Quantum-Inspired Audio Unlearning: Towards Privacy-Preserving Voice Biometrics

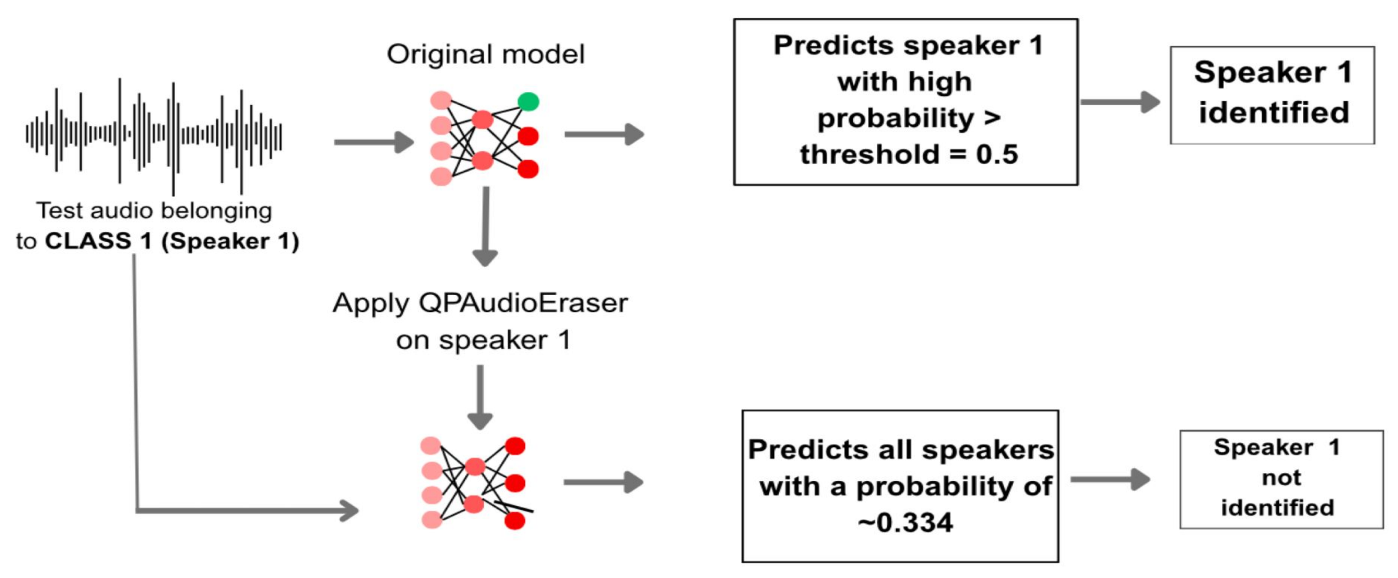
Shreyansh Pathak, Sonu Shreshtha, Mayank Vatsa and Richa Singh
Indian Institute of Technology Jodhpur, India

Motivation

The widespread adoption of voice enabled biometrics have significantly increased privacy vulnerabilities associated with speech data.

Existing Machine Unlearning methods designed for visual data inadequately handle the temporal and high dimensional nature of audio signals, to address this, we introduce QPAudioErase.

Goal



- **Pre-trained model:** Accurately recognizes speaker 1 with high confidence.
- **Unlearned model:** Produces nearly uniform probabilities across speakers.
- **Effect:** Leads to high uncertainty and failure to identify speaker 1.

Experimental Protocols

Evaluation metrics: We evaluate QPAudioEraser on metrics like Forget Accuracy (FA), Retain Accuracy (RA), ERB, PER, IL, FAR and FRR.

Baselines: QPAudioEraser was rigorously compared against well established baselines like:

- Synaptic Dampening: Uses Fisher Information to identify and "scrub" weights for forget data.
- Negative Gradient.: Applies negative gradients to remove data influence.

Proposed Framework

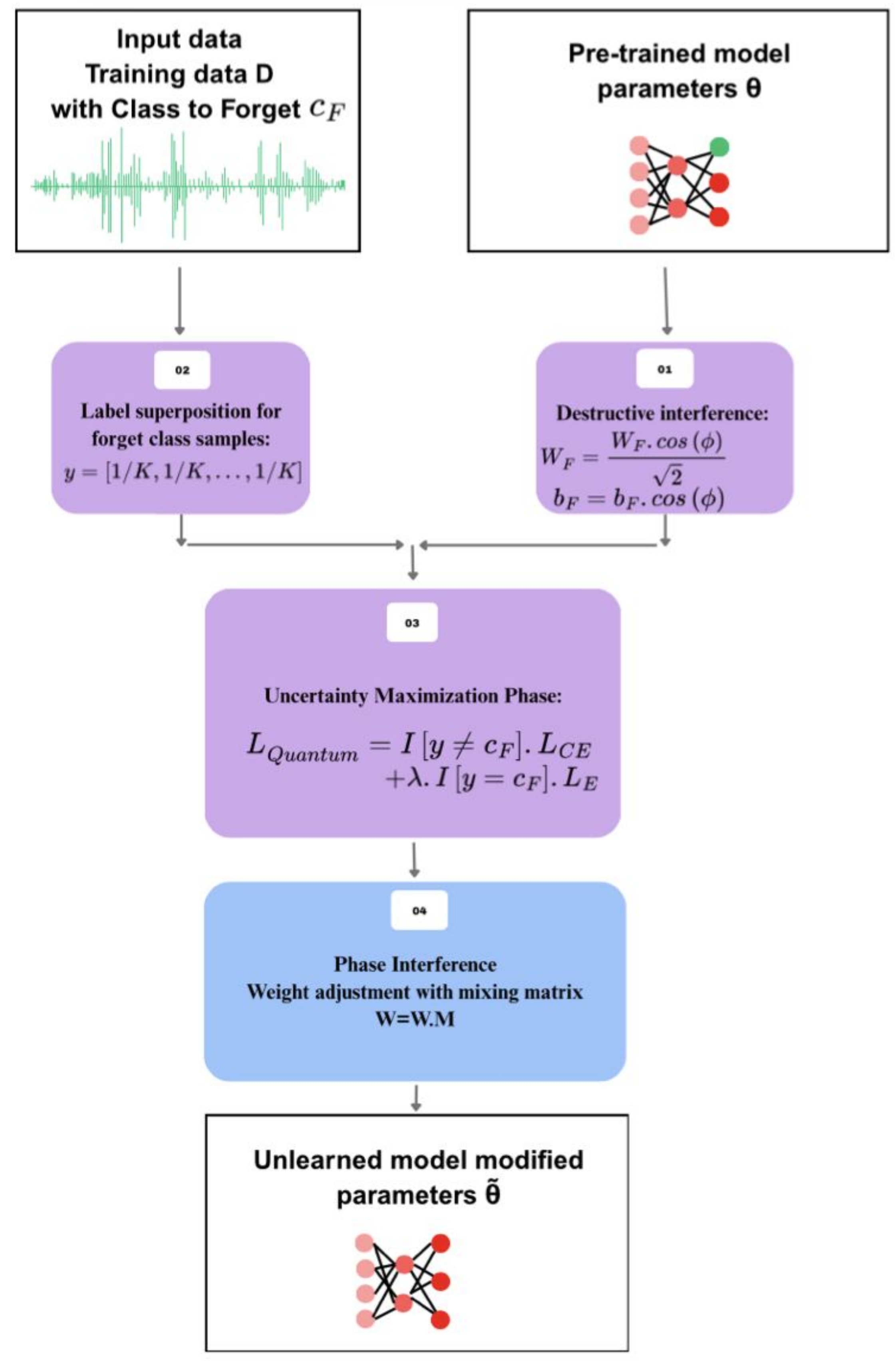
Audio Unlearning. It is a 4 phase approach:

- Destructive interference weight initialization−downscales forget-class weights to suppress its contribution.

$$\widetilde{W}_{ij} = \begin{cases} \frac{W_{ij}\cdot\cos\varphi}{\sqrt{2}} & j = F \\ W_{ij} & j \neq F \end{cases}$$

$$\widetilde{b}_j = \begin{cases} b_j \cdot \cos\varphi & j = F \\ b_j & j \neq F \end{cases}$$

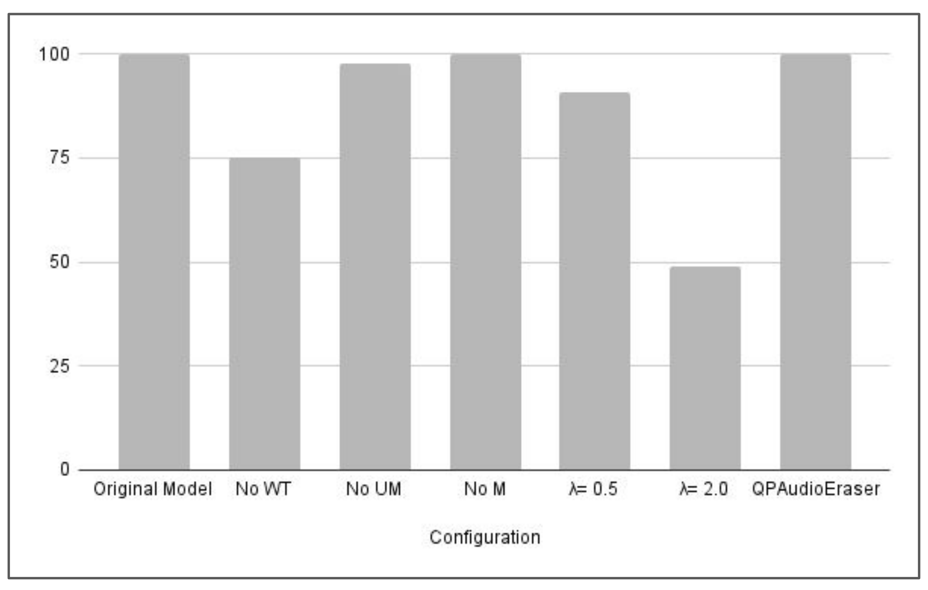
- Superposition-based label transformation replaces forget labels with a uniform distribution for maximum uncertainty.
- Uncertainty maximizing quantum inspired retraining: fine-tunes with an uncertainty-maximizing loss.
- Entanglement-inspired weight mixing interference – applies sparse mixing to disperse residual traces across classes.



Results

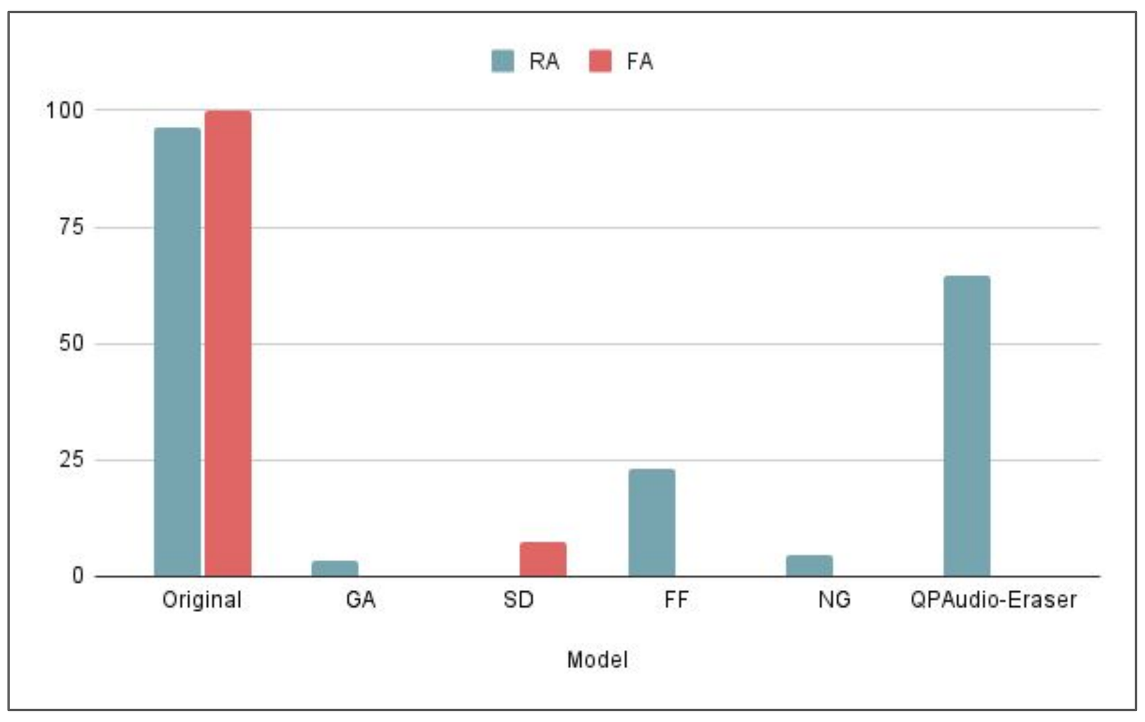
- **Single Class Unlearning**: QPAudioEraser successfully erases audio classes when tested on ViT-Tiny and ResNet-18.
- **Parallel Unlearning**: Removing forget classes' information from the models while erasing classes in parallel.
- **Continual Unlearning**: QPAudioEraser was successful on erasing information of classes sequentially on AudioMNIST.
- **Accent Unlearning**: a challenging task due to overlapping phonetic features across accents.

Ablation Study

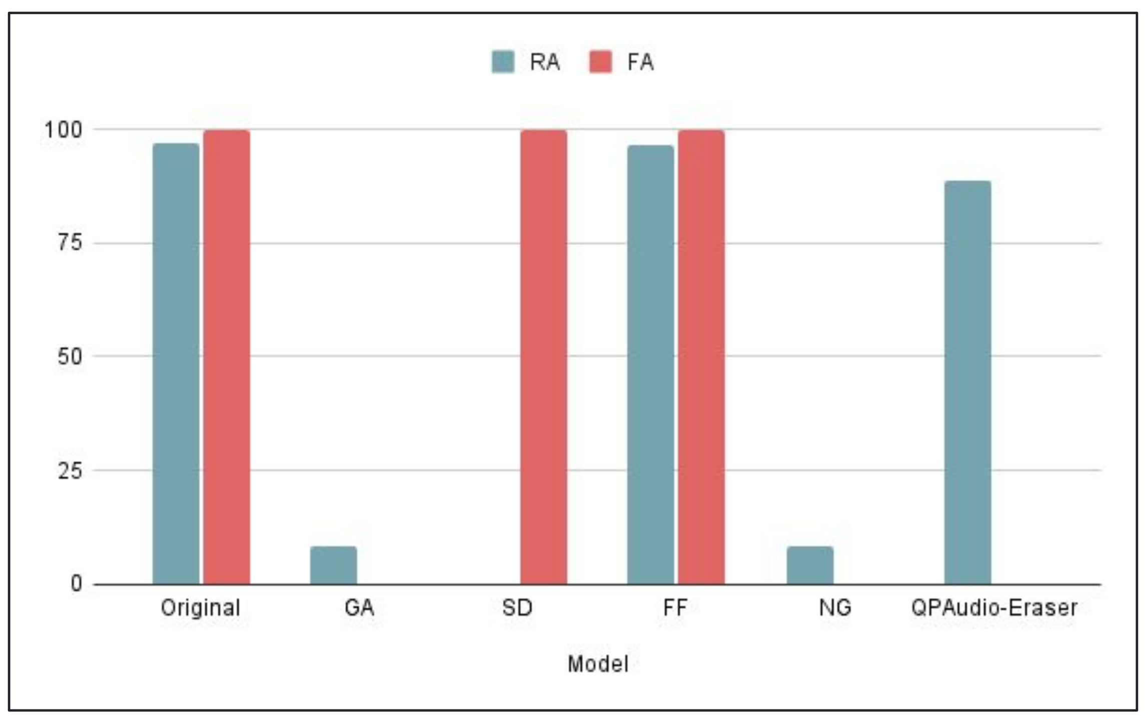


Ablation Study (Retain Accuracy) of our QPAudioEraser method on AudioMNIST using Resnet-18 backbone. FA is 0 for all cases. WT-> Weight Transform, UM-> Uncertainty Maximization.

Quantitative Results



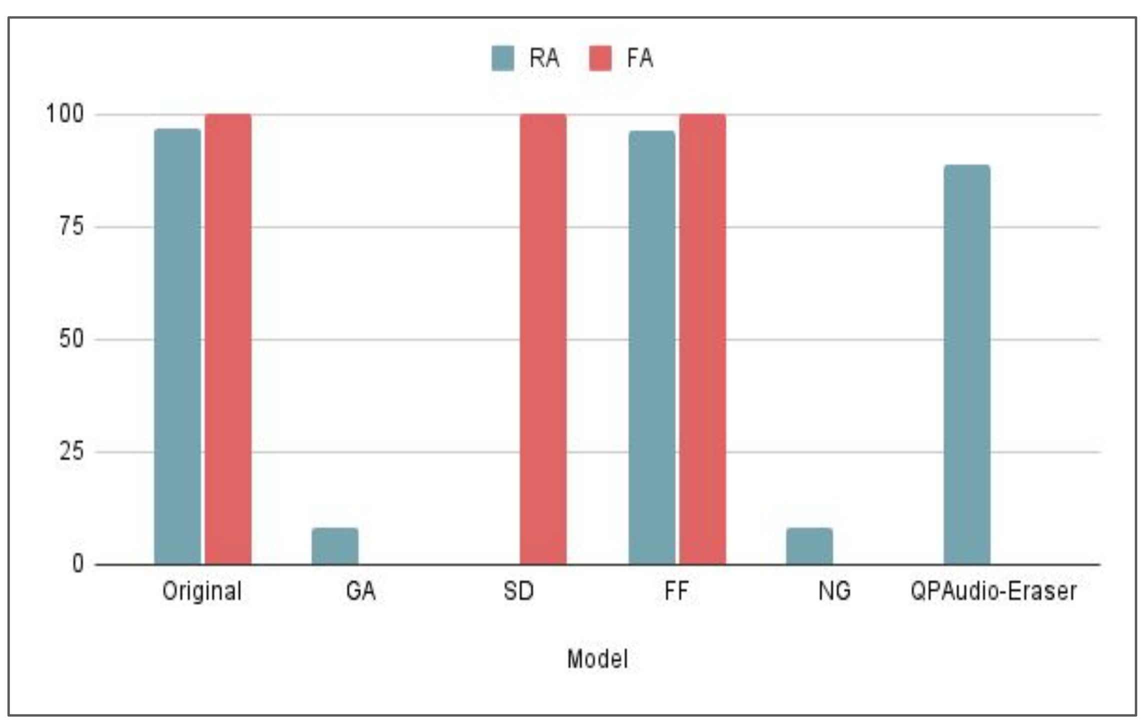
Continual Unlearning with 10% class removal on ResNet-18 (AudioMNIST).



Accent Unlearning Results on SpeechArchive Dataset using CNN classifier.

Retain Accuracy (RA) ↑
Forget Accuracy (FA) ↓

- QPAudioEraser is the only method that simultaneously drives Forget Accuracy to 0.00%
- The pipeline shows that is can scale and maintains the similar privacy–utility balance when erasing multiple classes at once or continual unlearning.
- Ablation analysis shows that every component is important.



Single Class Unlearning Results on ViT-Tiny (LibriSpeech).

Conclusion

- Proposed a quantum-inspired 4-phase audio unlearning framework that selectively erases targeted class information.
- Achieves complete removal of forget-class data while preserving accuracy on retained classes.
- QPAudioEraser was rigorously compared against well established baselines like Gradient Ascent, Fisher Forgetting, Synaptic Dampening and Negative Gradient.

IEEE International Joint Conference on Biometrics

Acknowledgment: IndiaAI Mission

Image Analysis & Biometrics Lab