Speaker Anonymization with Feature-Matched F0 Trajectories

VoicePrivacy Challenge 2022 Submission

Ünal Ege Gaznepoglu, Anna Leschanowsky, Nils Peters
uenal.ege.gaznepoglu@iis.fraunhofer.de

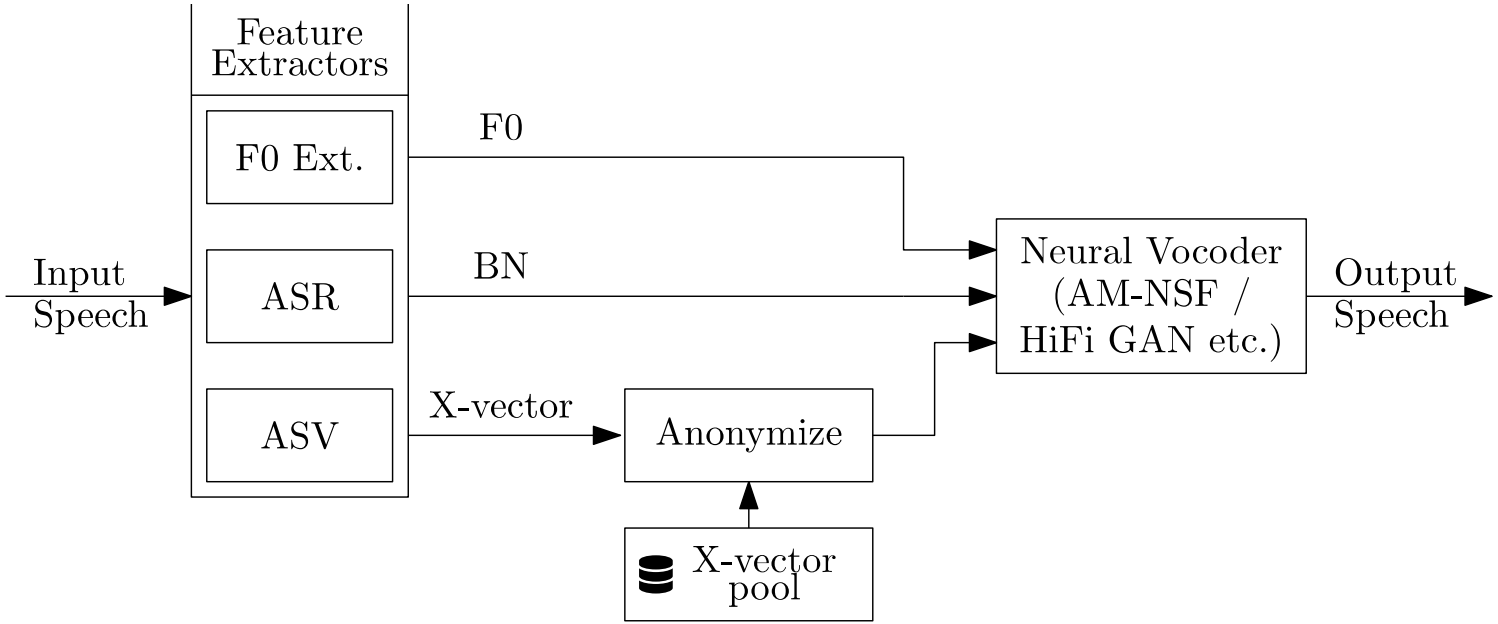
Friedrich-Alexander-Universität Erlangen-Nürnberg
Fraunhofer IIS

Introduction
Speaker Anonymization

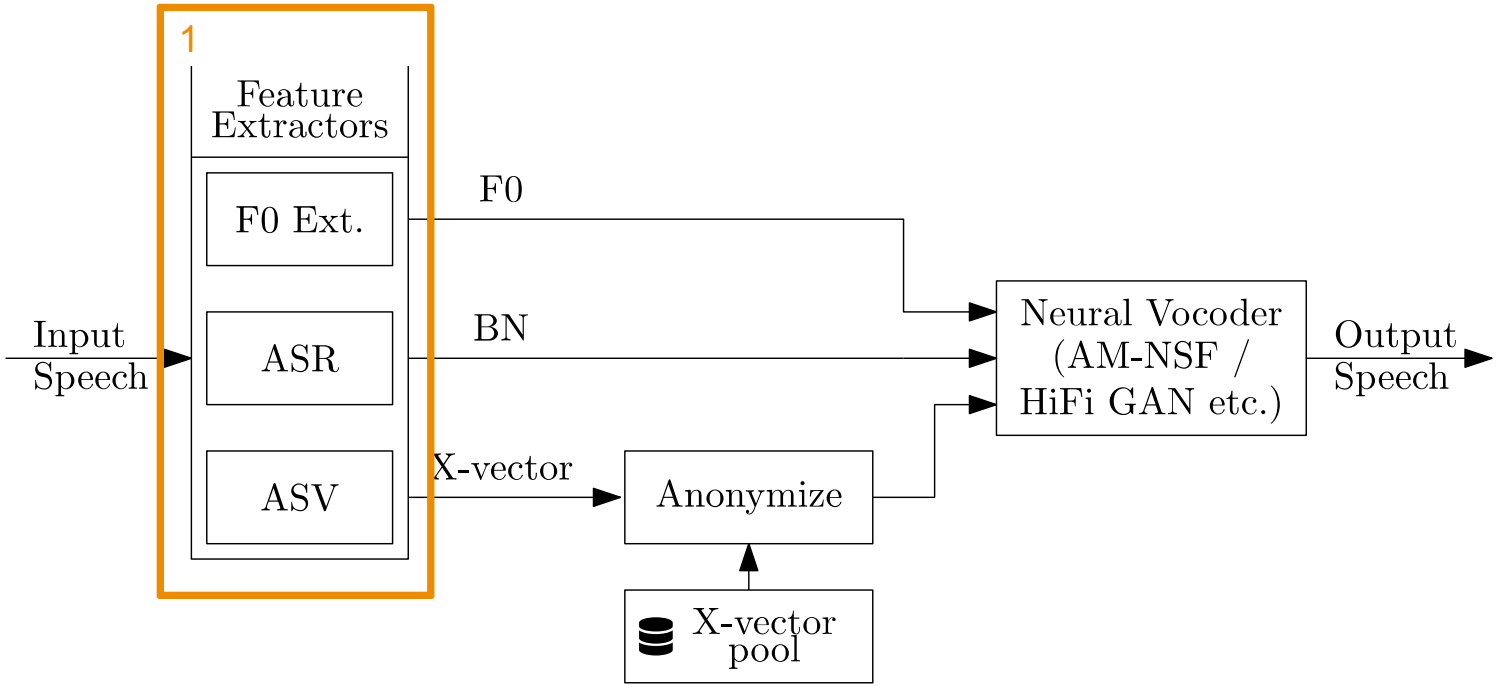
- **Context:** Internet hosts and smart devices process massive amount of speech signals
- **Problem:** Voice signals contain sensitive information (age, gender, health condition etc.) and could be cloned
- **Solution:** Modify speech signals such that it is not linkable to the original speaker anymore, but still useful for other tasks (e.g. ASR)

i.e., a voice conversion task but without a concretely stated target speaker

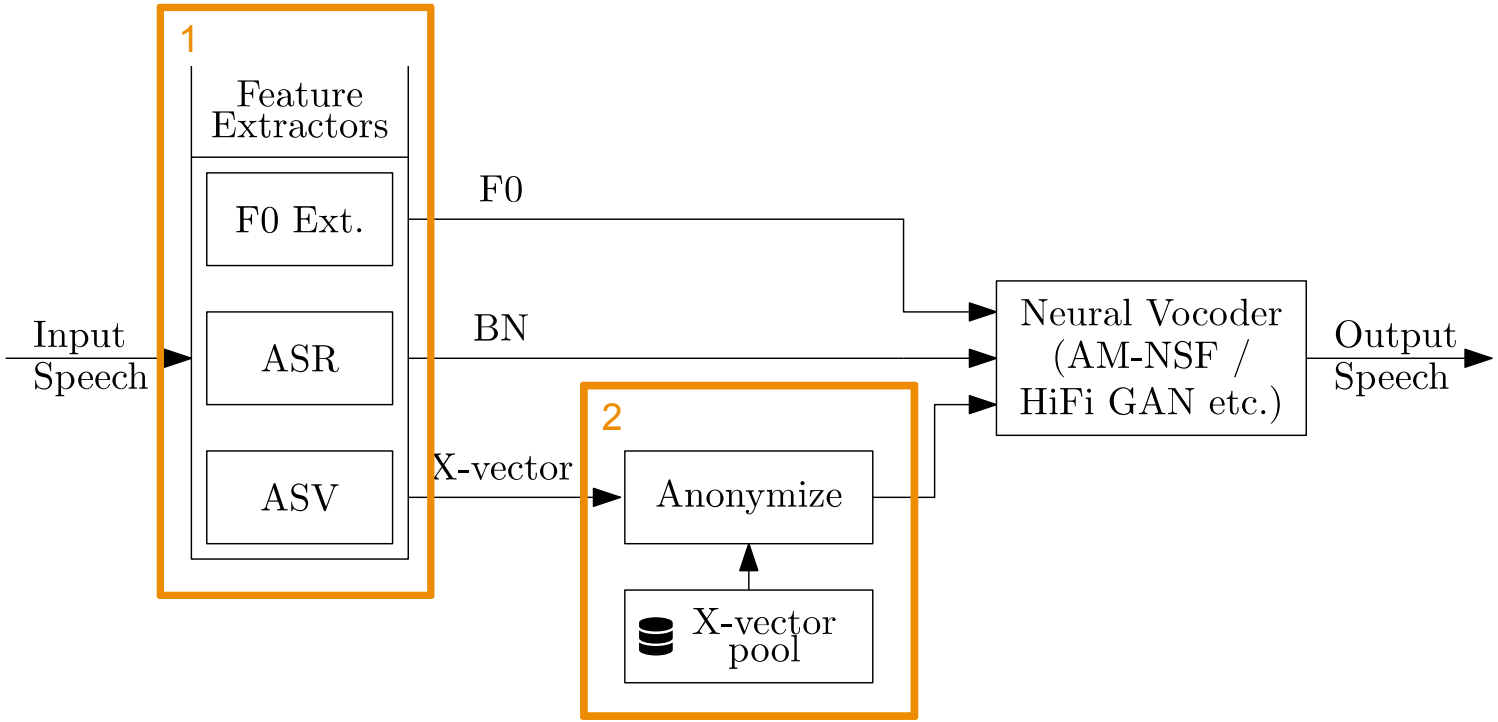
Introduction
VPC Baseline B1 – ML-based & modular



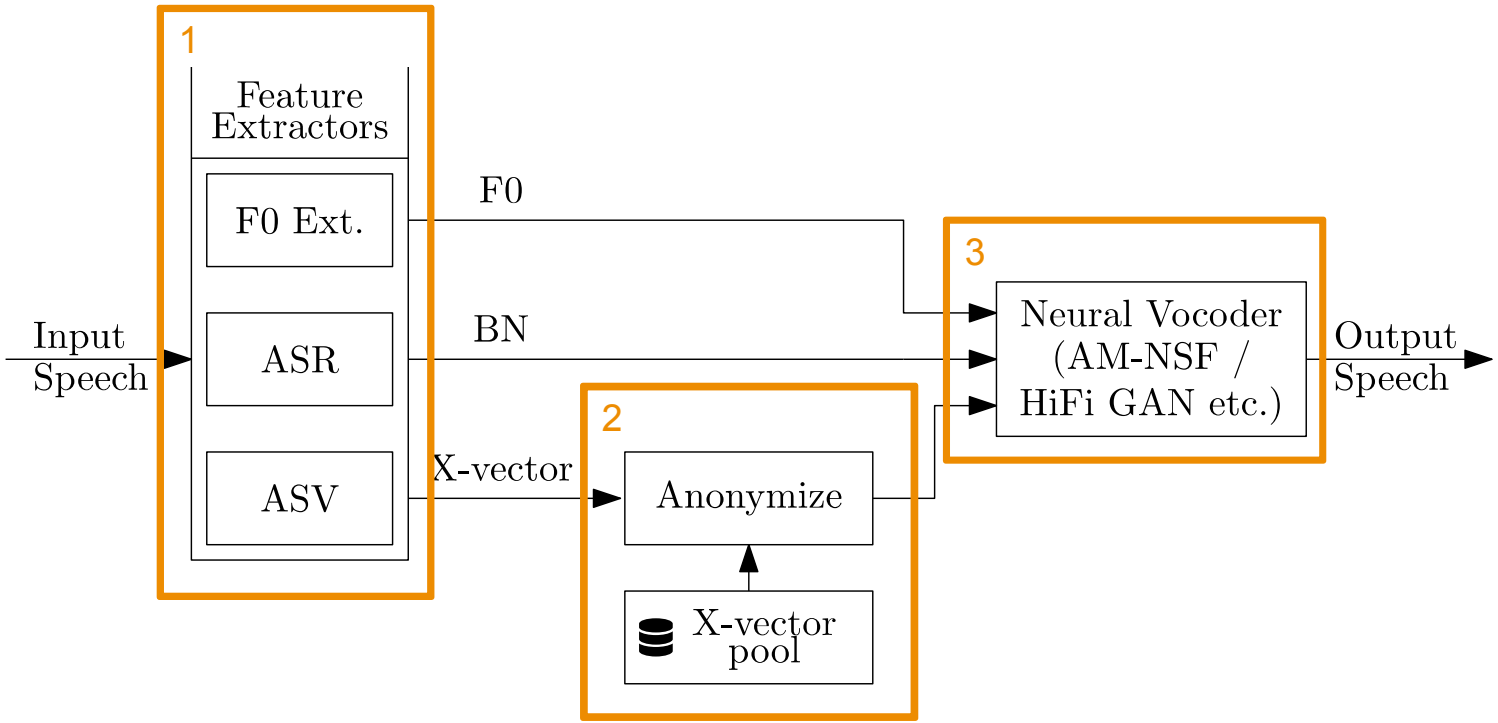
Introduction
VPC Baseline B1 – ML-based & modular



Introduction
VPC Baseline B1 – ML-based & modular



Introduction
VPC Baseline B1 – ML-based & modular



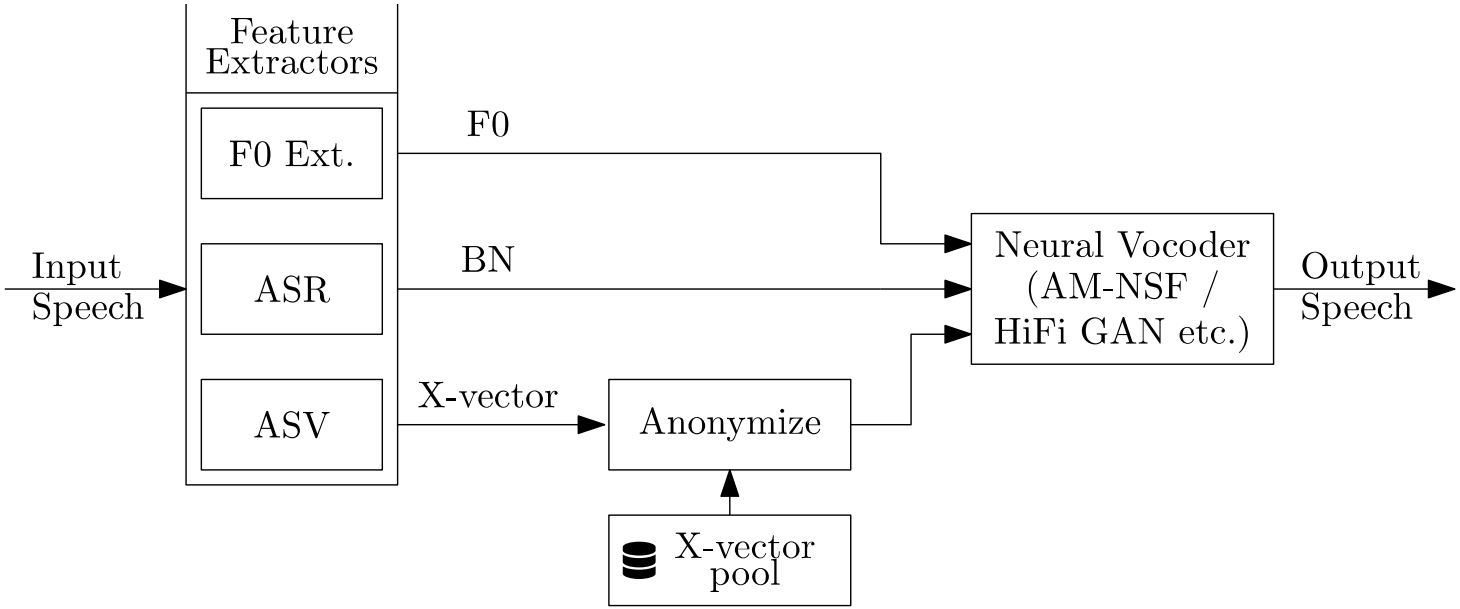
Facts: Fundamental Frequency F0

- F0 extractors (e.g., YAAPT) are signal statistics-based and slow:

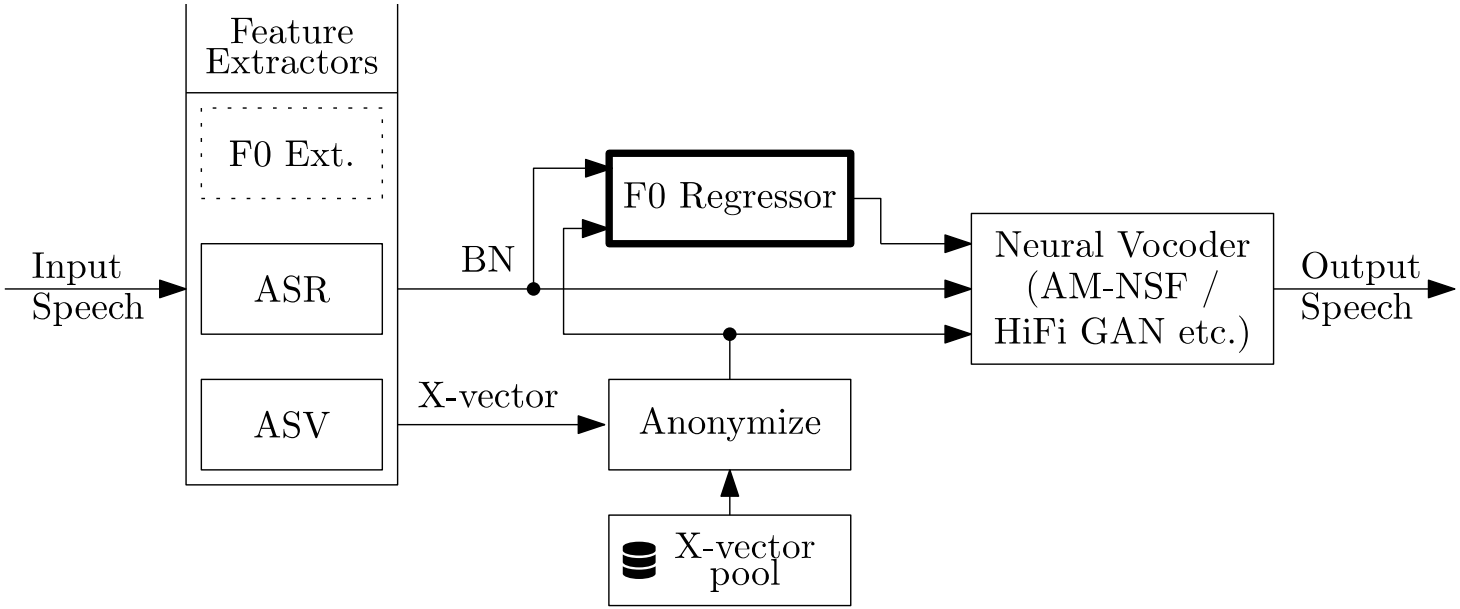
  70 minutes for processing *-dev and *-test data on our system with 96-core CPU
- F0 is affected by the identity, linguistic content, and prosody [1]
- Multiple works [2,3,4] concluded F0 modifications could provide an advantage for anonymization task

[1] S. Johar, “Psychology of Voice,” in Emotion, Affect and Personality in Speech: The Bias of Language and Paralanguage, S. Johar, Ed. Cham: Springer Intl. Publishing, 2016, pp. 9–15.
[2] P. Champion, D. Jouvet, and A. Larcher, “A Study of F0 Modification for X-Vector Based Speech Pseudonymization Across Gender,”
[3] U. E. Gaznepoglu and N. Peters, “Exploring the Importance of F0 Trajectories for Speaker Anonymization using X-vectors and Neural Waveform Models,” *MLSLP2021*, Sep. 2021
[4] L. Tavi, T. Kinnunen, and R. G. Hautamäki, “Improving speaker de-identification with functional data analysis of f0 trajectories,” *Speech Communication*, vol. 140, pp. 1–10, May 2022

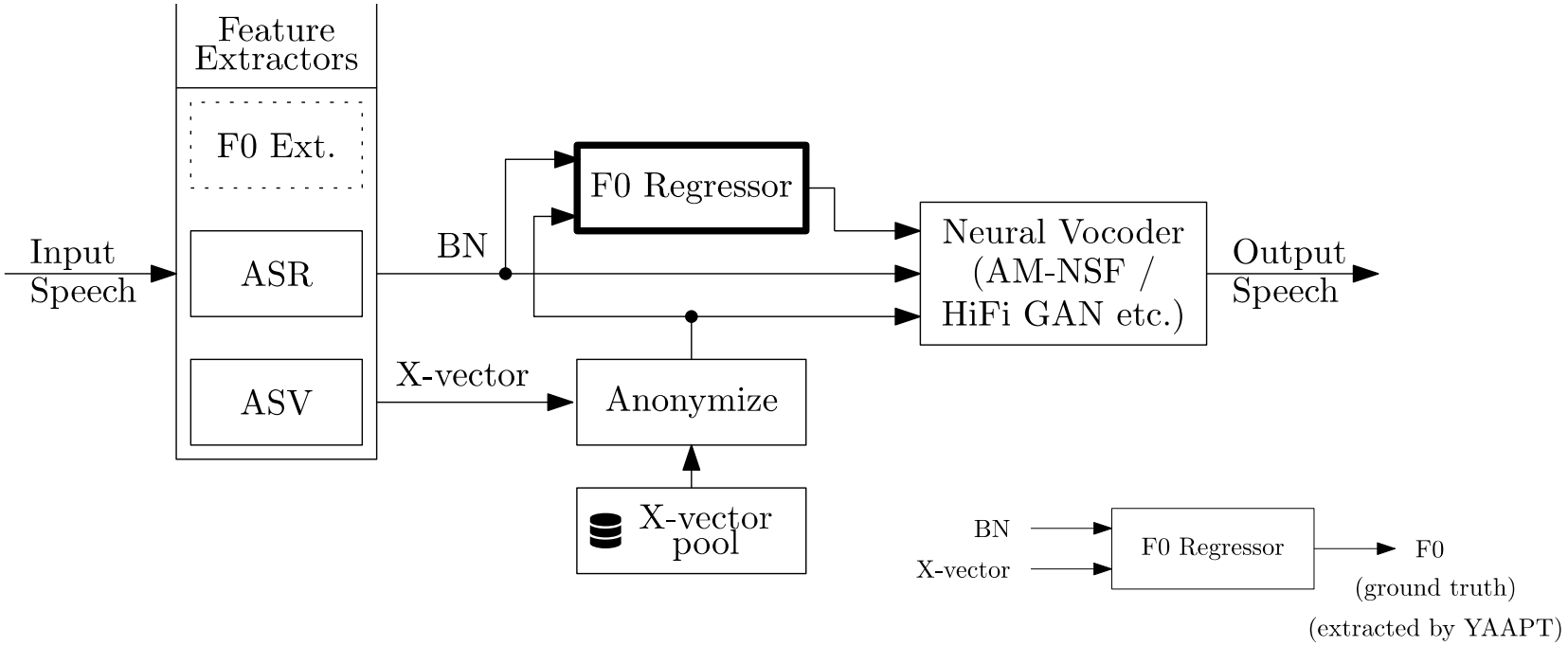
Modification: Estimate F0 from BN and anonymized X-Vector



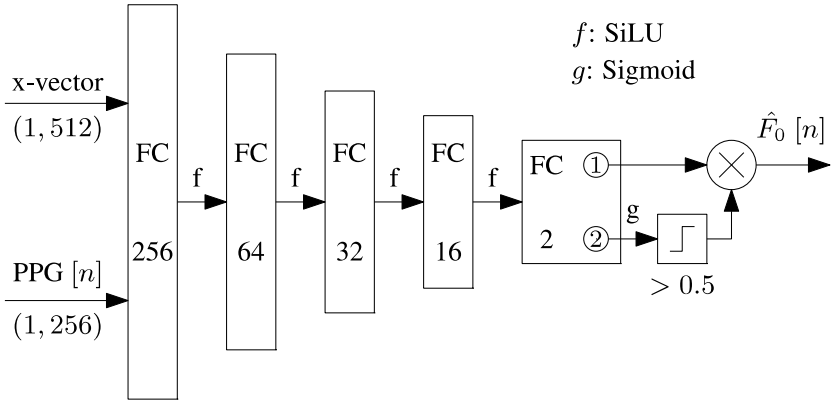
Modification: Estimate F0 from BN and anonymized X-Vector

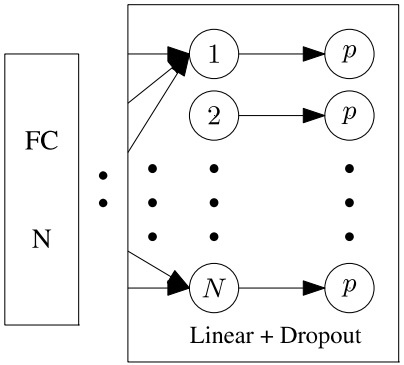


Modification: Estimate F0 from BN and anonymized X-Vector



Block: F0 Regressor

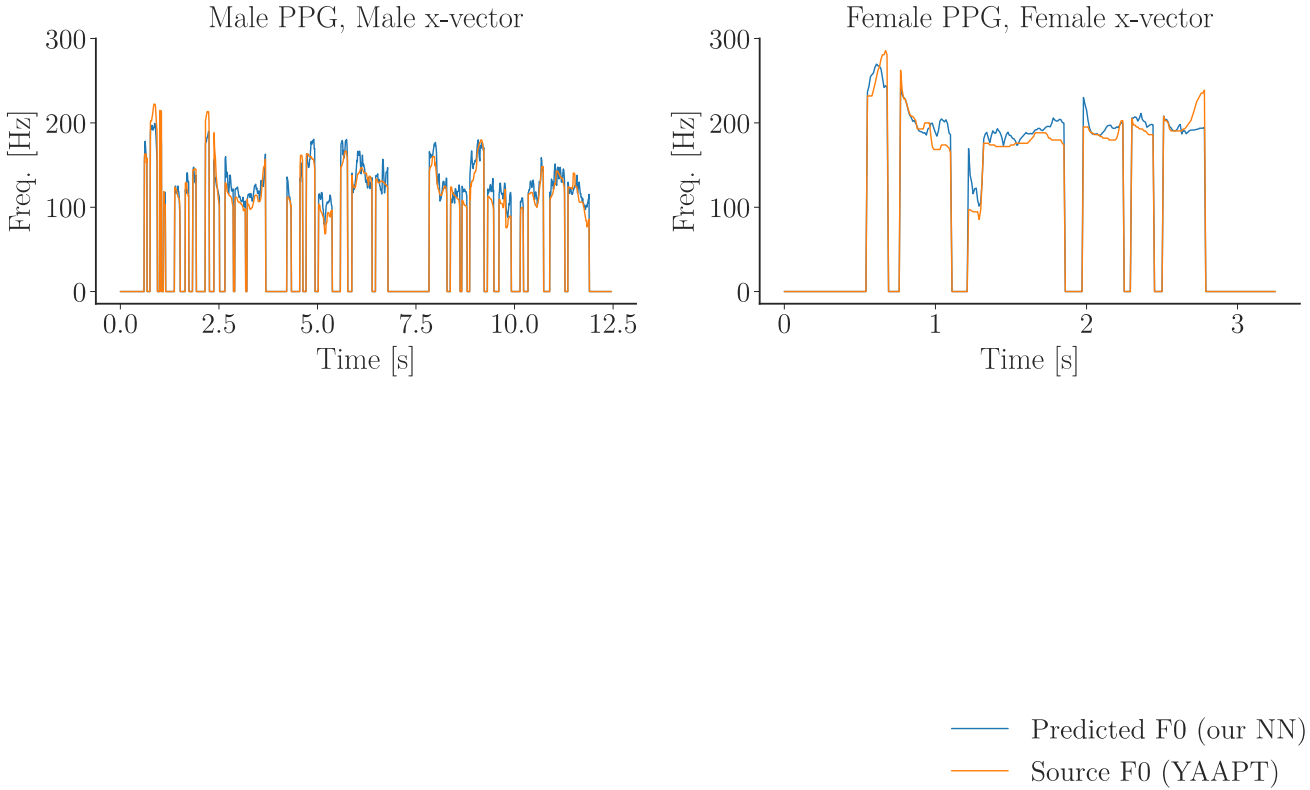




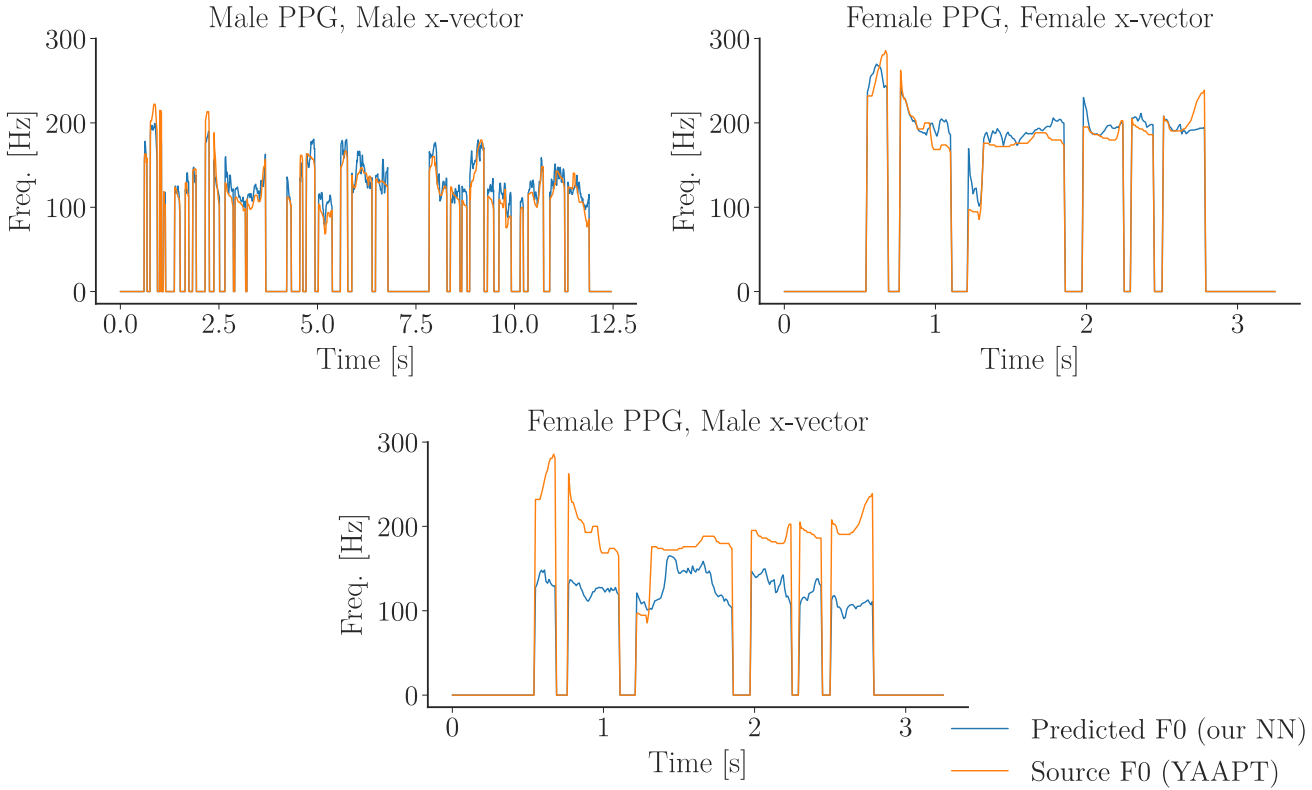
$$\mathcal{L}(F_0, \hat{F}_0) = \text{MSE}(F_0 - \hat{F}_0)^2 + \alpha \text{BCE}(p_v, v)$$

* Mean normalized log F0 is predicted.

Generalization Check: Cross gender F0 conversion



Generalization Check: Cross gender F0 conversion



Evaluation: VPC Framework

Dataset	Sex	EER [%]			WER [%]			ρ^{F_0}			G_{VD} [dB]		
		B1.b	Submitted	Fixed	B1.b	Submitted	Fixed	B1.b	Submitted	Fixed	B1.b	Submitted	Fixed
libri-test	F	8.39	22.63	**22.99**	**4.28**	4.43	4.45	0.83	0.81	0.82	-5.58	-6.16	-7.14
	M	6.46	19.38	**21.83**				0.68	0.60	0.59	-5.52	-5.54	-5.68
vctk-test	F	9.00	22.99	**22.99**	**10.44**	10.32	10.52	0.86	0.85	0.85	-8.21	-8.87	-12.42
	M	8.15	17.51	**17.51**				0.75	0.70	0.70	-8.18	-8.81	-10.42
vctk-test-com	F	11.56	19.65	**19.65**				0.84	0.82	0.82	-6.68	-7.34	-10.66
	M	7.63	12.99	**12.99**				0.69	0.63	0.64	-6.11	-6.14	-7.43
Ø (test)		8.10	20.24	**22.07**	**7.36**	7.38	7.49	0.78	0.74	0.74	-6.69	-7.14	-8.68

Table 2: *Results from Baseline B1.b variant* `joint-hifigan` *taken from [13] compared with the variant including our modifications. Better performing entries (either 'Fixed' or baseline) are highlighted for the primary metrics EER and WER. The column 'Submitted' indicates the results we have shared before the submission deadline. The column 'Fixed' indicates the results we obtained after fixing a bug within our system, counting as 'late submission'. Weighted average per challenge guidelines is denoted with* **Ø**.

- And it is fast: 2 minutes using a single RTX 3090 to process *-dev and *-test datasets.

Conclusion

- Novel low-complexity DNN-based F0 synthesis method
- Input features: BNs and the anonymized x-vector
- Better Anonymization:
  - Avoids leakage of original F0 information
  - Improves EER anonymization metric by 2.5 times
  - (Almost) No negative impact on the other 3 performance metrics
- Better Audio Quality:
  - Harmonizing F0 with anonymized X-Vectors
  - More natural sounding voice synthesis e.g., for cross-gender speech anonymization
- Faster Processing:
  - No time-consuming F0 extraction at runtime
  - F0 Regressor is 35x faster than YAAPT F0 extractor

Thank you for your attention!

Speaker Anonymization with Feature-Matched F0 Trajectories

Listening samples available

VoicePrivacy Challenge 2022 Submission

Ünal Ege Gaznepoglu, Anna Leschanowsky, Nils Peters
uenal.ege.gaznepoglu@iis.fraunhofer.de

Friedrich-Alexander-Universität Erlangen-Nürnberg
Fraunhofer IIS