A Study of F0 Modification for X-Vector Based Speech Pseudonymization Across Gender

Pierre Champion [1] Denis Jouvet [1] Anthony Larcher [2]

[1]Université de Lorraine, CNRS, Inria, LORIA, Nancy, France. [2]Le Mans Université, LIUM, France

Speech anonymization

Suppress identifiable personal information contained within speech signals.

Use cases:

- **Hide speaker identity** before sending signals to centralized servers.
- **Keep the spoken content intelligible** to share speech data for improved training.

Anonymization technique

The speaker identity (x-vector) and linguistic content (F0 and Phonetic features) from an input utterance are first extracted. Then, in the baseline, the x-vector is replaced by a pseudo-speaker x-vector, and F0 is unchanged. We added an F0 modification consistent with the chosen pseudo-speaker x-vector.

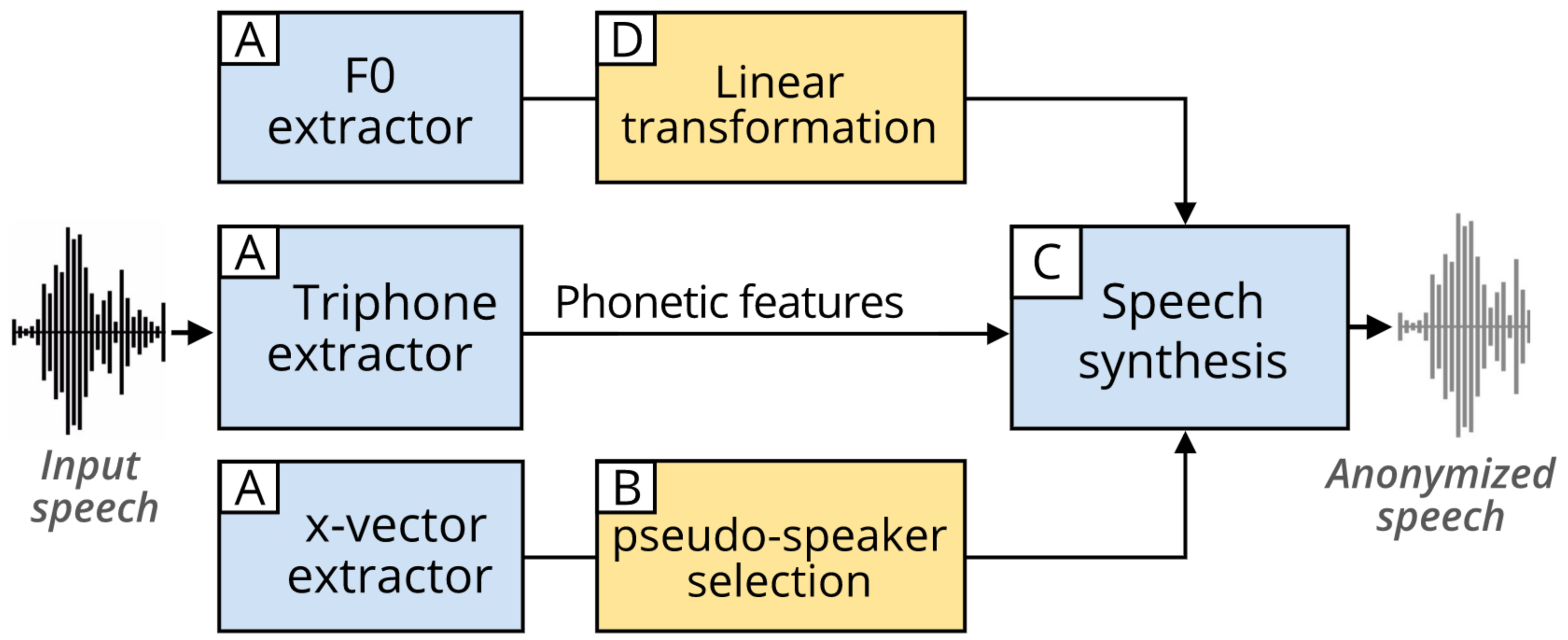

Figure 1. The speaker anonymization pipeline. We added module D.

F0 modification

Used Linear transformation:

$$\hat{x}_t = \mu_y + \frac{\sigma_y}{\sigma_x}(x_t - \mu_x)$$

x_t = F0 of the source speaker
μ_x and σ_x: statistics source speaker
μ_y and σ_y: statistics selected pseudo-speaker.

Experiment and evaluation

Results compared to the VoicePrivacy baseline system on **LibriSpeech** *test-clean*.

Privacy and Utility metrics:

1. **Equal Error Rate ($EER_{\%}$)**, measures the speaker's concealing capability through speaker verification. (verify whether an input speech corresponds to the claimed identity, score to maximize).
2. **Word Error Rate ($WER_{\%}$)**, measures speech intelligibility through speech recognition. (translate an input speech sequence into text, score to minimize).

Metrics	Without anonymization	(baseline) VoicePrivacy	(proposed) VoicePrivacy + F0
Utility ($WER_{\%}$)	4.1	6.7	6.7
Male Privacy ($EER_{\%}$)	1.0	36.7	**48.7**
Female Privacy ($EER_{\%}$)	7.1	32.1	**43.4**

Table 1. Scores obtained with the VoicePrivacy evaluation system.

Multiple pseudo-speaker selections were studied. Table 1 shows the best results obtained by selecting a pseudo-speaker from the *opposite* gender. The utility score stays as low as the baseline, while privacy enhanced.

Conclusion

State-of-the-art **speaker verification performance decreases** when input speech signals are anonymized. Modifying both the pseudo-speaker **and the log F0 with a linear transformation** yields **better privacy protection without a utility penalty**.

References

[1] Tomashenko, Srivastava, Wang, Vincent, Nautsch, Yamagishi, Evans, Patino, Bonastre, and Noé. Introducing the VoicePrivacy Initiative. *Proc. Interspeech*, 2020.