Evaluation of Speaker Anonymization on Emotional Speech

Hubert Nourtel[1] Pierre Champion [1,2] Denis Jouvet [1] Anthony Larcher [2] Marie Tahon [2]

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

Anonymization System

This work fits into the context of speaker anonymization, where the goal is to suppress the personally identifiable information from speech while maintaining the intelligibility of the spoken content. The baseline used is the *Voice Privacy Challenge 2020* (VPC) [1].
In the VPC, speaker identity (x-vector), fundamental frequency (F_0), and linguistic content (phonetic features) are extracted from speech segments. Then, the x-vector is altered to generate a new anonymized speech. Alongside, the F_0, which also contains information about the speaker, can be modified using different methods.

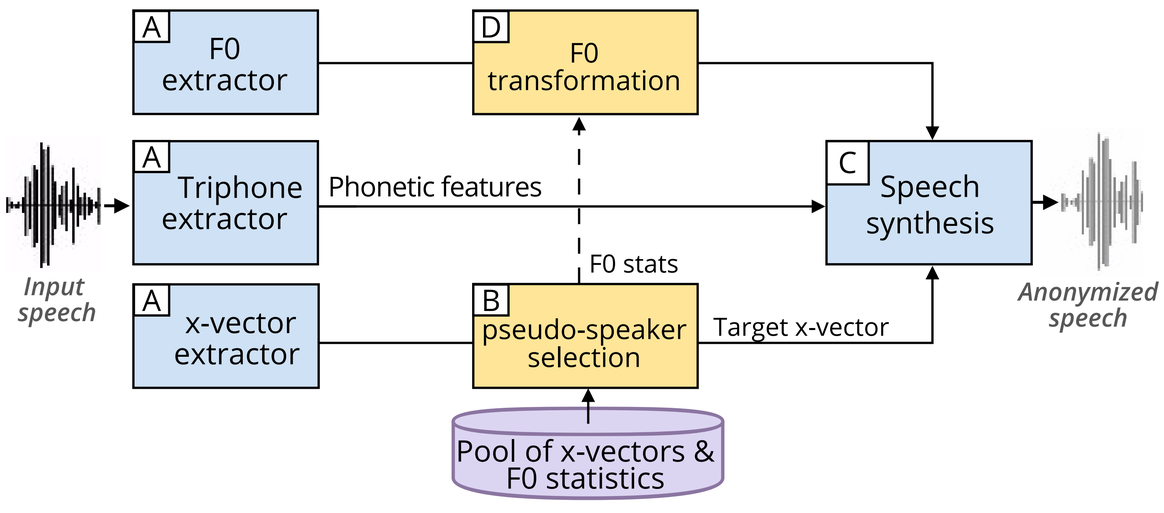

Figure 1. The speaker anonymization framework. F_0 transformation can be transformed with different methods

This study explores the impact of the speaker anonymization on the emotional information present in speech utterances.

Speech Emotion Dataset

The *IEMOCAP dataset* [2] is an emotional speech dataset commonly used for speech emotion recognition. It is composed of 12h of data from 10 different actors (5 female and 5 male) with scripted and improvised speech. It provides text transcription and emotion annotation among five classes: neutral, frustration, sadness, anger and happiness.

F_0 Modifications

Two F_0 modifications have been applied to evaluate their impact on emotion classification performances:

- **Linear transformation:** The F_0 features of the source speaker are transformed using a linear transformation.
- **Warping factor:** The contour of the F_0 is randomly modified to increase or decrease the F_0 variation using a warping factor, applied on top of the linear transformation.

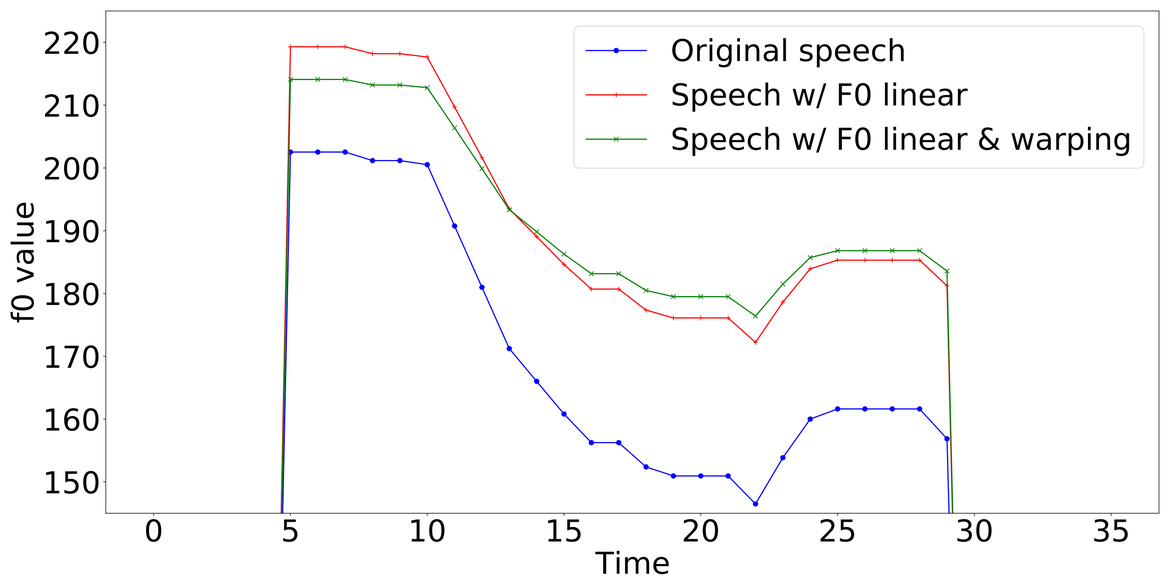

Figure 2. F_0 values for several transformations

Experiments and Evaluation

To evaluate the performance, two metrics were used:

1. **Utility metric:** Word Error Rate (WER) measures speech intelligibility through speech recognition by translating an input speech sequence into text. (score to minimize for high utility)
2. **Emotion metric:** Unweighted Average Recall (UAR) measures the emotion recognition performance of a classifier (score to minimize if one wants to hide emotional information).

Definition of UAR is given by:

$$UAR = \frac{\sum \text{Recall per class}}{\#\ class} \tag{1}$$

	$WER_{\%}$	$UAR_{\%}$
Original data	34.62	44.48
Anonymized data (baseline)	38.97	37.92
Anonymized data (baseline + F_0 linear)	38.64	38.08
Anonymized data (baseline + F_0 linear & warping)	38.65	38.04
Difference Anonymized / Original	13% degradation	15% degradation

Table 1. WER and UAR results on IEMOCAP dataset

Table 1 shows the WER and UAR scores on original and anonymized speech with the VPC anonymization system and various F_0 modification techniques. Emotion is classified using an SVM model. We can see that the UAR **degradation** is **similar** to the WER **degradation**.

Conclusion

Regarding UAR and WER degradation, we can point out two views of the situation :

- If we consider emotion as a **valuable information** to be kept in speech after anonymization, the UAR **degradation is acceptable** as the emotion information suppressed is of the same order of magnitude to the loss of speech intelligibility.
- If we consider emotion as a **personal information** to remove, adding **simple modifications** of F_0 to the VPC anonymization process **is not enough** to hide emotion in the speech.

References

[1] N. Tomashenko, B. M. L. Srivastava, X. Wang, E. Vincent, A. Nautsch, J. Yamagishi, N. Evans, J. Patino, J.-F. Bonastre, P.-G. Noé, and M. Todisco, "Introducing the VoicePrivacy Initiative," *Proc. Interspeech*, 2020.

[2] C. Busso, M. Bulut, C.-C. Lee, A. Kazemzadeh, E. Mower, S. Kim, J. N. Chang, S. Lee, and S. S. Narayanan, "Iemocap: Interactive emotional dyadic motion capture database," *Journal of Language Resources and Evaluation*, 2008.