Literature Survey: Digital Signal Processing Approaches for Voice Anonymisation and De-identification

Mohamed Ibrahim Behery — 24123761

November 3, 2025

Abstract

This report surveys five recent academic papers focusing on non-deep learning methods for voice anonymisation and speaker de-identification. The surveyed approaches primarily utilize Digital Signal Processing (DSP) techniques, offering alternatives to complex, data-intensive machine learning models. Key methodologies include the application of the McAdams coefficient for spectral envelope manipulation (Papers 1 and 2), Functional Data Analysis (FDA) for f0 trajectory concealment (Paper 3), combined formant-shifting and spectral swapping for deterministic pseudonymisation (Paper 4), and Phase Vocoder-Based Time-Scale Modification (PV-TSM) for pitch shifting (Paper 5). These methods are evaluated based on their effectiveness in degrading Automatic Speaker Verification (ASV) performance (measured by Equal Error Rate, EER) while maintaining speech utility (measured by Word Error Rate, WER). The findings highlight the continued relevance and competitive performance of DSP techniques in the field of voice privacy.

List of Abbreviations

- **ASV:** Automatic Speaker Verification
- **DSP:** Digital Signal Processing
- **EER:** Equal Error Rate
- **FDA:** Functional Data Analysis
- **FPCA:** Functional Principal Component Analysis
- **IFFT:** Inverse Fast Fourier Transform
- **ISCA:** International Speech Communication Association
- **LPC:** Linear Predictive Coding
- **PII:** Personally Identifiable Information
- **PV-TSM:** Phase Vocoder-Based Time-Scale Modification
- **STFT:** Short-Time Fourier Transform
- **VTL:** Vocal Tract Length
- **WER:** Word Error Rate

1 Introduction and Context

Voice data contains Personally Identifiable Information (PII), primarily through characteristics used for speaker recognition. Protecting this privacy is a critical area of research. While state-of-the-art methods often rely on complex deep neural networks, there is a strong need for simpler, more robust, and training-free solutions based on traditional Digital Signal Processing (DSP). This literature survey focuses on recent DSP-based methods designed to anonymise or de-identify speakers, with the dual objective of frustrating Automatic Speaker Verification (ASV) systems while preserving the linguistic content of the speech.

2 Literature Review: DSP-Based Anonymisation Techniques

2.1 Paper 1: Speaker anonymisation using the McAdams coefficient

ID: arXiv:2011.01130
Paper Title: Speaker anonymisation using the McAdams coefficient
Authors: Jose Patino, Natalia Tomashenko, Massimiliano Todisco, Andreas Nautsch, Nicholas Evans
Publication Venue: INTERSPEECH 2021 (Proceedings of the 22nd Annual Conference of the International Speech Communication Association (ISCA))
Core Problem Addressed: Degrading the reliability of Automatic Speaker Verification (ASV) while preserving speech intelligibility.
Core Methodology: Digital Signal Processing (DSP)-based approach requiring no training data.
Key Mechanism: The McAdams Coefficient (α) is used to manipulate the spectral envelope by affecting the angular position of the Linear Predictive Coding (LPC) filter poles.
Steps: 1. LPC Analysis (extracts filter coefficients and residual). 2. Pole Extraction (finds roots of the LPC filter). 3. Pole Transformation ($\phi \rightarrow \phi^{\alpha}$ is applied to the angle of the complex poles). 4. New Coeffs (new LPC coefficients derived from transformed poles). 5. Resynthesis (original residual passed through the modified filter).
Impact: Served as the Secondary Baseline for the VoicePrivacy 2020 Challenge, demonstrating high efficiency and competitive anonymization strength against deep learning methods.

2.2 Paper 2: Design of Voice Privacy System using Linear Prediction

ID: IEEEXplore:9306379
Paper Title: Design of Voice Privacy System using Linear Prediction
Authors: Priyanka Gupta, Gauri P. Prajapati, Shrishti Singh, Madhu R. Kamble, Hemant A. Patil
Publication Venue: APSIPA ASC 2020 (Proceedings of the 12th Annual Conference of the Asia-Pacific Signal and Information Processing Association's (APSIPA) Annual Summit and Conference)
Core Problem Addressed: Extension of the McAdams coefficient method (Paper 2.1) by incorporating manipulation of the pole radius to further improve speaker anonymity and address fairness issues in female speech.
Core Methodology: DSP-based approach utilizing Linear Predictive Coding (LPC), requiring no training data.
Key Mechanism: McAdams Coefficient (α) applied to the spectral envelope, extended to include manipulation of both the pole angle (ϕ) and the pole radius (r).
Advantages: Achieved a 18.98% higher Equal Error Rate (EER) for anonymity compared to the baseline and 5% lower Word Error Rate (WER) for utility on the Libri-test dataset, demonstrating effective improvement.
Impact: Demonstrated effective improvements on the widely-used McAdams baseline method.

2.3 Paper 3: Improving speaker de-identification with functional data analysis of f0 trajectories

ID: Elsevier:s2.0-S0167639322000498
Paper Title: Improving speaker de-identification with functional data analysis of f0 trajectories
Authors: Lauri Tavi, Tomi Kinnunen, Rosa González Hautamäki
Publication Venue: Speech Communication 2022 (Volume 140)

Core Problem Addressed: Addressing residual speaker-dependent cues present in intonational patterns (f0 trajectories) that persist after simple formant shifts are applied.
Core Methodology: A novel hybrid method combining existing DSP-based formant shifts with sophisticated manipulation of f0 using Functional Data Analysis (FDA).
Key Mechanism: Functional Principal Component Analysis (FPCA) is applied to the f0 trajectories (pitch) to identify and tamper with speaker-identifying characteristics by manipulating the principal component scores ($A_{\text{source},k}$).
Steps: 1. LPC Analysis and f0 Extraction. 2. Functional Conversion and Time Normalization. 3. FPCA Decomposition (Mean Function $\mu(t)$ and Eigenfunctions $\phi_k(t)$). 4. Score Calculation and Tampering. 5. f0 Synthesis and Vocoder Resynthesis.
Advantages: This simple, training-free approach improves formant-based speaker de-identification by up to 25%, providing an effective and irreversible manipulation of prosody.
Impact: High. Effective DSP method providing an irreversible manipulation of prosody.

2.4 Paper 4: Adjustable Deterministic Pseudonymisation of Speech (Idiap-NKI F03-9)

ID: Elsevier: s2.0-S0885230821X00058
Paper Title: Adjustable Deterministic Pseudonymisation of Speech
Authors: S. Pavankumar Dubagunta, Rob J.J.H. van Son, and Mathew Magimai.-Doss
Publication Venue: Elsevier's "Computer Speech & Language" [2022] — VoicePrivacy 2020 Challenge Submission
Core Problem Addressed: Speaker Pseudonymisation: Hiding identity while preserving linguistic and paralinguistic content using a deterministic and **reversible** DSP method.
Core Methodology: DSP-based approach utilizing formant-shifting and spectral swapping.
Key Mechanism: The F03-9 System, which uses Vocal Tract Length (VTL) simulation, fundamental frequency (F0) shifting, and targeted **Spectral Swapping** (via Hann filters) to move spectral energy.
Steps (F03-9 System): 1. Simulate VTL shift. 2. Spectral Swapping (moving F_{1-3} energy). 3. Modify Prosody (Shift F0). 4. High-Frequency Masking (swapping F_4/F_5 and replacing F_{6-9} with noise). 5. Vocoder Resynthesis.
Advantages: Demonstrated better ASV performance (higher EER) and better intelligibility (lower ASR WER) than the McAdams coefficient baseline. Its deterministic and reversible nature makes it ideal for pseudonymisation applications where original data recovery is required.

2.5 Paper 5: Speaker Anonymization by Pitch Shifting Based on Time-Scale Modification (PV-TSM)

ID: isca:mawalim22
Paper Title: Speaker Anonymization by Pitch Shifting Based on Time-Scale Modification
Authors: Candy Olivia Mawalim, Shogo Okada, Masashi Unoki
Publication Venue: 2nd Symposium on Security and Privacy in Speech Communication (SPSC 2022)
Core Problem Addressed: Suppressing Personally Identifiable Information (PII) using a simple, non-training, DSP approach with high-quality pitch manipulation.
Core Methodology: DSP-based approach utilizing a **Time-Scale Modification (TSM)** technique.
Key Mechanism: The **Phase Vocoder-Based TSM (PV-TSM)** algorithm. This method achieves pitch manipulation independent of duration by calculating phase propagation and adjusting the Synthesis Hop Size relative to the Analysis Hop Size.
Steps: 1. Short-Time Fourier Transform (STFT) Analysis. 2. Pitch Shift (scaling F0). 3. Phase Propagation and TSM (adjusting phase and setting the Synthesis Hop Size > Analysis Hop Size). 4. Signal Reconstruction (Inverse Fast Fourier Transform (IFFT) and Overlap-Add).
Advantages: Achieves a superior balance of privacy (EER) and utility (WER) compared to the McAdams coefficient baseline, specifically by utilizing high-quality phase vocoder techniques to preserve speech intelligibility.

3 Conclusion

The surveyed literature confirms that traditional Digital Signal Processing (DSP) methods are more than just historical techniques; they constitute a **vital foundational and reference framework** for all advanced voice anonymisation efforts. The intrinsic understanding of voice characteristics (like the source-filter model, represented by f0, formants, and spectral envelope) provided by these DSP methods is crucial for *preprocessing our dataset*. As our ultimate goal is to develop a machine learning (ML) algorithm that is interpretable by design, we must commit to an interpretability standard at every stage of the design. In ML, models can be categorized as post-hoc (complex systems where external tools are needed to interpret the trained "black-box" decisions) or explainable by design (architectures that intrinsically provide transparent, human-understandable reasoning). We explicitly aim for the second category, ensuring our solution is built upon a solid signal-processing foundation. The insights gained from the deterministic and mathematically explicit DSP techniques, such as the Phase Vocoder-Based Time-Scale Modification (PV-TSM) and McAdams manipulation, are indispensable when seeking to design an inherently transparent and understandable ML model for voice privacy.