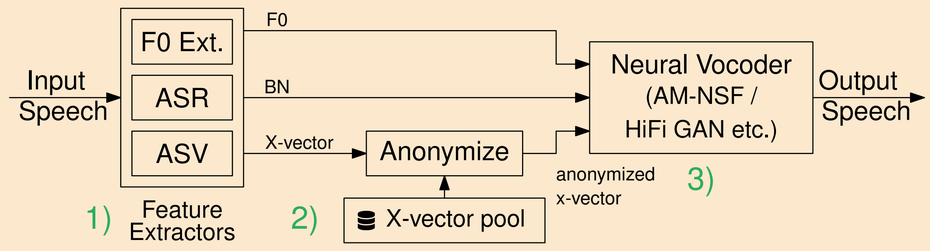
INTERNATIONAL AUDIO LABORATORIES ERLANGEN
A joint institution of Fraunhofer IIS and Universität Erlangen-Nürnberg

Deep learning-based F0 synthesis for speaker anonymization

Ünal Ege Gaznepoglu, Nils Peters (ege.gaznepoglu@audiolabs-erlangen.de)

1. Introduction



- Many systems build upon the VPC Baseline B1 [1]

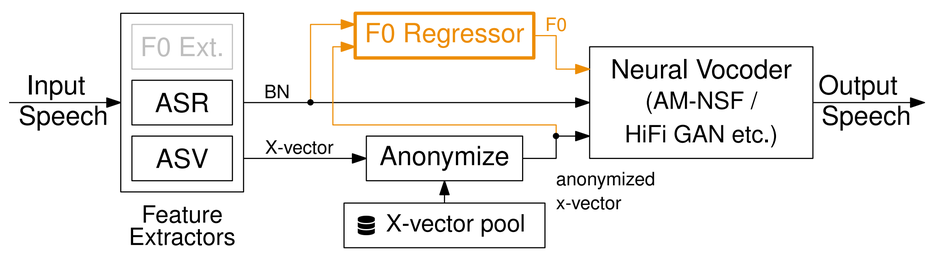
1) Speech represented by (F0, x-vector, bottleneck features)
2) Modify the speaker identity (x-vector)
3) Synthesize speech using a neural vocoder (AM-NSF, HiFiGAN etc.)

Shortcomings:

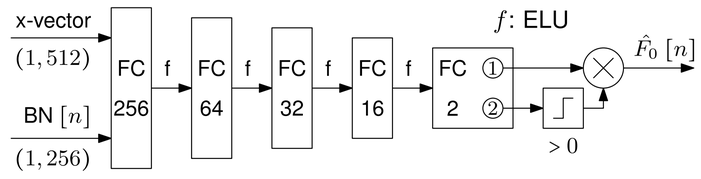
- New identity (esp. in cross-gender anonymization scenario) does not match with the F0 from original speech
- F0 contains personal data and is not sanitized
- F0 extraction happens on CPU and takes a long time
- Approaches in the literature (e.g. [2]) require pool-based anonymizers

2. Proposed Approach

We substitute the F0 extractor with a regressor, to framewise predict F0 values from other features [3]. We assume YAAPT extractions as training and evaluation ground truth.



The regressor, a fully connected (FC) network, is trained on eq. (1):



$$\mathcal{L}(F_0, \hat{F}_0, g, v) = \mathsf{L1}(F_0, \hat{F}_0) + 28.112\ \mathsf{BCE}(g, v) \quad (1)$$

3. Evaluation Methodology

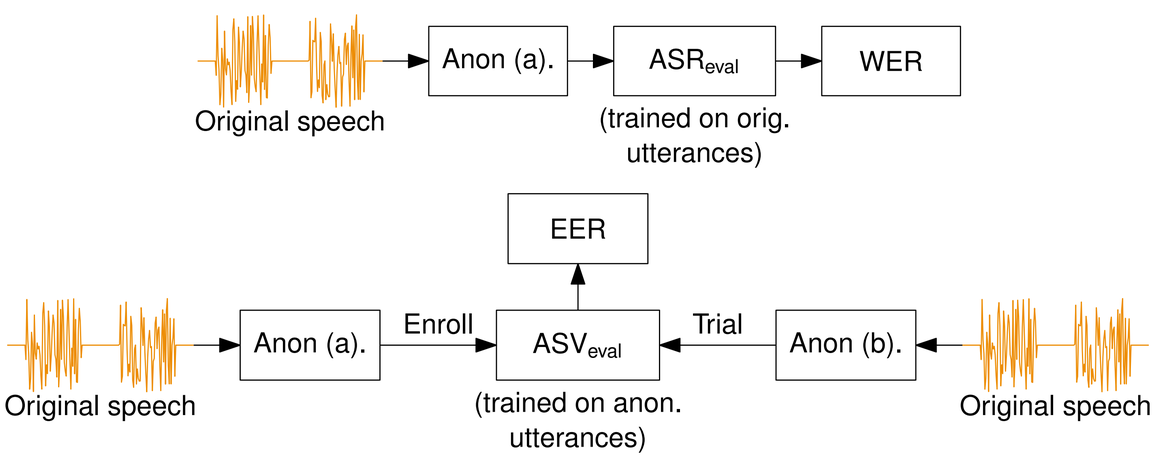
i) F0 Reconstruction Performance:

- Accuracy, precision, recall for voiced-unvoiced decision
- Gross and fine pitch error (GPE, FPE) for pitch regression

$$\textbf{GPE}: \frac{\text{num. of frames whose error} > 20\%}{\text{num. of correctly identified voiced frames}}$$

$$\textbf{FPE}: \frac{\text{num. of frames whose error} > 5\%}{\text{num. of frames whose error} < 20\%}$$

ii) Effects on Speaker Anonymization: We inherit the VoicePrivacy Challenge metrics and attack models [4].

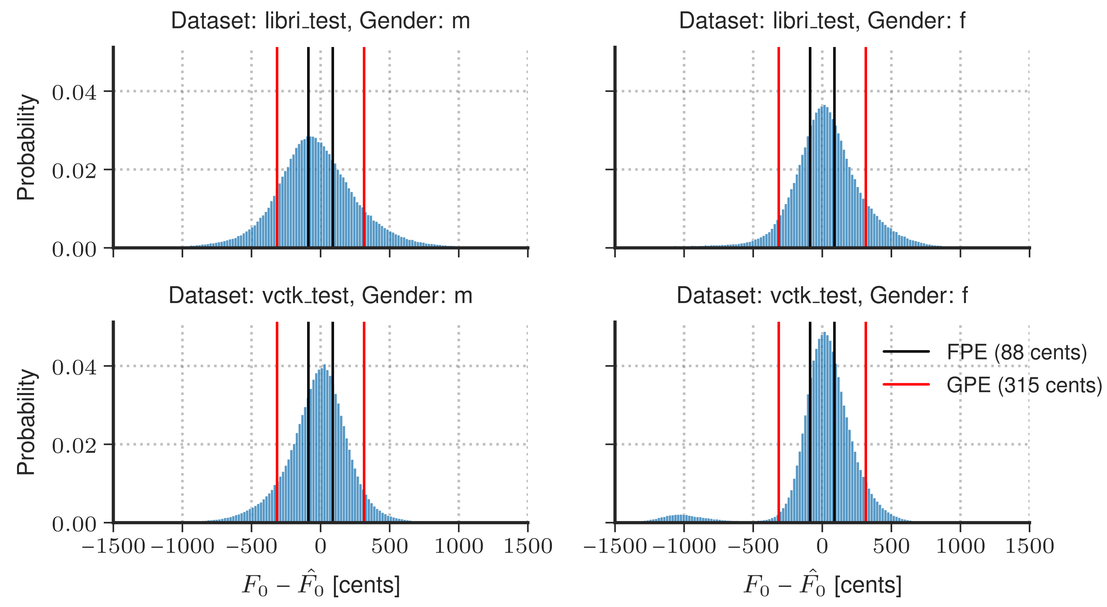


iii) Contrastive Study: We investigate the relative contributions of x-vector and F0 anonymization in our proposed system.

4. Evaluation Results

i) F0 Reconstruction Performance

Dataset	Gender	GPE(↓)	FPE(↓)	Acc.(↑)	Prec.(↑)	Rec.(↑)
libri-test	F	31.6	66.9	93.0	94.6	93.3
	M	41.8	71.8	92.5	93.0	93.0
vctk-test	F	24.6	63.9	95.1	94.1	93.5
	M	38.8	69.9	94.6	93.5	92.5



ii) Effects on Speaker Anonymization

Dataset	Weight	Gender (From → To)	EER [%] (↑) B1.b	EER [%] (↑) [2]	EER [%] (↑) Ours	WER [%] (↓) B1.b	WER [%] (↓) [2]	WER [%] (↓) Ours
weighted average / same gender			9.81	11.53	**12.54**	10.13	10.17	**10.03**
weighted average / cross gender			13.57	22.87	**25.71**	10.68	10.4	**10.23**

Detailed breakdowns (across datasets) and further experiments featuring contrastive systems are available in our paper (scan the QR code).

5. Conclusions

- **F0 reconstruction**
  - Our system attains voiced-unvoiced decisions comparable to YAAPT's reported accuracy [5]
  - Differences in those are often at the edges of voiced segments
  - Lack of temporal context from other frames, and not having perfect F0 annotations, causes suboptimal F0 value prediction
- **Integration into speaker anonymization**
  - Our system improves all VPC metrics, outperforming the state-of-the-art speaker-based F0 modification in the literature [2]
  - Our approach is complementary to x-vector anonymization
  - It also attained the best naturalness scores in VPC 2022 [6]
  - Our F0 regression is 35x faster than F0 extraction by YAAPT

References

[1] F. Fang *et al.*, "Speaker anonymization using x-vector and neural waveform models," in *Proc. 10th ISCA Speech Synthesis Workshop*, 2019.

[2] P. Champion, D. Jouvet, and A. Larcher, "A study of f0 modification for x-vector based speech pseudonymization across gender," in *2nd AAAI Workshop on Privacy-Preserving AI*, 2021.

[3] U. E. Gaznepoglu, A. Leschanowsky, and N. Peters, "Voiceprivacy 2022 system description: Speaker anonymization with feature-matched F0 trajectories," in *VoicePrivacy Challenge Submission*, 2022.

[4] N. Tomashenko *et al.* "2nd VoicePrivacy challenge evaluation plan." (2022).

[5] R. Vaysse, C. Astésano, and J. Farinas, "Performance analysis of various fundamental frequency estimation algorithms in the context of pathological speech," *The Journal Acoust. Soc. of America*, vol. 152, no. 5, 2022.

[6] N. Tomashenko *et al.* "The VoicePrivacy 2022 challenge results." (2022).

Friedrich-Alexander-Universität Erlangen-Nürnberg

Fraunhofer IIS