	0.05	0.00	0.06	0.00	0.39

Dysfluency prediction finetuned and tested on Data anonymized with StarGAN F0 Delta

	Block	Fillers	Garbage	Mod.	Prolong.	Sound Rep.	Word Rep.	No Disfl.
Block	0.60	0.03	0.01	0.11	0.07	0.05	0.00	0.13
Fillers	0.10	0.55	0.00	0.18	0.04	0.01	0.00	0.12
Garbage	0.44	0.00	0.12	0.00	0.00	0.00	0.00	0.44
Mod.	0.00	0.00	0.00	0.94	0.01	0.01	0.00	0.03
Prolong.	0.19	0.02	0.01	0.15	0.43	0.12	0.00	0.08
Sound Rep.	0.22	0.04	0.00	0.06	0.12	0.51	0.00	0.05
Word Rep.	0.22	0.06	0.00	0.22	0.11	0.17	0.00	0.22
No Disfl.	0.05	0.00	0.02	0.27	0.06	0.05	0.00	0.56

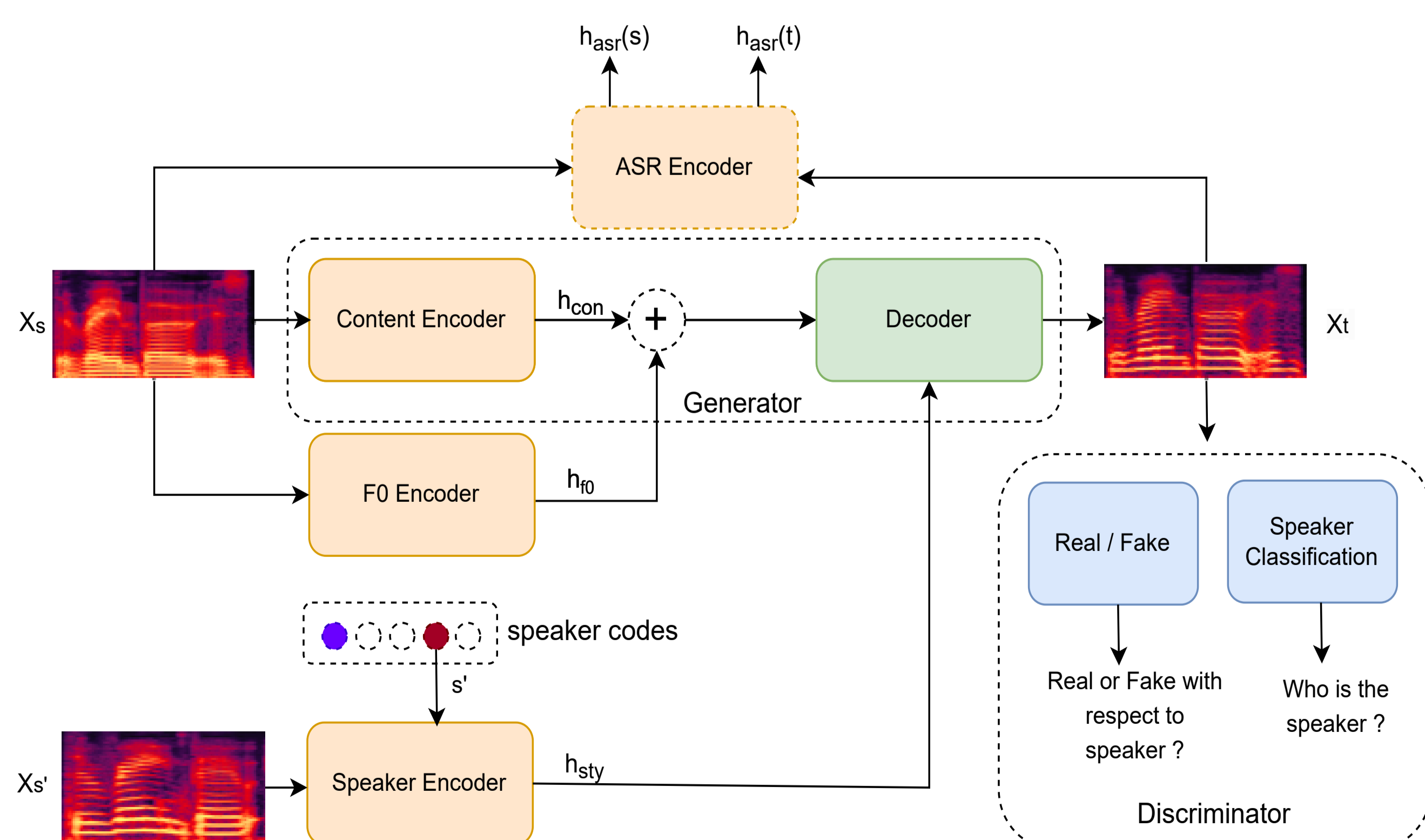
## Conclusions

### Re-synthesis:

- Grants high anonymity
- Difficult to convey information about the speech disorder
- Most modifications and prolongations are masked
- Phoneme based recognition and synthesis could lead to improvement

### Voice Conversion:

- Best trade-off between anonymization and preserving stuttering for Speech therapy purposes.
- Has to be confirmed for other speech pathologies.
- Possibility of anonymized data collection



Read the Paper!



ANONYM PREVENT

Get in touch!



INTERSPEECH 2023  
20th-24th August 2023 - Dublin, Ireland

ohm Technische Hochschule Nürnberg

