

Analyzing Multimodal Biometric Data with Occluded Facial Conditions

A data analysis project exploring performance metrics of multimodal biometric authentication systems when facial features are partially hidden.

By: Melvin Mathew

Introduction

Biometric systems use physiological/behavioral traits for authentication

Face recognition performance drops under occlusion (e.g., masks, sunglasses)

This project explores whether combining voice improves accuracy in such cases

We'll cover: methodology, data, fusion techniques, results, and ethical concerns



Prior Research – Key Findings from Literature



- Unimodal Systems (Face/Voice):
 - High accuracy in ideal conditions (clear face images, noise-free voice)
 - Performance degrades significantly with occlusion or background noise
- Multimodal Fusion:
 - Prior studies (Ross & Jain, 2003) show fusion improves robustness by combining complementary modalities
 - Feature-level fusion (like our approach) often outperforms score-level fusion



Prior Research – Gaps Addressed by Our Project



- Occlusion Handling:
 - Most research focuses on non-occluded faces; we target real-world scenarios (scarves, glasses)
- Precomputed Embeddings:
 - Prior work uses raw images (ArcFace); we innovate with precomputed landmarks/embeddings
- Synthetic User Mapping:
- Simulated identity linkage between voice/face datasets, a practical solution for data scarcity



Goals & Research Question:



1

Main Objective: Improve authentication under face occlusion using multimodal biometrics

2

Research Question:
Does incorporating a second biometric modality (e.g., voice) through feature-level fusion improve authentication accuracy under occluded face conditions?

3

Hypothesis: Fusion of voice and occluded face data yields better accuracy than either alone



+

•

○

Data Collection



- **Face Data:**
 - SOF (Surveillance Occluded Faces) Dataset
 - Face embeddings with occlusion (scarves, glasses)
 - Precomputed 68-landmark vectors
- **Voice Data:**
 - Mozilla Common Voice Dataset
 - Clean MP3 recordings converted to WAV
 - ECAPA-TDNN model used for voice embeddings
 - Users were simulated using randomized identity mapping between datasets



Methodology

Designed a multimodal authentication system using voice and face biometrics.

Used ECAPA-TDNN for voice embeddings and precomputed face embeddings from SoF dataset.

Implemented feature-level fusion to combine normalized voice and face features.

Classifiers: SVM (linear kernel) with 5-fold cross-validation for evaluation.



Results

Voice-only accuracy: 95%

Face-only accuracy: 15% (due to occlusion in SoF)

Fusion accuracy: 97.5%

Evaluation Metrics: ROC, DET, EER, d-prime, and Score Distribution

Fusion consistently improved performance across all metrics



Discussion

- Fusion significantly enhanced accuracy in occluded face scenarios.
- Supports our hypothesis that multimodal fusion improves robustness.
- Limitations:
 - Couldn't use models like ArcFace (precomputed embeddings).
 - Face-only underperformed due to heavily occluded dataset.
- Future Work:
 - Try raw face image pipelines (e.g., ArcFace).
 - Expand dataset diversity and add more modalities





Ethical Impacts

Privacy & Consent Risks

- Biometric data like voice and face is highly sensitive.
- Risk of collection without clear user awareness or permission.

Bias & Fairness Concerns

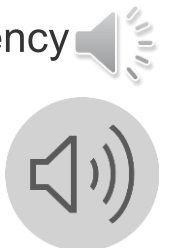
- Systems may underperform for certain age, gender, or racial groups.
- Can lead to misidentification or denial of access.

Trust & Function Creep

- Data collected for one purpose may be reused for another without consent.
- This undermines user trust in biometric technologies.

Ethical Safeguards

- Important to enforce informed consent, fairness, and secure data handling.
- Systems should be designed with transparency and accountability in mind.



Conclusion

✓ Research Question Recap

- Does adding voice to occluded face data improve authentication accuracy?
- Yes — feature-level fusion significantly improved performance.

📊 Key Results

- Voice-only accuracy: **95%**
- Face-only accuracy: **15%**
- Fusion accuracy: **97.5%**

🔍 Insights

- Voice embeddings from ECAPA-TDNN are highly robust.
- Fusion helps compensate for weaknesses in occluded face data.
- ROC, DET, and EER metrics confirmed improved performance.

⚠️ Limitations

- Small sample size
- Precomputed face embeddings
- Artificial user mapping between datasets

